ISSN 1574-1796

An International Journal on Grey Literature

Volume 15, Number 1, Spring 2019

'DATA LIBRARIANS SPEAK OUT ON GREY LITERATURE'

GreyNet

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

The Grey Journal

An International Journal on Grey Literature

COLOPHON

About TGJ

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

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

CIP
The Grey Journal (TGJ) : An international journal on grey literature / Dominic Farace (Journal Editor); Jerry Frantzen (Technical Editor) ; GreyNet International, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 15, Number 1, Spring 2019. – *TIB (DE), DANS-KNAW (NL), CVTISR (SK), EBSCO (USA), Inist CNRS (FR), ISTI-CNR (IT), KISTI (KR), NIS IAEA (AT), NTK (CZ), and the University of Florida (USA) are Corporate Authors and Associate Members of GreyNet International.*
ISSN 1574-1796 (Print)
ISSN 1574-180X (PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:
€240 institutional, including postage & handling

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

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

About GreyNet

The Grey Literature Network Services was established in order to facilitate dialog, research, and communication between persons and organizations in the field of grey literature. GreyNet further seeks to identify and distribute information on and about grey literature in networked environments. Its main activities include the International Conference Series on Grey Literature, the creation and maintenance of web-based resources, a moderated Listserv, and The Grey Journal. GreyNet is also engaged in the development of distance learning courses for graduate and post-graduate students, as well as workshops and seminars for practitioners.

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

Contents
'DATA LIBRARIANS SPEAK OUT ON GREY LITERATURE'

The data librarian: myth, reality or utopia? 7
Silvia Giannini and Anna Molino, Institute of Information Science and Technologies, ISTI-CNR, Italy

Research Data Management: What can librarians really help? 23
Yuan Li, Willow Dressel and Denise Hersey, Princeton University, United States

Data Management and the Role of Librarians 31
Plato L. Smith, Sara Gonzalez, and Jean Bossart, George A. Smathers Libraries; University of Florida, United States

Librarians' Role in GAO Reports 39
Meg Tulloch, U.S. Government Accountability Office, GAO, United States

Published electronic media are becoming Grey 45
Yui Kumazaki, Satoru Suzuki, Masashi Kanazawa, Katsuhiko Kunii, Minoru Yonezawa, and Keizo Itabashi, Japan Atomic Energy Agency

Measuring Reuse of Institutionally-Hosted Grey Literature 51
Ayla Stein Kenfield, University of Illinois at Urbana–Champaign; Elizabeth Kelly, Loyola University New Orleans; Caroline Muglia, University of Southern California; Genya O'Gara, Virtual Library of Virginia; Santi Thompson, University of Houston; Liz Woolcott, Utah State University, United States

Colophon 2

Editor's Note 5

On The Newsfront

AccessGrey Project - Securing Open Access to Grey Literature for Science and Society 60

GL21 Conference Call-for-Posters 61

Advertisements

Ebsco Library, Information Science & Technology Abstracts with Full Text (LISTA) 4

ISTI-CNR, Institute of Information Science and Technologies, Italy 6

TIB, German National Library of Science and Technology 59

Author Information 62

Notes for Contributors 63

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

EDITOR'S NOTE

Twenty-First International Conference on Grey Literature

Open Science Encompasses New Forms of Grey Literature, 22-23 October 2019

For more than a quarter century, grey Literature communities have explored ways to open science to other methods of reviewing, publishing, and making valuable information resources publicly accessible. This Twenty-First International Conference on Grey Literature seeks to demonstrate how the principles of science and advancements in information technology have impacted the field of grey literature and in turn how grey literature by implementing these has contributed to the open science movement.

Open science is defined as the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society. Grey literature by definition seeks to make publications produced on all levels of government, academics, and business openly accessible different from those controlled by commercial publishing. As such the open science movement incorporates the work carried out by grey literature communities and renders an even broader framework encompassing newer forms of grey in both textual and non-textual formats.

Open science encompasses the life and physical sciences as well the social sciences and humanities as does grey literature. Open science recognizes the value of grey literature in the process of knowledge generation and as such acknowledges contributions made by researchers, authors, and their communities of practice. Open science changes the way research is done and allows for convergence with the field of grey literature. It is within the open science movement that grey literature and its wealth of information resources are valued and properly exploited for society as a whole.

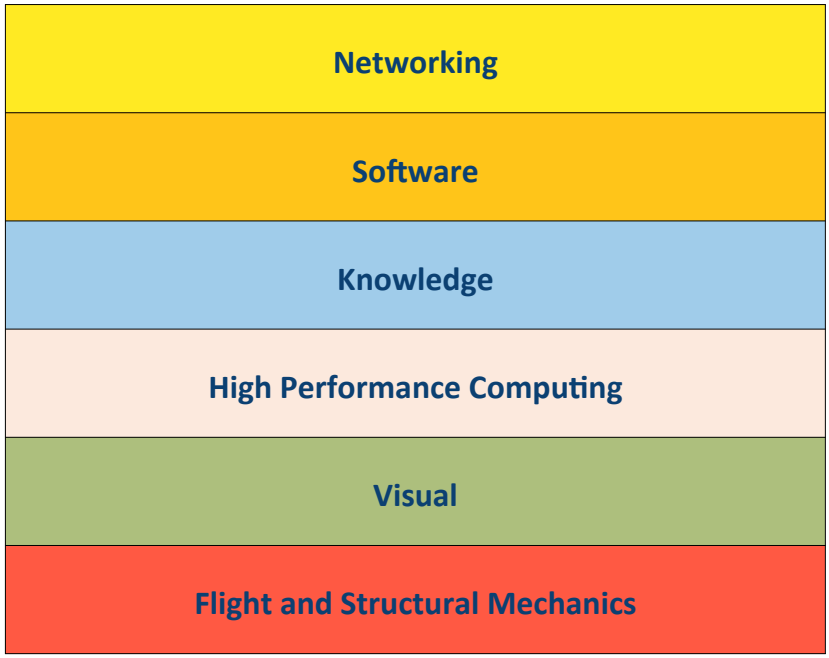
Dominic Farace,
Journal Editor

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

1. Introduction

It is undoubtedly true that we live in a more and more data-centric world (Cassella, 2016). As citizens and users of the Internet, we produce an enormous amount of data on a daily basis. This may be done consciously (e.g. when we deal with governmental and public organizations, universities and research centers, structured companies, etc.), as well as less intentionally, for instance when we use social networks, and, more generally, as users of the Internet for the widest variety of purposes.

In this scenario, data is becoming of greater importance not only for the average user when it helps making choices and decisions, but mainly for the growing number of companies and entities worldwide founding their activity in collecting and elaborating the exponentially increasing amount of information produced by the citizenry.

Therefore, we are now facing and coping with what has been defined by many commentators a "data deluge".

In the academic and, more generally, scientific world the need for management, long term preservation, and storage of research data has grown exponentially in the recent past. As the Turing award winner Jim Gray states: "a fourth paradigm [of science] is emerging, consisting of techniques and technologies needed to perform data-intensive science" (Gray, Szalay, 2007). Indeed, in almost all discipline areas "born digital" documents proliferate as files, spreadsheets, databases, digital notebooks, wikis, etc. As a consequence, the management, curation, and archiving of such data are becoming of crucial importance (Bell, Hey and Szalay, 2009).

In this context, the expressions *eScience* and *eResearch*[1] have been identified as umbrella terms describing converging sets of trends and technologies that are radically changing the way science is conduced. Librarians may bring important knowledge and skills complementary to the activities carried on by the eScience community, most notably in the management, preservation and archiving of information (Wright et al., 2007). Moreover, another essential component of eScience concerns the management of the scholarly communication lifecycle, being this one of the most prominent area of interest in the work performed by libraries.

The Open Science (OS) movement, in turn, is radically changing the perspectives adopted in the scientific production and dissemination, fostering new approaches to research and scholarly communication. The 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.

As far as OD is concerned, firstly it must be pointed out that after the regulation of the Open Access to scientific literature, the legal framework in Europe is recently supporting the approach to OA.

Indeed, last European Recommendations (April 25 2018)[2] ask Member States to ensure that data management planning becomes a scientific practice and since 2017 the European Commission has made mandatory to open research data for all participants in Horizon 2020 and for any subject area, provided this is allowed from a legal or ethical point of view.

Thus, not surprisingly various funding bodies (e.g. the European Commission, the Wellcome Trust and the RCUK in UK, the Australian Research Council in Australia, or the National Institutes of Health in the U.S.) have been making mandatory the submission of a Data

* First published in the GL20 Conference Proceedings, February 2019.

[1] *eScience* is the term preferred in Europe, while in other countries (e.g. Australia) the initiatives aiming at transforming the approach to science are labelled as *eResearch* (cfr. Wright et al., 2007).

[2] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018H0790&from=EN

Management Plan (DMP) together with the project proposal. For instance, in the article 29.3 of the Annotated Model Grant Agreement[3], the EU asks all consortia submitting a proposal in H2020 program to declare: the type of data that would be produced during the project; the strategies for their management in order to guarantee their short- and long-term preservation; how much of the produced data would be openly available.

In this perspective, academic libraries are indeed increasingly involved in the management of research data across the lifecycle (Schmidt and Shearer, 2016), 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 context, we may ask ourselves: who is currently responsible for the management, curation, and archiving of (research) data? Fearon et al. (2013) observes that:

"the data management space in US in higher education is predominantly owned by the libraries […], whereas in the UK it is much more dependent on individual institutional cultures and circumstances whether it is the librarians, the academics, or the administrators who take the lead".

The Research Data Alliance (RDA)[4] recognizes that: "Many academic libraries are now extending their century-long track record in the professional management of knowledge resources towards the area of research data and therefore seek to maximize research data skills among staff in their organizations". They identify five main routes to achieve such goal, consisting in: training, expert recruitment, learning-on-the-job, online-courses, and (academic) degrees.

Swan and Brown (2008) in their report commissioned by the UK Joint Information Systems Committee (JISC) recognize a strategic role for libraries in data management, identifying three main potential roles: increasing data awareness; providing archiving and preservation services; developing a new professional strand of practice as data librarianship.

Many commentators have argued that the background knowledge of librarians may be essential in this scenario. For instance, the management of repository's contents may be seen as a *collection management issue* (Genoni, 2004), while the expertise in classification and description through cataloguing and metadata, as well as the experience in the selection of the information may be crucial for data curation (Witt, 2008).

Starting from this framework, we can consider academic libraries as "aggregator, collector and curator of external scholarship, be it printed or online". For this reason, in our work we will try to get an idea of the competencies required to a Data Curator, giving an overview of the features of Research Data Management and its conversion into an effective service (RDS), both from a theoretical point of view and through the observation of some concrete examples of data management by librarians. The aim is to understand if librarians are the most accredited candidates to fill the role of Data Curator, giving possible answers and outlining specific qualifications required to those currently operating in academic libraries with the purpose of possibly identifying the figure of *data librarian*.

2. What is (research) data?

2.1. Data: definitions and types

As anticipated in the introduction, the contemporary world may be defined as a data-centric reality, due to the huge amount of data we use and produce in our everyday life.

But what is the meaning of the word *data*?

The broader definition of the term given by the Cambridge Dictionary[5] is: "information collected for use"; more specifically, data can be seen as "information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer".

The Oxford English Dictionary[6] generally speaks about: "facts and statistics collected together for reference or analysis", providing the more discipline-oriented sub-definition:

[3] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf.

[4] 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.

[5] https://dictionary.cambridge.org/dictionary/english/data

[6] https://en.oxforddictionaries.com/definition/data

"The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media".
Merriam-Webster[7] specifies three different connotations of the word, seeing data as: "factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation"; "information in digital form that can be transmitted or processed"; "information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful".
It is noticeable that all three dictionaries cited take into account the shades of meaning concerning information science and technology, making reference to what is relevant for data-analysis as specific area of interest.
It is possible to identify different types of data, based on their exploitation, purpose, etc., being them identified and categorized in a more or less discipline as well as use oriented.
In our work, we focus on a specific typology of data, namely *research data*. The following subsection is dedicated to the description of this category, trying to outline a possible definition embracing the quantity of information produced in this specific context.

2.2. Research Data

The definition of research data may be quite broad and open to different meanings, which can vary depending on the disciplinary field we consider. Indeed, what a researcher considers to be *research data* depends on the meaning of this data in the research process and this may differ for each scientific discipline. Therefore, different definitions of this concept have been suggested by various actors and entities operating in this field. Here we report those we find more significant, proposed by the "Essentials 4 Data Support" online course of the Research Data Netherlands (RDNL)[8].
The Queensland University of Technology in its *Manual of Policies and Procedures* defines research data[9] as: "data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or other research output is based. It relates to data generated, collected, or used, during research projects, and in some cases, may include the research output itself. Data may be numerical, descriptive, visual or tactile. It may be raw, cleaned or processed, and may be held in any format or media […]".
The UK Engineering and Physical Sciences Research Council (EPSRC)[10] consider research data as: "recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created".
Crossley (2004) in her "Introduction to managing research data"[11] argues that they are "collected, observed or created for the purpose of analysis to produce and validate original research results".
Indeed, it must be taken into account that research data may be presented in a variety of formats, both digital or physical, e.g. electronic text documents; spreadsheets; laboratory notebooks, field notebooks, and diaries; audiotapes and videotapes; specimens, samples, and artefacts; methodologies, workflows, standard operating procedures and protocols; metadata, and so on (Scott & Cox, 2016).
Overall, the definition we prefer for its conciseness and effectiveness is the more general one, saying: "Research data is the material underpinning a research assertion", making reference to all the outcomes produced in the course of the research, from statistics to field observations and answers to questionnaires, in spite of their formats or media.

[7] https://www.merriam-webster.com/dictionary/data
[8] https://datasupport.researchdata.nl/en/start-the-course/i-definitions/research-data/
[9] http://www.mopp.qut.edu.au/D/D_02_08.jsp
[10] https://epsrc.ukri.org/about/standards/researchdata/scope/
[11] https://www.scribd.com/presentation/138079216/Managing-Research-Data

2.3. Research Data as Open Data

In the context of Open Science, the possibility of making research data as open as possible becomes of crucial importance. Indeed, there are considerable advantages in sharing materials supporting the research: sharing the research outcomes encourages the cooperation between scientific communities and favors a faster and more efficient research process, as it avoids useless data duplication and stipulates the collaboration between institutions and with the citizenry.

This may be accomplished following a series of practices and principles helping the scientific community in the correct production and reuse of the research results. For instance, the main goal of FORCE 11[12], a community of scientists, librarians, archivists, publishers and funders, is the promotion of the FAIR data principles. The acronym means *Findable, Accessible, Interoperable, Reusable* and it corresponds to a set of guidelines that enables a better realization and sharing of the data.

This brings to light the crucial issue concerning digital and, more specifically, data curation. The affirmation of digital products and services has in fact brought with it a set of strategies, technological approaches and activities that have taken the name of *Digital Curation*.

Digital Curation can be considered a transversal activity to various fields consisting in the creation, the maintenance and the preservation of a digital object throughout its lifecycle. The active management of research data reduces threats to their research value and mitigates the risk of digital obsolescence, enhancing the long-term value of existing data by making it available for further high quality research.

Digital curation and data preservation are ongoing processes, requiring considerable thought and the investment of adequate time and resources. This is the reason why the DCC[13] in UK has identified some steps to be followed during what has been named *digital curation lifecycle*, as represented in the picture below:

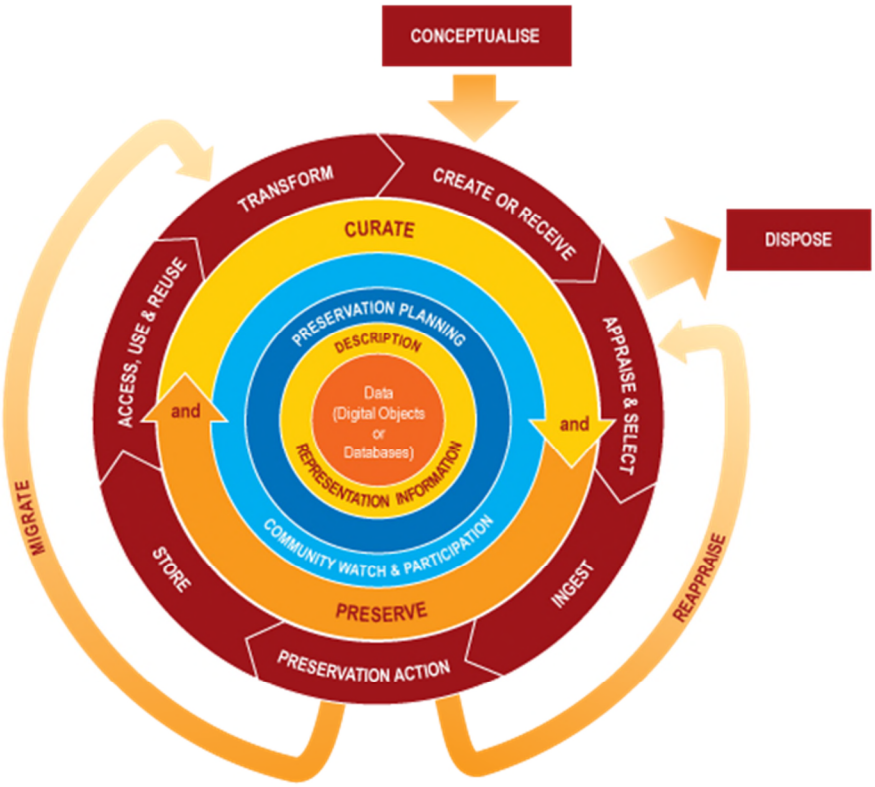


Figure 1: Digital Curation Lifecycle

[12] https://www.force11.org/group/fairgroup/fairprinciples
[13] http://www.dcc.ac.uk/digital-curation/what-digital-curation

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*. At the present time, this may not be considered a structured and well-defined character; as a consequence, the breadth of the definition implies that its reference community may include different actors and professional figures.
The concept of Data Curation can be seen as a sort of subset of the Digital Curation. Strictly connected with the academic and research world, it comes from the connection of the digital curation with the development and management of Open Access repositories. At the same time, the data curator becomes a specialization of the digital curator.
The reference community of Data Curation is very often limited to the researchers and the type of data taken into account are the research data associated to the scientific literature.
Thus, the professional figures move towards the librarians or, more exactly, towards the upcoming figure of the *data librarian*, for their wide experience in different disciplinary domains, their skills in the management of metadata sets, in the maintenance of collections and their involvement in the management of research information.
The following paragraph is dedicated to Research Data Management, giving particular attention to its transformation into a specific service granted by some established research entities, namely Research Data Service.

3. Research Data Management and Research Data Service

3.1. What is Research Data Management (RDM)?

Research Data Management (RDM) is a general term to indicate a set of good practices concerning collecting, storing, using, sharing and preserving research data in an effective and productive manner. It involves services, tools and infrastructures that support the management of research data, which may significantly differ across the lifecycle. (Schmidt and Shearer, 2016; Schmidt et al., 2016).
Whyte and Tedds (2011) in their Briefing Paper for the DCC clarify some terminological distinctions between research data management, preservation and curation, arguing that: "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. Preservation is about ensuring that what is handed over to a repository or publisher remains fit for secondary use in the longer term (e.g. 10 years post-project). Curation connects first use to secondary use. It is about ensuring that project results are fit to archive, and that valued research assets remain fit for reuse".
The various aspects of RDM should be seen as research support services distributed across various departments (e.g. Research Offices, IT Services, Libraries), as researchers need support in different areas, such as planning, organizing, documenting and sharing, preparing datasets for deposit and long-term preservation, not forgetting copyright issues (Schmidt and Shearer, 2016).
Therefore, RDM involves a wide range of activities across the data lifecycle, requiring a high level of interaction with both researchers and other support services (e.g. technical services and research officers), such as creating and collecting, processing, analyzing, publishing, archiving and preserving, and re-using data (Schmidt et al., 2016).
Figure 2 illustrates the major steps of the RDM lifecycle, highlighting the common points that might be shared by different scientific communities and realities worldwide:

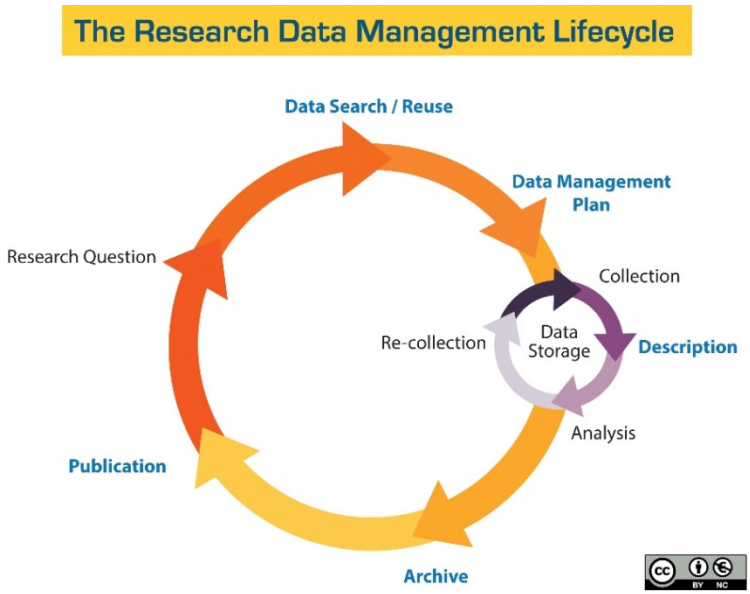


Figure 2: Research Data Management Lifecycle, diagram (University of California)

As for the Research Data Netherlands[14], those operating in RDM should guarantee support within the following influence spheres:

- Legislation and policy;
- Technological infrastructure;
- Culture (i.e. research practices and their exploitation);
- Knowledge of storing, managing, archiving and sharing research data by both researchers and support services;
- Skills (e.g. conversational and influence skills);
- Motivation for collecting and managing research data.

3.2. Libraries' competencies in RDM

As anticipated above, the traditional competencies of librarians may be well-versed in RDM, due to the broad experience in wide-ranging disciplinary domains, the practice in the management of metadata sets, as well as in the creation and preservation of collections. More in general, we can recognize that librarians have always been familiar in dealing with data.

In their Final Report on Research Data Management (2012), the LIBER Working Group on eScience identified *Ten recommendations for libraries to get started with research data management*, underlining the crucial role of librarians in offering support in data management, in the development of metadata and data standards, their participation in the elaboration of institutional policies for data administration, as well as in assisting in the exploitation of interoperable infrastructures granting access and storage to research data (e.g. via the application of persistent identifiers), being possibly involved in subject specific RDM practices.

Schmidt et al. (2016) identifies three main groups into which library services for RDM can be broadly categorized: provide access to data; awareness and support to students and researchers in handling data; managing data collections. Despite the possible overlaps between them, authors also identified distinctive roles for librarians inside each area. Indeed, the first one reflects more traditional library services (e.g. consultation and reference for datasets); the second involves hands on support for researchers across the data lifecycle, e. g. in policy and advocacy on RDM and data sharing, as well as training. The third and final category includes the preparation of data for deposit, the management of metadata, and data preservation activities (Schmidt and Shearer, 2016).

Moreover, libraries have also the opportunity to act as point of contact with the public audience, supporting public engagement with science and acting as a hub to collate links and

[14] https://datasupport.researchdata.nl/en/start-the-course/vi-data-support/influence-sphere/

information about citizen science activities, as well as researchers advocating the use of guidelines and templates. In addition to this, librarians are crucially involved in the training of scientists in data management and reuse (Lyon 2012).
This favors the evolvement of RDM practices into Research Data Services (RDS), which find numerous and different applications throughout academic and, more in general, research realities around the world.

3.3. Research Data Management as Research Data Service

The fundamental importance acquired by data collection and reuse in the research process, as well as the establishment of data management mandates by funding bodies have motivated research libraries to develop a set of services that may be generally labelled as Research Data Services (RDS).
Indeed, RDM acquires substance when it becomes a service, i.e. when an infrastructure made up of people and tools is established, providing assistance and advice for RDM practices supporting all the phases constituting a research project.
The results of the studies conducted by Tenopir et al. (2014) on U.S. and Canadian research libraries shows that the provision of RDS would augment institutions' research impact as well as the perception of the library in terms of relevance and prestige, RDS fitting the traditional role of librarians as "stewards of scholarship". At the time of their investigation, the most common services provided by academic libraries in U.S. and Canada could be seen as "extensions" of familiar reference services into the realm of data, dealing mostly with access to and citation of datasets (Tenopir et al., 2014).
We identified some common basic services provided nowadays by research libraries participating in RDS; generally speaking, they are primarily involved in:

- Support in Data Management Plan (DMP) development;
- Digital Curation, i.e. data selection, preservation, maintenance, and archiving;
- Metadata creation and transformation.

Another fundamental aspect that must be taken into account when speaking about established RDS in universities and research bodies worldwide, is the presence of a formal RDM Policy, being this at institutional level (as in the case of many universities in UK, e.g. University of Edinburgh) or presented as codes of conduct at national level, as in the case of the *Australian Code for the Responsible Conduct of Research*, which covers a wide range of topics associated with research including the management of research data and associated materials[15]. The availability of such policies endorses the formal establishment of the roles involved in RDS, regulating the smooth functioning of the research support in each institution and contributing to the definition of upcoming figures such as the *data librarian*.
Finally, we can argue that RDS may be reasonably conceived as the conversion of RDM concepts into concrete practices. In order to validate such assertion, we reviewed some literature about RDS experiences worldwide and selected three case studies that we propose in the following paragraph. This helped us to understand how this process turns into reality.

4. Case Studies

4.1. University of Edinburgh Research Data Service (RDS)

The first case study we analyzed is the Research Data Service provided by the University of Edinburgh[16]. The aim of the service is to provide tailored support to researchers dealing with the production and/or reuse of research data, offering tools, support and training to university staff and students.
The Research Data Service, led by a *data librarian* whose background is in library services, is part of the Information Services of the University, whose experts contribute in delivering specific tools and software components for the management of the data produced during the research lifecycle. The main goal is to provide tailored services for researchers aiming to achieve good practices in RDM, according to their specific needs. Moreover, at any time of

[15] https://www.ands.org.au/guides/code-awareness
[16] https://www.ed.ac.uk/information-services/research-support/research-data-service/about-the-research-data-service

such process people working with research data may ask support for training, which will be delivered following specific programs.
It must be pointed out that the University of Edinburgh has a formal Policy for RDM, establishing that data must be "managed to the highest standards as part of the University's commitment to research excellence", granting that data will be made available as open as possible, protecting those considered as sensitive and giving the widest outreach to those that may be of public interest. Indeed, the last document's clause clearly states that: "Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding" [17].
The support offered may be divided into three major phases: planning, active research project phase, project conclusion. These represent the milestones of the project lifecycle and, subsequently, of the assistance provided by RDS.

Following this pattern, the RDM services delivered at the University of Edinburgh may be schematized as follows:

1) *Before*: this consists of the identification of existing datasets; the planning for the data collection and storage; the identification of possible sensitive data, as well as methods and terms for data sharing. We can reasonably argue that the crucial part of this step regards the creation of a Data Management Plan (DMP), required by the majority of funding bodies and universities, which accompanies the research project during its lifetime. The RDS at the University of Edinburgh assist researchers in its development, providing either tools and templates (e.g. DMPonline[18]) or personal consultation to discuss the DMP in detail and obtain expert advice.
2) *During*: in the active development of the project, consultation is offered about finding existing datasets containing data that might be reused and re-elaborated, some being freely available, others behind paywall. The Data Library here plays a crucial role, giving advices about possibility for exploitation by users, as well as helping researchers in the selection of the data resources based on their type (e.g. surveys, censuses, databases, etc.). Other contributions offered by the RDS regard solutions for data storage during the project lifetime, for the control and safeguard of sensitive data, for the sharing and versioning of data, keeping track of the changes while working with other researchers or research teams.
3) *After*: after the conclusion of the research project, RDS grant assistance in recording, sharing and archiving research data for the long-term. This is made via a set of specific tools recording descriptive metadata, providing storage in an open repository for the online discovery and re-use through the association of a persistent identifier (DOI) to researchers' data resources, and securing long-term archiving in order to keep data safe from accidental deletion or inappropriate access, meeting possible funders' requirements.

Besides RDM consultancy and support, particular attention is given to training. The RDS at The University of Edinburgh offer a wide range of courses for those unfamiliar with the fundamentals of research data management and sharing, in the form of online courses, classroom-based workshops and seminars.
Indeed, people dealing with research data have the possibility to select a suitable training option among the variety proposed, depending on their specific necessities.
For instance, a free five-week MOOC - created by the Universities of Edinburgh and North Carolina – has been designed to reach learners of various types across disciplines and continents. The subjects covered are: understanding research data; data management planning; working with data; sharing data; archiving data, following the stages of a generic research project.

[17] https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy
[18] https://dmponline.dcc.ac.uk

In addition, a free, online course named MANTRA[19] has been realized with the purpose of understanding and reflecting on the management of the data collected throughout the research. This is particularly referred to post-graduate students, early career researchers, and also information professionals. It is composed of a series of interactive online units concentrating on the explanation of the terminology, key concepts, and best practices in RDM.

Inside this training path, a special focus is dedicated to librarians, underlining the central role of such figure in the RDM workflow. Indeed, a *Do-It-Yourself Research Data Management Training Kit for Librarians*[20] has been created in order to supply to the needs of academic liaison librarians. It is provided by EDINA and Data Library, University of Edinburgh, in association with the UK Data Archive, Digital Curation Centre (DCC), and Distributed Data Curation Center at the Purdue University Libraries. After an introductory "pre-training", the course is divided into five main sections, each containing a wide range of materials (e.g. podcasts, presentations, assignments, etc.) and specifically concentrated on: data management planning; organizing and documenting data; data storage and security; ethics and copyright; data sharing. Moreover, materials for post-training study are available, such as the *Data Curation Profiles,* which provide a complete framework for interviewing a researcher in any discipline about their research data and their data management practices, giving practical overviews on what a librarian involved in RDM would face in the daily practice.

Finally, also bespoke group training and face-to-face classes and workshops are planned upon requests or periodically.

4.2. The Research Data Netherlands Front Office – Back Office model

The second case study we have taken into consideration regards the Front Office – Back Office Model of the Research Data Netherlands (RDNL FO-BO Model).

It consists of a federated infrastructure handled by DANS, 3TU.Datacentrum and SURFsara - the three organizations constituting RDNL, a coalition joining three archives in the area of long-term archiving, also open to other third parties - and modelled into a four-layer structure:

1) a basic technical infrastructure under the computer centers responsibility;
2) back office data services, providing facilities for long-term archiving and accessibility;
3) front office services, granting support and training to researchers and students in responsible data management;
4) data generators and data users.

The model is graphically represented in figure 3 below:

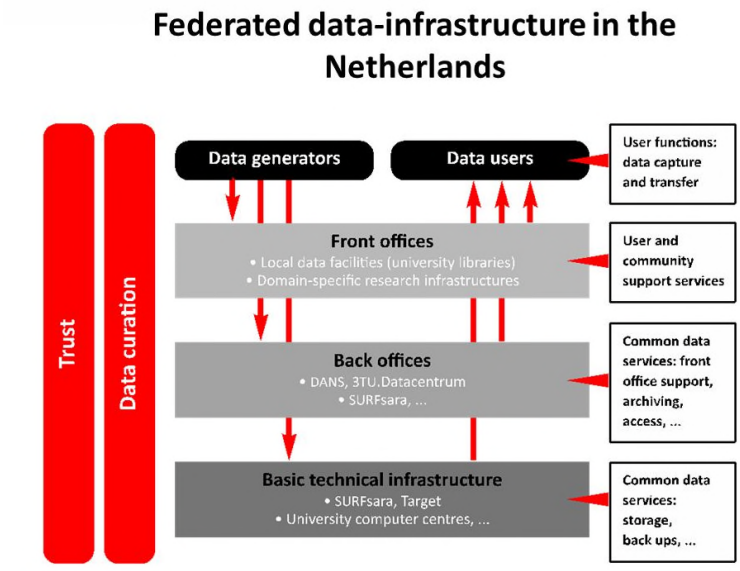


Figure 3: The federated data-infrastructure in the Netherlands

[19] https://mantra.edina.ac.uk/

[20] https://mantra.edina.ac.uk/libtraining.html

RDNL infrastructure moves from the model proposed in the EU *Riding the wave* report (2010), giving a general impression of how the various actors, data types and services should be interconnected in a global e-infrastructure for science. It must be also noticed that Netherlands have recently published the *Netherlands Code of Conduct for Research Integrity* (2018), which delineate specific directions for data management (pp. 20-21).

Globally, the services provided fall into three main categories: information provision, training, and data curation, management and storage.

More in details, the front-offices are situated mainly locally, in most of the cases in research libraries; their focus is on supporting their own research organizations, being primarily responsible for the quality assurance of the data produced.

The front-office is accountable for data collection and acquisition, as well as awareness raising in best practices for data management within its research community of reference, providing information and training to their research personnel. Moreover, the front office hosts so-called Virtual Research Environments or Data Labs, offering research tools and securing mid-term storage facilities for the organization's researchers. In consultation with the back office, the front office also facilitates the transfer of data to a trusted back-office digital repository after the research has been completed. Facilities that are shared by several universities can be hosted and supported by the back offices.

The back-offices are constituted mainly of computer experts, who provide data stewardship, guaranteeing long-term storage and accessibility to the collected data. Their functions are performed mainly by organizations such as DANS, 3TU.Datacentrum, and SURFsara, having a nationwide coverage and expertise on data from various discipline areas (humanities, social studies and sciences).

The back-office also provides consultation, training and support to front-office employees, acting as a center of expertise and innovation. Its fundamental duty is to ensure a sustainable and secure storage and retrieval upon completion of the research project.

It must be pointed out that since the division of labor between front-office and back-office is not always necessarily sharp, especially because organizations may differ in size, staffing capacity, etc. there might be institutes performing front-office tasks only, while outsourcing the back-office ones to a data archive. In this perspective, we can see how the figure of the librarian in this kind of RDM model is always present and plays a relevant role.

4.3. eResearch at Griffith University

The third and final case study we analyzed regards the activities conducted by the eResearch unit at Griffith University (Australia), where the Division of Information Services (INS) integrates what they have named e-research, library, and information and communication technology into a single organization (Brown et al., 2015).

As in the previous case studies, Griffith University responds to the *Australian Code for the Responsible Conduct of Research* (2007) - assigning researchers and their institutions a shared responsibility to manage research data and primary materials well – and to the *Griffith University Code for the Responsible Conduct of Research* (2012). As stated in this policy, researchers are required to manage their data – using methods appropriate to the discipline and to the nature of the data – to the highest standards, while the University is required to provide infrastructures, opportunities to develop professional skills, and access to advice and expertise that enable researchers to meet these standards. Finally, the *Best practice guidelines for researchers: Managing research data and primary materials* (Richardson, 2016) were published, aiming at expanding the Code for the Responsible Conduct of Research in relation to specific aspects of research data management; outlining practical steps that researchers can take; highlighting technology, advisory services and professional enabling development opportunities.

In this context, the first thing to point out is that faculty librarian roles have been modified to address data management. Indeed, while traditional librarian's duties are kept, and their core capabilities (e.g. structured thinking, knowledge of information management theory, etc.) are considered of great value for positioning them in the process of research data management (Brown et al., 2015), the librarians operating in the eResearch team are required to develop additional, more technical and discipline-oriented skills. More in detail, on the one hand librarians within the INS portfolios of Library and Learning Services and

Information Management support researchers in well-established areas such as acquisitions, collection development, copyright advice and information literacy training, and are moving into newer areas such as open access advocacy, publications repositories, research assessment exercises, and bibliometrics. On the other, Griffith's eResearch Services team operates within another portfolio in INS, building and managing technical research infrastructure for supporting researchers. Part of this infrastructure is targeted at specific research needs, while some is associated with university-wide management and discovery. Therefore, librarians in eResearch must explicitly demonstrate how their skills can be combined in productive ways with technical specialties, including software development and business analysis (Simons and Searle, 2014), as they work in close relationship with ICT experts and, for this reason, need to undergo specific and constant training to develop the specific and necessary skills to supply RDM demands.

Librarians working in this unit tend to consider themselves as "generalists", as they are required to have such a broad range of skills, knowledge and expertise that is difficult to acquire a specialization in any of these (Simons and Searle, 2014).

However, as anticipated above, a core set of skills and knowledge for librarians have been identified. As far as skills are concerned, they can be considered a sort of enhancement of the traditional librarian's competences. In fact, they deal with advanced metadata skills, high level communication skills, as in such context the role of the librarian foresees the "translation" of information between research groups. In addition, high level documentation skills are necessary for the production of documentation addressing a wide variety of audiences and purposes.

With respect to core knowledges, there must be a deep knowledge of the broader research environment where the librarian is acting, as well as of the mechanisms and processes of scholarly communication, and of the legal and regulatory framework, concerning mainly contract law and copyright issues, with a specific focus on licensing and data re-use.

Generic technical and managerial skills also play a distinctive role, as research teams usually work on goal-oriented projects, and since in eResearch project teams are comprised largely of software developers.

As far as we understand, the librarian acts as an advisor, even though technical skills make the difference in understanding how eResearch projects are run and in liaising with researchers and research managers.

5. The *data librarian* profile

In the light of the case studies, we tried to identify some basic characteristics of the *data librarian* profile.

Schmidt et al. (2016) briefly describe this upcoming figure in the research scenario as consisting of "Traditional librarian competences and skills into renewed organizational structures". Authors like Sada et al. (2013) highlight necessary technical skills, underlining the fact that many of the competencies of such professional are adopted from the ICT domain. In this perspective, the *data librarian* is seen mainly as responsible for the implementation of collaborative infrastructures for data access and reuse, fostering protocols for data interoperability and dedicating special attention to digital preservation.

Indeed, as argued by Cassella (2016) and anticipated previously, just a few steps separate the *digital curator* from the *data librarian*, the latest being a RDM specialist who constantly collaborates with other professionals.

In her presentation for the 2nd DCC/RIN Research Data Management forum, Rice (2008) defines the *data librarian*s as "people originating from the library community, trained and specializing in the curation, preservation and archiving of data". As a follow-up of the same event, the diagram reported below has been published in the DCC Data Management Forum[21]:

[21] http://data-forum.blogspot.com/2008/12/rdmf2-core-skills-diagram.html

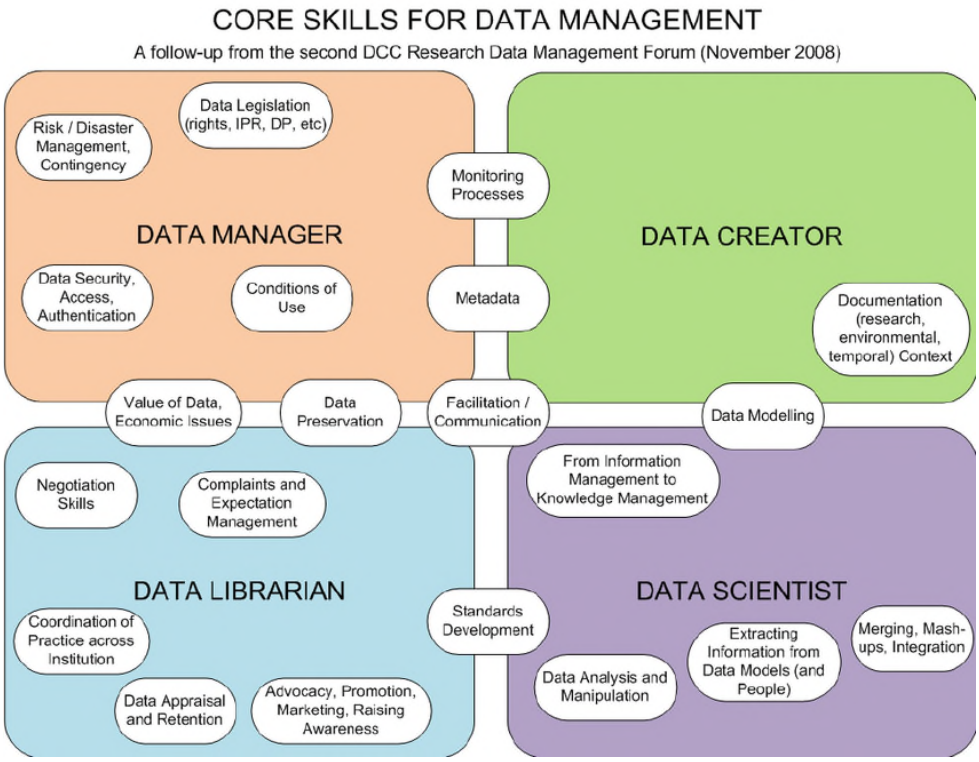


Figure 4: Core skills for data management. Chris Rusbridge and Martin Donnelly, November 2008.

The picture schematically describes the fundamental characteristics and tasks performed by the professional figures operating with data. Even though making a clear distinction between the four, a number of tasks overlap between them, as in the case of the preservation of data or the development of standards. It emerges from this model that at the time the major characteristics of the *data librarian* consisted in:

- Skills in communication and facilitation
- Standards development
- Data selection and evaluation
- Negotiation skills
- Advocacy, promotion and marketing
- Economic issues related to data value
- Preservation
- Complaints and expectations management

Despite its clarity and effectiveness, this model has undergone some critics regarding, for instance, the singular choice of placing the specialization in metadata at the boundary between the roles of data manager and data creator. Another remark concerns the absence of any reference to training for all data professionals and for *data librarian*s in particular (Cassella, 2016), which, as we reported in the previous paragraph, is now considered as crucial, thus representing a fundamental aspect of any RDS.

In this perspective, Carlson et al. (2011) identified a list of twelve core educational objectives for a data information literacy program, which might be of interest also for librarians approaching RDM:

- Databases and Data Formats
- Discovery and Acquisition of Data
- Data Management and Organization
- Data Conversion and Interoperability
- Quality Assurance
- Metadata
- Data Curation and Re-use
- Cultures of Practice
- Data Preservation
- Data Analysis
- Data Visualization
- Ethics, including citation of data

In the following years, the debate as well as the reality concerning professional roles operating with data has had various developments, as we illustrated in the case studies. Indeed, RDM policies and practices have evolved through time, allowing the growth of RDS in many academic realities.
In such scenario, Schmidt et al. (2016) identified core competencies for *data librarian*s. Besides having a basic understanding of the specific disciplinary landscape, as well as being aware of norms and standards, they would be in charge of:

1) Provide access to data
2) Advocacy and support for managing data (e.g. knowledge of DMP templates and tools, data sharing options, licenses, data citation and reference practices, etc.)
3) Manage data collections

More in details, librarians are required to have a good knowledge of existing data centers, repositories and data discovery mechanisms, funders' policies and publication requirements of journals, metadata standards and schemas, data formats, domain ontologies, discovery tools, and so on.
In addition to this, *data librarian*s would cooperate in related services such as collections' development and curation, assistance in OA and copyright policies, information literacy, digital curation and preservation, etc.

Cassella (2016) identifies five major areas of reference for the role of the *data librarian*:

1) Library science
2) Scholarly communication
3) Technology
4) Disciplinary law, copyright and licenses
5) Communication and management

Another fundamental aspect emerged from the analysis of the case studies is the importance of domain-specific competencies. Indeed, the knowledge of the research mechanisms of specific scientific areas would represent an advantage, as the *data librarian* always operates in teams, being part of a staff composed by different researchers and technical figures.
In this respect, Brown et al. (2015) recognize the increasing complexity of the roles of the librarians supporting what they have named eResearch, underlining again that it is in the network of specialists they closely work with that *data librarian*s acquire the domain specific competencies differentiating them from other library professionals.
Thus, *data librarian*s would present advanced skills in those areas where their colleagues not operating in eResearch units have general or basic knowledges.

More in details:

- Advanced understanding of discipline-based research process, outputs and scholarly communication (e.g. data types and formats)
- Advanced knowledge of ethics, intellectual property, copyright and licensing
- Advanced 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
- High level communication and documentation skills, project management and business analysis skills

In these regards, we are facing a composite reality, being the role of the *data librarian* one of the most complex to define and identify in such composite scenario, as it emerges firstly from the case studies we observed.
We may reasonably argue that at the present time there is still no overall agreement on the competencies and, most of all, on the tasks that a *data librarian* operating in RDS should perform, this being reflected also in the terminology currently in use to describe and identify such role. As a matter of fact, the *data librarian* may identify a wide range of different context-related professionals. However, most of the commentators agree on stating that *data librarian*s usually work in team with other specialists and for this reason having or acquiring some basic domain knowledge would be an advantage. Moreover, the traditional librarians' ability to count on both

existing capabilities and newly acquired skills favors their establishment as core members of a research support team.

6. Conclusions

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. For instance, many funder bodies now request Open Access to scientific publications and require the presentation of a Data Management Plan along with the project proposal.

In addition, it has been widely recognized that data sharing would bring major benefits to the scientific community, avoiding useless duplications and saving the researchers' time and resources.

In order to pursue such aim, an efficient management of research data has become essential. This is the reason why in the recent years an increasing number of institutions has been adopting specific policies dedicated to the effective management of the data produced during the research process, leading to the creation and clear definition of dedicated services, namely Research Data Services.

In this scenario, perspectives and concrete realities are quite different, as we observed in the case studies, even though some common aspects may be identified, as the importance of data access, curation, preservation, and, last but not least, the fundamental importance of training either for the professionals operating in this field, or for researchers and students producing and collecting data in their daily work.

In such context, in the recent years the figure of the *data librarian* is acquiring importance, as it may represent one of the possible evolutions of the traditional librarian in the contemporary academic world. Many definitions describing this role are available in the literature, making quite difficult outlining unique skills, knowledge, competences, and tasks.

However, it is quite clear that it is a role that would not develop based on the classic skills of the librarian only, although these represent an extreme value and a concrete base for the development of a constantly evolving career. Indeed, it must be pointed out that, due to the recent advancements in science, technology and scholarly communication, almost all professionals working in the research field had to reshape their attitude towards the research processes and scholarly communication. In such prospect, librarians occupy a privileged position either for their conventional background, or their successful adaptation to the transformations and evolutions their profession has undergone over the years.

Finally, we can argue that the traditional competences of the librarian should not be idealized, assuming that they are sufficient for becoming a *data librarian*. On the other hand, the *data librarian* should not be seen as a utopia, since numerous academic experiences show how this role is becoming a concrete reality for many professionals around the world.

As a conclusion, we may say that the *data librarian* is neither myth, nor utopia, but a composite reality.

Bibliography[1]

1. AGA – Annotated Model Grant Agreement, version 5.1, 6 December 2018. *Horizon 2020 Programme, European Commission*, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf.
2. Australian Government (2007). *Australian Code for the Responsible Conduct of Research*, http://www.nhmrc.gov.au/index.htm.
3. Australian National Data Service. *Research data policy and the Australian Code for the Responsible Conduct of Research*, https://www.ands.org.au/guides/code-awareness.
4. Ball A. (2012). *Review of Data Management Lifecycle Models* (version 1.0). REDm-MED Project Document redm1rep120110ab10. Bath, UK: University of Bath.
5. Bell G., Hey T., Szalay A. (2009). *Beyond the Data Deluge. Science,* 323 (5919) 1297-1298; http://science.sciencemag.org/content/323/5919/1297.
6. Brown R.A., Wolski M., Richardson J. (2015). *Developing new skills for research support librarians.* The Australian Library Journal, 64 (3), 224-234.
7. Carlson J., Fosmire M., Miller C., Sapp Nelson M.R. (2011). *Determining data information literacy needs: a study of students and research faculty.* Portal: Libraries and the Academy, 11 (2), 629–657.
8. Cassella M. (2016). *Dal digital curator al data librarian*. Biblioteche oggi, 34 (4) 13-21.
9. *Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information*. Official Journal of the European Union*,* https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018H0790&from=EN.
10. Crossley J. (2014). *An introduction to managing research data*, https://www.scribd.com/presentation/138079216/Managing-Research-Data.
11. Digital Curation Centre. *What is digital curation?* http://www.dcc.ac.uk/digital-curation/what-digital-curation.
12. Digital Curation Centre. *DMPonline,* https://dmponline.dcc.ac.uk.
13. European Union (2010). *Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission*, https://www.fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data.
14. Fearon D. Jr., Gunia B., Pralle B.E., Lake S., Sallans A.S. (2013). *Research Data Management Services*, SPEC Kit n. 334.
15. FORCE11, *The FAIR data principles*, https://www.force11.org/group/fairgroup/fairprinciples.
16. Genoni P. (2004). *Content in institutional repositories: a collection management issue*. Library Management, 25 (6/7), 300-306.
17. Gray J., Szalay A. (2007). *eScience – A Transformed Scientific Method*. Presentation to the Computer Science and Technology Board of the National Research Council, Mountain View, CA, 11 January 2007; http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt.
18. Griffith University (2012). *Griffith University Code for the Responsible Conduct of Research*, https://policies.griffith.edu.au/.
19. KNAW; NFU; NWO; TO2-federatie; Vereniging Hogescholen; VSNU (2018): *Nederlandse gedragscode wetenschappelijke integriteit*. DANS, https://doi.org/10.17026/dans-2cj-nvwu.
20. LIBER (2012). *Ten recommendations for libraries to get started with research data management. Final Report of the LIBER Working Group on E-Science/Research Data Management*, https://www.fosteropenscience.eu/content/ten-recommendations-libraries-get-started-research-data-management.
21. Lyon L. (2012). *The Informatics Transform: Re-Engineering Libraries for the Data Decade.* The International Journal of Digital Curation, 7 (1), 126-138.
22. McMillan D. (2014). *Data Sharing and Discovery: What Librarians Need to Know*. The Journal of Academic Librarianship, 40, 541-549.
23. OECD (2007). *Principles and Guidelines for Access to Research Data from Public Funding*, http://www.oecd.org/sti/inno/38500813.pdf.
24. Osswald A., Strathmann S. (2012). The Role of Libraries in Curation and Preservation of Research Data in Germany: Findings of a survey, https://www.ifla.org/past-wlic/2012/116-osswald-en.pdf.
25. Perrier L., Blondal E., MacDonald H. (2018). *Exploring the experiences of academic libraries with research data management: A meta-ethnographic analysis of qualitative studies*. Library and Information Science Research, 40 (3-4), 173-183.
26. Queensland University of Technology (2015). *Manual of Policies and Procedures. D/2.8 Management of research data*, http://www.mopp.qut.edu.au/D/D_02_08.jsp.
27. 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.
28. Research Data Netherlands. *Essentials 4 Data Support*, https://datasupport.researchdata.nl/en/start-the-course/i-definitions/research-data/.
29. Research Data Netherlands (2018). *A federated data infrastructure for the Netherlands: the front-office – back-office model. UK web version,* https://researchdata.nl/fileadmin/content/RDNL_algemeen/Documenten/RDNL_FOBOmodel-UK-web.pdf.
30. Rice R. (2008). *Roles and Responsibilities for Data Curation: the Data librarian.* RDMF2: Roles and Responsibilities for Effective Data Management, Manchester, Chancellors Hotel and Conference Centre, 26-27 November 2008. Slides, http://www.dcc.ac.uk/events/research-data-management-forum/roles-and-responsibilities.
31. Richardson J. (2016). *Best practice guidelines for researchers: Managing research data and primary materials.* Griffith University, https://www.griffith.edu.au/library/research-publishing/best-practice-guidelines-for-researchers.
32. Sada E., Gregori L., Sirito P. (2013). *Un bersaglio mobile: l'evoluzione dei profili degli "information professionals" alla luce dei nuovi scenari accademici.* AIB studi, 53 (1), 92-99.

33. Schmidt B., Calarco P., Kuchma I., Shearer K. (2016). *Time to Adopt: Librarians' New Skills and Competency Profiles. Positioning and Power in Academic Publishing: Players, Agents and Agendas.* Proceedings of the 20th International Conference on Electronic Publishing, Loizides F., Schmidt B. (eds.), IOS Press, 1-8.
34. Schmidt B., Shearer K. (2016). *Librarians' Competencies Profile for Research Data Management.* 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/.
35. Scott M., Cox S. (eds) (2016). *Introducing Research Data.* University of Southampton, fourth edition, https://eprints.soton.ac.uk/403440/1/introducing_research_data.pdf.
36. Simons N., Searle S. (2014). *Redefining 'The Librarian' in the Context of Emerging eResearch Services.* Paper presented at VALA 2014, http://www.vala.org.au/vala2014-proceedings/vala2014-session-15-simons.
37. Swan A., Brown S. (2008). *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs*. Truro: Key Perspectives. http://www.jisc.ac.uk/publications/documents/dataskillscareersfinalreport.aspx.
38. Tammaro A.M. (2016). *Libraries in Digital Age (LIDA): la trasformazione delle biblioteche in era digitale*, https://annamariatammaro.com/2016/06/21/libraries-in-digital-age-lida-la-trasformazione-delle-biblioteche-in-era-digitale/.
39. Tammaro A.M. (2018). *IFLA Global Vision Project: Barcelona Kick-Off Workshop*, https://annamariatammaro.com/category/ifla-global-vision/.
40. Tammaro A.M. (2018). *L'evoluzione della biblioteca digitale: dall'accesso alle risorse elettroniche alla creazione e alla cura dei dati*, https://annamariatammaro.com/2018/05/09/levoluzione-della-biblioteca-digitale-dallaccesso-alle-risorse-elettroniche-alla-creazione-e-alla-cura-dei-dati/.
41. Tenopir C., Sandusky R.J., Allard S., Birch B. (2014). *Research data management services in academic research libraries and perceptions of librarians.* Library & Information Science Research, 36 (2), 84-90.
42. UK Engineering and Physical Sciences Research Council. *EPSRC policy framework on research data – Scope and benefits*, https://epsrc.ukri.org/about/standards/researchdata/scope/.
43. University of Edinburgh *Research Data Service*: https://www.ed.ac.uk/information-services/research-support/research-data-service/about-the-research-data-service.
44. University of Edinburgh. *Research Data Management Policy,* https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy.
45. University of Edinburgh – EDINA. *MANTRA. Research Data Management Training*, https://mantra.edina.ac.uk.
46. University of Edinburgh – *EDINA*. *Do-It-Yourself Research Data Management Training Kit for Librarians,* https://mantra.edina.ac.uk/libtraining.html.
47. Whyte, A., Tedds, J. (2011). *Making the Case for Research Data Management. DCC Briefing Papers*, Edinburgh: Digital Curation Centre, http://www.dcc.ac.uk/sites/default/files/documents/publications/Making%20the%20case.pdf.
48. Witt M. (2008). *Institutional repositories and research data curation in a distributed environment*. Library Trends, 57 (2), 191-201.
49. Wright M., Sumner T., Moore R., Koch T. (2007). *Connecting digital libraries to eScience: the future of scientific scholarship.* International Journal on Digital Libraries, 7 (1-2), 1-4.
50. https://dictionary.cambridge.org/dictionary/english/data.
51. https://en.oxforddictionaries.com/definition/data.
52. https://www.merriam-webster.com/dictionary/data.

[1] URL last access: December 2018.

Research Data Management: What can librarians really help?*

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

Abstract
As national government agencies continue to mandate specific data management requirements and the need for research data management (RDM) grows, many libraries are developing RDM services to help with the research mission of their institution. Research libraries' mission and expertise have always included a variety of research services. What can the library's role be in RDM services? This paper describes the possible roles that libraries and librarians can play throughout the data lifecycle. The Princeton University Library is presented as a case study to demonstrate these roles and the development of RDM services including advocacy, awareness, education, advisory services, data management plan development, and data repository development and promotion. In addition, this paper also discusses the challenges and opportunities associated with RDM services development in libraries and the future plans at Princeton, including the development of a RDM mini course for graduate students and a robust RDM program.

Introduction

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 (Holdren). At the same time 74% of libraries surveyed by the Association of Research Libraries indicated they were already engaged in research data management services (Fearon et al.). 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.

Literature Review

There are two notable large scale surveys of research data management services in libraries. In 2012 Carol Tenopir et al. wrote a white paper for ACRL titled "Academic Libraries and Research Data Services". The following year David Fearon et al published an Association of Research Libraries (ARL) Spec Kit titled "Research Data Management Services". Tenopir's survey included 221 academic libraries from the United States and Canada. Key findings of the survey showed that at the time less than a third of academic libraries were engaged in research data services, but many more planned to develop services in the future. The most common service presently offered or planned was creation of web guides to help researchers find data sets. Staffing of research data services was most commonly approach by re-assigning existing staff (Tenopir et al.). The ARL Spec Kit surveyed 73 of 125 ARL libraries. Emerging trends identified in this study were research data management support

* First published in the GL20 Conference Proceedings, February 2019.

at the grant stage for data management plan requirements and data archiving support. One area the authors regretted as a limitation of their study was the lack of "new, comprehensive case studies" (Fearon et al. 12).

Several case studies from the library perspective were included in *Research Data Management: Practical Strategies for Information Professionals* edited by Joyce M. Ray in 2014. These case studies feature overviews of the development of infrastructure and services at four institutions with fairly well established programs: Cornell University, Purdue University, Rice University, and the University of Oregon. The Cornell University Library (CUL) case study examines their long history of supporting research data needs at Cornell starting in 1982 by collaborating with the Cornell Institute for Social and Economic Research data archive, then providing infrastructure for the other disciplinary repositories in the 1990s, managing the "eCommons" institutional repository launched in 2002, and finally developing the Datastar staging repository for research data. In addition to infrastructure, CUL has also made efforts to develop other research data management services for faculty starting with a Data Working Group from 2006-2008, then participating in the ARL E-science Institute in 2010-2011, and subsequently launching the cross-departmental Research Data Management Service Group. Looking to the future Cornell hopes to improve existing infrastructure and provide more support while refining the scope for data services (Steinhart). Purdue University Library began researching research data management needs of scientists at Purdue in 2004. Their research led them to create the Data Curation Profiles Toolkit to aid in understanding a dataset with the aim of curation, then eventually the Purdue University Research Repository (Brandt). At Rice University, the Library's Digital Scholarship Services team was seeing a lot of use from the emerging field of digital humanities, so they developed the Rice Digital Scholarship Archive and work closely with researchers to curate locally created digital humanities collections (Henry). To develop its research data management program, the University of Oregon Library performed a needs assessment study of researchers. The needs assessment identified several areas for development including helping researchers assess data management tools such as electronic notebooks, data management plan assistance, institutional repository, and data management training (Westra).

In 2017 the library cooperative OCLC released a four part report series titled *Realities of Research Data Management* aimed at exploring the "context, influences, and choices research universities face in building or acquiring RDM capacity" (Bryant et al. 5). The series starts with a small survey of RDM services before moving on to three separate reports examining scoping, incentives for, and sourcing/scaling of RDM services. The survey identified three categories of service that were practiced at some level in most institutions: education, expertise, and curation (6). To delve into the different phases of RDM service development, the series uses case studies from four different institutions: University of Edinburgh (UK), the University of Illinois at Urbana-Champaign (US), Monash University (Australia), and Wageningen University & Research (the Netherlands).

The case studies showed a distinct variety in approaches to RDM capacity building from these institutions. Despite these differences they identified several key findings. Among other things, the Scoping report exemplified the finding that RDM services at different institutions are shaped by a range of internal and external factors and will thus emphasize different service categories (28-29). A key takeaway from Part Three on Incentives report is that RDM service development is driven by local incentives that usually fall into one of four categories, compliance, evolving scholarly norms, institutional strategies, or researcher demand (18). The variety of internal and external factors as well as local incentives at each institution become clear in Part Four on Sourcing and Scaling as each institution employs a different strategy for developing services in house or utilizing outside resources.

In introducing the case studies produced for the OCLC reports, the authors note that there is no template for research data management service development and adoption will vary across institutions. However, they believe "...the *aggregated* experiences of research universities that have deployed various forms of RDM capacity provide useful markers, lessons and decisions points for other institutional contexts" (10). They point to several notable research data management case studies and large scale surveys (including Tenopir

et al. and Fearon et al. mentioned above) produced so far but indicate that they believe more case studies are needed, "But the value of these case studies increases with the number available: in this way, the details of the individual case studies are complemented by a diversity of circumstances and contexts from which to draw them" (10). It is the authors' hope that this case study of RDM service development at an institution which began by providing support on a case by case basis with plans to grow into a formalized program will add to this diversity.

Librarians' role in Research Data Management

Librarians have a history of collecting, organizing, describing, providing access to and preserving information. Since data is a type of information, it makes sense for librarians to be involved in similar work relevant to the data management lifecycle. While there are a number of models that define the stages of the research data management lifecycle, the UK Data Archive's Research Data Management Lifecycle model is a good place to start.

Pictured above, this data lifecycle includes the stages of planning, collecting, processing and analyzing, publishing and sharing, preserving, and re-using research data (UK Data Archive). Librarians can play a role in each stage of the lifecycle; regardless of the discipline they help support. While the types of data created and curated may vary based on fields of study (i.e. humanities, social sciences, sciences, biosciences), the stages remain constant.

Planning

Currently, librarians that support data management are most active in the planning and archiving stages. Data management planning is often the direct result of funder mandates and data management plans now often accompany grant funded research. Librarians can raise awareness of the benefits of data management planning and can also help guide researchers through the specific requirements of different funder mandates (Bryant et al. 1). Very often, they work to support the creation of formal data management plans (DMP). They do this through educational initiatives such as online tutorials, webinars, in person workshops and personalized consultations. The DMP Tool, created and curated by the University of California Curation Center of the California Digital Library and others (https://dmptool.org), has been a popular template amongst librarians who use it to support researchers developing a data management plan of their own. During this process, librarians have the opportunity to educate researchers on good data management practices such as guidance on naming protocols, preferred file formats and storage options.

Collecting

Collecting data is the stage in the lifecycle when researchers gather their data, including performing experiments or locating existing data. Librarians often provide data reference at this point of the process, helping researchers collect, purchase and/or license third party

data sets that may be necessary to their research. If researchers are generating their own data, librarians can provide advice on documenting the collection process as well as file organization.

Processing and Analyzing

Processing and analysis of data may seem to be a stage during which librarians would be of little use. However, it is at this phase that librarians can identify and license or purchase data analysis tools for both quantitative (examples: Qlucore, MetaCore, etc.) and qualitative analysis. (examples: REDCap, NVivo) In addition, libraries can offer training on how to most efficiently use vendor provided analyses resources, or provide training in statistical analyses and coding methods that assist in data analysis such as SAS, Python, and R. It is at this point that librarians can provide guidance on data documentation and appropriate metadata to help ensure that the data is discoverable by other researchers (AAU-APLU).

Publishing and Sharing

In addition to the research outputs that will end up in a published paper, researchers may want to share or publish the data underlying the research results. Heather Piwowar showed that open data led to higher citation rates for the published article (Piwowar et al.), and many funders are requiring that data be made publicly available, assuming it does not contain sensitive information. Librarians can recommend options for publishing and sharing data including relevant repositories. Many institutions now offer an open access repository, often created and maintained by their library. Some repositories can be used to host data as well as other file formats. In addition to providing storage for the data, librarians can also provide tools for synchronizing/sharing data with collaborators. Open access repositories also promote the dissemination of research by making materials more easily discoverable and accessible to the public and other researchers. It is also useful at this stage for librarians to instruct researchers on data rights management.

Preserving

Preservation strategies include maintaining a long-term archival space and migration of data when needed. This is not always the same as publishing data, especially if the data is not able to be made publicly available in an open repository. Librarians who support data curation can ensure that file formats are usable or migrate data to new file formats as technologies evolve. They can also help generate permanent identifiers for data so it remains continually available.

Re-Using

The phase of re-using research data brings the research data life cycle around full circle. All steps prior to this make re-use possible. Librarians can instruct on the need for proper data citation when researchers' use others' data, particularly since it will affect citation counts and impact calculations which are meaningful metrics to many researchers.

Case study at Princeton University

Similar to many other research libraries, Princeton University Library (PUL) started Research Data Management (RDM) services as a response to funders' data management plan requirements. The library had already been providing a few long-standing services in economics and the social sciences that support data collection, analysis, sharing, and reuse. However, when the NSF announced its plans to begin requiring data management plans with grant applications a need for RDM services in the sciences was recognized. Initially science librarians worked together to create a template for NSF data management plans. Then as more and more funders started requiring data management plans a science librarian was tasked with creating research data management services for the sciences as a half-time eScience librarian. These services were based on services being developed by peers and included data management plan assistance, data management education, and assistance depositing in repositories. When the eScience librarian was promoted to lead the Engineering Library in 2015, library administration did not replace her but worked to develop a more robust response to campus needs. In absence of a dedicated librarian for the role, the Scholarly Communications Librarian and Engineering Librarian have continued the services initially developed on an ad-hoc and responsive basis. This has enabled the library to continue to engage in the campus response to research data management needs. It is our

goal to develop our RDM services to a more robust program along the data lifecycle. The following are services that we provide at each stage of the research data lifecycle.

Planning

Raising awareness. Since the White House directive on public access to federally funded research in 2013 (Holdren), many funders have developed distinct and varying public access policies. In 2014, a few projects were developed to raise awareness on campus about the new policy and the importance of RDM. A website and a Libguide were created to focus on general information and resources on RDM. The RDM website and Libguide[1] serve as very useful resource on campus. A handout with information about the public access policies on publications and data from the top 5 funders[2] at Princeton University was also developed with a focus on informing faculty about the new requirement from funders. The library worked with the Office of the Dean for Research and distributed the handout to faculty. Our Population Research Librarian also created a guide specifically on NIH public access policy[3] and detailed in steps how to comply with the mandate.

Educating the campus on research data management. In addition, as an important part of our RDM services, we educate the campus about the importance of data management and data management best practices by providing individual consultations and group workshops. Whenever we receive a question or request regarding data management, the Research Data Management (RDM) team, which consists of two librarians and one OIT (Office of Information Technology) staff, will schedule a meeting to talk about the details of the data management needs and offer customize solutions. For group workshops, we offer both tailored workshops for individual departments, such as a workshop specifically on best practices in data management; and a general session for the campus. A recent example is the Data Management Plan workshop offered for in March 2018. The workshop was a collaborative effort with colleagues across campus, including Library, OIT, the Office of Dean for Research, Research Computing, and Princeton Institute for Computational Science and Engineering (PICSciE). It had a great turnout with over ninety attendees, including faculty, graduate students, and grant managers.

Assisting in creating data management plan. Many funders with a public access policy also require a data management plan as part of the grant application. This requirement presents challenges to some faculty who have never written a data management plan before despite guidelines provided by funders. Therefore, there are rising needs for assistance in creating such a plan on campus. To address this emerging need, the Library formed a RDM Team (consisting of two librarians and one OIT staff), to develop services that could assist campus in creating data management plans. The first project that the team engaged was with OIT to adopt the DMPTool at Princeton. Following the DMPTool adoption, the team created promotional materials, such as postcard and a services brochure, to promote the DMPTool and RDM services. It didn't take long for the team to get requests to review other data management plans for grant applications.

In 2014, the Department of Energy released their public access plan in response to the 2013 White House directive. As a Department of Energy (DOE) national laboratory, the Princeton Plasma Physics Laboratory (PPPL) was tasked with writing a data management plan for data produced at PPPL under DOE funding as well as making the data underlying publications publicly available. The RDM Team consulted on the data management plan PPPL wrote in response to DOE's public access requirements.[4]

Customized support in data management plan creation. It's not often but we have gotten requests for guidance developing data management plans for a specific purpose other than for a grant requirement. In 2017, The Molecular Biology department requested help

[1] Research Data Management at Princeton https://libguides.princeton.edu/rdm

[2] Research Funder Open Access Mandates http://library.princeton.edu/sites/default/files/FunderRequirements-April14-2015.pdf

[3] NIH Public Access Policy Guide https://libguides.princeton.edu/c.php?g=84087

[4] For further details see Dressel, Willow. *Helping a Department of Energy Laboratory Respond to Public Access Requirements*. https://escholarship.umassmed.edu/escience_symposium/2016/posters/1. University of Massachusetts and New England Area Librarian e-Science Symposium.

developing data management guidelines for use of shared equipment such as new microscopes that generate large amounts of data. The RDM Team and the liaison librarians worked together to develop recommendations for the department.

Collecting

Data Reference. Social Science Librarians at Princeton have been assisting researchers in determining their data needs and locating data sources such as using ICPSR and financial data sources like Bloomberg over decades. There has been an increase in research and teaching involving text mining across the social sciences, humanities and sciences at Princeton and liaison librarians of all disciplines are providing assistance finding appropriate corpora ranging from using the Scopus API or understanding licensing limitations of other databases to free corpora such as Project Gutenberg.

Purchasing and licensing datasets and data analysis tools. PUL has a long history of assisting researchers in licensing existing 3rd party data, such as statistical dataset- Data-Planet, ICPSR datasets, iPOLL databank, social science electronic data library-Sociometrics, integrated public use microdata series-IPUMS, and Wharton research data services-WRDS. Other databases recently acquired or licensed outside of social science include Biocyc a collection of genome database and LexisNexis Web Services Kit for text mining. As to the data analysis tools, the statistical packages are R/R Studio, Stata, and SPSS.

Processing and Analyzing

Offering Training on data analysis and visualization. The library's Data and Statistical Services[5] (DSS) has a long history providing data and statistical consulting. Experts are available to advise Princeton University students, faculty, and staff on choosing appropriate data, application of quantitative research methods, and the interpretation of statistical analyses, data conversion, and data visualization. DSS provides statistical and software assistance in quantitative analysis of electronic data as part of independent research projects, such as junior papers, senior theses, term papers, dissertations, and scholarly articles. Students can schedule individual consultation through their website or just walk in the DSS lab to get help during the walk-in hours. DSS also offers online tutorials and group workshop on a variety of topics related to data analysis.

The Center for Digital Humanities[6] (CDH) is an interdisciplinary research center and academic unit within the Library. CDH offers many services regarding data, including training on data visualization, grant on data curation (for humanities data sets, and consultation on issues related to humanities data curation), digital humanities software tools, and events (Year of Data 2018-2019).

Publishing and Sharing

Data sharing, publishing, and preservation. Princeton's Office of Information Technology (OIT) provides a data repository called DataSpace. When the service was launched, OIT consulted with cataloging librarians at Princeton on metadata for discovery. Librarians also work with OIT to promote DataSpace as a local solution for data sharing, publishing, and preservation. When Princeton Librarians consulted on the Princeton Plasma Physics Library's Data Management Plan, they also assisted OIT with setting up the DataSpace collection for public access compliance, including advising on providing study level documentation as well as linking data to the published article. Librarians also direct researchers to appropriate and available repositories for their disciplines to maintain grant compliance (i.e. PubMed Central for NIH Open Access Policy).

GIS data and services. The Map and Geospatial Information Center is located in the Lewis Science Library. The Center provides access to paper maps, geospatial data, digital maps and geographic information systems (GIS) services. In addition, the GIS center provides ongoing reference, research consultation and instruction to users with all levels of experience. The Center has eight workstations with 30-inch monitors loaded with GIS and satellite image processing software package, and commonly used geographic data. Faculty, students, staff and others in the Princeton University community are welcome to contact or visit the Center

[5] Princeton Data and Statistical Services https://dss-princeton-edu.ezproxy.princeton.edu/

[6] Princeton Center for Digital Humanities https://cdh.princeton.edu/

for additional information. At the center, users can access over 300,000 paper maps, charts, aerial and satellite photographs; create, modify, and print custom maps by using GIS software packages and third party data; analyze data graphically by conducting "what-if" scenarios[7].

Re-using

The library is in full support of data reuse through licensing/purchasing data, providing workshops on legal and ethical considerations when reusing the data, providing guidance on citing data, and supporting campus workshops on generating reproducible data.

Conclusion and Next Steps

Challenges and Opportunities

Developing a comprehensive RDM program is not an easy job for a research library, even a large academic one. It is not only costly in terms of staffing and technical infrastructure (e.g. a data repository), but it also takes time to build relationships with faculty, researchers, students, service and resources providers (e.g. grant administrators), and to educate and engage the campus through the data management lifecycle. However the opportunities that RDM services bring to the library are too compelling to ignore. RDM services increase the library's relevance to research, strengthen the relationship with campus primary stakeholders, confirm the library's expertise in information management, demonstrate the library's central role in supporting the institution's mission, proves library's ability in innovation and staying abreast of change, and affirms the leading role in open access.

Future plans

PUL was able to start small (for a particular purpose) and build up from there. From one science librarian, working half time, developing a service to help faculty comply with the funder's DMP requirement, to an RDM Team providing multiple services, including DMPTool, DMP assistance, education and training, data storage, data publishing and sharing, and data preservation.

The development of a week long course for graduate students. In summer 2018, the library began collaborating with Research Computing, OIT, and PICSciE to develop a week long mini course on RDM for graduate students, which is to be offered in January 2019. The associate CIO of OIT initiated and led this effort based on his own experience as a doctoral student at Princeton. The mini course will be the first program to comprehensively teach skills and knowledge on RDM through the data lifecycle at Princeton. The mini course involves the presenters from several partners on campus, including Research Integrity and Assurance, Office of Research and Project Administration, OIT, and Research Computing. The long term goal is to develop it as a transcripted course at the Graduate School, which will be required for every graduate student.

Building towards a more robust RDM program. Over the last few years, Princeton's University Librarian has been collaborating with the Dean for Research and the CIO of OIT to propose to the university administration an RDM program to meet the increasing needs on campus. The proposed program includes multiple dedicated staff with specific expertise in data management/curation and an infrastructure for data storage, discovery, publishing, and preservation. The proposal was approved last fall. The work on the staff recruitment and the development of an infrastructure for data storage and curation are currently underway. With all the good faith and work, we look forward to a more robust RDM program offered on campus in the near future.

[7] GIS Center http://library.princeton.edu/collections/pumagic

References

AAU-APLU. *AAU-APLU Public Access Working Group Report and Recommendations*. Association of American Universities and Association of Public and Land-grant Universities, 2017, https://www.aau.edu/key-issues/aau-aplu-public-access-working-group-report-and-recommendations.

Brandt, D. Scott. "Purdue University Research Repository: Collaborations in Data Management." *Research Data Management : Practical Strategies for Information Professionals*, edited by Joyce M. Ray, West Lafayette, Indiana: Purdue University Press, 2014. *catalog.princeton.edu*, http://www.jstor.org/stable/10.2307/j.ctt6wq34t .

Bryant, Rebecca, et al. "The Realities of Research Data Management." *OCLC Research*, https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html. Accessed 28 Dec. 2018.

Fearon, David Jr, et al. *Research Data Management Services, SPEC Kit 334 (July 2013)*. Association of Research Libraries, 2013. *publications.arl.org*, http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/.

Henry, Geneva. "Data Curation for the Humanities: Perspectives from Rice University." *Research Data Management : Practical Strategies for Information Professionals*, edited by Joyce M. Ray, West Lafayette, Indiana: Purdue University Press, 2014. *catalog.princeton.edu*, http://www.jstor.org/stable/10.2307/j.ctt6wq34t.

Holdren, John. *Increasing Access to the Results of Federally Funded Scientific Research*. Memo, Executive Office of the President Office of Science and Technology Policy, 22 Feb. 2013, https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Steinhart, Gail. "An Institutional Perspective on Data Curation Services: A View from Cornell University." *Research Data Management: Practical Strategies for Information Professionals*, edited by Joyce M. Ray, Purdue University Press, 2014.

Tenopir, Carol, et al. *Academic Libraries and Research Data Services Current Practices and Plans for the Future*.

UK Data Archive. "Research Data Lifecycle." *UK Data Service* , 2019 2012, https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx .

Westra, Brian. "Developing Data Management Services for Researchers at the University of Oregon." *Research Data Management : Practical Strategies for Information Professionals*, edited by Joyce M. Ray, West Lafayette, Indiana: Purdue University Press, 2014. *catalog.princeton.edu*, http://www.jstor.org/stable/10.2307/j.ctt6wq34t .

Data Management and the Role of Librarians*

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

Abstract
'Research Data Science' is defined by Committee on Data of the International Council for Science Research Data Alliance (CODATA-RDA) as an ensemble of (a) Open Science principles and practices (FAIR) and research data management and curation skills, (b) the use of a range of data platforms and infrastructures, (c) large scale analysis, (d) statistics, (e) visualization and modeling techniques, (f) software development and annotation, and (g) more. Data management and the role of librarians must now include developing expertise and training with 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, campus-wide focused UF Data Management and Curation Working (DMCWG) and ARCS work in collaborative partnerships with the campus units such as 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. This paper will discuss the developing outreach activities, inter-departmental collaborations, some initial outcomes, and future goals of leveraging capacity, infrastructure, and resources to develop data management efforts across communities of practice within an institution's current organizational culture. One aim of this paper is to highlight the importance and significance of developing good library and faculty partnerships built on character, integrity, and humility as the cornerstones for the roles of librarians as collaborators in promoting socio-technical data management programs.

Introduction

All data is not Findable, Accessible, Interoperable, and Reusable (FAIR). Researchers face many data management challenges resulting from the need to comply with funding agencies' data management and sharing requirements. While major funding agencies such as the American Heart Association (AHA), National Institutes of Health (NIH), NSF, and United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) have data management and sharing requirements guidelines and/or statements, researchers seeking funding struggle with complying, fulfilling, and negotiating data lifecycle management due to existing capacity, infrastructure, resources, and support. Data lifecycle requirements include:

a) Key components of a data management plan [6] and data lifecycle processes [7]
b) Data management life cycle [1]
c) Research data management services for scalability, support, & sustainability [8].

The coordination, collaboration, and connection of diverse stakeholders (See Fig 3) across multiple departments, organizations, and units in support of good data management requires consistent organization, technology, and resources to maintain stability [2] over time.

In January 2011, the National Science Foundation began requiring researchers to include a research data management (RDM) plan in all of their funding proposals prompting academic libraries to develop data management tools and roles [9]. The mandate was further expanded through executive orders and legislation in 2013, when many grant applications were required to include a RDM plan for the purpose of making the research data more openly available [10]. Up until this mandate, most academic libraries were offering only limited RDM services. This mandate provided the catalyst for new services and challenges for

* First published in the GL20 Conference Proceedings, February 2019.

academic libraries [11, 12]. Universities and academic libraries responded with developing their own institutional framework to comply with the mandates [14]. Even as funding agencies and policy makers supported making research data available to others, data sharing is still often met with resistance and the data sharing process is complex [14].

What are some of the barriers to data sharing? Tenopir *et al* examined differences in data sharing practices across researchers' age groups, geographic regions, and subject disciplines [15]. Variations in data management practices are often seen across academic disciplines [16]. Deposition and sharing practices vary across journal publishers, repositories, and universities [17]. The data itself seems to be a barrier to data sharing. Data tends to be stored in local domains and have minimal structures and documentation making shared data difficult to fully be utilize by other researchers [18]. Data newly generated as part of a grant is often easier to organize and store than legacy datasets which offer additional challenges [19]. A 2016 article in PLoS One by VanTuyl and Whitmire evaluated how effective data sharing has been and the usability of the data that was shared [20].

One way to increase data sharing is to reward those who share data and those who utilize shared data. In peer reviewed literature, rewarding researchers who create well-documented data sets is one method of encouraging good data curation practices. The scholarly communication community is encouraging data publication [21]. Studies have shown that scholarly publications which used a publicly available data set had a higher number of citations than similar studies without publicly available data [22].

While there are many tools and recommendations for creating a good RDM plan [23], there are arguments for and against research libraries taking on the role as research data managers as there are many competing agencies in the data curation field [24]. Pinfield *et al* examined the roles and relationships involved in RDM and the major drivers and components of developing an RDM plan [25]. As libraries take on a more active role in RDM, a shift in the libraries' role from support to partnership with researchers is emerging [26]. Librarians who are becoming more involved with research data are being asked to assist with preparing RDM plans and insertion of data into data repositories [27]. Since data management is a new service many academic libraries are asked to provide, there are challenges both intellectual and economic related to offering these new services [28]. Although more researchers are aware of the need for management of data sets, RDM is still primarily conducted at the institutional level using a locally developed framework [29] in need of a socio-technical systems thinking [3] approach.

Socio-technical data management collaborations

This paper introduces the concept of socio-technical data management collaborations as one approach in developing the data management and curation skills required for librarians to understand and contribute to research data science as collaborators and consultants. The socio-technical systems thinking approach framework (See Fig. 1) was selected for this paper as the preferred model to begin investigating the interrelated changing dynamics of a socio-technical system in the context of data management collaborations. "Social-technical systems theory has around 60 years of development and application internationally by both researchers and practitioner [3]". Understanding and analyzing data management and the required collaborations from a socio-technical systems thinking perspective at an academic research institution is beyond the scope of this paper. However, the application and use of *steps in analyzing and understanding an existing socio-technical system* (See Table 1) of data management can aid in the development of socio-technical data management collaborations and a Business Model Canvas (See Fig 2) on which to build benchmarks, metrics, and success.

Table 1.
The steps involved in analyzing and understanding an existing socio-technical system [3].

Step	Task description
1	Gather relevant data from appropriate sources, including key actors, stakeholders, subject matter experts, and internal and external documents.
2	Analyze and classify data, using techniques such as template analysis (King, 2004). Initial template consists of the socio-technical framework.
3	Identify and group key system factors. Visually represent the groups of factors on each node of the framework.
4	Consider the implication of the external environment in which the system is embedded within the node which it relates.
5	Systematically consider relationships between each set of factors, and identify contingencies and direction of relationships.
6	Visually inspect the hexagon framework and assess underexplored or related areas, and reappraise evidence or seek input from colleagues and subject matter experts (e.g., with expertise in buildings and infrastructure).
7	Add any additional relevant factors that emerge from the data during analysis or following previous step.
8	If appropriate: Generate a timeline of key factors leading up to the event or scenario, grouped by six factors. Classify as: (long-standing issues (3+ months); issues immediately preceding the event (0-3 months); and factors involved on the day.
9	Test analysis on key stakeholders for accuracy, omissions and interpretations, and modify as necessary after discussion.
10	Generate key inferences regarding the system and how it works.

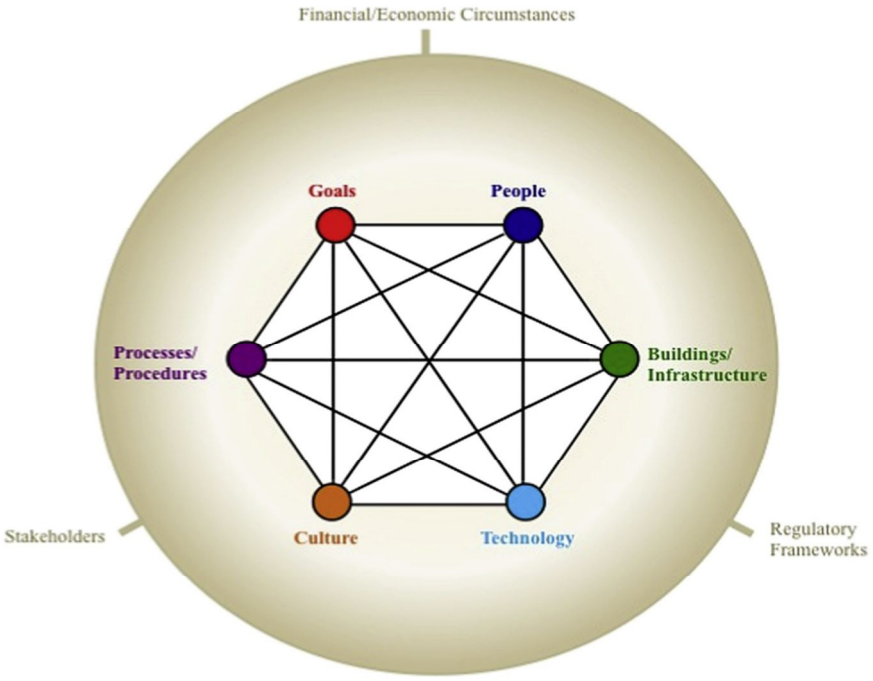


Figure 1. Socio-technical system, illustrating the interrelated nature of an organizational system, embedded within an external environment [3].

Figure 1 illustrates the conceptual framework of key interrelated nodes that figure prominently within an existing organizational system. Analyzing and understanding each of the nodes and their impact within an organization system using data management and cyberinfrastructure as variable for analysis can provide useful data in support of capacity, infrastructure, and resources necessary for socio-technical data management collaborations within and across departments, funded projects, institutions, labs, organizations, and units. The socio-technical framework was used a frame of reference for this paper. Below are a few

developing socio-technical data management collaborations via the DMCWG at the University of Florida.

- **UF Marston Science Library (MSL)** – provides DMCWG monthly meetings venue, STEM research, resources, support, training and partnership in 1st Data Symposium event
- **UF Office of Research** – provides access to senior stakeholders' guidance and support
- **UF Informatics Institute (UFII)** – provides venue, promotion, and lunch for the DMCWG training workshops and membership to the UFII Faculty Affiliates
- **UF Research Computing** – provides university-wide capacity, infrastructure, resources, support, and membership to the Research Computing Advisory Committee (RCAC)
- **UF College of Engineering** – provides data management plan referrals and partnership
- **UF Clinical and Translational Institute (UF/CTSI)** – provides data management and analysis and informatics consultations, expertise, and support for funded projects
- **UF Institute of Food and Agricultural Sciences (UF/IFAS)** – provides DMP projects
- **UF Biomedical Informatics (UF BMI)** – provided invitation to 5 yrs. strategic plan retreat

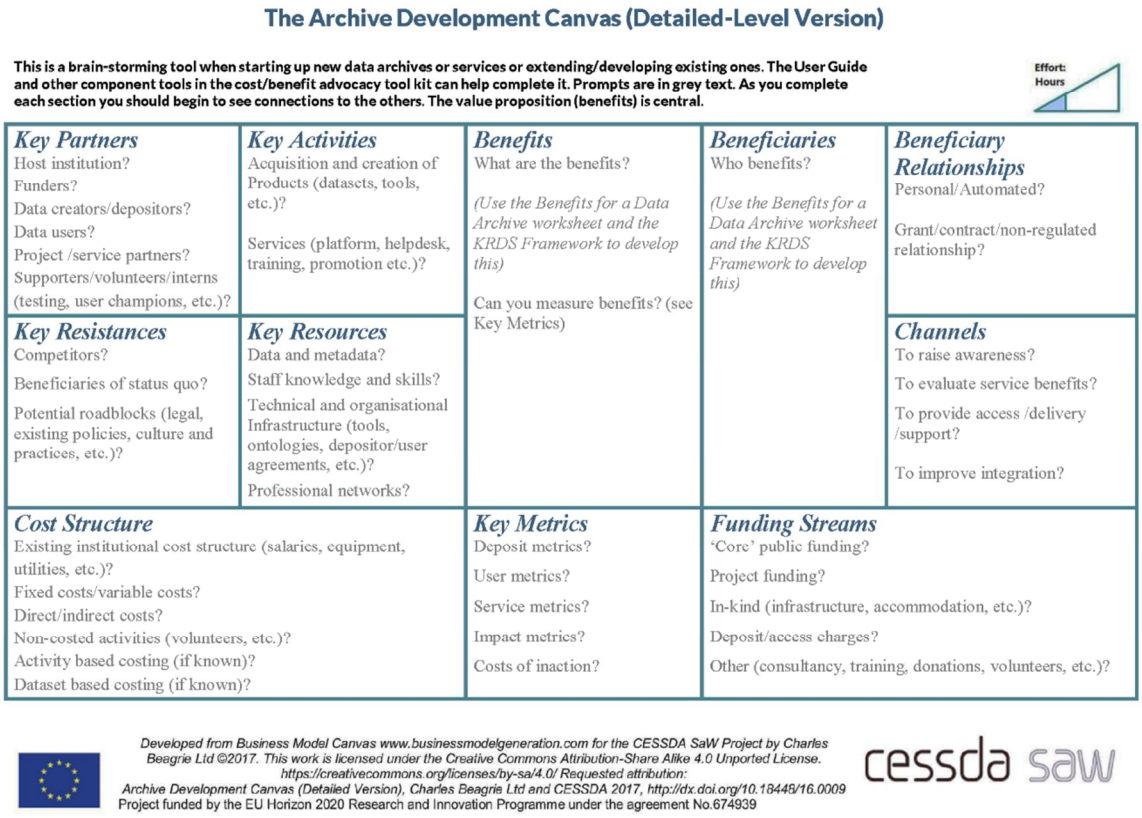


Figure 2: The Archive Development Canvas (Detailed-Level Version)[1]

Figure 2 is an illustration of the Consortium of European Social Science Data Archives (cessda) Strengthening and widening (saw) the European infrastructure for social science data archives brain-storming tool when starting up a new data archives or services or extending/developing existing ones as part of a cost/benefit advocacy tool kit. Fig 2 is introduced as a complimentary tool for use in conjunction with Fig 1 to explore ways of developing sustainability for socio-technical data management collaborations. Oftentimes researchers are required to provide data management plans without the funding resources to sustain data lifecycle management. Developing a data management plan that includes the economic and financial considerations for long-term preservation require investments from multiple stakeholders (See Fig 3). Researchers must now consult, collaborate, and work with campus groups addressing data management in order to effectively and successfully comply with funding agencies' mandates.

[1] Charles Beagrie Ltd. (2017). CESSDA SaW Archive Development Canvas (Detailed Version). http://tinyurl.com/ybpy94hu.

Data Management and Curation Working Group (DMCWG)

The Data Management and Curation Task Force (DMCTF) (2012 – 2016) was formed to develop a culture of data management at UF. The task forces developed data management initiatives, projects, presentations, and the Data Management Librarian faculty position. The Data Management Librarian hired in 2016 developed a DMCWG Charge[2], Year End Reports[3], and organized for the 1st Annual Data Symposium. The DMCWG has 53 subscribers, comprised of chairs, directors, faculty, IT, and staff, and meets every 4th Monday in Marston Science Library.

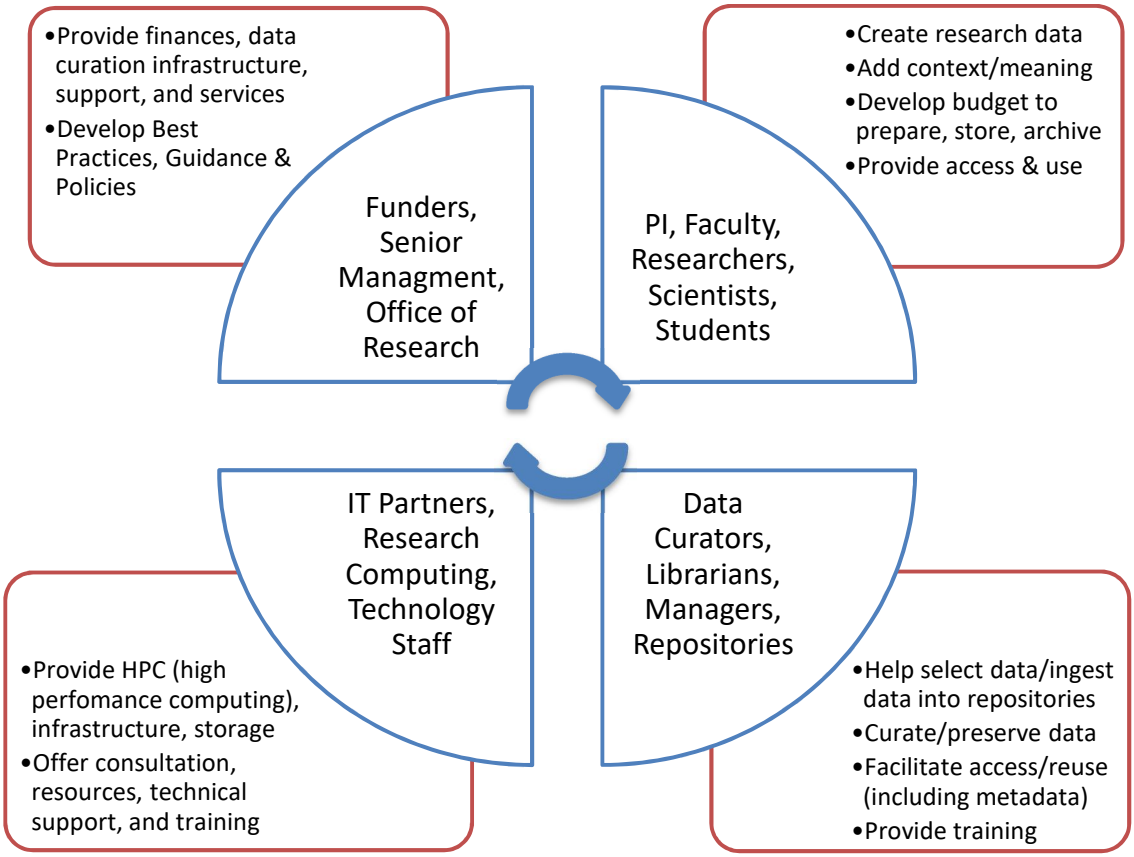


Figure 3: Stakeholders' roles aligned with Data Curation Lifecycle Responsibilities[4]

Figure 3 was developed from the DCC Curation Lifecycle Model mapped to key stakeholders involved in data curation lifecycle process as illustrated by the DAF implementation guide. Figure 3 is an example of some key stakeholders and their overlapping roles in the curation, management, and planning of research data during its lifecycle. In order to ensure compliance with evolving funding agencies data sharing requirements, stakeholders must collaborate, communicate, and aggregate best practices, guidelines, and recommendations across domains.

The DMCWG developed and distributed a small pilot data management survey to researchers at UF in spring 2017. The survey recorded 159 starts and 139 completes. The survey participants included faculty 52.94% (81), staff 16.99% (26), graduate/professional student 22.22% (34), postdoctoral fellow 3.92% (6), resident 1.31% (2), undergraduate 1.31% (2), and other 1.31% (2). Of the 159 respondents, 156 indicated that they would participate in the survey, although not all participants answered every question. Due to limited scope of this paper, only one question from the survey is included to support socio-technical data management collaboration.

[2] DMCWG/DMCTF. (2016). DMCWG. http://ufdc.ufl.edu/AA00014835/00076.

[3] DMCWG Year End Reports. (2019). DMCWG Year End Report (2018) http://ufdc.ufl.edu/AA00014835/00135; DMCWG Year End Report (2017) http://ufdc.ufl.edu/AA00014835/00117; DMCWG Year End Report (2016) http://ufdc.ufl.edu/AA00014835/00085.

[4] JISC, University of Glasgow – HATII, & DCC. (2009). DAF Implementation Guide, p. 3. (Adapted)

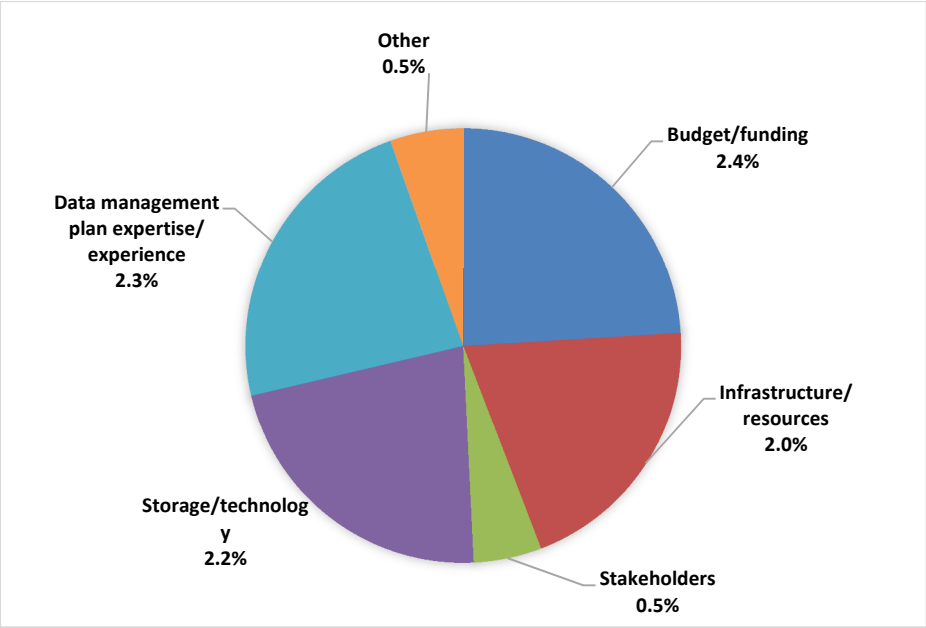


Figure 4. Data management and storage barriers (Q15)

Figure 4 identifies some barriers to data management and storage that are also included in the nodes in the socio-technical framework in Figure 1. When asked to identify barriers regarding management and storage of data, the researchers were given six possible choices and could choose more than one response, if it applied. Of the 258 responses by the 108 researchers that answered this question, 78 researchers chose more than one answer. As shown in Figure Q15, the largest barrier was budget/funding at 24.0% followed closely by data management expertise/experience at 23.3%, storage/technology at 22.1%, and infrastructure and resources at 20.2%. Thus, a socio-technical framework with an embedded business model can help librarians assist researchers in addressing and solving some data management issues.

Some DMCWG accomplishments include but not limited to: several general data management training workshops with an average attendance of 12 participants; several discipline-specific training workshops; several DMP support projects resulting in successful awards totaling $6.2 million dollars (US); 1st Annual Data Symposium[5]; two sub-awards to the UF Libraries; lead on a UF Data Management and Analysis Core (DMAC) for a National Institutes of Health (NIH) National Institute of Environmental Health Sciences (NIEHS) grant proposal.

The Carpentries @ UF

The Carpentries @ UF was established in 2016 to provide data and software carpentries workshops. There were 18 workshops in 2018 (up to 10/24/18) that included R for Geospatial Data; One Day R workshop; Research Bazaar; Data Carpentry (R); Software Carpentry (Python); Data Carpentry Workshops for the 1st Annual Data Symposium; Instructor Training; and Multi-Week Data Carpentry. The Carpentry @ UF premium membership is $15k annually and includes extended collaborations with UFII, UF Biodiversity Institute (UFBI), UF Marston Science Library, iDigBio, UF Florida Museum of Natural History (FLMNH), Wildlife Ecology, and the White Lab.

UF Data Science and Informatics (DSI)

The UF DSI is a recognized UF undergraduate student organization that is creating a data science community through training workshops. Training workshops include use of Jupyter notebooks and the anaconda distribution, use of R Studio for R workshops, use of MySQL for

[5] 1st Annual Data Symposium. (2018). Enabling Data Reproducibility and Sustainability. March 19, 2018. http://cms.uflib.ufl.edu/envisioning-data-symposium/Index.aspx.

SQL workshops, and developing learning pathways for students. UFII fellows and students attend of help at workshops. DSI is funded by UFII and collaborates with UF Libraries/MSL.

Academic Research Consulting & Services (ARCS)

Academic Research Consulting and Services (ARCS) offers a wide range of research support services to the University of Florida community. Hosted by the George A. Smathers Libraries, ARCS has been developed to provide expert services throughout the research process – from data collection through publication. The team also has expertise in the several aspects of data science, including: Statistical analysis (SAS, SPSS); Writing of computer code in using R and Python for analysis of data; Bioinformatics and genomic data analysis using Galaxy, HiPerGator, and R; Geospatial analysis (ArcGIS); Data management and archiving; 3D printing/visualizations.

Testimonials

"Participating in DMCWG gave me chance to work with so many different experts in developing and promoting Data Management service, librarians, research administration and IT managers. Also, I have learnt a lot from the DM training and workshops. At first so many topics (DM planning, Data Sharing and Storage, data security and backup and so on) sound overwhelming, but fortunately, there are more practical tools than ever before to help to achieve the goal. Thanks to the workshop, I know where to start and how to manage the data and made them more accessible." – UF Hough Graduate School of Business, Information Systems and Operations Management graduate (former UF Libraries' student employee), February 4, 2018

Concluding thoughts

"Long-term curation and preservation represent a complex set of challenges, which are exceptionally difficult for data centres and institutions to address individually. They will require a step change in current investment and approaches, and concerted effort on fundamental research, development of shared services, expertise and tools to assist organisations in this work" (JISC Circular 6/03 (Revised), 2003) [5].

References

1. University of New South Wales (UNSW). (2017). Data Governance Policy, Appendix 1 - Data Management Life Cycle. p. 7. Available from: https://www.gs.unsw.edu.au/policy/documents/datagovernancepolicy.pdf.
2. Joint Information Systems Committee (JISC), University of Glasgow Humanities Advanced Technology & Information Institute (HATII), and Digital Curation Center, 2009. Data Asset Framework [formerly Data Audit Framework] Implementation Guide. Retrieved May 30, 2013 from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf.
3. 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.
4. Charles Beagrie Ltd. (2017). CESSDA SaW Archive Development Canvas (Detailed Version). Available from http://tinyurl.com/ybpy94hu.
5. JISC. (2003). JISC Circular 6/03 (Revised). An invitation for expressions of interest to establish a new Digital Curation Centre for research into and support of the curation and preservation of digital data and publications. Accessed January 19, 2019 online from http://tinyurl.com/y95rq5wp.
6. Digital Curation Centre. Checklist for a Data Management Plan [Internet]. Version 4.0. Edinburgh; 2013 [cited 2017 Jan 23]. Available from: http://www.dcc.ac.uk/resources/data-management-plans
7. United States Geological Survey. USGS Data Management [Internet]. Data Lifecycle Overview. 2013 [cited 2017 Jan 20]. Available from: https://www2.usgs.gov/datamanagement/why-dm/lifecycleoverview.php
8. Jones S, Pryor G, Whyte A. How to Develop Research DataManagement Services: A Guide for HEIs'. DCC How-to Guides [Internet]. Edinburgh; 2013. Available from: http://tinyurl.com/hpf6nod
9. Antell K, Foote JB, Turner J, Shults B. Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. Coll Res Libr. 2014;75(4).
10. Diekema AR, Wesolek A, Walters CD. The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories. J Acad Librariansh. 2014;40(3–4).
11. Cox AM, Pinfield S. Research data management and libraries: Current activities and future priorities. J Librariansh Inf Sci. 2014;46(4).
12. Saunders L. Academic Libraries' Strategic Plans: Top Trends and Under-Recognized Areas. J Acad Librariansh. 2015;41(3).
13. Hickson S, Poulton KA, Connor M, Richardson J, Wolski M. Modifying researchers data management practices: A behavioural framework for library practitioners. IFLA J. 2016;42(4).

14. Fecher B, Friesike S, Hebing M. What drives academic data sharing? PLoS One. 2015;10(2).
15. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PLoS One. 2015;10(8).
16. Weller T, Monroe-Gulick A. Understanding methodological and disciplinary differences in the data practices of academic researchers. Libr Hi Tech. 2014;32(3).
17. MacMillan D. Data sharing and discovery: What librarians need to know. Vol. 40, Journal of Academic Librarianship. 2014.
18. Wallis JC, Rolando E, Borgman CL. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS One. 2013;8(7).
19. Marshall B, O'Bryan K, Qin N, Vernon R. Organizing, contextualizing, and storing legacy research data: A case study of data management for librarians. Issues Sci Technol Librariansh. 2013;74.
20. Van Tuyl S, Whitmire AL. Water, water, everywhere: Defining and assessing data sharing in Academia. PLoS One. 2016;11(2).
21. Kratz JE, Strasser C. Researcher perspectives on publication and peer review of data. PLoS One. 2015;10(2).
22. Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. PeerJ. 2013;1.
23. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol. 2015;11(10).
24. Jørn Nielsen H, Hjørland B. Curating research data: the potential roles of libraries and information professionals. J Doc. 2014;70(2).
25. Pinfield S, Cox AM, Smith J. Research data management and libraries: Relationships, activities, drivers and influences. PLoS One. 2014;9(12):1–28.
26. Cox AM, Verbaan E. How academic librarians, IT staff, and research administrators perceive and relate to research. Libr Inf Sci Res [Internet]. 2016;38(4):319–26. Available from: http://dx.doi.org/10.1016/j.lisr.2016.11.004
27. Tenopir C, Sandusky RJ, Allard S, Birch B. Research data management services in academic research libraries and perceptions of librarians. Libr Inf Sci Res [Internet]. 2014 Apr [cited 2017 Aug 7];36(2):84–90. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0740818814000255
28. Kennan MA, Corrall S, Afzal W. "Making space" in practice and education: research support services in academic libraries. Libr Manag. 2014;35(8/9).
29. Patel D. Research data management: a conceptual framework. Libr Rev. 2016;65(4/5).
30. Greene JC, Caracelli VJ, Graham WF. Toward a conceptual framework for mixed-method evaluation designs. Educational evaluation and policy analysis. 1989 Sep;11(3):255-74.
31. Creswell JW, Klassen AC, Plano Clark VL, Smith KC. Best practices for mixed methods research in the health sciences. Bethesda (Maryland): National Institutes of Health. 2011 Aug 1;2013:541-5.

Librarians' Role in GAO Reports*

Meg Tulloch
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 discuss 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.

Abstract:
Working at the U.S. Government Accountability Office (GAO), librarians know that their research and documentation are part of an important process that can provide background information, help refine a researchable question that the agency is answering for the U.S. Congress, or support a finding in a GAO report. Carefully documented literature searches form the backbone of this process. The librarians with a team of analysts and other specialists, such as methodologists, economists, and scientists, define the criteria for the searches and the evaluation of the results. The librarian performs the searches and participates in the evaluation process often with the methodologist. The librarian is responsible for documenting searches and their returns in the Record of Research. As well, she captures the results for evaluation in a way that can be linked back to the searches. At the end of the process, the librarian has produced a Record of Research and a results document or Data Collection Instrument. Sometimes, the librarian also writes analysis about the results to support a finding.

Introduction

The United States Government Accountability Office (GAO) is an independent, nonpartisan agency within the legislative branch of government. "Our mission is to support the Congress in meeting its constitutional responsibilities and to help improve the performance and ensure the accountability of the federal government for the benefit of the American people. We provide Congress with timely information that is objective, fact-based, nonpartisan, non-ideological, fair, and balanced."[1]

GAO's work aids congressional oversight in several ways—auditing agency operations, including government programs and policies; investigating allegations of governmental wrongdoing; outlining policy options for Congress to consider; and issuing legal decisions and opinions. As the supreme audit institution for the United States, one of GAO's most important roles is to investigate and understand how the federal government spends taxpayer dollars. For this reason, GAO has been nicknamed the "congressional watchdog."

GAO is headquartered in Washington, D.C. and has eleven field offices around the country to support its work. GAO has over three thousand employees; the majority of whom work for the mission teams who create GAO reports. Thirteen of the fourteen teams audit specific subject areas such as defense, contracting, health care, education, homeland security, physical infrastructure, natural resources, financial management, information technology,

* First published in the GL20 Conference Proceedings, February 2019.
[1] "What GAO Is." *GAO Web site,* 3 Jan. 2019, https://www.gao.gov/about/what-gao-is/.

international affairs, and many other topics.[2] In fiscal year (FY) 2018[3], GAO produced six hundred thirty-three reports as part of its auditing function.[4]

GAO librarians work within the fourteenth mission team, Applied Research and Methods (ARM). ARM mainly supports the thirteen other teams in creating GAO reports.

> ARM offers expertise in many areas including cost analysis, engagement design, economics, data analysis, evaluation, library research and literature review, science, statistics, surveys, technology, engineering, and IT security. In addition, five technical chiefs provide expertise in statistics, economics, technology, actuarial science, accounting, and cutting-edge science.[5]

Library Services provides GAO with research products and services, with access to accurate and reliable information, and with proficiency in organizing and preserving information in all formats. Librarians provide direct support to the mission teams responsible for GAO reports, including as stakeholders. Librarians work with the teams to develop research and search strategies and the results of those strategies. They also purchase access to accurate and reliable information and data sources and provide access to information through interlibrary loan and document delivery. Librarians work with other ARM Specialists to develop guidance on topics such as literature searches and reviews and information reliability.

In FY2018 GAO librarians conducted two hundred nineteen literature searches and fifty-four background searches in support of GAO reports.[6] This paper will discuss the librarians' process and role in creating GAO reports.

Literature Searches and Reviews Guidance

GAO has several types of internal guidance that applies to literature searches and reviews. For the engagement process GAO staff follows the *Electronic Assistance Guide for Leading Engagements* (EAGLE II). "It contains GAO guidance, consistent with GAO policies and protocols and GAGAS [GAO's Government Auditing Standards] requirements, and is a definitive source for implementing GAO's Engagement Management Process."[7] Within EAGLE II under "Plan & Proposed Design," GAO staff can refer to a section on obtaining background information (2.3.2). This section says that, "an engagement team conducts and documents a sufficient review of existing literature or other appropriate research to assure themselves that they understand the nature and background of the program or agency under review."

The section goes on to describe the literature searches more specifically. It says that "the engagement team conducts literature searches, or has ARM Research Librarians conduct the searches for them, as appropriate. Literature searches can provide contextual sophistication or can be the basis of a GAO product, such as an evaluation synthesis."[8] ARM also has three internal guidance papers-- *Using Studies Conducted by Outside Researchers for GAO Engagements, Guidelines for Assessing Web-based Sources of Information*, and *Sample Objectives, Scope, and Methodology (OSM) Language for Literature Searches and Reviews*. The first guidance paper describes the literature search process as follows:

> The literature search identifies studies that are relevant to the researchable questions and that meet specified criteria. ARM specialists and research librarians can help determine what the criteria should be in order to conduct a comprehensive yet specific literature search. [9]

[2] "Our Teams." *GAO Web site,* 3 Jan. 2019, https://www.gao.gov/about/careers/our-teams/ .

[3] The fiscal year for the U.S. Government is October 1st through September 30th. FY2018 is from Oct. 1, 2017 through September 30th, 2018.

[4] "About GAO, Performance." *GAO Web site,* 4 Jan. 2019, https://www.gao.gov/about/what-gao-is/performance.

[5] "ARM Overview." *GAO Intranet,* 3 Jan. 2019, ARM Home page (internal web site).

[6] "ITSM Library Subject Statistics Between FY2015 and FY2018." *ITSM Reports*, Generated 1 Nov. 2018, (internal document).

[7] "EAGLE II Overview." *GAO Intranet,* 3 Jan. 2019, EAGLE II Home page (internal web site).

[8] "EAGLE II "Planning & Proposal." *GAO Intranet,* 3 Jan. 2019, EAGLE II Home page (internal web site).

[9] "Using Studies Conducted by Outside Researchers for GAO Engagements." *ARM Guidance*, p. 4, Accessed 3 Jan. 2019, ARM Guidance Web site (internal web site).

Guidelines for Assessing Web-based Sources of Information helps analysts determine the sources that might be used as part of a background or literature search are authentic, authoritative, and reliable (see Appendix 4 for definitions). For information to be used as evidence, the information must be authentic and authoritative, and almost always it must also meet the test for reliability. As the guidance says,

> Reliability Assessment
> Information from U.S. government and military Web sites should not be assumed to be reliable even if the source is authentic and authoritative. The information could be inaccurate, outdated, or incomplete. Depending on the nature of the information and the auditors' use of the information in relation to the audit's objectives, findings, and conclusions, the auditors may need to corroborate the information. The auditor should always use professional judgment when assessing the value of information for use as GAO evidence.[10]

The following are examples of web-based content that is considered low risk and generally usable as evidence: Congressional Budget Office Reports, Code of Federal Regulation or U.S. Code from the Government Publishing Office or an approved third party such as Westlaw or LexisNexis, peer-reviewed/scholarly articles, professional and technical standards from the authoring web site, U.S. federal, state, and local government web sites.

Sample Objectives, Scope, and Methodology (OSM) Language for Literature Searches and Reviews provides examples and recommendations on how a literature search and review are addressed in the formal written part of a GAO report. It suggests that five key elements should be included to adequately address how the literature review was conducted and how the results were reviewed. The five elements are 1) how the articles or reports were identified 2) a list of the databases searched 3) criteria used to select articles for the literature review 4) number of articles identified 5) description of the methodical review.[11] For the second element, the librarians provide detailed search strings for the databases searched, which are entered in the Record of Research as is shown in Appendix 2. When a literature review provides evidence for an objective a librarian may be asked to write the OSM for the literature review or review what an analyst or methodologist has written (see Appendix 5).

Literature Searches and Reviews Tools and Process

In order to follow the EAGLE II and ARM guidance, GAO librarians developed different tools to document the research process. The Record of Research is filled out by the mission team asking for the literature search. The team outlines their information needs, whether the information or studies will be used as background information, evidence or to answer a researchable question, and documents their own searching for information. The team also lists limiting factors, such as date range and desired document types (see Appendix 1). The librarian then records her searches and the number of results into the Record of Research (see Appendix 2). For a background search, the librarian will use a select set of databases for searching. For a literature review, the librarian will be as comprehensive as possible in her searching. In either case, the librarian will often send an initial set of results to make sure that she has understood the research request. Once she has feedback, the librarian may continue searching. For a literature review or a complex background search, the searching process can be iterative, with several exchanges between the librarian and the analysts.

The results from the searching are exported to RefWorks, de-duped, organized, and then exported as citations with abstracts to a results document or Data Collection Instrument (see Appendix 3) for analysis by the librarian, ARM Specialist, or team analyst. The analysis is recorded in the Data Collection Instrument (DCI) to include evaluation of the materials using pre-determined criteria.

[10] "Guidelines for Assessing Web-based Sources of Information." *ARM Guidance*, April 2017, p. 7, Accessed 4 Jan. 2019, ARM Guidance Web site (internal web site).

[11] "Sample Objectives, Scope, and Methodology (OSM) Language for Literature Searches and Reviews." *ARM Guidance*, March 2011, p. 1-2. Accessed 3 Jan. 2019, ARM Guidance Web site (internal web site).

The Library Services members will also help the team find full-text of any or all of the results for the background or literature search. This process can involve finding full-text in existing library materials, borrowing or requesting full-text from outside organizations, or purchasing the materials or data. The librarians are constantly evaluating materials to make sure that the right mix of library materials are readily available, easily accessible, and licensed so that the materials can be saved with the engagement workpapers as evidence in the audit process.

As noted earlier, librarians can be involved in writing or reviewing the OSM. The sample in Appendix 5 was written by a librarian.

Conclusion

Working at the U.S. Government Accountability Office (GAO), librarians know that their research and documentation are part of an important process that can provide background information, help refine a researchable question that the agency is answering for the U.S. Congress, or support a finding in a GAO report. As members of the ARM team, they can be involved with any report that the mission teams are working on. Librarians were listed as contributors in twenty GAO reports in FY2018.

Appendix 1 : Sample Record of Research, Searching Details

2. Purpose

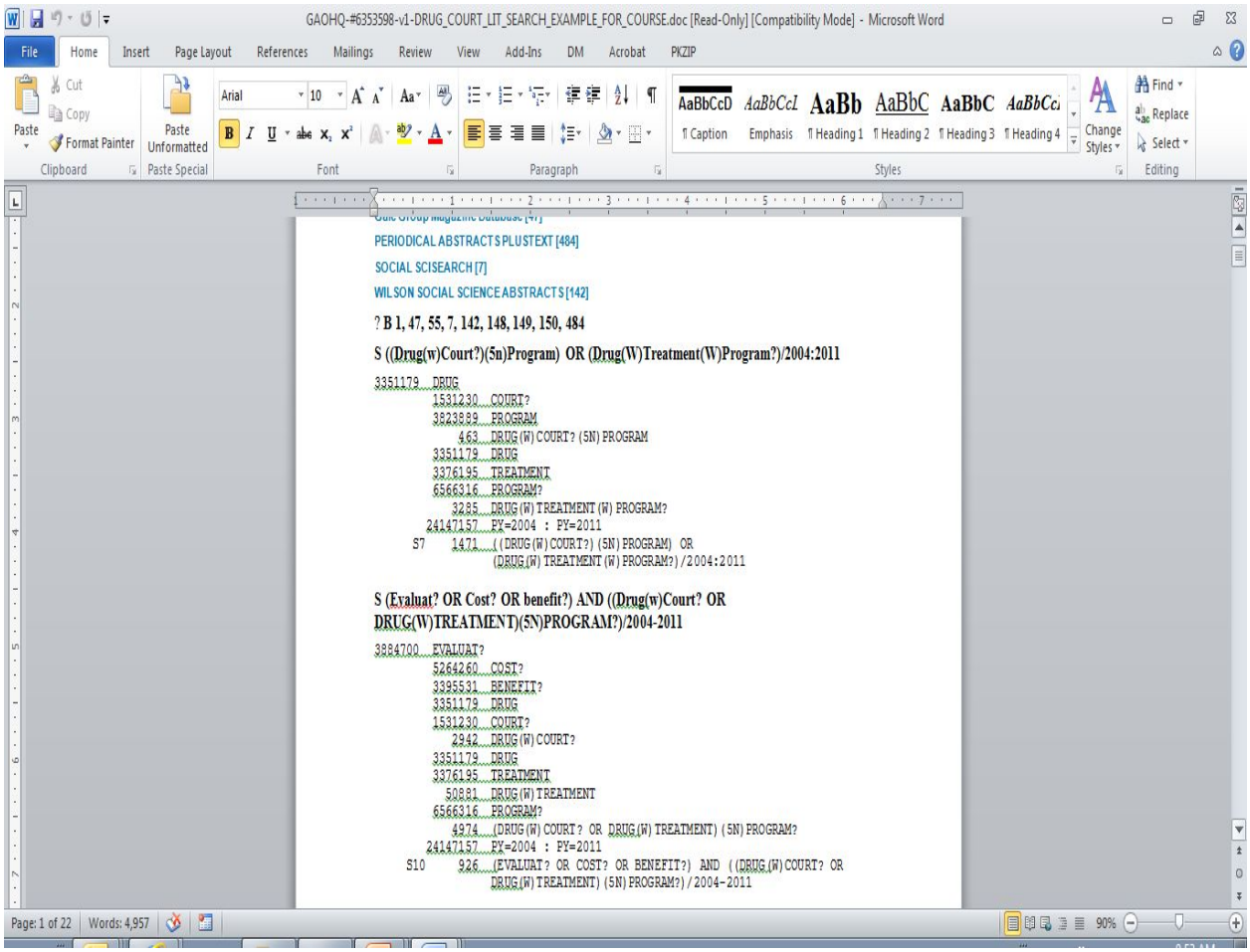
2.1 Intended Use of Literature Search (check all that apply) *Note:* There is ARM Guidance on Literature Searches and Studies Conducted by Others If the intended use of the literature search changes over the course of the job, you and your librarian may make changes to this form documenting updated searches.	
	Initial background search to help understand engagement topic and/or to provide context for background section of report
	Provide a source of evidence for one or more objectives/researchable questions (examples include open source projects and literature searches conducted to identify articles for a review of studies) • Please provide details on how the search results will be used _____ Results will be systematically presented as a section in the report _____ Results will be referenced as contextual information in the report _____ Results will be used for identifying and/or developing criteria or identifying examples or case studies for further examination *Note:* Your methodologist can help you identify how the literature search can fit into your methodology.
	Identify subject matter experts or other stakeholders for interviews
	Other (please explain):

3. Search Preferences

3.1 Types of materials desired (check all that apply):			
	Scholarly/Peer Reviewed Material		Government Reports
	Conference Papers		General News
	Dissertations		Trade and/or Industry Articles
	Working papers		Association/Nonprofit/Think tank Publications
	Books		Legislative materials (Congressional transcripts, bills, etc.)
NOTE: Contact your General Counsel stakeholders regarding legal material.			

3.2 Publication date range: *Note:* Discuss with your librarian if you're not sure how far to go back.	
	Previous 5 years
	Previous 10 years
	Other (please specify):

Appendix 2 : Sample Search Strings



GAOHQ-#6353598-v1-DRUG_COURT_LIT_SEARCH_EXAMPLE_FOR_COURSE.doc [Read-Only] [Compatibility Mode] - Microsoft Word

PERIODICAL ABSTRACTS PLUSTEXT [484]

SOCIAL SCISEARCH [7]

WILSON SOCIAL SCIENCE ABSTRACTS [142]

? B 1, 47, 55, 7, 142, 148, 149, 150, 484

S ((Drug(w)Court?)(5n)Program) OR (Drug(W)Treatment(W)Program?)/2004:2011

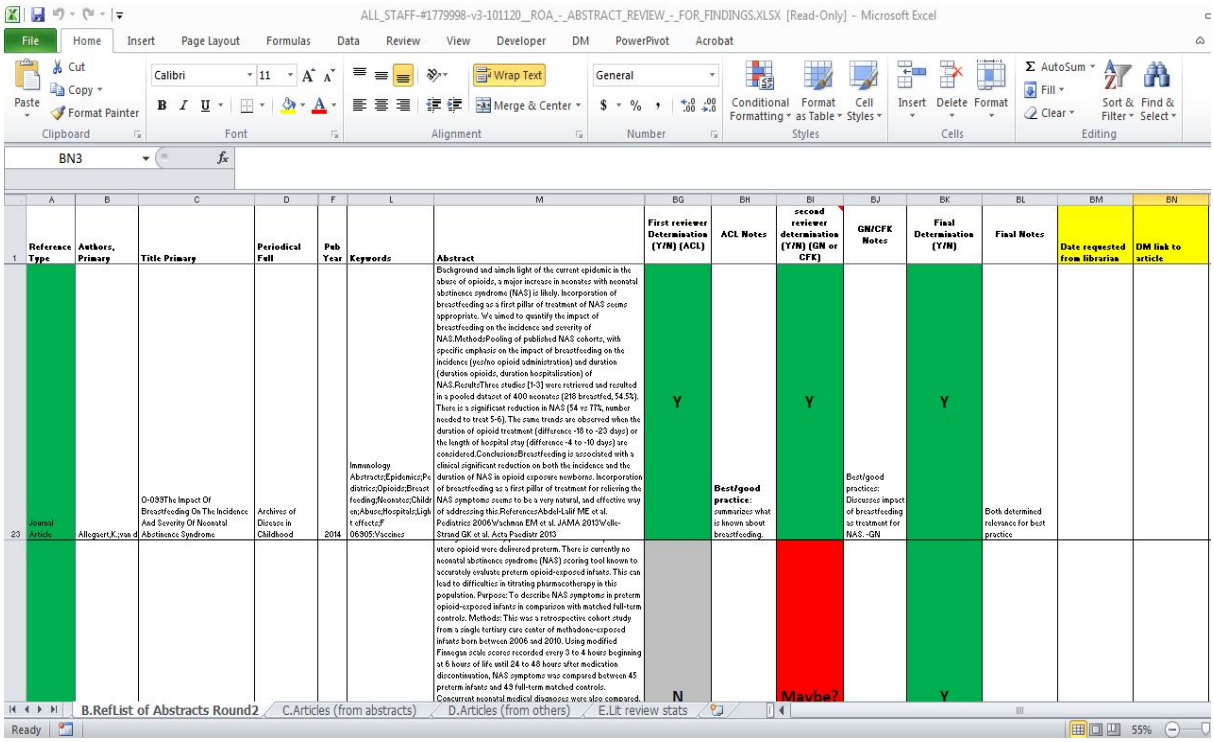
3351179 DRUG
1531230 COURT?
3823889 PROGRAM
463 DRUG(W)COURT?(5N)PROGRAM
3351179 DRUG
3376195 TREATMENT
6566316 PROGRAM?
3285 DRUG(W)TREATMENT(W)PROGRAM?
24147157 PY=2004 : PY=2011
S7 1471 ((DRUG(W)COURT?)(5N)PROGRAM) OR (DRUG(W)TREATMENT(W)PROGRAM?)/2004:2011

S (Evaluat? OR Cost? OR benefit?) AND ((Drug(w)Court? OR DRUG(W)TREATMENT)(5N)PROGRAM?)/2004-2011

3884700 EVALUAT?
5264260 COST?
3395531 BENEFIT?
3351179 DRUG
1531230 COURT?
2942 DRUG(W)COURT?
3351179 DRUG
3376195 TREATMENT
50881 DRUG(W)TREATMENT
6566316 PROGRAM?
4974 (DRUG(W)COURT? OR DRUG(W)TREATMENT)(5N)PROGRAM?
24147157 PY=2004 : PY=2011
S10 926 (EVALUAT? OR COST? OR BENEFIT?) AND ((DRUG(W)COURT? OR DRUG(W)TREATMENT)(5N)PROGRAM?)/2004-2011

Page: 1 of 22 Words: 4,957 90%

Appendix 3 : Sample Data Collection Instrument (DCI)



ALL_STAFF-#1779998-v3-101120__ROA_-_ABSTRACT_REVIEW_-_FOR_FINDINGS.XLSX [Read-Only] - Microsoft Excel

	Reference Type	Authors, Primary	Title Primary	Periodical Full	Pub Year	Keywords	Abstract	First reviewer Determination (Y/N) (ACL)	ACL Notes	second reviewer determination (Y/N) (GN or CFK)	GN/CFK Notes	Final Determination (Y/N)	Final Notes	Date requested from librarian	DM link to article
23	Journal Article	Allegaert,K;van d	0-0387The Impact Of Breastfeeding On The Incidence And Severity Of Neonatal Abstinence Syndrome	Archives of Disease in Childhood	2014	Immunology Abstracts;Epidemics;Pediatrics;Opioids;Breastfeeding;Neonates;Childhood;Abuse;Hospitals;Lights effects;F06905:Vaccines	Background and aimsIn light of the current epidemic in the abuse of opioids, a major increase in neonates with neonatal abstinence syndrome (NAS) is likely. Incorporation of breastfeeding as a first pillar of treatment of NAS seems appropriate. We aimed to quantify the impact of breastfeeding on the incidence and severity of NAS.MethodsPooling of published NAS cohorts, with specific emphasis on the impact of breastfeeding on the incidence (yes/no opioid administration) and duration (duration opioids, duration hospitalisation) of NAS.ResultsThree studies [1-3] were retrieved and resulted in a pooled dataset of 400 neonates (218 breastfed, 54.5%). There is a significant reduction in NAS (54 vs 77%, number needed to treat 5-6). The same trends are observed when the duration of opioid treatment (difference -18 to -23 days) or the length of hospital stay (difference -4 to -10 days) are considered.ConclusionsBreastfeeding is associated with a clinical significant reduction on both the incidence and the duration of NAS in opioid exposure newborns. Incorporation of breastfeeding as a first pillar of treatment for relieving the NAS symptoms seems to be a very natural, and effective way of addressing this.ReferencesAbdel-Latif ME et al. Pediatrics 2006;Wachman EM et al. JAMA 2013;Welle-Strand GK et al. Acta Paediatr 2013	Y	Best/good practice: summarizes what is known about breastfeeding.	Y	Best/good practice: Discusses impact of breastfeeding as treatment for NAS. -GN	Y	Both determined relevance for best practice		
							utero opioid were delivered preterm. There is currently no neonatal abstinence syndrome (NAS) scoring tool known to accurately evaluate preterm opioid-exposed infants. This can lead to difficulties in titrating pharmacotherapy in this population. Purpose: To describe NAS symptoms in preterm opioid-exposed infants in comparison with matched full-term controls. Methods: This was a retrospective cohort study from a single tertiary care center of methadone-exposed infants born between 2006 and 2010. Using modified Finnegan scale scores recorded every 3 to 4 hours beginning at 6 hours of life until 24 to 48 hours after medication discontinuation, NAS symptoms was compared between 45 preterm infants and 43 full-term matched controls. Concurrent neonatal medical diagnoses were also compared.	N		Maybe?		Y			

B.RefList of Abstracts Round2 C.Articles (from abstracts) D.Articles (from others) E.Lit review stats

Ready 55%

Appendix 4 : Definitions of Authentic, Authoritative, and Reliable

From GAO ARM Guidance Paper: ***Guidelines for Assessing Web-based Sources of Information***

Authentic: The accessed Internet source is, in fact, what it declares to be or the Internet source's provenance can be verified.

Authoritative: The Internet source is the official site for an organization (public or private sector) or is recognized as an legitimate source for information on the subject matter.

Reliable: The information presented in the internet source is sufficiently unbiased and reasonably complete, timely, and accurate with regards to the analyst's intended purposes; as a result, it is associated with a low risk of the analyst making incorrect or improper conclusions based on that information (GAGAS 6.56-6.72).[12]
GAGAS = generally accepted government auditing standards (GAGAS) or GAO's Government Auditing Standards (Yellow Book)

Appendix 5 : Sample OSM[13]

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.

[12] United States, Government Accountability Office. *Government Auditing Standards*, Government Accountability Office, 2011, p. 150-156. Accessed 4 Jan 2019. https://www.gao.gov/products/GAO-12-331G.
[13] United States, Government Accountability Office. *Child Care: Multiple Factors Could Have Contributed to the Recent Decline in the Number of Children Whose Families Receive Subsidies*, Government Accountability Office, May 5, 2010, p. 37-38. Accessed 4 Jan 2019. https://www.gao.gov/products/GAO-10-344.

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

Abstract
The library of the Japan Atomic Energy Agency (JAEA) focuses on collecting conference proceedings and technical reports which are very important and referred to as materials for our users in the field of nuclear science and technology. Increasing numbers of cases are published in electronic media such as CD/DVD-ROM and flash memory. Meanwhile, problems exist as to how to manage such electronic media in a library such as long-term preservation and permanent access, because such electronic media are updated or changed in terms of specifications. For this reason, it is well known that respective electronic media would be unavailable. Therefore it's necessary to have appropriate playback equipment and software corresponding to the electronic media. The lifetime of electronic media is much shorter compared with paper and microform. The authors would like to present the JAEA Library's current activities on long-term preservation and use of electronic media. They conclude the fact that even published electronic media are becoming grey literature in certain environment.

1. Introduction

Japan Atomic Energy Agency (hereinafter, this is called "JAEA") (ref 1) is a comprehensive R&D institute dedicated to nuclear energy in Japan. The JAEA Library is one of the largest nuclear information centers in Japan. It provides technical information services to researchers and engineers in the nuclear science and technology.
The primary roles of the Library (ref 2) are:

- To collect and make academic information in nuclear field available for library users
- To repose and disseminate documented results of research and development activities

The following practical work of the Library is detailed:

- General work of library (acquisition and technical service, network system, photocopy service, interlibrary loan and so on)
- Dissemination of the R&D results/achievements via its own institutional repository, JAEA Originated Papers Searching System (JOPSS)
- Publishing the JAEA Reports (technical reports of JAEA) and the "JAEA R&D Review" (annual publication that introduces JAEA's R&D results/achievements in a form of digest)
- Collection and dissemination of Fukushima Accident information via Fukushima Nuclear Accident Archive (FNAA)
- Taking part in the International Nuclear Information System (INIS) program, academic information sharing engaged in International Nuclear Library Network (INLN)

In this article, the electronic media are the ones ranged and limited to materials electronically published, recorded and passed into electronic media tangible, in accordance with the definition found in the survey report issued by the National Diet Library, Japan (NDL) (ref 3).

2. Facts of materials of electronic media at JAEA Library

2.1. Collection of electronic media

As of today, the JAEA Library includes some 50,000 books, 2,000 journal titles, 800,000 research and/or technical reports and about 365,000 US nuclear licensing documents of US Nuclear Regulatory Commission (DOCKET). The breakdown of the Library collections are as follows (Table 1, Fig. 1).

* First published in the GL20 Conference Proceedings, February 2019.

Table 1: JAEA Library Collections

	Print	Microforms	Electronic media	Online	Total
Books	46,667	0	1,201	2,285	50,153
Technical reports	74,783	705,035	4,469	0	784,287
DOCKET	0	365,502	0	0	365,502
	121,450	1,070,537	5,670	2,285	1,199,942

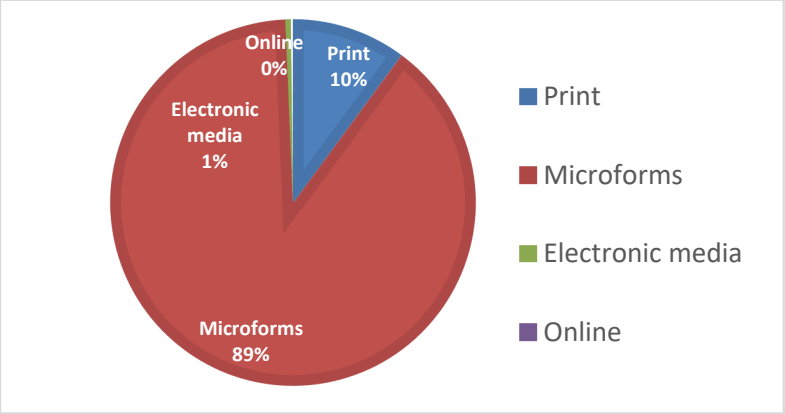


Fig. 1: JAEA Library Collections

The proceedings specific to nuclear science take priority in the library's collections. Notable is the collection of the grey literature having strived over decades since the establishment of the library. We have investigated the type of JAEA achievement announcements over the past three years. Of JAEA's achievement announcements, 76% were proceedings of papers with presentations. From this point, it can be said that the proceedings are important for the field of nuclear science and technology (Fig. 2).

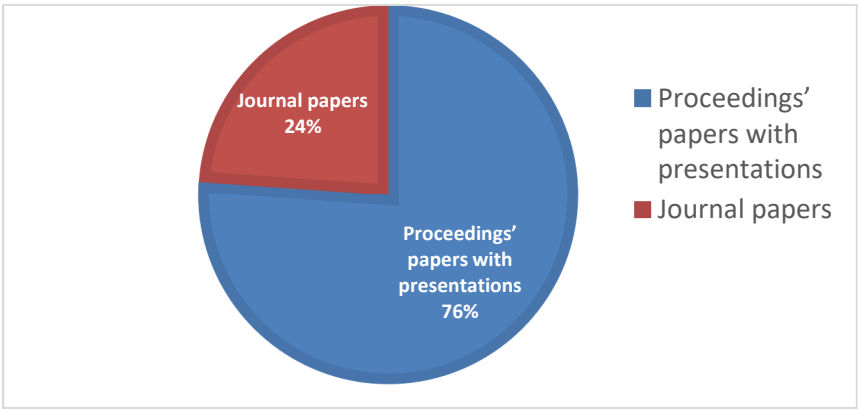


Fig. 2: Submission of JAEA R&D results over the latest three years (Apr 2015 to Mar 2018)

Furthermore, it is important for the Library to stay and collect proceedings as many proceedings as possible. Occasionally, for example, the library looks for JAEA personnel who took part in conferences and requests to grant the proceedings to the library to enrich the collection, courtesy of them. That is why the Library can often satisfy users' requests, for example, such requests as those via Document Delivery Service from other libraries in Japan.

Proceedings are increasingly delivered and distributed at conferences in the form of electronic media such as CD/DVD-ROM, flash memory (Fig. 3). And they can even be downloaded by participants from websites controlled by ID/PW.

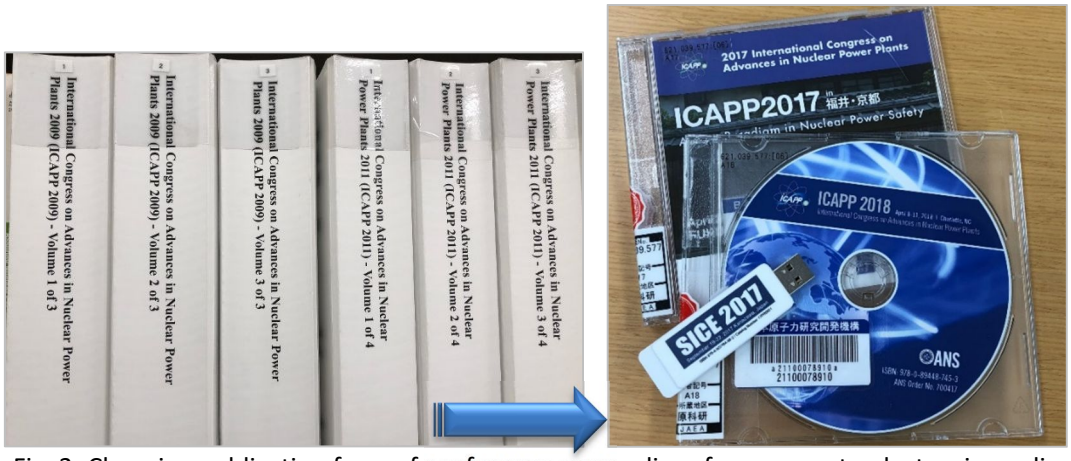


Fig. 3: Changing publication form of conference proceedings from paper to electronic media

The collection status of books and conference proceedings by the Library by such media are, shown in Table 2. The books in Table 1 are separated with books and conference records.

Table 2(a): Breakdown and purchase time of electronic media in books of JAEA Library

	CD/ DVD	Flash Memory	Floppy Disk	Video Cassette	Magneto -Optical Disc	Total
1990s	2	0	0	38	0	40
2000s	132	0	0	18	1	151
2010s	45	2	5	0	0	462
Total	589	2	5	56	1	653

Table 2(b): Breakdown and purchase time of electronic media in proceedings of JAEA Library

	CD/ DVD	Flash Memory	Floppy Disk	Video Cassette	EBook reader*	Total
1990s	16	0	0	0	0	16
2000s	194	0	0	0	0	194
2010s	303	33	0	0	1	337
Total	513	33	0	0	1	547

*EBook reader: Proceeding stored in TRECSTORE ebook reader 3.0

3. Current situation and issues

3.1. Current situation of the JAEA Library

One of the serious and often urgent concerns for libraries would be coping with the unavailability of documents held in the form of electronic media. It would be neither simple nor cost effective for libraries to make and keep the hardware and software corresponding with such media available for reading at any time. These items are not only of case-by-case but of year-by year following updates to procedures or routine maintenance. Meanwhile, alongside such inconveniences, there is the other concern that the Library cannot take an enough time to maintain electronic media and take measures such as emulation. Therefore, lifetime of electronic media is much shorter than paper and microform.

Below are some challenges the JAEA Library has faced so far:

- most of books and technical reports, electronic medium-based, collected and held since 1990's cannot be read out at present without corresponding and suitable software
- whereas the Library has held old personal computers (PCs) with old versions of Operating Systems (OSs) like Windows 95 to read out the respective electronic media. Such PCs with OSs are mostly limited to be used to applications. They cannot be connected to any electronic devices from of the current PCs related under the in-house security rules
- some of articles of proceedings are displayed on suitable monitors and fortunately might have been printed out if properly connected to usable printers. However, some printers

are too old, even worse without toners, and do not have the corresponding electronic format needed for use (Fig. 4)[1]

Fig. 4: Windows 95 machines and their printers kept in JAEA Library, purchased before 2000

In this regard, the Library has experienced difficulty managing electronic media, to collect, hold, and provide accessibility in the long term. Improving the collection of electronic medium documents is necessary at the Library in near future.
The Library attempts to share important and useful information regarding how to deal with electronic media stemming from its own practical experiences with other libraries worldwide, as likely measures we could take even if not definite . This article guides the following two topics of discussion:

1. associated issues having arisen for years at the Library in practice,
2. some measures having been taken and sought to be taken afterward against such issues. It is believed such concerns would be not only of ours but yours as well and would be shared, as a matter of "grey literature being coming, anytime."

3.2. Investigating on long-term preservation and use of electronic media by NDL, Japan

National Diet Library of Japan (NDL, Japan) has implemented surveys on how to effectively preserve electronic media in the long term since several years ago (ref 3), which they mention in their 2018 report. It also focuses on how to deal with media packaged.
In the 2018 report, NDL, Japan has investigated the interoperability by migration from such media to the other media or something like databases. The experiences and findings remain as follows:

- such migration is usable and comparatively easier to be conducted
- however, for electronic media like CD/DVD-ROM, it is necessary to be applied with applications to read out. Electronic configurations applying virtual machine like VMware (ref 4), should be installed and used. In addition, further investigation is necessary for how to apply the methods of migration.

Measures for using electronic media materials at JAEA Library

3.3. Policies toward necessary measures

In order to ensure preservation and permanent access to all the electronic media of the collection of the JAEA Library, the following surveys to check such media and seek for measures for those:

- Could the media be read out or not?
- What are the conditions regarding hardware and software to have the media read out?
- Are effective methods of migration of the media for long-term preservation foreseen?

On the other hand, resources like manpower, budgets, and documents/data management skills are insufficient at the Library. Taking account of this practical and uneasy situation, procedures prioritized as the measures are:

1. to list proceedings before they are requested by library users or through photocopy service and to check if they can be read out or not,

[1] In Japan by the domestic law of copyright and its guideline, libraries are neither allowed to electronically duplicate materials nor to electronically provide them to users. Therefore, the user who wants to copy the proceedings of the electronic medium needs to print the relevant part and it is necessary for my library engaged in Japan to use the printer.

2. to seek for dealing with proceedings of electronic media which are readable out only by Windows 95/98 as of today,
3. to deal with proceedings of electronic media which are not able to be read out as of today.

In view of the above, we decided to take countermeasures against this situation. Our final goal should not be to migrate the electronic media, but should be to figure out the ways and measures for long term preservation and accessibility. Preserving the media is not the challenging but rather making them available for library users, i.e., to make them readable out by any measures, which we have not explored.

3.4. Measures via OSs

The Library has possessed a PC compatible with Windows 95 and made it available as of today to make PDF files available to be read out, for most of electronic files in proceedings published in the form of CD/DVD-ROM and of PDF files. Doing so highly depends on the versions of Adobe Acrobat or Adobe Acrobat Reader.
However, we have problems with this as well:

- The PC with Windows 95 was purchased before 2000, and the state is unstable and incompatible. For example, the touch pad is temporarily unusable.
- It cannot print articles because it does not have toner for the printer connected to the Windows 95 PC is produced anymore.

In this regard, the solution to succeed in reading out some electronic files has been identified. To avoid challenges arising from deeply limited use of PC compatible with Windows 95, its counter-measure is to provide a PC compatible with Windows 10 for 32 bit with NTVDM (NT Virtual DOS Machine)(Fig. 5).

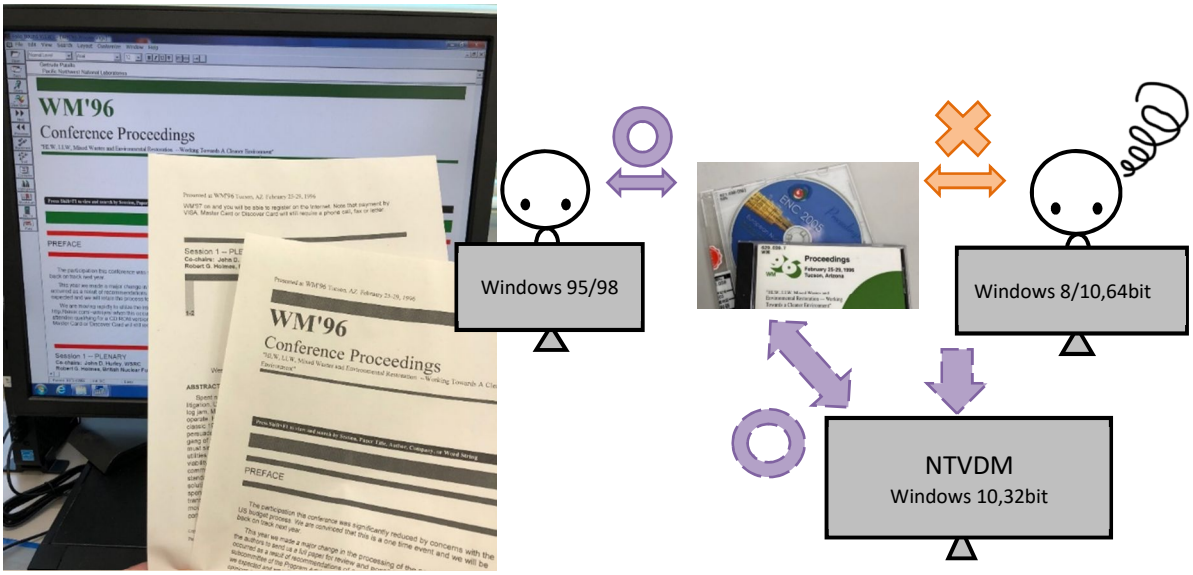


Fig. 5: Read and print proceedings using Windows 10 for 32 bit

The following, as references, are trials with the consideration, sought for and worked out:

- the fault to let the respective programs run via the present PCs occurs due to its incompatibility with the 16 bit based ordering commands for the respective programs. The programs were previously run via PCs with Windows 95/98
- NTVDM can make such programs in days of Windows 95 usable by activating the virtual 86 mode. This mode enables the orders of old CPU to run via the present CPU. This program is available on Windows 10 for 32 bit but not on Windows 10 for 64 bit.

Considering this, we decided to purchase Windows 10 for 32 bit newly and enable the playback of the old version file. However, not all Windows 10 for 32 bit can play back all the past conference proceedings. We are investigating now.

3.5. Example of migration - migration simply and easily -

Here is one example for easily and easily migrating electronic media stored in an electronic book reader made in Germany, "TRECSTORE (ref 5) ebook reader 3.0" at the Library. It was either difficult for library users to find out some documents of them by the reader due to the proceeding with more than 1000 pages, nor was it impossible for them to connect to printers and print out due to the reader like "stand-alone". A sort of simple migration was done for it, by reading out the electronic file via the reader and saved in a DVD-ROM for use.

4. Closing remarks

We believe that the most important and necessary is not only to distribute/share proceedings to/with participants but also to publish them to make them accessible for the potential users through libraries. Proceedings distributed to participants in the form of electronic media are one area of concern of grey literature. Those are unlikely to be available for potential users for a long time and will disappear, to some extent.
The JAEA Library has a serious concern as entitled "Published materials are becoming Grey". We have a wish "Published materials should not become Grey." and would like to ask conference organizers or academic associations who possess the electronic media of proceedings to recognize the concern. To avoid such situation loomed up, becoming Grey, it is necessary to take action requesting the academic associations to do as follows:

- Not only distributing/sharing the proceedings in the form of electronic media between participants but publishing those in print and/or online accessible for users in general
- In addition, having agreement on the publishing with authors regarding copyright
- Furthermore possibly being with DOI in the case of online publishing

It is hoped, acknowledging such request, little by little, the academic associations would change their standpoints on the concern with "becoming Grey".

Proceedings in the form of CD/DVD-ROM and flash memory are easy to carry around, easy to search, and very convenient for users. Whether to publish or distribute those only depends on the conference's scope, in which any users cannot be involved, we, libraries neither. Therefore it's true those convenience and scope of the disclosure are not matters we deny adversely.
The Library persists to improve and enhance the proceedings' collection in particular the skills to read out as measures against the issues and difficulty associated with, mentioned above. The hope is to be able to share important and useful information stemming from practical experiences of the Library with other libraries worldwide, believing such concerns would be not only of ours but shared as well.

References

1) Japan Atomic Energy Agency: https://www.jaea.go.jp/english/ ,(accessed Dec. 27th 2018).
2) Library of Japan Atomic Energy Agency: https://tenkai.jaea.go.jp/english/library/ (accessed Dec. 27th 2018).
3) National Diet Library, Japan : http://www.ndl.go.jp/jp/preservation/dlib/index.html (in Japanese), (accessed Dec. 27th 2018).
4) VMware: https://www.vmware.com/jp.html (in Japanese), (accessed Dec. 127h 2018).
5) TRECSTORE: http://www.trekstor.de/home-en.html , (accessed Dec. 27th 2018).v

Measuring Reuse of Institutionally-Hosted Grey Literature*

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

URL: https://reuse.diglib.org

Abstract

Grey literature is often hosted in digital libraries such as institutional repositories (IRs) 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 repository materials are utilized and repurposed, can bridge this gap for IRs. Reuse of digital repository collections and materials is an important assessment measurement because, unlike use, reuse shows engagement with collections and impact of digital repository resources in a more meaningful fashion.

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 repository 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 repository 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 stakeholders such as 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 ended in June 2018, the research team 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.

Introduction

Reuse of digital repository works and collections is an important assessment measurement because, unlike use, reuse shows engagement with collections and impact of digital repository resources in a more meaningful fashion. This paper focuses on the findings of a needs assessment from the "Developing a Framework for Measuring Reuse of Digital Objects" (Measuring Reuse) grant project that are directly pertinent to grey literature and institutional repositories (IRs).

The Measuring Reuse grant focused on carrying out an in-depth needs assessment of the digital repository community, spanning the Cultural Heritage and Knowledge Organization (CHKO) spectrum, to determine desired functionality of a reuse assessment toolkit. Via surveys and focus groups, the project team elucidated from this needs assessment the

* First published in the GL20 Conference Proceedings, February 2019.

features, functionalities, and use cases the digital repository community prioritized in a reuse assessment toolkit.

During this year-long needs assessment, the team gathered data using two surveys, five in-person focus groups and five virtual focus groups. The researchers derived use cases from the coded focus group data, which were, along with features of a future toolkit for assessing reuse, prioritized by the digital repository community in the second survey. Measuring reuse of grey literature and IR collections has emerged as a complex issue, with all the complications of assessing reuse of published material with additional challenges unique to grey literature and its myriad forms.

The paper will conclude with avenues of future research from the Measuring Reuse project centering assessment of IRs and grey literature, and the research team's next steps. The remaining sections of this paper are as follows: Grant and Project Background, Literature Review, Focus Group Data, Discussion, and Conclusion.

Grant and Project Background

This work is a result of the 2017-2018 Institute of Museum and Library Services (IMLS) National Leadership/National Forum (NL/NF) grant project, Developing a Framework for Measuring Reuse of Digital Objects[1] (hereafter referred to as "Measuring Reuse"), which began on July 01, 2017 and ended June 30, 2018. The grant itself was written based on recommendations from the 2015 white paper, "Surveying the Landscape: Use and Usability Assessment of Digital Libraries", which identified multiple areas in the digital library assessment literature where more research and work would benefit the digital repository community. The Measuring Reuse project team were original members of the User Studies Working Group (USWG) of the Digital Library Federation (DLF) Assessment Interest Group (AIG), which wrote the "Surveying the Landscape" white paper. After the release of the white paper and in order to best tackle its recommendations, the USWG further divided itself into several subgroups. The members of that Content Reuse Subgroup became the project team[2].

In the following section, the literature review will provide an overview of the professional scholarship on assessment of grey literature reuse overall with a special interest on institutionally-hosted grey literature.

Literature Review

For the purposes of this paper, grey literature will be defined to include all of the document types identified in the GL Survey 2004, the results of which are listed on the GreyNet website ("Grey Literature - GreySource"). Previous research by the project team was conducted in conjunction with other members of the DLF AIG and culminated in the previously mentioned 2015 white paper, "Surveying the Landscape: Use and Usability Assessment of Digital Libraries," which identified three areas of focus within digital library assessment: use and usability studies, return on investment, and content reuse (Chapman et al. 2015). The white paper outlined the paucity of research surrounding digital content reuse, particularly with non-imaged-based objects and identified challenges in assessing content reuse. Further research by the Measuring Reuse project team in 2017-2018 explored more recent reuse assessment literature, some of which expanded the boundaries of the original research project's definitions of digital repository content and which is particularly relevant to assessing the reuse of grey literature. Dataset reuse, in particular, is the subject of new research, as is the reclassification of digital collections themselves as datasets (O'Gara et al. 2018).

Reuse of digital repository collections and items is an important assessment measurement because, unlike use, reuse shows engagement with collections and impact of digital repository resources in ways that are more valuable to stakeholders. Reuse data can be looked at to show the success of digitization and born-digital hosting efforts as well as to distinguish positive reuse, such as text-mining digitized books for a digital humanities (DH) project, from negative reuse, such as plagiarism. Buchanan and McKay discuss the use of plagiarism-checking software and manual examination to find instances of plagiarism from

[1] Developing a Framework for Measuring Reuse project website: https://reuse.diglib.org/

[2] Content Reuse Subgroup OSF Repository: https://osf.io/36npw/

items in digital library collections to analyze connections between open access publications, prestige, and plagiarism (2017). Similarly, Biagioni et al. assessed reuse of grey literature through citations to find evidence that open access publishing, such as that of conference proceedings, may increase the likelihood of scholarly publications citing grey literature (2017). Ferreras-Fernández et al. also found that open access theses in IRs received significant numbers of citations in academic research (2016). Finally, reuse assessment of technical and markup language documentation may also point researchers in the digital humanities to examples of other researchers producing similar projects using theirs as templates (Warwick et al. 2009).

There are challenges to assessing reuse, as identified previously by the research team and then further examined in literature regarding grey literature reuse. A lack of persistent identifiers in grey literature prevents capitalizing on the potential of altmetrics tools for more accurate reuse assessment (Cancedda and De Biagi 2017; Schöpfel and Prost 2016). Big data tools could allow grey literature repositories to better analyze (or help others analyze) their collection materials, but a dearth of sufficient resources and skilled personnel prevent widespread use among the grey literature community (Crowe and Candlish 2013). Existing infrastructure, such as the COUNTER standards for assessing use of electronic resources, focus primarily on subscription resources and do not filter out web robots, making them difficult to be modified for grey literature reuse assessment (Greene 2017). And open data can be much more difficult to track reuse of than data that requires permission for reuse (Parsons and Summers 2016).

Some possible solutions in the literature provide further areas of exploration for the project team. The use of linked data and publication codes such as ISBN, ISNI, ISIL, and ISSN could combat inconsistency in persistent identifiers for grey literature (Cancedda and De Biagi 2017). The COUNTER Robots Working Group is striving to recommend spider detection techniques for measuring the use of open digital content, including grey literature such as research data and institutional repository materials (Greene 2017). IRUS-UK, the Institutional Repository Usage Statistics United Kingdom service, is engaging in similar attempts by adding a short “Tracker Protocol” code to its repository that strips robots from usage data and converts the remainder to COUNTER-compliant statistics (MacIntyre and Jones 2016). The UK Data Service Communications team tracks the impact of research that uses their data in part by requiring researchers to register with the UK Data Service in order to use some datasets, and by requesting that researchers contact them with notifications of any publications that cite their collections (Parsons and Summers 2016).

Several new and recently proposed systems in the Open Science community in particular may provide needed infrastructure for reuse assessment of grey literature. The OpenAIRE-Connect project has exciting implications for reuse assessment. The “one-stop-shop” lets researchers "publish" portions of their research that are not published on their own traditionally, such as research methods, alongside their data and articles, and links all the related objects together easily. It also provides a notification service so researchers know when their research materials have been used and reused (Manghi et al. 2017). The Socionet researcher information system, too, proposes a semantic setup that facilitates researcher publication of the entire grey literature life-cycle. One function, Socionet Statistics, provides statistics on research artifact reuse (Parinov, Kogalovsky, and Lyapunov 2014).

Project Overview

In-depth explanations of the Measuring Reuse data collection and analysis methods are published elsewhere and will not be reiterated in full here (O’Gara et al., 2018; Kelly et al. 2018a). For clarity’s sake, the data collection methods will be covered briefly in the remainder of this section.

The research team employed several methods of data collection throughout the course of the Measuring Reuse project, including two surveys; one at the beginning of the project to gauge the digital repository community’s interest in certain topics (O’Gara et al. 2018); and the second at the end of the project that asked the community to indicate support and prioritize potential use cases, features, and functionality of a future toolkit for assessing reuse of digital repository materials (Kelly et al. 2018b).

The second method of data collection generated the bulk of the data from the project. The research team held three rounds of both in-person and virtual focus groups. The first two rounds consisted of two group sessions for each focus group venue, and the last round consisted of only one session for both the in-person and virtual focus groups, for a total of 10 focus group sessions over the course of the year-long project.

From the focus group data, the research team identified specific use cases from the focus group session notes, compiled them, then synthesized the individual use cases into generalized thematic use cases (Kelly et al., 2018b). Additionally, for the purpose of provoking discussion, the research team developed working definitions that differentiated between use and reuse. Use was defined as, "Discovering and browsing objects in a digital library, often described as "clicks" or "downloads," without knowing the specific context for this use". Reuse was characterized as "how often and in what ways digital library materials are utilized and repurposed. In this definition, we do know the context of the use". The research team fully expected these definitions to evolve, informed by the findings of the needs assessment, and actively solicited feedback from project participants and members of the digital repository community (Stein et al. 2018). It should be noted that as of the time of writing, the research team is still in the process of solidifying what to consider use and reuse or the differences between them.

In the next section, the data analysis process and findings for this paper are stated.

Focus Group Data

Data from the three rounds of in-person and virtual focus groups was re-analyzed in order to pull out specific examples and issues relating to content reuse of grey literature discussed by participants.

Reuse: What and Why

Throughout the focus groups, the project team sought to refine its definition of reuse. Of particular interest was the goal of expanding the definition beyond scholarly types of reuse (such as citations) to also include non-scholarly types of reuse that could be identified by altmetrics, such as web-based genealogy or news. Participants differentiated between scholarly and non-scholarly reuse of their assets, noting that both should be measured but would be weighed differently depending on institutional priorities and mission. Similarly, the impact measures of IRs, data repositories, and digitized special and archival collections might differ. In conjunction with each other, compelling reuse cases for digitized special and archival collections could potentially be used to convince administration that scholarly repositories should also be invested in at institutions that do not currently have or prioritize them.

The focus groups also served to provide additional examples of what the participants considered reuse that the project team had not previously identified in its foundational survey results. These included the following, which the research team has since extended with illustrations applicable to grey literature:

1. Repurposing metadata in aggregated consortial or membership repositories
    - Inclusion of data in aggregated data repositories (like The UK Data Service) would serve as reuse
2. Indexing records with abstract and record location information, and selling it on portable storage media
    - Health data could potentially serve as an example of this
3. Building a database from opened, digitized vital records and connecting that to living people
    - Seen frequently with genealogical data, such a database, e.g., Ancestry.com, would be a type of reuse

Collecting reuse data for grey literature could serve not only to demonstrate the value of these documents, but also to help understand where gaps exist in these collections. Some potential use cases for grey literature administrators include the ability to connect reuse of grey literature collections to community or industry values; and to generate reports of collection item reuse to share internally, making all staff aware of the types of materials being reused. Gathering data on what content was not reusable in order to develop

deaccessioning criteria was voiced as a potential benefit of assessing reuse of data collections. Repositories with datasets wished that they could see when Data Management Plans referred to their institutional data repositories so they could assess deposit and eventual reuse of these datasets. One participant noted that IRs were increasingly being looked to as official repositories for what they called "data bundles including images, hyperlinks, pdf files, and databases." Grant agencies were then using the IR to verify grant applicant claims, and focus group participants did not know how to measure the value of providing such a service.

Reuse Data Collection - Focus Group Use Cases and Examples

Methods for collecting grey literature reuse data include the following:

- citations (including Google Scholar) to grey literature in scholarly works and non-academic works (like Wikipedia)
- web analytics to trace reuse via URLs
- reports within a department and then sharing across institutional units
- mentions of grey literature on external blog posts
- historic data that's been transformed and extracted into a digital dataset
- consuming/sharing of IR objects and differences in how the same objects are consumed/shared from other repositories

Limitations to collecting reuse data, especially for grey literature, range from:

- lack of sharing of use/reuse data from aggregators and outside partners; scarcity of multi-genre grey literature aggregators
- deficit of standardized citation formats, particularly for vital and genealogical records, and incorrectly formatted citations of genealogical records
- absence of available reuse data as a result of poor search engine optimization and, therefore, diminished discovery
- separation of digitized data formats from their analog originals and any corresponding codebooks or documentation
- Outdated and colonial vocabulary in collection materials, metadata, and technical documentation

Several participants identified what they called the "problem" of electronic theses and dissertations (ETDs). Inadequate oversight or culture of attribution were indicated as problems leading to IR-hosted theses and dissertations with faulty citations, or no citations at all. These pose a challenge to digital repository staff hoping to find citations to their collections in IR-hosted ETDs.

Reuse data analysis

Once reuse data has been collected, further challenges may prevent stakeholders from analyzing the data in ways that can be beneficial to their institutions. Transforming quantitative data into meaningful stories that are suitable for consumption by a variety of audiences is difficult. Aggregated data in particular create complications for capturing individual-specific data, because aggregated data can obscure the important contextual information that reuse data emphasizes. Some aggregators, for example genealogical research databases, may have differing metadata priorities than others. Finally, legal limitations on data protection and privacy, such as the European Union General Data Protection Regulation (GDPR), may (and reasonably so) limit the ability of digital repository stakeholders to analyze reuse data without exposing individuals' personal information.

Reuse data reporting

Institutions collecting reuse data vary in how that data is shared. For public or federal grant-receiving institutions it may be shared via periodic reports to government officials. For other organizations, reports on new services and programs developed from collected data may be sent to their administration. Lack of best practices on what to include in these reports further complicates reuse assessment, however. There are no guidelines on how to measure quantity of reuse. For example, is repeated reuse of an object indicative of more meaningful or valuable objects than those with a single reuse? Are some types of reuse "better" or more "important" than others?

Facilitating reuse assessment

Focus group participants noted a number of common or recommended practices that should be followed in collecting reuse data. The ensuing examples are of particular interest to grey literature repositories.

First, as citations (or lack thereof) have been identified as barriers to assessing reuse of grey literature, bibliometrics should, and traditionally do, include quantitative measures like the number of times an object has been formally cited. In addition, qualitative measures, such as whether the reuse crosses disciplines, should also be noted, as interdisciplinary reuse may have a greater impact overall.

Several documentation features have also been identified as recommended practices. Documentation on how to measure information in referatory services (for example, a data repository aggregator that directs users to your collection like OpenDOAR or re3data) that also enables tracing a users' journey to or from the referencing service is needed. Guidelines on reuse of specific content types with special properties, like theses and dissertations, is needed. And finally, codes of practice for meeting the needs of indigenous and/or marginalized groups that may have specific reuse requirements is essential.. While there were not any specific examples of grey literature from indigenous communities discussed in the focus groups, cultural appropriation and the need for general codes of ethics for use and reuse of indigenous and marginalized cultural heritage materials emerged as an important theme that could be relevant to grey literature as well; for example, data sets of location data if sacred sites. Participants suggested some system features that could accommodate reuse measurement and, in particular, these best practices, consist of rights statements included with all digital materials; and auto-generated citations for digital library objects.

Discussion

The scope of the Measuring Reuse project was not limited to grey literature; the purview encompasses any and all materials in digital repositories hosted and curated by CHKOs, inherently including digital grey literature collections. However, because the boundaries of the project are so broad, many of the findings can be related to assessing reuse of grey literature; therefore, this paper only covers use cases and findings explicitly mentioning a grey literature document type ("Grey Literature - GreySource"). The majority of focus group participants did hail from academic libraries, despite the project team's best efforts, and the majority of data we collected on measuring reuse of grey literature materials involved ETDs and research data.

Akin to the breadth of materials in digital repositories, the focus group feedback regarding features and functionality for a future toolkit supporting measuring reuse appears overwhelming at first. Ultimately, the community needs can be distilled to a handful of high level use cases. During the data analysis phase of the project, the research team reduced the many specific and sundry user stories and examples into 18 generalized use cases which were divided into three categories: Data Collection, Analysis, and Reporting; Collection Development; and Privacy, Rights Management, and Ethics (Kelly et al. 2018). The overwhelming majority of participants indicated the need for best practices on all aspects of reuse assessment: data collection, analysis, reporting, and, albeit more implied than stated, how to *enable* reuse of their digital grey literature materials.

Considering the latter, traditional publications in the Information Age typically receive some sort of unique and persistent identifier, with Digital Object Identifiers (DOIs) being the most popular. Standardized identifiers such as DOIs, Archival Resource Keys (ARKs), and Handle System identifiers (Handles) facilitate assessment work, especially for reuse. The opportunity for the grey literature field to vastly improve data gathering practices on the impact of their collections is tangible, but unless the community is able to seize it through ubiquitous, concerted assignment of persistent unique identifiers, the chance will become completely unattainable Schöpfel and Prost argue in their 2016 article. The results of the Measuring Reuse project reinforce their conclusions and further add that this is the opportune moment for tracking and evaluating transformative uses, rather than passive consumption, of grey literature works as reuse data.

The increasing pattern of making grey literature available via IRs not only aids in discovery of the materials, but can also facilitate reuse and citation, since institutional

repository software may be set up to automatically generate Handles for each deposited item.

Depositing into IRs also helps address another discovery challenge for grey literature: the lack of a major aggregator, meaning a potential source of reuse data that is available to digital repository collections aggregated via services (such as Google Scholar or SHARE[3]) is not available to grey literature collections. Obstacles to the creation of a unified grey literature aggregator are exacerbated by the sheer breadth of grey literature content types. Certain types of grey literature, such as research data, have their own difficulties with discovery, aggregation, and citation, even with the proliferation of standards such as DataCite. Making collections of grey literature available via institutional scholarly repositories, which are often set up with OAI-PMH servers, makes it possible for aggregators to harvest their metadata, thus feeding grey literature records to genre or thematically focused harvesting services.

Discovery systems powered by metadata aggregation are often limited in scope to certain content types, such as DataCite for research data and the Global ETD Search by the Networked Digital Library of Theses and Dissertations (NDLTD) for ETDs. They have their own challenges, typically relying on institutional membership dues, which automatically precludes participation from financially-vulnerable organizations. A full exploration of the pros and cons of metadata aggregation discovery systems is out of scope for this paper.

Findings such as these are important because the research team can further synthesize the results to develop use cases, features and functionalities for the future toolkit that are designed and included expressly for the needs of the grey literature and institutional repository communities.

Conclusion

The Measuring Reuse grant was a year-long project wherein the research team conducted a needs assessment of the digital repository, spanning across CHKO domains. The intent of gathering and prioritizing the use cases, features, and functionalities is to build a toolkit that will guide digital library professionals in assessing the reuse of the collection materials. Grey literature collections pose unique challenges in assessing reuse, some of which may be ameliorated by using IRs to collect, preserve, and provide access to grey literature.

Perhaps the biggest limitation of the project was the overrepresentation of academic libraries among the focus group and survey participants despite strident efforts to recruit from a wide variety of institutions and CHKO domains. The special difficulties facing administrators and maintainers of grey literature repositories wishing to collect reuse data point to a new area of research and development.

The next logical stage of the project is to begin development of the reuse assessment toolkit, which the project team has applied for funding to pursue. Regardless of whether or not the grant request is successful, the research team intends to move forward with creating a reuse assessment toolkit. The goal of the toolkit will be to enable any digital repository professional to gather, analyze, and report on assessment data that show how their digital materials are innovatively and transformatively reused.

[3] http://www.share-research.org/

References

Biagioni, Silvia Giannini Stefania, Sara Goggi, and Gabriella Pardelli. 2017. "Grey Literature Citations in the Age of Digital Repositories and Open Access." *Grey Journal (TGJ)* 13 (1): 23–31.

Buchanan, George, and Dana McKay. 2017. "The Lowest Form of Flattery: Characterising Text Re-Use and Plagiarism Patterns in a Digital Library Corpus." In 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 1–10. https://doi.org/10.1109/JCDL.2017.7991570.

Cancedda, Flavia, and Luisa De Biagi. 2017. "International Identification and 'White and Grey Literature': Identities, Retrieval, Reuse and the Certainty of Knowledge While Sharing and Connecting Information." *Grey Journal* (TGJ) 13 (1): 32–36.

Chapman, Joyce, Jody DeRidder, Megan Hurst, Elizabeth Joan Kelly, Martha Kyrillidou, Caroline Muglia, Genya O'Gara, Ayla Stein, Santi Thompson, Rachel Trent, Liz Woolcott, and Tao Zhang. (2015), "Surveying the landscape: use and usability assessment of digital libraries," working paper, Digital Library Federation Assessment Interest Group, User Studies Working Group, December. https://doi.org/10.17605/OSF.IO/9NBQG.

Crowe, June, and J. R. Candlish. 2013. "Data Analytics: The next Big Thing in Information." *Grey Journal* (TGJ) 9 (3): 157–59.

Ferreras-Fernández, Tránsito, Francisco García-Peñalvo, José A. Merlo-Vega, and Helena Martín-Rodero. 2016. "Providing Open Access to PhD Theses: Visibility and Citation Benefits." *Program* 50 (4): 399–416. https://doi.org/10.1108/PROG-04-2016-0039.

Greene, Joseph. 2017. "Developing COUNTER Standards to Measure the Use of Open Access Resources." In *9th International Conference on Qualitative and Quantitative Methods in Libraries* (QQML2017), Limerick, Ireland, 23-26 May 2017. https://researchrepository.ucd.ie/handle/10197/8464.

"Grey Literature - GreySource, A Selection of Web-based Resources in Grey Literature." (n.d.). Retrieved October 7, 2018, from http://www.greynet.org/greysourceindex/documenttypes.html.

Kelly, Elizabeth Joan, Caroline Muglia, Genya O'Gara, Ayla Stein, Santi Thompson, and Liz Woolcott. 2018a. "Measuring Reuse of Digital Objects: Preliminary Findings from the IMLS-Funded Project." In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries - JCDL '18*, 351–52. Fort Worth, Texas, USA: ACM Press. https://doi.org/10.1145/3197026.3203871.

Kelly, Elizabeth Joan, Ayla Stein Kenfield, Genya O'Gara, Caroline Muglia, Santi Thompson, and Liz Woolcott. 2018b. "Setting a Foundation for Assessing Content Reuse: A White Paper From the Developing a Framework for Measuring Reuse of Digital Objects Project." https://doi.org/10.17605/osf.io/bqjvr.

MacIntyre, Ross, and Hilary Jones. 2016. "IRUS-UK: Improving Understanding of the Value and Impact of Institutional Repositories." *The Serials Librarian* 70 (1–4): 100–105. https://doi.org/10.1080/0361526X.2016.1148423.

Manghi, Paolo, Michele Artini, Claudio Atzori, Miriam Baglioni, Alessia1 Bardi, Sandro La Bruzzo, Michele De Bonis, et al. 2017. "OpenAIRE: Advancing Open Science." *Conference Papers: International Conference on Grey Literature* 19 (January): 107–12.

O'Gara, Genya, Liz Woolcott, Elizabeth Joan Kelly, Caroline Muglia, Ayla Stein Kenfield, Santi Thompson. (2018), "Barriers and solutions to assessing digital library reuse: preliminary findings", *Performance Measurement and Metrics*, Vol. 19 Issue: 3, pp.130-141. https://doi.org/10.1108/PMM-03-2018-0012.

Parinov, Sergey, Mikhail Kogalovsky, and Victor Lyapunov. 2014. "A Challenge of Research Outputs in GL Circuit: From Open Access to Open Use." Grey Journal (TGJ) 10 (2): 87–94.

Parsons, Rebecca, and Scott Summers. 2016. "The Role of Case Studies in Effective Data Sharing, Reuse and Impact." *IASSIST Quarterly* 40 (3): 14–19.

Schöpfel, Joachim, and Hélène Prost. 2016. "Altmetrics and Grey Literature: Perspectives and Challenges." In *Eighteenth International Conference on Grey Literature: Leveraging Diversity in Grey Literature*. GL Conference Series. New York, NY, USA: TextRelease. https://hal.univ-lille3.fr/hal-01405443/document.

Stein, Ayla, Santi Thompson, Elizabeth J. Kelly, Caroline Muglia, Genya O'Gara, and Liz Woolcott. 2018. "Developing a Framework for Assessing Use and Reuse: ALA Midwinter 2018." OSF. February 28. https://doi.org/10.17605/OSF.IO/YWH6K.

Warwick, Claire, Isabel Galina, Jon Rimmer, Melissa Terras, Ann Blandford, Jeremy Gow, and George Buchanan. 2009. "Documentation and the Users of Digital Resources in the Humanities." *Journal of Documentation* 65 (1): 33–57.

TIB
LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

"AS AN INFORMATION CENTRE FOR THE DIGITISATION OF SCIENCE AND TECHNOLOGY, OUR OBJECTIVE IS TO SUPPORT RESEARCHERS AT ALL STAGES OF THEIR WORK BY PROVIDING THEM WITH OUR SERVICES."

Professor Dr. Sören Auer

WWW.TIB.EU

AccessGrey Project – For Stakeholders in Grey Literature

Securing Open Access to Grey Literature for Science and Society

Persistent identifiers such as a DOI for a publication and an ORCID for an author/researcher can be approached from both the demand-side as well as supply-side of information. It appears however that the former attracts more attention. Here emphasis lies in the access to and preservation of research output. Yet, it is on the supply-side regarding the acquisition of research output that persistent identifiers may by the same token have influence in identifying and populating prospective data archives and repositories. This study will look at the influence persistent identifiers have in securing the acquisition of grey literature for public access.

The goal of this project is twofold. First, to carry out a survey within the grey literature community as to the opinions, uses, and applications of persistent identifiers. And second, to initiate a project geared to populate a new collection housed in the GreyGuide Repository by using the DOI as an incentive. Resources in Grey Literature (RGL) is as a generic, multidisciplinary collection that will serve for this purpose.

Using GreyNet's distribution channels and social media, stakeholders in the field of grey literature are invited to enter one or more of their publications in the RGL collection. Each new entry will receive a DOI minted by GreyNet International and further stored and preserved in the DataCite registry. Also, a system generated citation will be added to each new entry in order to facilitate record use. The types of grey literature documents eligible for entry in the RGL collection are numerous and can be found at http://www.greynet.org/greysourceindex/documenttypes.html.

Brief guidelines for record entry require that it be self-archived using the existing online-template and that both the metadata record and accompanying full-text document(s) are in English. An additional descriptive field does allow for entry in another language. And, a translation of the document can also be uploaded in the repository. Finally, it is understood that by submitting the metadata record and file(s), they become open access compliant under Creative Commons license CC-BY-SA.

The initial phase of the project commences in April 2019 and closes in September 2019. Records harvested during this period along with the results of the survey will be analyzed in its second phase. In the final phase, the project's outcome will be published. Results should indicate whether the AccessGrey Project be extended to other collections in the GreyGuide, and if this project would be of value to other communities of practice in the field of grey literature.

Project Team

Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands

Stefania Biagioni and Carlo Carlesi, NeMIS Research Laboratory, ISTI-CNR, Italy

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

Call for Posters

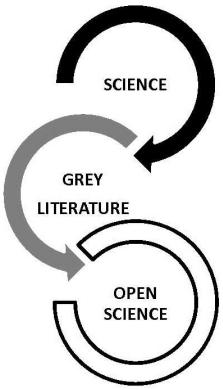
Title of Poster:	**Conference Topic(s):**
Author Name(s):	**Phone:**
Organization(s):	**Email:**
Postal Address:	**URL:**

Guidelines

Persons who seek to present a poster during GL21 are invited to submit an English abstract between 250-300 words. The abstract should describe the project, activity, information product or service. The abstract should likewise include a title, name(s) of the author(s) and their full address information. Abstracts are an important source of information available prior to the conference, which is accessible to both conference delegates and participants.

Related Conference Topics

- ☐ **Research Sharing relies on Open Data**
- ☐ **Open Source Software benefits Grey Literature**
- ☐ **Publishing Grey Literature opens up the Review Process**
- ☐ **Confronting Obstacles and Challenges to Open Access**
- ☐ **Open Resources for Education in Library and Information Science**
- ☐ **Open Science Principles promote the field of Grey Literature**
- ☐ **Other Related Topic:**



Due Date, Format, and Poster Prize 2019

Timely registration is a guarantee for your placement on the conference program. The conference venue is able to accommodate a limited number of physical posters (maximum size *A0 (841 x 1189 mm / 33.1 x 46.8 in). Posters* will be mounted on display panels allowing for optimal viewing. Abstracts in MSWord should be emailed to conference@textrelease.com on or before **September 20th 2019**. Those submitting poster abstracts will receive verification upon receipt accompanied by further guidelines for posters, submission of PowerPoint slides, Conference Registration, etc.

Those presenting conference posters are eligible for the Poster Prize 2019 that will be awarded during the GL21 Closing Session. Posters will be judged on their innovative content, relevance to the conference, graphic design, and presentation.

TextRelease
GL21 Program and Conference Bureau

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

Author Information

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

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.

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

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).

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

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. Email: muglia@usc.edu

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.

Plato L. 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

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

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

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. Email: liz.woolcott@usu.edu

Notes for Contributors

Non-Exclusive Rights Agreement

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

Submission Requirements

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

REFERENCE GUIDELINES

General

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

Specific

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

Examples

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

Crowe, J., G. Hodge, and D. Redmond (2010), Grey Literature Repositories: Tools for NGOs involved in public health activities in developing countries. – In: Grey Literature in Library and Information Studies, Chapter 13, pp. 199-214. – ISBN 978-3-598-11793-0

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

Review Process

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

Journal Publication and Article Deposit

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

the grey Journal

15 YEARS

Victorine van Schaick Prize 2008

International Journal on Grey Literature

'DATA LIBRARIANS SPEAK OUT ON GREY LITERATURE'

The data librarian: myth, reality or utopia? 7
Silvia Giannini and Anna Molino, Institute of Information Science and Technologies, ISTI-CNR, Italy

Research Data Management: What can librarians really help? 23
Yuan Li, Willow Dressel and Denise Hersey, Princeton University, United States

Data Management and the Role of Librarians 31
Plato L. Smith, Sara Gonzalez, and Jean Bossart, George A. Smathers Libraries; University of Florida, United States

Librarians' Role in GAO Reports 39
Meg Tulloch, U.S. Government Accountability Office, GAO, United States

Published electronic media are becoming Grey 45
Yui Kumazaki, Satoru Suzuki, Masashi Kanazawa, Katsuhiko Kunii, Minoru Yonezawa, and Keizo Itabashi, Japan Atomic Energy Agency

Measuring Reuse of Institutionally-Hosted Grey Literature 51
Ayla Stein Kenfield, University of Illinois at Urbana–Champaign; Elizabeth Kelly, Loyola University New Orleans; Caroline Muglia, University of Southern California; Genya O'Gara, Virtual Library of Virginia; Santi Thompson, University of Houston; Liz Woolcott, Utah State University, United States

Colophon 2

Editor's Note 5

On The Newsfront

AccessGrey Project - Securing Open Access to Grey Literature for Science and Society 60

GL21 Conference Call-for-Posters 61

Advertisements

Ebsco Library, Information Science & Technology Abstracts with Full Text (LISTA) 4

ISTI-CNR, Institute of Information Science and Technologies, Italy 6

TIB, German National Library of Science and Technology 59

Author Information 62

Notes for Contributors 63