

## Contents

### Special Issue: Research Data Services

**Guest Editors:** Michael Witt and Wolfram Horstmann

### Guest Editorial

- Libraries tackle the challenge of research data management 3  
*Wolfram Horstmann and Michael Witt*

### Articles

- Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs 5  
*Pierre-Yves Burgi, Eliane Blumer and Basma Makhoul-Shabou*
- A brief assessment of researchers' perceptions towards research data in India 22  
*Manorama Tripathi, Mahesh Chand, S. K. Sonkar and V. K. J. Jeevan*
- 'We have to make an effort with it': Exploring the use of stages to help understand the personal information management needs of humanities and social science doctoral students managing dissertation information 40  
*Amber L. Cushing and Odile Dumbleton*
- Managing research data at an academic library in a developing country 51  
*Shamin Renwick, Marsha Winter and Michelle Gill*
- Building professional development opportunities in data services for academic librarians 65  
*Suzanna Conrad, Yasmeen Shorish, Amanda L. Whitmire and Patricia Hswe*
- Open access and open data on natural disasters 81  
*Marisa R. De Giusti, Gonzalo L. Villarreal, Carlos J. Nusch, Analía V. Pinto and Ariel J. Lira*
- Building a Research Data Management Service at the University of California, Berkeley: A tale of collaboration 89  
*Jamie Wittenberg and Mary Elings*
- Quality evaluation of data management plans at a research university 98  
*James E. Van Loon, Katherine G. Akers, Cole Hudson and Alexandra Sarkozy*
- Team-based data management instruction at small liberal arts colleges 105  
*Ryan Clement, Amy Blau, Parvaneh Abbaspour and Eli Gandour-Rood*

### Abstracts 119

### Aims and Scope

IFLA Journal is an international journal publishing peer reviewed articles on library and information services and the social, political and economic issues that impact access to information through libraries. The Journal publishes research, case studies and essays that reflect the broad spectrum of the profession internationally. To submit an article to IFLA Journal please visit: [journals.sagepub.com/home/ifl](http://journals.sagepub.com/home/ifl)

## IFLA Journal

Official Journal of the International Federation of Library Associations and Institutions  
ISSN 0340-0352 [print] 1745-2651 [online]

Published 4 times a year in March, June, October and December

---

### Editor

Steve Witt, University of Illinois at Urbana-Champaign, 321 Main Library,  
MC – 522 1408 W. Gregory Drive, Urbana, IL, USA. Email: [switt@illinois.edu](mailto:switt@illinois.edu)

### Editorial Committee

Barbara Combes,  
School of Information Studies, Charles Sturt University, Wagga Wagga, NSW Australia. Email: [bcombes@csu.edu.au](mailto:bcombes@csu.edu.au)

Milena Dobрева-McPherson,  
Library, Information and Archive Sciences Department, University of Malta, Malta. Email: [Milena.dobрева@um.edu.mt](mailto:Milena.dobрева@um.edu.mt)

Anne Goulding,  
School of Information Management, Victoria University of Wellington, New Zealand. Email: [Anne.goulding@vuw.ac.nz](mailto:Anne.goulding@vuw.ac.nz)

Ben Gu,  
National Library of China, Beijing, People's Republic of China. Email: [bgu@nlc.cn](mailto:bgu@nlc.cn)

Dinesh Gupta,  
Vardhaman Mahaveer Open University, Kota, India. Email: [dineshkg.in@gmail.com](mailto:dineshkg.in@gmail.com)/[dineshkumargupta@vmou.ac.in](mailto:dineshkumargupta@vmou.ac.in)

Mahmood Khosrowjerdi,  
Allameh Tabataba'i University, Tehran, Iran. Email: [mkhosro@gmail.com](mailto:mkhosro@gmail.com)/[mkhosro@atu.ac.ir](mailto:mkhosro@atu.ac.ir)

Jerry W. Mansfield (*Chair*)  
Congressional Research Service, Library of Congress, Washington, DC. Email: [jmansfield@crs.loc.gov](mailto:jmansfield@crs.loc.gov)

Ellen Ndeshi Namhila (*Governing Board Liaison*)  
University of Namibia, Windhoek, Namibia. Email: [enamhila@unam.na](mailto:enamhila@unam.na)

Seamus Ross,  
Faculty of Information, University of Toronto, Toronto, Canada. Email: [seamus.ross@utoronto.ca](mailto:seamus.ross@utoronto.ca)

Shali Zhang,  
University of Montana, Missoula, Montana, United States. Email: [Shali.Zhang@mso.umt.edu](mailto:Shali.Zhang@mso.umt.edu)

### Publisher

SAGE, Los Angeles, London, New Delhi, Singapore, Washington DC and Melbourne.

Copyright © 2017 International Federation of Library Associations and Institutions. UK: Apart from fair dealing for the purposes of research or private study, or criticism or review, and only as permitted under the Copyright, Designs and Patents Acts 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the Publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency ([www.cla.co.uk/](http://www.cla.co.uk/)). US: Authorization to photocopy journal material may be obtained directly from SAGE Publications or through a licence from the Copyright Clearance Center, Inc. ([www.copyright.com/](http://www.copyright.com/)). Inquiries concerning reproduction outside those terms should be sent to SAGE.

Annual subscription (4 issues, 2017) Free to IFLA members. Non-members: full rate (includes electronic version) £304/\$561. Prices include postage. Full rate subscriptions include the right for members of the subscribing institution to access the electronic content of the journal at no extra charge from SAGE. The content can be accessed online through a number of electronic journal intermediaries, who may charge for access. Free e-mail alerts of contents listings are also available. For full details visit the SAGE website: [sagepublishing.com](http://sagepublishing.com)

Student discounts, single issue rates and advertising details are available from SAGE, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP, UK. Tel: +44 (0) 20 7324 8500; e-mail: [subscriptions@sagepub.co.uk](mailto:subscriptions@sagepub.co.uk); website: [sagepublishing.com](http://sagepublishing.com). In North America from SAGE Publications, PO Box 5096, Thousand Oaks, CA 91359, USA.

Please visit [journals.sagepub.com/home/ifl](http://journals.sagepub.com/home/ifl) and click on More about this journal, then Abstracting/indexing, to view a full list of databases in which this journal is indexed.

Printed by Henry Ling Ltd, Dorset, Dorchester, UK.



# Libraries tackle the challenge of research data management

**Wolfram Horstmann**  
University of Goettingen, Germany

**Michael Witt**  
Purdue University, USA

International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 3–4  
© The Author(s) 2017  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216688787  
journals.sagepub.com/home/iff



Two special issues of the International Federation of Library Associations' *IFLA Journal* are required to adequately represent the quality submissions of a call for papers for research data activities in libraries. This underlines that libraries are indeed tackling the challenge of research data management (Witt and Horstmann, 2016). Libraries are not doing this alone: they collaborate closely with the researcher as well as with other organizations such as IT services, research offices and research funders.

The central motivation for the two special issues of *IFLA Journal* is to provide concrete examples of research data activities in libraries. The first issue already indicated that these examples align with a certain pattern of questions. One or more of these questions are being answered in a given article:

1. Requirement analysis: How exactly can libraries help researchers?
2. Skills development: What skills do library staff need to provide this help?
3. Service provision: How do libraries design and deliver this support?
4. Data literacy: How can libraries foster informed and professional research data activities?

These questions are addressing different dimensions: (a) *geographically* they can be at local, at a national or even international scale, (b) *disciplinarily* they can be generically spanning subjects, be interdisciplinary or relate to a single discipline, and (c) *functionally* they might range from a simple reference services to indexing and archiving or even complex analytic services.

Depending on which question is posed and how the geographic, disciplinary and functional dimensions

are addressed, a plethora of research data activities in libraries is hypothetically possible. This second issue shows another selection of what is actually being done. With respect to the question of requirement analysis, we find two national approaches: one that is looking at sustainability of research outputs in Switzerland and another that is assessing researchers' needs in India. A perspective of institutional requirements is presented in studies about personal information management behaviour in the social sciences and humanities in Ireland and another perspective looks at research data practices on a campus in the Caribbean. With respect to requirement analysis, we are thus providing national as well as local cases and provide exemplars of different parts of the world. How the second question, skills development in libraries, can be addressed, is exemplified with a paper from the Association of College and Research Libraries about a survey to identify the types of librarians most needed. Examples of service provision that answer the third question range from a highly constraint-driven case, in which a flood required the setup of an information service on local water quality, to an institutional example and an example addressing a national funder's policy, the National Science Foundation requirement for research data management plans in the United States. Finally, the fourth question, how libraries can foster data literacy, is represented by a case in which libraries are collaborating to develop a shared service that provides training and outreach activities for researchers.

---

## Corresponding author:

Wolfram Horstmann, Goettingen State and University Library,  
University of Goettingen, Goettingen, Germany.  
Email: horstmann@sub.uni-goettingen.de

Research data services in libraries are a growing and evolving field. The contributions in the two issues indicate that tackling the challenge is required and executed along several dimensions: geographic, disciplinary and functional. Thereby, libraries successfully exploit their long-established strengths of forming a multi-dimensional system among different actors at all levels, from local to global, spanning all subjects and comprehensively regarding the diverse aspects of information management – to carefully choose where libraries can be helpful in supporting the research data of today to become the cultural and scientific heritage of tomorrow.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **Reference**

Witt M and Horstmann W (2016) International approaches to research data services in libraries. *IFLA Journal* 42(4): 251–252. DOI: 10.1177/0340035216678726.



# Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs

International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 5–21  
© The Author(s) 2016  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216678238  
journals.sagepub.com/home/iff



**Pierre-Yves Burgi**

University of Geneva, Switzerland

**Eliane Blumer**

University of Geneva, Switzerland

**Basma Makhoul-Shabou**

Geneva School of Business Administration, Switzerland

## Abstract

In this article, the authors report on an ongoing data life cycle management national project realized in Switzerland, with a major focus on long-term preservation. Based on an extensive document analysis as well as semi-structured interviews, the project aims at providing national services to respond to the most relevant researchers' data life cycle management needs, which include: guidelines for establishing a data management plan, active data management solutions, long-term preservation storage options, training, and a single point of access and contact to get support. In addition to presenting the different working axes of the project, the authors describe a strategic management and lean startup template for developing new business models, which is key for building viable services.

## Keywords

Business model, data life cycle management, preservation of digital data, research data management services, value proposition canvas

Submitted: 16 May 2016; Accepted: 28 September 2016.

## Introduction

The ongoing technological developments in digital content give rise to new ways to collect, capture, store, manipulate and transmit large volumes of data and stimulate communication and collaboration between researchers. Research data, one important class of digital content, is defined by the Organisation for Economic Co-operation and Development (OECD, 2007) as 'factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings'.

The information contained in research data thus generally constitutes valuable assets for researchers.

The US National Science Board (NSB, 2005) distinguishes three types of data: observational, computational or experimental. This distinction is crucial in the choices made for archiving and preserving information: In the cases where observations will not repeat (for instance in astronomy, earth science, etc.) data are consequently unique and priceless. Conversely, for computational data, and if comprehensive information about the model is available (typically contained in the metadata), preservation in a long-

## Corresponding author:

Eliane Blumer, University of Geneva, Rue du Général Dufour 24, Geneva, 1204, Switzerland.  
Email: eliane.blumer@unige.ch

term repository may not be necessary because the data can be reproduced. In between, data from experiments that can be accurately reproduced need not be stored indefinitely; yet in practice it may not be possible to reproduce precisely all of the experimental conditions, and/or the costs of reproducing the experiment are prohibitive. Apart from these kinds of data, others can be regrouped under the term of 'record' (Borgman, 2015) covering general documentation (of government, business, etc.), to which we might add electronic laboratory notebooks and scanned artefacts in the research context.

Commonly, researchers do not pay attention in particular to the importance of the life cycle of their data, a concept that encompasses many facets (as exemplified in CEOS, 2011), some of which, as illustrated previously, might depend on the relevant discipline. The data life cycle begins with the acquisition of raw data. Subsequently these data are analysed to test hypotheses and validate models. The results of these studies usually lead to publications, an important milestone in the life of those processed data, as this can correspond to the end of the project and/or its funding. At this stage, data are often neglected, typically abandoned on the researcher's self-bought storage, improvised servers and/or institutional servers (in the best case), to be quickly forgotten by the failure to promote them in other contexts. Disciplines with huge data amounts, i.e. life sciences, depend on their respective institutional or external proposed infrastructure. Also, data management know-how is usually transferred among research teams and does not come from an outside service with data expertise. Even if back-ups at a minimum are advised and used, loss of data could be a serious issue independent of research discipline.

Preservation of digital information is nowadays rather well mastered (Schumacher et al., 2014), even if it remains challenging to keep bit streams unaltered over long periods of time (Rosenthal, 2010). Indeed, information is subject to the second law of thermodynamics, that is, the disorder (or entropy) tends to increase permanently, with the result that any data necessarily become corrupted with time. This implies that to preserve information this natural disorder must be prevented, for instance through the ability of matter to perform calculations to restore a corrupted information. This mechanism is limited, however, by the required energy, and therefore a selection of information becomes a necessity. It is established that the exponential growth of production data, the multiplication of computer uses, the obsolescence of objects, etc. will increase demand for energy (Bihouix, 2014).

While preservation is key to saving research data, it is by itself useless if the different stages of the data life cycle are omitted. Researchers should indeed make explicit at an early stage of the project what they intend to do with their data during the project and afterwards, whether they plan to share them (or keep them confidential), how long they need to preserve them, etc. They should also document their datasets so as to be able in the future to understand them, for themselves but also for other researchers mainly within the same discipline, but not only (Goodman et al., 2014; Wallis, 2014; Wallis et al., 2013). Such data management plans (DMP) are increasingly asked for by funding agencies and research institutions.

One important part of the DMP concerns data formats and their eventual obsolescence. In principle before ingesting data into a repository, researchers should comply with recommended formats so that migration cycles can be ensured in the long term. Such compliance is however not always possible for various reasons, which include performance and volumetry, and also simply researchers' willingness to do it. In such cases, data description should be further developed so as to become self-described, a complex and time-consuming process, which also has its limits (Lee, 2011). Researchers will rarely comply with doing it, precisely because they prefer using their time for tasks they deem more pertinent for advancing their research work. To support preservation of any data regardless of format, Klindt and Amrhein (2015) define two levels of preservation: passive and active. At the passive level, data are preserved at the bit-level only, providing researchers with the possibility of storing data at minimal cost while complying with the editors' and/or funders' constraints. Obviously, this is not what information professionals are seeking, but this corresponds to a field demand. Conversely, at the active level all the necessary preservation mechanisms are applied to ensure data remain interpretable throughout the migration cycles, yet this necessitates more intensive preparatory work from the data producers. In both cases, preserved data should remain as accessible as possible for further uses, and not kept as dark archives. Access should consequently be facilitated, opened as much as possible following standard formats, with the aim of serving other researchers in the near and far future. Periodic value assessments must be performed so that the costs for preserving bit streams match the intrinsic value of the preserved datasets, given that this value can lessen with time.

How many researchers have in mind all those management issues when acquiring their (increasingly) large volumes of data? Preservation of digital



information is clearly a complex and costly process but cannot anymore be circumvented. Researchers need to be at the least rendered aware of these issues through training and assistance dispensed by information professionals and/or data curators to maximize digital curation of their data. Digital curation is to be understood in simple terms and following the Digital Curation Centre's (2007) definition as 'the active management and appraisal of data over the life cycle of scholarly and scientific interest' (see also Kim et al., 2013, for more extensive definitions). The Swiss project that we describe in this paper has such an ambition to cover all main issues contained in data life cycle management (DLCM), taking for granted that preservation of digital information over the long term represents the core element of this ambitious enterprise.

In the following sections we expose the main methodological considerations, the need analysis for DLCM guidelines and tools in an academic environment, and present the Swiss DLCM project along with its related dimensions and focus on long-term preservation.

## Methodological considerations

This section presents the general methodological approach, the main data collection and analysis techniques pertaining to the DLCM field and an overview of the major results expected by the end of the project. Given the exploratory nature of this work, the realization of these objectives is based on a qualitative approach. Regarding the gathering of data and their analysis, three steps were performed:

1. A large document analysis aimed mainly at performing an exhaustive academic and professional literature review with which the DLCM project can build a set of best national and international practices in terms of DMP and policies.
2. Semi-structured interviews to confirm the researcher's needs; more than 50 such interviews were performed with researchers from different departments in a variety of disciplines (see below), providing a deep knowledge of the diverse research data practices in Swiss universities.
3. The analysis of the data, collected both from the literature review and semi-structured interviews, to lead to an overview of the main considerations that should be taken into account to offer a general framework/model for rational research DLCM with related guidelines, tools, competences and

technologies to allow its effective operationalization in the academic environment.

## Expected results

The following outcomes are expected by the end of the project:

- a DMP adapted to the Swiss research communities and compliant with the main funders (e.g. Swiss National Science Foundation, Horizon 2020, etc.);
- guidelines for researchers and information professionals on research data management (including guidelines for data and metadata dissemination) to encourage the use of best practices in data management;
- a policy template for the higher education institutions (HEI), which can be used to establish sound policies to manage research data with all related issues (such as data privacy, intellectual property, storage costs, etc.);
- a National Portal on research data management with recommended tools and practices for researchers and information professionals;
- a toolbox for building SIP, AIP and DIP OAIS packages from subsets of research data (including graphic user interfaces adapted to different tools);
- a prototype of a scalable OAIS-compliant infrastructure;
- business models for the delivery of viable long-term preservation services;
- an inventory of existing data management training modules, including expert networks in collaboration with other Swiss and international projects;
- specific data management training modules and teaching modules for integration into Library Information Systems (LIS) courses;
- publications on data management in proceedings and journals.

## Research validity

To ensure the quality of our studies, the DLCM project team investigated scientific validity by:

- establishing the state of the art regarding the work performed in data management with local, national and international field experts;
- being aware of the main theories, standards, projects in link with research data governance;
- sharing and publishing intermediate results at peer-reviewed international conferences and journals.

**Table 1.** Summary of the need analysis interviews.

Institution	Number of interviews	Disciplines
University of Geneva (UNIGE)	8	Theology, Informatics, Linguistics, German, Cognitive Neuroscience, Educational Sciences, Geomatics, Archaeology, Vulnerability, Political Sciences, Medicine (Child Cancer Research)
ETH Zurich (ETHZ)	8	Biosystems Science and Engineering, Seismic Networks, Sociology, Consumer Behaviour, Quantum Optics Group, Scientific Computing/Photon Science, Physics
University of Lausanne (UNIL)	15	Social Medicine, Social Sciences, Digital Humanities, Genomics, System Biology, Bioinformatics, Public Health, Imaging and Media Lab, Cancer Research
EPF Lausanne (EPFL)	5	Transport and Mobility, Quantum Optoelectronics, Supramolecular Nanomaterials and Interfaces, Audiovisual Communication Laboratory, Virology and Genetics.
University of Basel (UNIBAS)	7	Biology research (Biozentrum, 2), Biology (Core facilities, 2), Molecular Psychology, Public Health (STPH), Digital Humanities
University of Zurich (UNIZH)	5	Law Science, Biology/Microscopy, Biology/Proteomics, University Hospital, Geosciences
Total	49	

To check practical relevance, the DLCM team met several professionals in public institutions and private companies with the intent of collecting their feedback experience in the field of long-term data preservation.

The interdisciplinary character of this work (data curation, information sciences, computer engineering, researchers' practices, etc.) brings a variety of competences that have an important impact on the quality of the outcomes. To ensure, however, a fluent collaboration between those several partners and researchers from different disciplines, a terminological glossary was defined and shared in regular meetings, workshops, etc.

### Need analysis

As mentioned before, an exploratory need analysis was conducted in order to get a deeper knowledge about the researchers' needs and the solutions in place. For this, every partner institution conducted semi-structured interviews during two months (September and October 2014). The structure of these interviews contained four major parts, namely: (1) initial data and workflow, (2) analysis and data exploration, (3) publication, archiving and long-term data management, and (4) research data in the future: challenges, risks, perspectives. Table 1 presents the compilation of all interviewed disciplines.

As Table 1 shows, interviewed disciplines are various. In a second step, every interview was entered into a summary table, organized by discipline and dispatched in the four main interview parts with a finer classification based on similarities in the answers when applicable (see Table 2 in Appendix).

The main outcomes of those four interview parts are the following.

*Initial data and workflow.* First and foremost, generally no formal DMP are being used, unless the funding instances require it at the time of the project application. As a consequence, data loss and description difficulties are often mentioned as a main issue.

Another challenge, concerning data description and storing, is that there are no common guidelines between disciplines and thus data exist in a plurality of formats (vector, video, audio, image, text, graph, raw bit streams, and so on), proprietary or/and open, depending on the software application. Those formats are tailored to the needs of the research projects (and team) and rarely in view of data preservation.

Even if common description standards exist in some internationally well-organized disciplines, such as in Geography, in Humanities and Educational Sciences, no standards are used, with sometimes even the question of what exactly represents a 'datum', which surprisingly can in some cases remain difficult to answer.

As for data storing, in most of the cases self-bought improvised servers are used, as institutional IT departments are often slower in providing solutions than the rate of data produced by researchers. Independent of research discipline, researchers are aware of the need to back up their data, as loss of data is a recognized worrying issue. However, the organization of backups is not always seen as the exclusive task of the institution, but also of the individual.

*Analysis and data exploration.* The biggest challenge in this part seems to be the freedom within data



organization. Every research project, department, sometimes even researcher has her/his own habits and it is difficult to harmonize any of them. As for sharing and preserving, at the moment, Dropbox<sup>1</sup> remains a useful solution, even if it is subject to the American Law for privacy and copyright.

#### *Publication, archiving and long-term data management.*

The notion of long-term preservation is generally absent and to the question of how long should data be kept their answers are either elusive or indicate that 10 years might be representative of such a long-term time period. Further questions without answers include what should be the best strategy for long-term preservation and whether there are any existing guidelines or rules.

Another difficulty is related to copyright, as there is no clear view on who the owner of specific data is or what has to be done in order to publish data in respect with the law in force.

*Research data in the future: Challenges, risks, perspectives.* Even if the opinions of the importance for long-term preservation differ, one point is mentioned regularly: researchers indicate that there is no adequate answer to the question of what could and should happen to research data after the end of a project and/or the successful publication of the scientific results. Often, data vanish in the wilderness of offices or on more or less well-cared servers. While the survival of these data is of concern, their proper management is cited as a more difficult matter.

Interviewed researchers repeat the importance of an incentive for data sharing. Such an incentive might be, for example, data citation or new ways of peer-review (as mentioned for instance in Tenopir et al., 2015). In order to do so, data sets have to be permanently identified by a DOI. Having a supplementary person who manages their data would be appreciated as well as the possibility to get more cited in the literature through data citation.

Finally, disciplines with huge amount of data see a risk as well in the rising costs of storage place with the related question of who will pay for it.

*Main needs.* An analysis of the main trends shows that data management clearly depends on the institutional strategies and/or research habits in the specific considered disciplines. Most of the interviewed disciplines are confronted with various challenges during different stages of their research projects. As a matter of fact, today's research cycle depends on ad hoc solutions as well as specific habits within research departments. Also, based on the above-mentioned

results, the following researchers' needs have been identified as development axes to be further proposed as deliverables:

- guidelines and support for helping researchers properly manage their data;
- ad hoc data storage, computing and analysis solutions;
- solutions for active data management with storage of research progress based on periodical snapshots of defined datasets;
- development and/or maintenance of online research data long-term repositories.

## **The DLCM project**

This section begins with an overview of the five major axes that compose the project and then focuses on the preservation topic and the viability of the proposed services.

### *The DLCM project organization*

The DLCM project started in September 2015 and is planned to last three years. It is organized into five tracks (or subprojects), headed by a specific partner institution, and which focus on: (1) Guidelines and policies; (2) Active research data; (3) Long-term preservation; (4) Consulting, training and teaching; and (5) Dissemination. The following sections briefly present the objectives of each of these axes.

*Guidelines and policies.* This part aims at defining guidelines based on an exhaustive academic and professional literature reviews with which the DLCM project can build a set of best national and international practices in terms of DMP and policies applicable in the HEI. Rather than dictating closed policies, a research data management policy template has been elaborated and presented at the steering board of one of the highest HEI political instances, i.e. the Swissuniversities research delegation (whose members are Rectors/Presidents from nine HEI, and the directors of the Swiss National Science Foundation and of the Commission for Technology and Innovation). The aim of such a top-down approach is to promote/recommend the development of policies based on a common national framework, while being flexible enough to allow adaptations for local specificities.

*Active research data.* Active research data refers to the stage of data collection, processing, and analysing early in the life cycle. Based on the researchers'

interviews, three main distinct scenarios could be identified: (1) 'Single Endpoint', consisting either in raw data, or a number of data processing steps before archiving datasets; (2) 'Open-Ended Work', which is representative of active data management that has no definite end and whose data can continuously evolve, and/or even refer to other data contained in the cloud (linked data) with obvious implication for properly archiving them; and (3) 'Times Series Data', data continuously collected, possibly pre-processed and which have to be archived before being further processed. These three scenarios give a hint of how complex can be the articulation between active data management and preservation, as in most case it is not a linear process but rather one with a large number of cycles and sub-cycles. Thus this part aims at providing to a broad spectrum of researchers concrete technological solutions and best practices to properly manage their active data according to the three identified scenarios, with particular focus on collecting, processing and analysing those data. Along this line, three main domains are being considered: (1) Electronic Laboratory Notebooks (ELN) and Laboratory Information Management Systems (LIMS) solutions and support; (2) Virtual Research Environment (VRE) for Digital Humanities; and (3) a range of working solutions for scientific facilities and software storage solutions for active data in variety of disciplines (with however a focus on life sciences).

**Long-term preservation.** This axis aims at establishing a bridge between active data and long-term preservation and publication solutions. For doing so, we consider well-established concepts, such as the Curation Life Cycle (Ball, 2010; Pouchard, 2015) and the OAIS Model (ISO 14721:2012). While this axis is described in more detail below, the main parts composing it are: (1) the establishment of a gap analysis of the repositories currently in use in the project partners' institutions; (2) the development of the required toolbox for building the OAIS information packages (SIP-AIP-DIP); (3) the design and tests of a prototype of a scalable OAIS-compliant infrastructure; (4) the elaboration of a business model with associated costs for preserving (large amounts of) data on the long term.

**Consulting, training and teaching.** The main targets for this axis are: (1) the establishment at national level of consulting, training and teaching within the DLCM field with both general and specific needs; (2) the creation of a consulting service at each partner institution, coordinated by a central desk; and (3) in parallel, the research data management knowledge will

be integrated in Bachelor's and Master's courses in the field of Information Science, so that the freshly trained librarian can ensure the sustainability of this knowledge for the future generations.

**Dissemination.** This axis aims at promoting the results of the DLCM project and establishing relevant contacts and collaborations with other institutions and projects not directly involved or linked to this one. For this, several sensitizing campaigns are planned during the whole project, including the organization of one-day annual workshops targeting researchers and information experts. Participating to international conference and workshops, as well as publishing in proceedings and journals constitute other important activities of this track.

### **Preservation**

Preservation of research data in the long term is one of the core objectives of the project by the fact that it serves at establishing a bridge between active data management and the mean to reliably store data (in the long term) so as to give access to the dataset that resulted into a publication, or simply allow to reuse it for furthering research.

**Gap analysis.** The project partnership is made of seven institutions, some of them with solutions for long-term preservation, but mainly used for archiving publications. Consequently, we initiated the project with a gap analysis to assess the different stages of maturity and compliance of these solutions with the Open Archival Information System (OAIS) standard. Also, a large panel of technologies was encountered among the participating institutions: the University of Zurich (UNIZH) operates a repository based on the Eprints software.<sup>2</sup> The Swiss Institutes of Technology are using Rosetta from Ex Libris<sup>3</sup> (in Zurich – ETHZ), and a repository based on Invenio, with the intention of linking it to Zenodo<sup>4</sup> in the future (in Lausanne – EPFL), while the Universities of Geneva and Lausanne are both using Fedora Commons<sup>5</sup> for their repository. To perform this assessment with most objectivity, we applied the evaluation tools conceived of by the Digital POWRR team<sup>6</sup> (Preserving digital Objects with Restricted Resources – Schumacher et al., 2014). The methodology defines five functional areas – storage and geographic location, file fixity and data integrity, information security, metadata, and file formats – and four levels of digital preservation – protect, know, monitor, and repair of data. This results in an evaluation grid (Figure 1), which represents the intersection of the Digital Curation Centre's digital

curation life cycle<sup>7</sup> steps and the OAIS Reference Model (ISO 14721:2012) specifications. From this grid it is apparent that no institution completely fulfils the required compliance, though some (e.g. ETHZ) are close to the target. Moreover, most repositories do not comply with the ‘Reliable, Long-Term Bit Preservation’ feature, which is a requisite for long-term preservation, yet a challenging one. Indeed, it is a known fact that disk storage based for instance only on RAID<sup>8</sup> systems does not detect all kind of errors and is subject to ‘silent data corruption’ (Bairavasundaram et al., 2008; Rosenthal, 2010), which thus requires other higher-level checksum mechanisms to guarantee errors are promptly detected.


*Toolbox for building OAIS information packages (SIP-AIP-DIP).* At some stage in their research work, researchers will want to select a specific dataset from the active data storage area for ingesting it into a longer-term storage repository. Motivations for researchers to accomplish this step (passing from the active to semi-active or passive status) are various and mainly include: publishers asking for sustainable access to the data used to get the results in the published paper, need for a Digital Object Identifier (DOI) (or any other permanent identifier) for openly sharing the dataset, or simply archiving data at the end of a research project. To make this step as flexible and transparent as possible, researchers should have the possibility to push their selected data from the active storage area into a data repository (semi-active status) and/or a long-term preservation system (passive status) ‘by a click’. Following the OAIS Reference Model requires the preparation of a Submission Information Package (SIP). In the process, several micro services are usually activated and include checksum calculations, format detection, virus check, etc. with at the end the assembly of an Archival Information Package (AIP). Tools exist to provide such micro services and for automatizing AIPs (for instance Archivematica, Curator’s Workbench, iRods, etc.). From the consumer side, specific user interfaces must be designed to allow users to retrieve information from the repository and/or from the archived AIP. In this latter case, this will be accomplished by generating a Dissemination Information Package (DIP) delivered to the user who has requested the information.

*Scalable OAIS-compliant infrastructure.* Another important constraint of the OAIS Reference Model concerns physical storage. Storage must be highly redundant, self-correcting, resilient, and must consist of multi-copies geographically distributed, while maintaining integrity and traceability of the stored information.

SAFE PLN is on ongoing project for building such an OAIS-compliant based on the LOCKSS infrastructure<sup>10</sup> (Maniatis et al., 2005) and regroups several partners in Europe and Canada.<sup>11</sup> LOCKSS is based on the Byzantine fault tolerance concept, originally known as the Byzantine Generals’ Problem (Lamport et al., 1982). To be able to defend against Byzantine failures, in which components of a system fail with symptoms that prevent some components of the system from reaching agreement among themselves, redundancy is required to form a voting poll. For example, for tolerating two node failures, at least seven nodes are necessary. The current implemented version offers a limited storing capacity (less than 1TB) over seven geographically distant nodes. Also, and even if in the future SAFE PLN could be extended to tens of terabytes (with many issues yet to be solved, such as bit stream transfers between nodes), to manage even much larger volumes, typically of the order of several petabytes, we have to consider more scalable architectures. One particularly promising solution is based on Archivematica for the ingest part, iRods for automatizing the archival storage (mainly for managing two replications, one being geographically distant), and Fedora Commons for exposing the DIP through a dedicated user interface. The Zuse Institute Berlin (ZIB) has developed such an infrastructure for managing more than five petabytes of data on disk (Klindt and Amrhein, 2015). Other solutions relying on object-based storage (such as Ceph<sup>12</sup>) are currently being investigated. Among the still many open questions yet to solve to render the infrastructure scalable, one is technical: how to check and replicate petabytes of data possibly distributed over hundreds of millions of files? Another issue, dealt with in the following section, is related to the financing of such large-scale infrastructures on the long term.

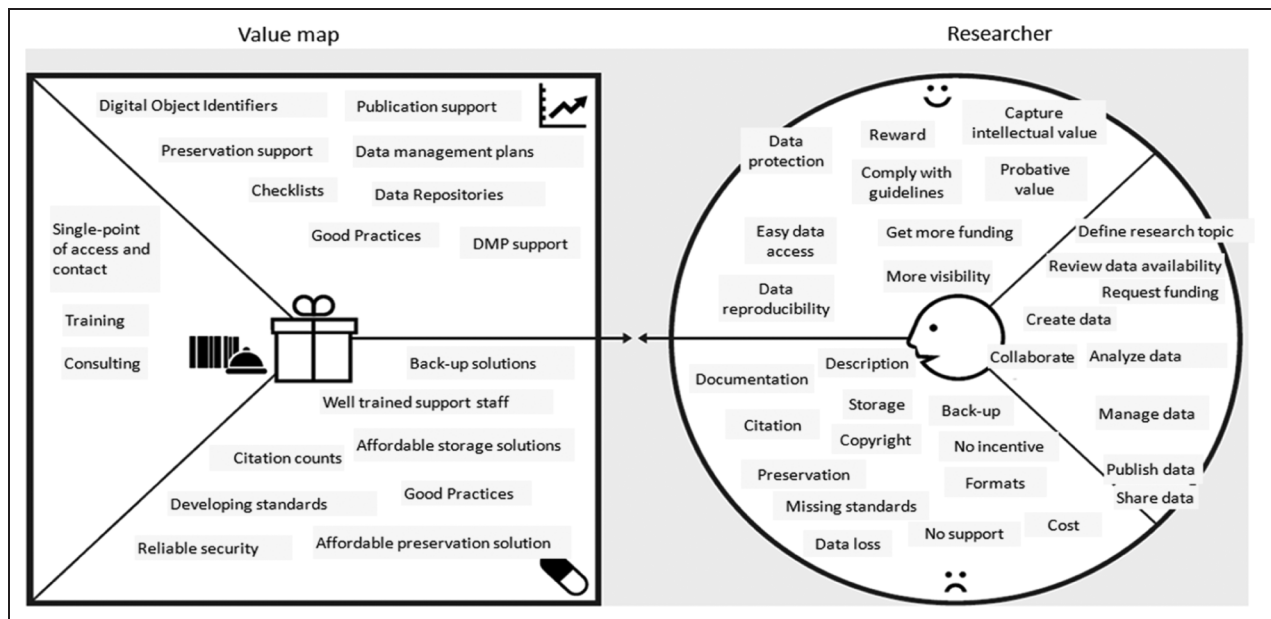
### Viability

Given the DLCM project is limited in time and resources (as all projects are), a viability methodology based on strategic management and lean startup templates for guaranteeing the sustainability of the services in the long term has to be considered. For this, the Business Model Canvas (BMC, see Osterwalder and Pigneur, 2010) and Value Proposition Design methodology (Osterwalder et al., 2014) are being applied. The methodology provides tools to simulate how an institution can potentially make, supply and earn value, while the BMC provides the essential elements for developing the accompanying business models. One important component of the Value Proposition is to place the customers (i.e. researchers) at

	Copy	Fixity Check	Virus Scan	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Auto SIP Creation	Public Interface	Auto DIP Creation	Auto AIP Creation	Reliable, Long-Term Preservation	Redundancy	Geographically Dispersed Data Storage Model	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation	Cost
digitalpowrr.nl/edu																								
ETH Zürich	(X)	X	X		X	X	X	X	X	X	X	(X)	X	X	X	X	(X)	X	X	X		X	(X)	
UNIGE Archive/ Fedora 3	X				X	X	X	X	X	X	X	X	X	X		X		X				X	Free	
UNIL Serval/ Fedora3	X	(X)	X	X	X	X	X	X	X	X	X	X	X	X		X	(X)	X	(X)		X	X	Free	
UZH/ZORA Eprints	(X)		X		X		X	X	X	X		X			(X)	X	X				X	X	Free	
Zenodo (EPFL)	(X)	X	X	(X)	X			X	X	X	(X)	X	(X)	(X)	X	X		X		X	X	X	Free	

**Figure 1.** Evaluation grid for Swiss partner institutions.  
(X) indicates partial or similar function. Based on the tool grid of Digital Powrr (2014).<sup>9</sup>





**Figure 2.** A Value Proposition Canvas, based on Pigneur and Osterwalder (n.d.),<sup>13</sup> Corresponding to three LTP services (value map at left) with on the right (circle) the Gains, Pains, and Customer Jobs indicated.

the centre, and develop services in relation to their real needs. This should in principle avoid developing services that will not be used. Yet, this does not solve the sustainability issues, which are related to the business model. Currently, the project members familiarize themselves with such a methodology (which per se is not an intuitive approach) and start to institutionalize it, with the help of a business analyst (who is also a member of the project team).

**Value Proposition and Business Model Canvases.** Within the Value Proposition Canvas (see Figure 2) one typically describes the considered customer practices, including their everyday tasks (job), potential gains and incumbent pains. In a further step, one attempts to find gain creators, pain relievers and transform them into products and services. This way, every created service answers an existing and preliminarily identified need and provides a viable value in reducing pains. One example of value proposition corresponding to three service developments for long-term preservation (LTP) is illustrated in Figure 2.

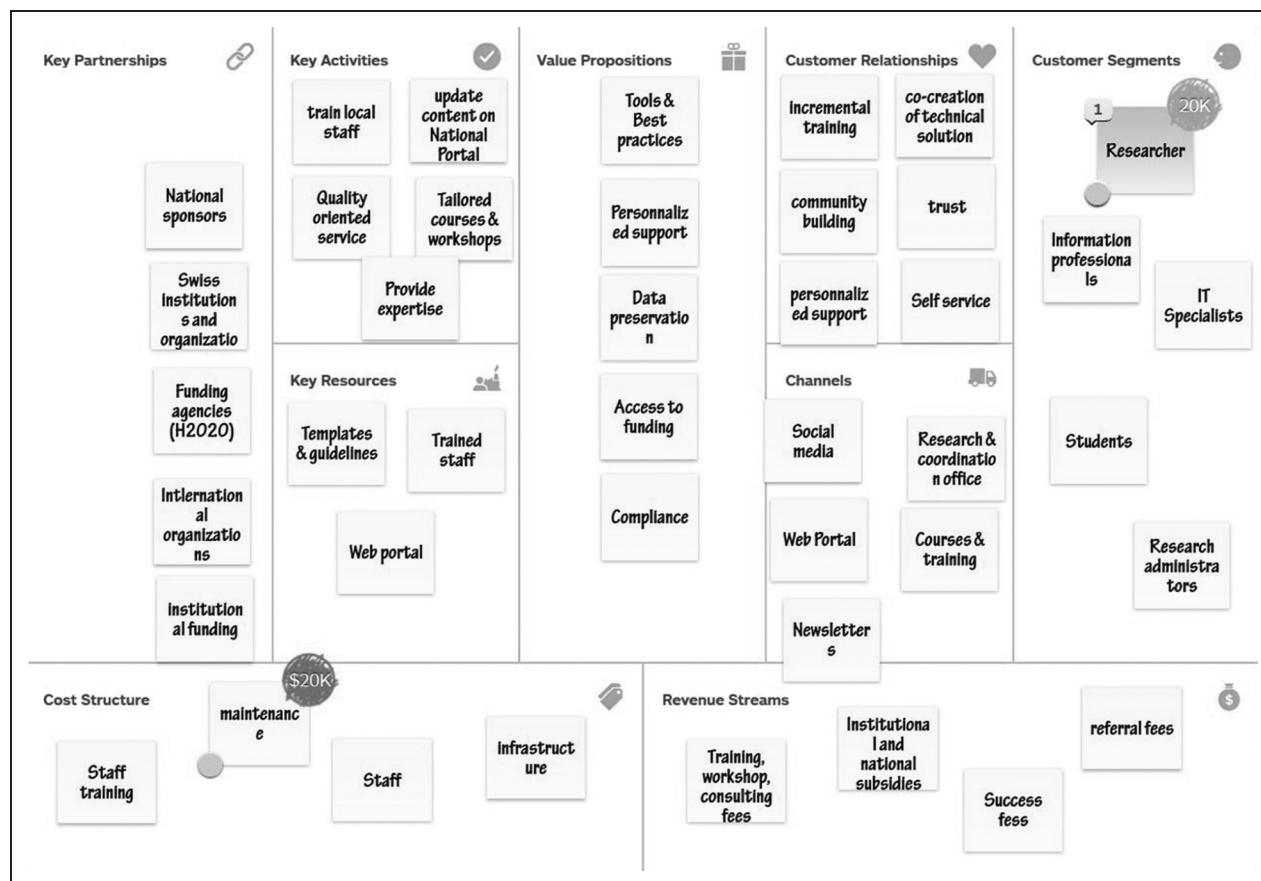
In a further step, the BMC is completed by first specifying the ‘Customer Segment’ part, by integrating the Value Proposition Canvas into the ‘Value Proposition’ field, and then by filling the other sections (customer relationship, key activities, key resources, key partnerships, revenue streams, and cost structure) in order to describe and finalize the targeted services. An example of BMC corresponding to a National Portal service is illustrated in Figure 3. In

this case several customer segments have been identified (specified in the outer right column of the canvas).

### Cost models

One key component of the BMC is Cost Structure (see Figure 3), which should match the Revenue Streams. The issue of the LTP costs has been grappled with through a number of international studies, most notably the KRDS (Keeping Research Data Safe) project (Beagrie et al., 2008, 2010). This project focused on UK case studies to draw conclusions on the main cost drivers, and on this basis a costing framework containing two major elements, economic and service adjustments, was proposed. Economic adjustments consist of inflation (e.g. salary), deflation (e.g. storage media), depreciation of the assets (historical/purchase costs across its useful life) and cost of return for financing and investment. Service adjustments consist of costs related to (1) staff, (2) acquisition, disposal and ingest phases, (3) archive storage, preservation planning and data management. Interestingly, back-up and long-term file storage represent only a tiny portion of the total costs. According to this study, acquisition and ingest count for up to 42%, access for 35%, and storage for 23%. Even more extreme is a typology of the activities that considers data creation, which counts for 73%, with curation and storage accounting respectively for 24% and 3%.

Apart from the KRDS study, another notable work in this domain is the comparative study of several cost



**Figure 3.** Business Model Canvas for a National Portal service.

A population of 20,000 Researchers have been evaluated as potential customers (blue circle), and maintenance costs of the infrastructure was estimated to 20,000 US\$ (red circle).<sup>14</sup>

models realized in the context of the 4C European project.<sup>15</sup> However, among all these works, one model in particular retained our attention, the Total Cost of Preservation (TCP) from the UC Curation Center (Abrams et al., 2012), which encompasses the full economic costs associated with long-term preservation of digital assets. This model takes into account 11 preservation activities (System, Services, Servers, Staff, Producers, Workflows, Content types, Storage, Monitoring, Interventions, Oversight), and considers two price models: pay-as-you-go and paid-up. If at first it seems considering such a comprehensive cost model is beneficial to determining financing load, in practice it brings more questions than it answers them. Indeed, to be applicable, first one has to know in advance the number of 'Producers'. And as the number of producers increases, the associated costs diminish through economy of scale, which is to say that a producer does better to come late to benefit from cheaper prices. Second, evaluating all 11 preservation activities is far from being trivial and/or necessarily reliable. Consequently, the DLCM project team developed a simplified TCP cost model independent of the number of producers, based on the 'pay-up price model'

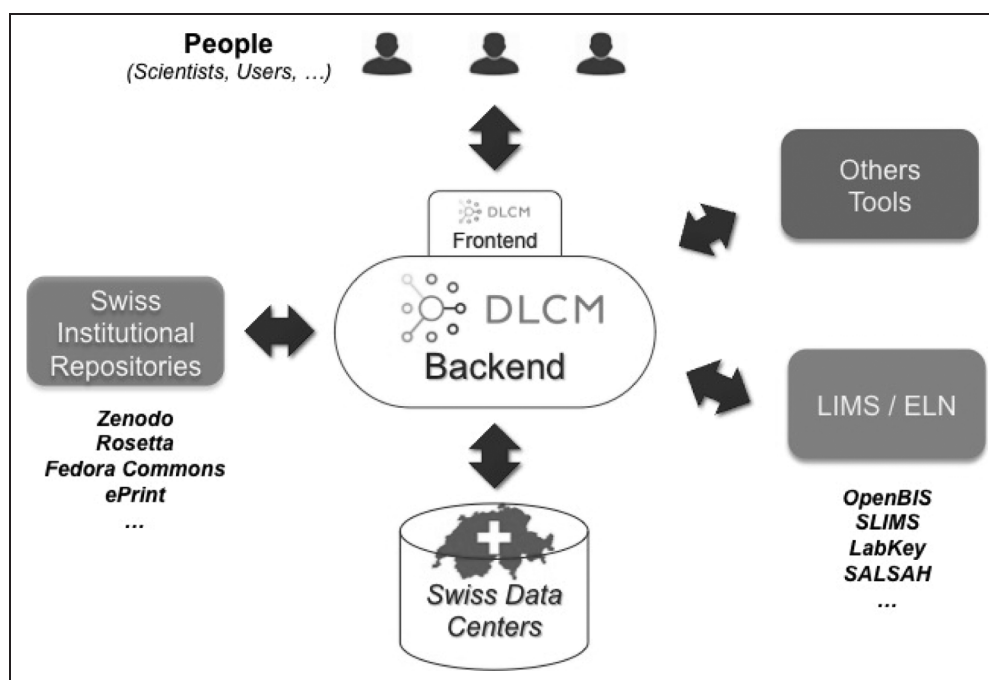
without assuming any investment return, but which takes into account a global cost for maintaining the infrastructure, including staff stipends and an annual percentage rate of price decrease for the hardware. It is acknowledged that other important costs related to services such as creation and curation of datasets exist, but to evaluate basic costs associated to safe storage of information, it was assumed that other services could be billed separately in function of the level of services customers ask for. And for researchers, having reasonable costs for storing their data in the long term is clearly an important incentive for avoiding letting data end their life on inappropriate storage infrastructure (such as USB keys or private computer).

### *Main realizations, future steps and challenges*

This Swiss DLCM project started in September 2014 with a pre-study, followed by a first implementation phase initiated in September 2015. The concrete outcomes after one year are the following:

**Guidelines, DMP and policies.** A website with relevant resources, tools and guidelines for researchers has





**Figure 4.** The DLCM LTP Ecosystem.

been set up ([www.dclm.ch](http://www.dclm.ch)). A data management policy template for guiding institutions in establishing research data management policies has been written and presented to the Swissuniversities Research Delegation, which is one of the highest Swiss assemblies for HEI. This top-down process is still ongoing with the intent of homogenizing the Swiss policies in this domain in coordination with the Swiss National Science Foundation, whose mission is to finance research at national level through competitive grants. Finally, a data management plan checklist has been elaborated based on previous experience of two partner institutions (EPFL and ETHZ), with the idea of collaborating with DMPOnline<sup>16</sup> developed by the UK Digital Curation Centre to produce an online tool adapted to the Swiss needs.

**Active research data.** This part of the project is particularly complex, as it must deal with the working environment of researchers from various disciplines. Also, one of the main axes concentrates on LIMS and ELN, a topic of concern for researchers willing to document their research processes and data. A relevant market and gap analysis concerning Swiss LIMS software (Slims,<sup>17</sup> OpenBIS,<sup>18</sup> ViKM<sup>19</sup>), and other tools mainly used in life science (e.g. Labkey<sup>20</sup>), has been accomplished along with video tutorials to facilitate their use. In a further step how these tools can be applied to other fields than life science will be assessed. For Digital Humanity (DH), a virtual research environment (Salsah/Knora) is currently being evaluated on several DH projects.<sup>21</sup>

**Long-term preservation.** During the first year of the project a gap analysis using the methodology (Schumacher et al., 2014) was conducted in order to identify the relevant gaps in institutional repositories for becoming OAIS compliant (see section ‘Preservation’). The main outcome of this work is an inventory of the tools and repositories expected to constitute the future ecosystem of the LTP national service (see Figure 4). Next steps will include specifications for interoperability between institutional repositories and research tools, and of the SIP, DIP and AIP (in progress), which will lead to the development of APIs for interfacing the national service.

**Consulting, training and teaching.** Computer engineers and information specialists in different institutions have already dispensed research data management (RDM) workshops and trainings to librarians. In parallel, an extensive catalogue of RDM training modules is being created along with a need analysis for Bachelor training in Swiss Universities of Applied Sciences. Another important aspect being worked on is the development of a generic consulting service for Swiss HEIs. One expected outcome of this initiative is a central and focal coordination desk integrated into a large network of trained RDM specialists representing their respective academic institutions and scientific communities.

**Dissemination.** At this early stage of the project, one of the main outreach milestones was to organize an

annual national event on the RDM thematic targeting the Swiss researchers community. This type of event (the first having taken place in November 2016) is intended to gather international keynote speakers, and representatives from academic direction boards (e.g. Swiss National Science Foundation, Swissuniversities, etc.), and leave space for discussions through parallel workshop sessions whose topics typically encompass the main DLCM facets (DMP, active data management tools, infrastructures, policies, etc.).

**Business models.** The DLCM project faces different challenges, such as the viability and sustainability of the proposed national services, collaboration between Swiss data centres and other RDM initiatives, and decision making of the HEI steering boards. The viability of the services has been worked on since the very beginning of the project in order to develop a business culture among the partners (which in an academic environment is not given), and to get enough time to analyse, validate the hypotheses and adapt iteratively the business models by interviewing the users (or 'customers') on their usages of the new services.

## Conclusions

Targeting rational and optimized management of research outcomes is challenging, particularly when it means being able to use securely, fluently, promptly and durably research data during the whole research life cycle. Researchers face several demanding situations regarding data management, but tend to use ad hoc solutions. Based on face-to-face interviews with researchers from a variety of disciplines, the present work identified the major needs, which reduce to: guidelines and support in managing data, computing and analysis solutions, and solutions for storage of research progress and repositories. At the term of the project, Swiss-wide services are expected to respond to those needs and beyond them through the implementation of RDM best practices. Key to the viability of the future proposed services is the elaboration of sound business models, and for that the project members have been imparted knowledge on a methodology, usually confined to economists, based on two concepts, the Business Model Canvas and the Value Proposition Design (Osterwalder and Pigneur, 2010; Osterwalder et al., 2014). Further advantages of applying this methodology stem from the fact that the DLCM project members develop a common culture and use the same vocabulary beyond their institutional limits and collaborate in a creative and constructive way.

The experience gathered so far has shown that there are additional challenges in addition to providing suitable and sustainable services. Indeed, the national dimension of the project, and the attributes and needs of specific partner institutions require good coordination and proper governance in an academic context in which researchers are not necessarily accustomed to the application of binding rules. Consequently, finding the right incentives for having researchers comply with a minimum set of RDM policies will determine the success of this enterprise.

The near future will show how the proposed services and the project results will be used. In any case, the increasing pressure from publishers and funding agencies necessitates the sharing of best practices and resources across all HEIs to meet these challenges.

## Acknowledgements

We would like to express our sincere gratitude and appreciation to Swissuniversities' programme SUK P2, which mandates our project, as well as to the whole DLCM project team for their fruitful collaboration.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Notes

1. Dropbox ([www.dropbox.com](http://www.dropbox.com)) is a file hosting service, operated by the American company Dropbox, Inc., which offers services like cloud storage and file synchronization.
2. <http://www.eprints.org>
3. <http://www.exlibrisgroup.com/category/RosettaOverview>
4. <http://zenodo.org>
5. <http://fedorarepository.org>
6. <http://digitalpowrr.niu.edu>
7. <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
8. Redundant Array of Independent Disks
9. <http://digitalpowrr.niu.edu/tool-grid/>
10. <http://www.lockss.org>
11. <http://www.safepln.org>
12. <http://ceph.com/ceph-storage/object-storage/>
13. <https://strategyzer.com/canvas/value-proposition-canvas?url=canvas/vpc>
14. Based on Strategyzer Tool by Pigneur and Osterwalder (n.d.), <https://strategyzer.com/>
15. See D3.1—Evaluation of Cost Models and Needs & Gaps Analysis, available at: <http://4cproject.eu>
16. <https://dmponline.dcc.ac.uk/>

17. <http://www.genohm.com/slims>
18. <https://wiki-bsse.ethz.ch/display/bis/Home>
19. <https://www.vital-it.ch/research/software/ViKM>
20. <http://www.labkey.com>
21. For more information on Salsah/Knora and related DH projects, see <https://dhlabs.unibas.ch>

## References

- Abrams S, Cruse P, Kunze J, et al. (2012) *Total Cost of Preservation (TCP): Cost Modeling for Sustainable Services*. UC Curation Center, California Digital Library.
- Bairavasundaram LN, Goodson GR, Schroeder B, et al. (2008) An analysis of data corruption in the storage stack. In: *FAST'08: 6th USENIX conference on file and storage technologies*, Boston, MA, USA, 22–27 June 2008, pp. 223–238.
- Ball A (2010) *Review of the State of the Art of the Digital Curation of Research Data (version 1.2)*. Bath, UK: University of Bath.
- Beagrie N, Chruszcz J and Lavoie B (2008) *Keeping Research Safe. A Cost Model and Guidance for UK Universities. Final Report*. Salisbury: Charles Beagrie Limited.
- Beagrie N, Lavoie B and Woollard M (2010) *Keeping Research Safe 2. Final Report*. Salisbury: Charles Beagrie Limited.
- Bihouix P (2014) *L'Âge des low-tech*. Paris: Seuil.
- Borgman CL (2015) *Big Data, Little Data, No Data*. Cambridge MA: MIT Press.
- CEOS (2011) *Data Life Cycle Models and Concepts*. Working Group on Information Systems and Services Data Stewardship Interest Group. 26 September 2011. CEOS.WGISS.DSIG.TN01.
- Digital Curation Centre (2007) *What is Digital Curation?* Available at: <http://www.dcc.ac.uk/digital-curation/what-digital-curation> (accessed 15 September 2016).
- Digital Humanities Lab Basel. Available at: <http://dhlabs.unibas.ch/#/> (accessed 10 October 2016).
- Digital POWRR (2014) *Preserving Digital Objects with Restricted Resources*. Available at: <http://digitalpowrr.niu.edu/> (accessed 10 October 2016).
- Goodman A, Pepe A, Blocker AW, et al. (2014) Ten simple rules for the care and feeding of scientific data. *PLoS Comput Biol* 10(4). DOI:10.1371/journal.pcbi.1003542.
- Kim J, Warga E and Moen WE (2013) Competencies required for digital curation: An analysis of job advertisements. *International Journal of Digital Curation* 8(1): 66–83.
- Klindt M and Amrhein K (2015) One core preservation system for all your data. No exception! In: *12th International conference on preservation of digital objects*, Chapel Hill, NC, USA, 2–6 November 2015, pp. 101–108.
- Lamport L, Shostak R and Pease M (1982) The Byzantine generals problem. *ACM Transactions on Programming Language and Systems* 4(3): 382–401. Available at: <http://research.microsoft.com/en-us/um/people/lamport/pubs/byz.pdf> (accessed 17 November 2016).
- Lee CA (2011) A framework for contextual information in digital collections. *Journal of Documentation* 67(1): 95–143.
- Maniatis P, Roussopoulos M, Giuli TJ, et al. (2005) The LOCKSS peer-to-peer digital preservation system. *ACM Transactions on Computer Systems* 23(1): 2–50.
- NSB (2005) *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*. Report no. NSB-05-40. National Science Foundation. Available at: <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf> (accessed 10 October 2016).
- OECD (2007) *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Available at: <http://www.oecd.org/science/sci-tech/38500813.pdf> (accessed 10 October 2016).
- Osterwalder A and Pigneur Y (2010) *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken, NJ: John Wiley and Sons.
- Osterwalder A, Pigneur Y, Bernarda G, et al. (2014) *Value Proposition Design: How to Create Products and Services Customers Want*. Hoboken, NJ: John Wiley and Sons.
- Pigneur Y and Osterwalder A (n.d.) *Strategyzer*. Available at: <https://strategyzer.com/canvas> (accessed 10 October 2016).
- Pouchard L (2015) Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation* 10(2): 176–192.
- Rosenthal DSH (2010) Bit preservation: A solved problem? *International Journal of Digital Curation* 5(1): 134–148.
- Schumacher J, et al. (2014) *From Theory to Action: 'Good Enough' Digital Preservation Solutions for Under-Resourced Cultural Heritage Institutions*. A Digital POWRR White Paper for the Institute of Museum and Library Services. DOI: <http://hdl.handle.net/10843/13610>.
- Tenopir C, Dalton ED, Allard S, et al. (2015) Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE* 10(8): e0134826. DOI:10.1371/journal.pone.0134826.
- Wallis JC, Rolando E and Borgman CL (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE* 8(7) DOI:10.1371/journal.pone.0067332.
- Wallis JC (2014) Data producers courting data reusers: Two cases from modeling communities. *International Journal of Digital Curation* 9(1): 98–109.

## Author biographies

**Pierre-Yves Burgi** is Deputy CIO, in charge of the Information, Communication, and Educational Technology (ICET) Unit at the Information System Department of the University of Geneva. The ICET unit focuses on Teaching and Learning, Research, and Collaboration. He received a BS degree in Computer Engineering from the Swiss Federal Institute of Technology, Lausanne, Switzerland, and a

PhD in Computer Science from the University of Geneva, Geneva, Switzerland. His doctoral studies were followed by a five-year post-doctoral period in Visual Neurosciences, mainly with the Smith-Kettlewell Eye Research Institute, San Francisco, CA. From 1997 to May 2003, he was with the Advanced Microelectronics Division of the Swiss Center for Electronics and Microtechnology, Neuchâtel, Switzerland, where he contributed to the development of fast silicon vision systems. Since June 2003, he has headed the ICET unit, University of Geneva. Pierre-Yves Burgi's interests include digital copyrights in education (he is an active member of the Swiss Competence Center in Digital Law), research data life cycle management (he is directing the Swiss Data Life Cycle Management CUS P-2 project), and digital humanities (he is part of the Bodmerlab team).

**Eliane Blumer** is Research Associate at the University of Geneva and is coordinating the Swiss Data Life Cycle Management CUS P-2 project and helping to implement a local research data management service within the university. She is also part of the continuing education programme committee of BIS (Bibliothek Information Schweiz), the Swiss Union for Information Professionals. She received her BSc in Information Science at the Geneva

School of Business Administration and is currently enrolled in a Master's Programme in Business Information Systems at the University of Applied Sciences and Arts Northwestern Switzerland.

**Basma Makhoul Shabou** is a full Professor at the University of Applied Sciences and Arts Western Switzerland. She specializes in Archival Science at the Department of Information Studies. She holds a PhD in Information Sciences obtained from the University of Montréal (School of Library and Information Sciences), and a Master's degree in Public Records Management. Her research focuses on Infonomics and information business value; information quality measurement; records appraisal and the value of electronic records; information governance. She is involved in the RIC project (<http://www.recordsinthecloud.org/>) collaborating with North American researchers. She leads a European project on Information Governance Maturity in European Public Administrations (InterPARES <https://interparestrust.org/trust/aboutus>). She is also working on Research Data Governance (DLCM <https://www.dlcm.ch/datacycle/>). She was senior records manager at National Archives of Tunisia and has realized many important records management projects in Montreal, Tunis and Geneva.

## Appendix

**Table 1.** Interview results: DLCM obstacles and challenges.

	Initial data and workflow	Analysis and data exploration	Publication, archiving and long-term management	Research data in the future: challenges, risks, perspectives
Theology	No DMP Back-up	Freedom of methodology to organize data	Freedom of organization	No local storage Copyright issues
Informatics	No DMP	Lack of description	Lack of description	No storage for sharing and versioning Incentive
Linguistics	No DMP Description rules Back-up	Freedom of methodology to organize data	Conservation rules Data is obsolete	Format issues
German	Lack of documentation No DMP Description rules Back-up	Freedom of methodology to organize data	Conservation rules	Citation rules No archive Incentive
Cognitive Neuroscience	DMP for US-projects Data loss Local back-up and storage	–	Conservation rules No reproducibility	Versioning Standardization
Educational Sciences	No DMP Local storage and back-up	Freedom of methodology to organize data	Copyright issues	Institutional storage on local level with access rules Standardization
Geomatics	No DMP	–	–	Need for a data archive
Archaeology	Description rules Data storage No DMP	Freedom of methodology to organize data	Conservation rules Copyright issues	Interdisciplinarity
Vulnerability	No DMP No description rules Local storage	–	Data loss Data publication	–
Political Sciences and International Relations	Naming Preserving	–	–	Data repository Twice production of same data Incentive
Medicine (Child Cancer Research)	–	–	–	–
Quantum Optoelectronics	Local storage	–	–	Link between data and publication Cost Guidelines Storage Incentive

(continued)



**Table 1.** (continued)

Initial data and workflow		Analysis and data exploration	Publication, archiving and long-term management	Research data in the future: challenges, risks, perspectives
Transport and Mobility Laboratory	Local back-up	Dropbox	Conservation rules	Checklists Institutional repository Incentive Lack of documentation Copyright issues Standardization Metadata
Supramolecular and Interfaces Laboratory	No DMP Back-up	–	–	Guidelines Reproducibility Cost National perspective Metadata
Genetics	Standardization in description	Dropbox	–	Incentive Reproducibility Best practices Incentive
Audiovisual communication laboratory	–	–	–	Cost National perspective Metadata
UZH	Missing standards	–	–	Incentive Reproducibility Best practices Incentive
Unispital ZH	Missing standards	Ethics	No long-term preservation strategy	Cost Centralized storage with access management Incentive
Proteomics	–	Centralized sharing tool	–	DMP Storage cost Long-term preservation Cost Incentive Politics
Photon Science/Scientific Computing	–	–	–	Easily accessible web front end for data publication Infrastructure
Biosystems Science and Engineering	–	–	Long-term preservation Conservation rules	Automatic snapshots Storage facilities
Seismic Networks	–	–	Sharing (others can see, but not process)	

(continued)



**Table 1.** (continued)

	Initial data and workflow	Analysis and data exploration	Publication, archiving and long-term management	Research data in the future: challenges, risks, perspectives
Social Sciences, Modelling, Simulation	No DMP	–	–	Indexing large volumes of data including hidden annotations Framework for management support Capture defined states of data periodically Differential record of algorithms' versions Comprehensive Swiss cloud solution Real time streaming of big data Reliable citation counts Real time data input Support and staff Managing data all in one Local storage Incentive Periodical snapshots of specific data Persistent link to published data Time Formats Incentive Repository Interoperability Long term storage is not funded –
Consumer Behaviour	No DMP Data loss	–	–	
Quantum Optics Group	Reproducibility –	–	–	
Physics, IT Services Group	No DMP Local Storage	–	–	
<b>Biocenter</b> Epidemiology	No DMP –	– –	– –	
Digital Humanities UNIL	Data loss	–	Copyright	



International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 22–39  
© The Author(s) 2017  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216686984  
journals.sagepub.com/home/ifi



# A brief assessment of researchers' perceptions towards research data in India

**Manorama Tripathi**

Jawaharlal Nehru University, New Delhi, India

**Mahesh Chand**

Jawaharlal Nehru University, New Delhi, India

**S. K. Sonkar**

Babasaheb Bhimrao Ambedkar University, Lucknow, India

**V. K. J. Jeevan**

Indira Gandhi National Open University, New Delhi, India

## Abstract

The present paper dwells upon the importance of raw data for the development of science and research. The study includes an interview of 40 researchers and faculty members to understand their perception towards the raw data. It has suggested that the libraries can play a pivotal role in extending support to the researchers for organizing, archiving and preserving raw data for future use. Libraries may evolve a system at the university level wherein the researchers and faculty members be encouraged to deposit their raw research data in the institutional repositories, which most of the university libraries have developed.

## Keywords

Data collections, data services, preservation and conservation

Submitted: 20 May 2016; Accepted: 6 December 2016.

## Introduction

The research in the 21st century is more data intensive as with the advent of ICTs, more computation, simulation and modelling are done to evaluate, interpret and arrive at the new findings, unlike in the past when it was more theoretical in nature. The overwhelming use of computational means has also led to the collaborative research across the nations and disciplines. This has also led the generation of large data, which has to be reused, shared with others for re-analysis and reinterpretation. It is also being advocated that if the research is publicly funded, the data and the results should be available for one and all to see (PARSE.Insight, 2009; Royal Society, 2012; Tenopir et al., 2011). The data sharing practices must be followed as they help verify the research findings, data of the publicly funded research must be accessible to one and all to see, verify, reinterpret and re-analyse;

this will lead to the broadening of the frontiers of knowledge (Witt, 2008). It has been reiterated that there should be a continuum of accessible and inter-linked research and scientific information right from the raw data (the term 'raw data' is used in this paper to represent primary data generated out of research; a term which is gaining currency among scholarly circles) to the final publication which is based on the analysis of the raw data. It can be further analysed and interpreted for arriving at new findings and conclusions. Other researchers in the field can consider and evaluate data from a different perspective thus more knowledge and theories may be generated

## Corresponding author:

Manorama Tripathi, Jawaharlal Nehru University, New Mehrauli Road, New Delhi, 110067, India.  
Email: manoramatripathi27@gmail.com

**Table 1.** University level institutions in India as per MHRD website (Government of India, 2016).

Central universities	State universities	State private universities	Deemed to be universities	Institutions of national importance	Total
45	318	185	129	51 [Indian Institutes of Technology (IITs) – 16 National Institutes of Technology (NITs) – 30 Indian Institutes of Science Education and Research (IISERs) – 5]	718

conserving the time spent for data generation. It may identify errors, contain academic misconduct like fabrication, falsification and wrong reporting of the facts. They are also required for developing and validating study methods, analysis techniques and software implementation. The timely and transparent pre-publication sharing of data and results especially in public health emergencies has to become the norm of scholarly communication.

The focus is to make all the protocols, analysis, methods of study and raw data publicly available. Concerted efforts need to be made to train the research workforce in documenting their research as per the standards of the scholarly communication (Ioannidis et al., 2014). The Data Revolution Group (2014) has also elaborated on the sharing of data, technologies, innovations, developing principles and standards in the interest of one and all. In this context, the present paper attempts to study the researchers' perceptions towards the augmented use and reuse of raw data in India.

### Higher education landscape in India

Education holds utmost importance in developing countries as a good certification would automatically lead to a bright future. Education is the only hope for have-nots to expect a decent living in countries with wider socio-economic imbalances across the population. The government agencies are fully sensitive to these aspects and it is amply reflected in the higher education scene in India. The number of universities has increased 34 times from 20 in 1950 to 677 in 2014 of which 45 are Central Universities (40 under the purview of the Central Ministry of Human Resource Development (MHRD)), and the number of colleges has also registered a manifold increase of 74 times with just 500 in 1950 growing to 37,204, as on 31 March 2013 (Government of India, 2016). See Table 1. Recent times have also increasingly witnessed the entry of private agencies in providing professional education in the form of universities and colleges. Due to UGC's enhanced thrust on quality research in higher

education institutions and their quantification in the form of APIs (Academic Performance Indicators), research, computing and information infrastructure is getting a significant boost in higher education.

The International Comparative Research Base (2009–2014) for India reflects that the country has contributed remarkably in terms of research publications. The report shows that there is growth by 68% in research publications from 62,955 to 106,065. An active Indian researcher published 14 papers per head against the world average of five. The same report shows that the research publications from India are attracting and accruing more citations. The report also highlights the fact that the country's contributions span basic research and application-oriented research. Chemistry and pharmacology are the focus research areas while citations are in the subject areas of material science, pharmacology and toxicology.

It is quite natural that such a large higher education system undertakes, promotes and sustains research activities which are very significant for the society and academia as well as for the country. Research activities of such magnitude also generate sizable data which need to be effectively tapped in an interdisciplinary and multidisciplinary intent for future use and sharing. The Government has already acknowledged the importance of data and mandated that it be made available to one and all for access and use as per one's needs. It has formulated very many policies and deployed the physical computing infrastructure and other facilities for archiving and preserving research data and facilitating their access and use. The different policies and initiatives which support free data sharing in the country are discussed below.

### National data sharing and accessibility policy

The Government of India has framed a national data sharing and accessibility (NDSA) policy which facilitates access to Government of India owned shareable data in machine readable form through a wide area network all over the country. The policy applies to all

data created, generated, collected and archived using public funds (which still is a major source for research funding in the country apart from a large number of institutions for research and education fully managed by the Government) provided by the Government of India directly, or through its authorized agencies by various ministries/departments/organizations/agencies and autonomous bodies. The Department of Science & Technology serves as the nodal agency for monitoring and implementation of policy across the country (Government of India, 2012).

The suo moto disclosure under Section 4 of RTI Act, 2005 mentions that a large amount of information must be put into the public domain on a proactive basis to make the functioning of the public authorities more transparent and accountable.

### **Policies of University Grants Commission regarding open access**

The University Grants Commission (UGC) was set up in 1956 as a statutory body of the Government of India. It supervises and coordinates activities, determines, funds and maintains universities in India. It is responsible for framing policies and standards and ensuring their adherence by the universities in the country. UGC, India (<http://www.ugc.ac.in/page/UGC-Regulations.aspx>) has mandated electronic submission of PhD theses vide UGC Regulation 2009 (Minimum Standards and Procedure for award of MPhil/PhD degree). According to this, all universities are supposed to set up an e-theses repository and facilitate e-submission, archiving, maintenance and access to this repository at university level. All universities are required to submit the metadata for all the PhD theses submitted to them from inception, to the INFLIBNET Centre (Information and Library Network Centre, ([www.inflibnet.ac.in](http://www.inflibnet.ac.in)), an inter-university centre (IUC) of the UGC for networking and e-enabling of university and college libraries in the country). It also provides infrastructural support to the universities for setting up ETD archives. The centre also hosts e-resource consortia, supports creating infrastructure for sharing library and information resources and services, and provides anti-plagiarism software to the universities which ensures that whatever content is uploaded is authentic and free from plagiarism. These measures no doubt improved the quality and quantity of publications of researchers from universities and colleges which could be a major reason for the significant improvements mentioned in the International Comparative Research Base report discussed earlier.

### **Shodhganga**

The ETD repository created and maintained by INFLIBNET centre is popularly known as Shodhganga (reservoir of knowledge) (<http://shodhganga.inflibnet.ac.in/>). This repository acts as a platform for researchers to deposit their PhD theses and make them available to the scholarly community in open access. UGC, India has entrusted INFLIBNET with the responsibility of creating and maintaining Shodhganga. The repository is Dspace enabled and captures, stores, indexes and disseminates electronic theses submitted by the researchers to the Indian universities. The repository has a collection of 50,000+ theses which can be globally accessed in open access mode (INFLIBNET Centre).

### **Shodhgangotri: Repository of Indian research in progress**

Under this initiative, research scholars are supposed to deposit electronic versions of the approved synopsis which they have submitted to the universities for getting registered for doctoral programmes. The synopses are made available in open access and may be browsed globally. This repository aims to highlight the research trends followed across Indian universities and thus prevents duplication of efforts in research. Once the full text of a particular thesis is submitted and uploaded, a link to the full text (in Shodhganga) from synopsis in Shodhgangotri (<http://shodhgangotri.inflibnet.ac.in/>) is provided. At present, there are 2131 synopses in the online repository.

### **Other institutions**

The Indian Council of Social Science Research (ICSSR, <http://www.icssr.org/>) is a funding body for projects in the field of Social Sciences. In order to release a grant for projects, it requires the project investigators to make suitable arrangements for preservation of data such as filled-in schedule, tabular sheets and manuscripts relating to the projects. It reserves the right to demand raw data relating to the projects.

The Registry of Open Access Repository (ROAR) (<http://roar.eprints.org/>) lists 112 institutional repositories (IRs) which have been set up by universities and other educational institutions in India. These repositories are populated with full texts of theses and dissertations, journal articles and in-house publications technical reports, etc. They are yet to focus on archiving and curating the raw datasets for sharing.

The Central Drug Research Institute (CDRI) under the Council of Scientific and Industrial Research

(CSIR) maintains a clinical strain repository of 116 sequences of *M.tuberculosis* clones and a protein repository of important clones of *M.tuberculosis*.

The Directorate General of Hydrocarbons, under the Ministry of Petroleum and Natural Gas, Government of India (<http://www.dghindia.org/DataManagement.aspx>) has developed a National Data Repository (NDR). The directorate is collecting the cultural, geological, petro physical, seismic, well, production, and reservoir data; it also archives unstructured data like reports and documents related to the oil fields of India.

### Initiatives at the international level

The journals like *Nature*, *Science*, *PLOS ONE* and publishers like Elsevier and Springer have mandated that all the raw data underlying the manuscripts submitted to them should be made available for the public to access and use. *Nature* has decided to give wider coverage to the methodology, technical statistics and raw datasets to ensure scrutiny and reproducibility by others. The journals also urge authors to share their computer codes which they have used in models, simulations and data analysis. It encourages its authors to submit their protocols to the Protocol Exchange.

The Bermuda principles stress and advocate on sharing of the sequence data. The Brussels Declaration (STM, 2007) also advocates that the research data should be freely available. A February 2013 Memo of the White House Office of Science and Technology Policy (2013) advocates that all scientific journal articles which are federally funded should be available for anyone to consult after a year of publication. It also stresses that all the unclassified research datasets should be archived, curated for future use, re-analysis and reinterpretation. The National Institutes of Health (NIH), National Science Foundation, USA requires that all researchers should incorporate their data management and sharing plans in their research proposals before submitting them for seeking grants.

The Wellcome Trust along with 30 other global health bodies have decided that all the research data gathered during the Zika virus outbreak and future public health emergencies should be made available as rapidly and openly as possible (Wellcome Trust, 2016)

Re3data (<http://www.re3data.org/>) is a global registry of research data which indexes 1500 data repositories from the different subject areas. The registry lists 29 data repositories which either India has set up solely or in collaboration with other research organizations of the world. The coverage and content of the

data repositories of India are given in a table in Appendix A.

The general public has placed a lot of faith in the researcher community. In the same vein, the researchers need to show integrity beyond doubt by making their datasets available in the public domain for peer scrutiny and building further research.

### Review of literature

#### *Advantages of data sharing vis-à-vis researchers' perceptions*

Traditional science involved experimentation and theoretical aspects, but with the advent of ICTs a new component of 'computation' has been added which has an overwhelming impact on how science is being done. The ICTs have played a significant role in the emergence of e-science on the information landscape. E-science has facilitated researchers across the different disciplines and geographic boundaries to collaborate and find solutions to common problems. As a result of all these factors, a voluminous amount of data is being generated which needs to be captured, organized and preserved for future use.

The widespread sharing of data may lead to their analysis and interpretation by scientists from different fields, thus enabling an interdisciplinary approach to education, training and research.

Open Access to Research data expedites research and new discoveries. This is particularly the case in epidemics and medical emergencies when the outbreak of the diseases has to be arrested as early as possible. When the Ebola epidemic broke out, the researchers sequenced 100 Ebola genomes and the data was uploaded to the GenBank. The easy access leads to collaboration among researchers across the disciplines which resulted in finding out about its spread (Yozwiak et al., 2015).

Recently, the ZEST (Zika Experimental Science Team) who have infected monkeys with the zika virus have made their data public. They have released them online for anyone to view and are updating their data on a daily basis. It is assumed that the publication of the data will accelerate the research into the nature and cause of the virus that has spread across the Americas (Butler, 2016). Major challenges like climate change and global warming which loom large across the globe can be addressed by sharing the datasets across disciplines (Witt, 2008).

There are no clear standards and guidelines for sharing the raw research data. There is a need to develop standards and rules at the international level, which will facilitate easy and immediate access and sharing in the case of medical emergencies. For



instance in 2006, a consortium of researchers, GISAID, established a framework for good practices. This resulted in sharing of H1N1 data when the disease broke out in 2009.

There are ambiguities concerning the ownership of data. There is confusion about whether the datasets generated in publicly funded research belong to the state or to the researcher, especially in medical or clinical research where permission has to be sought from the patients before sharing the data. It also does happen that the researchers lack the skills for organizing, categorizing, tagging, annotating the datasets which they generate. This may lead to low access and use by others over time. The researchers are also handicapped by the paucity of time and funds and do not pay the required attention to the issue of organizing the datasets.

The PARSE.Insight study has highlighted that scientists do not share data because of the legal problems and fear its misuse. Scientists do not report responsibly and in a transparent manner the experimentation work and analyses they conduct on animals in medical sciences (Eisen et al., 2014). Tenopir et al. (2011) reported that scientists do not share their data because of the shortage of time and funds. Savage and Vickers (2009) have also indicated that journals have mandated the submission of datasets along with the manuscripts, but they do not ensure their strict implementation. The Working Group on Digital Data has emphasized that the preservation and access of research data accelerate progress in science and society.

In spite of all the mandates and mandatory guidelines for implementing data submission for future use, it appears that sharing is yet to become the norm in scholarly communication. In this context, libraries have a significant role to play. They must deploy facilities for data storage, use and sharing. The libraries of the University of Illinois and University of Purdue are providing data storage and curation services on a grand scale.

In the present study, considering the different studies discussed above, two questions (Questions 8 and 9 in Appendix B) were posed to find out what researcher feels about data sharing culture and practice.

#### *Discipline specific impact on data sharing*

Kim and Stanton (2016) surveyed 1317 scientists in 43 different disciplines of Science and Engineering to find out how the institutional environment and individual motivational factors influence the data sharing behaviour of the researchers. They advocated that

scientific data sharing can be promoted by the collaborative efforts of the funding agencies, publishers and research institutions by developing an incentive system, standardizing the data sharing protocols, providing data curation and management plans and also by promoting an altruistic culture in the researcher community. Borgman (2012) observed that data sharing practices vary across the different disciplines, though they are very common in astronomy and genomics. It is not a common practice in education and research activities (Piwowar and Chapman, 2010).

Stamatoplos et al. (2016) studied the data management needs of a teaching intensive organization as compared to the research-intensive organizations. They observed that the data management needs of the researchers at the teaching intensive institutions are the same as those of the research intensive organizations. Similar services could be deployed to cater to the data management of both the groups.

Faniel et al. (2015) studied the social science researchers' satisfaction with the data reuse. Their study spotlighted how attributes like completion, credibility, accessibility and ease of operation of data influence the reuse of data. Their study also reported a positive relationship between the documentation quality and researchers' satisfaction while reusing the data.

Cheah et al. (2015) investigated and studied the perception of different stakeholders towards the concept of data sharing. They reported that the stakeholders had a positive attitude towards data sharing. They concluded that data sharing practices ensure better analyses and interpretation, optimum use of resources, greater accountability and transparency. Their study also spotlighted that the areas of 'data standardization, appropriate consent models and governance' need to be seriously attended to for ensuring sharing of data responsibly.

Discipline specificity in research may also lead to different methods and issues in handling research data. Three questions in the present study assess what type of data is generated by the researcher and their organization and storage problems.

#### *Education and training for library staff*

Tenopir et al. (2013) have investigated the research data services (RDS) practices of various libraries in the USA and Canada and whether the library staff has requisite skills and training to provide the RDS services satisfactorily. The authors reported that 75% of the respondent library staff did not provide RDS services though they had the requisite skills to provide the same.



Buys and Shaw (2015) surveyed and investigated how the researchers at Northwestern University in Evanston and Chicago campuses managed their raw data. They reported that the researchers needed support services for data management and sensitization towards the mandatory policies of funding bodies and publishers and for raw data submission and reuse.

Henderson and Knott (2015) elaborated about RDS services to the researchers at the Virginia Commonwealth University libraries. The staff liaised with the different stakeholders of the university to sensitize, educate and train them in preparing data management plans for reuse and sharing. The institutional policies and support in the form of repositories with set format and metadata for submitting raw data, mandates of the funding agencies and publishers play a significant role in influencing the researchers' data sharing behaviour and practices. The libraries should provide information about the importance of data documentation and use of appropriate metadata. The metadata has a pivotal role in data management and sharing. It helps the researchers describe their datasets appropriately and facilitate others to search and retrieve them (Wiley, 2014).

Libraries may offer a centralized, well-utilized service. There is need to identify champions who further the data services initiatives on the university campuses by talking about the services with their colleagues offering to pilot new potential services and collaborate as partners in research. There is need to deploy skilled manpower for providing appropriate research data management (RDM) services (Wright et al., 2014).

Library staff needs to be trained in data management, organization, preservation and curation skills. The library professionals need to be well conversant with the needs and requirements of their researchers, the kind of raw data they generate and how they can be archived for future use (Peters and Dryden, 2011). Bresnahan and Johnson (2013) observed that there is need to train the liaison librarians for providing RDM services effectively. Charbonneau (2013) has surveyed the researchers and faculty members in order to understand their data management and sharing practices and accordingly suggested changes in the Library Science curriculum to train the professionals for providing RDS efficiently. The present study evolved a question about researchers' view of the role of the library in managing research data out of the experiences of these many studies which underlined the significant roles libraries have to play in archiving and reuse of research data.

## Data curation

Data curation entails all the procedures and activities which are adopted and followed for organizing, managing and preserving the digital data. Studies by Dearborn et al. (2014) and Witt (2008, 2012) have reported about the Purdue University Research Repository (PURR). The HUB Zero (a platform used to create dynamic websites for scientific research and educational activities) has been customized to publish, organize, preserve and share the datasets in PURR (Dearborn et al., 2014).

Libraries have expertise in classifying and cataloguing of information: the same can be used for classifying, indexing, organizing and preserving the datasets generated in the research activities of their universities for future use and sharing. The universities support the curation and archiving of the research data through their institutional repositories (IRs), though the levels of services provided vary in the different levels of architecture and models. The IRs play a pivotal role in stewardship of datasets which can be used in interdisciplinary research endeavours (Witt, 2008).

Poole (2015) has dwelt upon the issues of cyberinfrastructure, research communities, collaboration, policies, standards, best practices, provenance, selection appraisal, appropriate metadata and risk management for data curation. It has been observed that the institutional support for data and metadata management varies across the disciplines even within a single institution. In order to support curation of research data the institutions must support heterogeneous kinds of projects data and metadata (Buys and Shaw, 2015; Mayernik, 2015).

Library staff needs to be trained in data management, organization, preservation and curation skills. They need to be well conversant with the needs and requirements of their researchers, the kind of raw data they generate and how the same can be archived for the future use (Brown et al., 2015; Peters and Dryden, 2011). Libraries will have to play a proactive role and partner with the researchers, scientists and faculty members in archiving, preserving and curating their datasets for future reuse and sharing. They should be crystal clear concerning the characteristics of the data like size and scope, varied form, the backlog of data, non-standard data and data formats. They will have to learn new skills and strengthen their computing infrastructure; otherwise, they will trip over themselves in trying to extend research data services to the researchers (Salo, 2010). The researcher community cannot be left on its own to manage the large datasets which they generate in

**Table 2.** The designation of the interviewees.

Designation	Research scholars	Post-doctoral research fellow	Assistant Professor	Associate Professor	Professor	Total
Number	06	01	15	09	09	40

their research activities (Scaramozzino et al., 2011). The library at the Johns Hopkins University is very meticulously engaged in curating the datasets across the diverse fields, astronomical data as well as medical manuscripts (Monastersky, 2013).

The essence of the programmes, projects and movements of open research data is to make research data openly and widely accessible to one and all across the globe. These also aim at promoting and fostering transparency and accountability and thus facilitate the discovery of new solutions to contain the outbreak of epidemic and diseases, especially in medical sciences. Another idea that is underlined in successful projects is 'it is always good to do what one is good at doing'. The researchers should concentrate on their research endeavours while the chores of organizing, preserving and curating the datasets for future use should be left for the librarians and libraries. Question 10 (see Appendix B) used in the present study seeks the opinion of researcher about free for all access of research data and the next question is about data repositories.

## Method of study

The interview method was used to understand the perception of the researchers towards the research data. The authors experienced some difficulties in convincing the researchers and faculty members to share their views and provide their feedback with regard to the raw research data. Initially the interviewees showed their extreme reluctance in sparing time for the interviews. A set of 12 questions was used to get the feedback. The first four questions deal with identifying the researcher. The next three questions assess what type of data is generated by the researcher and their organization and storage problems. Questions 8 and 9 attempt to find out what the researcher feels about data sharing culture and practice. Question 10 seeks the opinion of the researcher about free for all access of research data and the next question is about data repositories. The last question is about researcher's view of the role of the library in managing research data. The answers were noted by the authors while interviewing them. The interviews were not recorded. The interviews were conducted in April 2016. In total 40 research scholars (six from

JNU) and faculty members (16 and 18 faculty members from Jawaharlal Nehru University (JNU) and Babasaheb Bhimrao Ambedkar University (BBAU) respectively) were interviewed for the present study. Table 2 shows the designation of the interviewees. The number of participants was 40 which is a small number. Still, it has helped in identifying and revealing the patterns and themes with regard to the researchers' perceptions to raw data. Each interview lasted for 15 to 45 minutes.

The authors interviewed the research scholars and faculty members from the following two universities:

Babasaheb Bhimrao Ambedkar University (<http://www.bbau.ac.in/>) set up in 2001, is located in Lucknow, Uttar Pradesh. It offers instructional and research facilities in science, technology and allied subject areas.

Jawaharlal Nehru University (<http://www.jnu.ac.in/>) was ranked at the third position in the top ten universities of the country. This is as per the 'India Ranking 2016' which has ranked 3500 private and public institutions. The National Bureau of Accreditation based on the five parameters of teaching and learning resources, graduation outcome, perception, outreach and inclusivity and research productivity. All the institutions were judged based on the information which they provided.

The senior library professionals of many prominent research institutions like CSIR labs and the Indian Statistical Institute were interacted with to know if they had devised some guidelines or services for extending research data services for their researchers. The professionals responded that the researchers at their organizations did not generate any sizeable raw research data, so their organizations do not have any system to store research data.

The researchers prepared a list of 12 questions (see Appendix B) on which the information was collected from the researchers. Questions 1–4 focused on the general aspects of the respondents like their designation, centre and areas of research. Questions 5–8 intended to know the various kinds of data generated by the respondents in their research activities; the

ways of organizing, maintaining and preserving them; if data sharing practices were prevalent in their field and if the respondents shared their data with their colleagues and students. Questions 9–12 elicited feedback from the respondents with regard to the free availability of the raw data, their level of awareness of different data repositories of their field, existence of data storage problems and if the researchers sought some support regarding data management from the Central Library of the university.

## Discussion

The subject areas and research interests of the 40 interviewees are presented in a table in Appendix C and the different types of data they generated in a table in Appendix D. It is evident that most of them work in the premier areas of their specialized research.

It shows that the researchers and faculty members generated a broad spectrum of data types but did not use any metadata for their organization. Of the researchers, 18 (45%) depended on experimental data and a close 17 (42.50%) used data from sample surveys. Four of them used either previous data or coding data. It corroborates the findings of previous studies (Averkamp et al., 2014; Steinhart et al., 2012) which spotlighted that the researchers and faculty members generated different types and sizes of data but were not bothered about the metadata. They were, of course, concerned about the storage of their data. The researchers believe that they need to focus on their research work and the work of organizing and preserving data for sharing, and reuse gets relegated (Wiley, 2014). Tenopir et al. (2011) reported that the researchers are content with their short-term data storage but not satisfied with the storage options available for the long term.

### Organization and preservation of research data

The researchers and faculty members prepared notes and kept them in computer storage options such as CDs, DVDs, portable hard disks and pen drives for further use. They maintained their data through graphs, tables, SPSS spreadsheets, .CSV and .Xls files. The majority of them (31, 77.50%) resorted to computer and allied storage technology for maintaining, organizing and preserving research data. Some of them (9, 22.50%) were solely satisfied with print publications generated out of research. Four of the researchers were not adopting any options at this stage for preserving their raw research data. Some researchers gave more than one option and hence the total number of responses is more than the total 40. The

researchers and faculty members were not very clear with regard to how to organize, store and preserve data for future use.

One of them said:

Not very good at it (organizing the data for future use).

Another faculty member said:

Follow no scientific way of organizing or maintaining raw data for future use. Once the data is utilized and paper is published, we don't care, we don't preserve.

The research scholars and faculty members in Sanskrit studies submit and store their data in the Java-based server of the centre. They also submit their raw data to the Indian Language Corpus Initiative Server.

Another faculty from Plant Sciences said:

Data are recorded in students' note books (Data) are analysed by working at multiple experiments to generate a sequence of events, also called a story or experimental findings. Each experimental finding is entered in a text or graphical format by the student or researcher and e-mailed to the supervisor. The backup of data is stored in hard drives.

The aforementioned faculty has stored all the small unfinished data since 2004. The faculty joined the university in 2004.

### Data sharing practices

Some of the researchers and faculty members (12, 30%) responded that there was no culture of sharing data in their fields. A majority (24, 60%) of them agreed about the culture of raw research data sharing in their fields. In the case of sharing data with other researchers, 25 (62.50%) of them said yes while 10 (25%) of the researchers said no. Many (27, 67.50%) of them said that they selectively shared their data with their students. The interviewees said that confidentiality was very important in their research endeavours. The research scholars said that they could share their data after submitting their PhD dissertations. It implies that they can only share after they have derived maximum dividends from the data which they have generated and collected. The researchers in Computer Science and International Relations said that they would selectively share their data. They would readily share their data once they had utilized and published their findings in journals and books.

The faculty member from Plant Sciences did not explicitly agree with the concept of data sharing practices. He said that he had no problem in sharing the published data. He also mentioned that he had shared

his unpublished raw data with other researchers who were writing review articles. He underlined that when sharing unpublished research data, he wanted to be duly acknowledged. He emphasized that:

Premature data gives wrong conclusions.

Another researcher from Biotechnology echoed the same feeling:

No, only subsequent to publication share data with the students.

The researcher from the field of machine Translation (Sanskrit Studies) said that he did not share the data of the work-in-progress. But in certain situations he would partially share some samples of the data.

One faculty member from Social Psychology said that:

Raw research data implies concerted efforts made by one individual researcher; one will be reluctant to share one's efforts with others. Raw data is just like an individual's salary. I will not enquire about anybody's salary and will expect no one to enquire about my salary.

(I) share research ideas, methods and findings with my students, raw data are not shared.

The researchers may want to have exclusive rights to the datasets which they took years to collect and generate. Other researchers said they had no objections to sharing their data with others. They shared their data with their colleagues and students.

Some researchers observed that the publication of data in journal articles was sufficient for data sharing. This shows a lack of awareness and understanding of mandatory policies of funding bodies and publishers for data sharing. The scholarly articles or book chapters do not include adequate raw data which may allow replication of the results. Maybe researchers are also reluctant to build their research on someone else's data due to the increasing strictness with which journals and institutions conduct plagiarism check on scholarly articles. The researchers and the faculty members need to be educated on the importance of providing access to the raw data on which the final work has been built and reported as a journal article or book chapter. Cragin et al. (2010) interviewed scientists across different fields to explore their data sharing behaviour. They reported that majority of scientists wanted to share their data with others on a very restrictive basis. Their study also highlighted the scientists' apprehension about depositing their datasets in the public repositories as others may misuse them.

### *Availability of raw data in public domain*

Most (22, 55%) of the researchers perceived the need for free access of research data to all researchers interested in the same. Some of them felt that free availability of data leads to improvement in the quality of research conducted and avoids cases of research duplication and plagiarism. Only 13 (32.50%) of researchers opposed this noble idea. One of the faculty members said:

For the development and growth of science, for verification of the findings reported, as data fudging is very common, data should be in public domain, but it should not be misused.

Another faculty member said:

Yes (it should be available) but with the limitations for fair use.

Another respondent said,

Yes, but not before publication.

### *Data storage problems*

Out of the 40 interviewees, 19 (47.50%) said they faced data storage problems while 16 (40%) said that they did not face any problem pertaining to the storage. They were three respondents (7.5%) who appeared to not to have a clear stand on the data storage problem, said, 'Can't say'.

Averkamp et al. (2014) observed that researchers were concerned about the storage of their data. But Stamatoplos et al. (2016) have reported that researchers generally do not face data storage problems. They said that researchers encounter problems of inadequate physical storage, using specialized software and remotely accessing their research files. Researchers do not require generally large storage spaces. On an average, researchers need less than one terabyte of data storage (Akers and Doty, 2013; McLure et al., 2014). Buys and Shaw (2015) have observed that the researchers generally store their data in local drives and departmental servers which prevent data sharing and long-term preservation. It is the duty of library professionals to address and handle the issues related to the organization, preservation and curation of research data and to develop, evolve and maintain tools, services and computing infrastructure to support the researchers' storage needs.



### Awareness about data repositories

Out of the 40 interviewees, 22 (55%) were aware while 12 (30%) were unaware of the subject-specific data repositories. Only 17 (42.5%) had submitted their data to such repositories out of which 4 (10%) said that they submitted only their final data to the repositories. Sixteen (40%) of the researchers so far have not archived their data in any repository. One of the researchers said that he uploaded his research data on Researchgate. The faculty members from Biotechnology generally submitted their sequence and culture to the National Centre for Biotechnology Information (NCBI)/AICC.

Wilson et al. (2013) observed that the researchers did not hold a common view with regard to data sharing and were completely unaware of the RDS being provided by the Oxford University Libraries.

### Library support

In response to the question on library support for organizing, maintaining and preserving their raw research data, a majority of them 29 (72.50%) said that they wanted library support which is a strong appreciation to many librarians who single handedly advocate for this provision as an activity under library research support; 7 (17.50%) stated that they did not want any help from the library and they could manage on their own. Two researchers (5%) were not very clear on this aspect. The researchers and faculty members from the subject areas of International Relations, Social Sciences, Language and Literature and Culture Studies and Computer Science said that they wanted some support while the faculty members from Plant Science and Biotechnology stated that they could manage their research data on their own. One of the faculty members said that they would expect the library to support them by providing access to scholarly journals and other information resources.

Scarmozzino et al. (2011) surveyed the faculty members of the College of Science and Mathematics, California Polytechnic State University and found that the teachers lacked skills to manage their data. But they certainly did not want any help from the libraries for their data management. Toups and Hughes (2013) explored and assessed the data needs of the researchers at the Trinity University; they concluded that the researchers did not want the library to collaborate with them for data management. Si et al. (2015) surveyed 87 libraries listed in the World's Best Universities released by the *USA News* in October 2012 and reported that only 57.5% of the libraries studied provided RDM services like the basics of

research data, guidelines and training for data management, data curation and storage. Corral et al. (2013) surveyed 140 libraries in Australia, New Zealand, Ireland and the UK and found the low level of engagement of libraries in providing data management services.

### Conclusion

The absence of a mechanism for organizing, maintaining, preserving and ensuring access for reuse of raw data often results in raw research data slipping in to oblivion. Libraries have a pivotal role to play in this regard. They have always been at the forefront in embracing change and showing the way by introducing new services as per the needs and expectations of the users.

Libraries need to work to collaborate with faculty, research and academic staff members in order to understand their way of functioning, how they initiate, conduct their research activities and arrive at the findings and in the process what kind of data are generated and how they are processed, organized and maintained.

The university libraries in India are in the very early stage of providing support for archiving, organizing and maintaining raw data; though a similar trend has been reported from other parts of the globe. Cox and Pinfield (2014) surveyed the UK universities to find out how the libraries were providing RDM services. They highlighted that the libraries offered very basic RDM services; while research intensive institutions had awakened to the need to provide RDS to their researchers and scientists. Libraries are still in the process of developing institutional RDM policies and services.

The present study has highlighted the efforts made in India to put research data in open access, which can be reused, re-analysed, reinterpreted for further study and thus add to the already existing knowledge. It has also studied and explored the researchers' perceptions towards the availability of raw data in open access so that others can browse access and use for further research and analysis and its sharing with others. It has underlined that the researchers and faculty members believe that the data should be freely available for anyone to use but they are themselves reluctant to share. Research data at the two universities under study, is currently stored in any of the several ways- in individual PCs, print files, institutional servers, Google drives, in national or international subject specific repositories, etc. No institution has yet been selected as a common, national repository for storing, preserving and curating the research data. It suggests

that the libraries need to play a proactive role in offering RDS to their researchers. It also needs to undertake awareness campaigns in order to dispel the ignorance of the researchers who are reluctant to release their data for others to use and reinterpret.

What is required is a national level policy on RDM involving the different stakeholders, which may be compulsorily followed for awarding promotions and project grants. Researchers can take the support of computer staff to convert their existing data into those formats which survive long term and archive the same in their library's institutional repositories. National level data archives may be thought of in a specialized manner for preserving, organizing and hosting these data. Researchers may be rewarded both for archiving their research data in such national archives and for rebuilding new research on other researchers' data archived in such repositories with API points as a due incentive to the cause of conserving and preserving research data generated out of intensive scholarly effort.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### References

- Akers KG and Doty J (2013) Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation* 8(2): 5–26.
- Averkamp S, Gu X and Rogers B (2014) *Data Management at the University of Iowa: A University Libraries Report on Campus Research Data Needs*. Available at: [http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1246&context=lib\\_pubs](http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1246&context=lib_pubs) (accessed 9 December 2016).
- Borgman CL (2012) The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63(6): 1059–1078. DOI: 10.1002/asi.22634/full.
- Brown R, Wolski M and Richardson J (2015) Developing new skills for research support librarians. *Australian Library Journal* 64(3): 224–234. DOI: 10.1080/00049670.2015.1041215.
- Bresnahan MM and Johnson AM (2013) Assessing scholarly communication and research data training needs. *Reference Services Review* 41(3): 413–433.
- Butler D (2016) Zika researchers release real-time data on viral infection study in monkeys. *Nature*. Available at: <http://www.nature.com/news/zika-researchers-release-real-time-data-on-viral-infection-study-in-monkeys-1.19438> (accessed 9 December 2016).
- Buys C and Shaw P (2015) Data management practices across an institution: Survey and report. *Journal of Librarianship and Scholarly Communication* 3(2): eP1225. DOI: 10.7710/2162-3309.1225.
- Charbonneau DH (2013) Strategies for data management engagement. *Medical Reference Services Quarterly* 32(3): 365–374. DOI: 10.1080/02763869.2013.807089.
- Cheah PY, Tangseefa D, Somsaman A, et al. (2015) Perceived benefits, harms, and views about how to share data responsibly: A qualitative study of experiences with and attitudes toward data sharing among research staff and community representatives in Thailand. *Journal of Empirical Research on Human Research Ethics* 10(3): 278–289.
- Corrall S, Kennan MA and Afzal W (2013) Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends* 61(3): 636–674.
- Cox AM and Pinfield S (2014) Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science* 46(4): 299–316.
- Cragin MH, Palmer CL, Carlson JR, et al. (2010) Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 368(1926): 4023–4038. DOI: 10.1098/rsta.2010.0165.
- Data Revolution Group (2014) *A World that Counts: Mobilising the Data Revolution for Sustainable Development*. Available at: <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf> (accessed 9 December 2016).
- Dearborn C, Barton A and Harmeyer N (2014) The Purdue University Research Repository: HUBzero customization for dataset publication and digital preservation. *OCLC Systems & Services* 30(1): 15–27.
- Eisen JA, Ganley E and MacCallum CJ (2014) Open science and reporting animal studies: Who's accountable? *PLoS Biol* 12(1): e1001757. DOI: 10.1371/journal.pbio.1001757.
- Faniel IM, Kriesberg A and Yakel E (2015) Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology* 67(6): 1404–1416. DOI: 10.1002/asi.23480.
- Government of India (2012) *National Data Sharing and Accessibility Policy-2012 (NDSAP-2012)*. Available at: <http://ogpl.gov.in/NDSAP/NDSAP-30Jan2012.pdf> (accessed 9 December 2016).
- Government of India (2016) *University and Higher Education: Overview*. Available at: <http://mhrd.gov.in/university-and-higher-education> (accessed 9 December 2016).
- Henderson ME and Knott TL (2015) Starting a research data management program based in a university library. *Medical Reference Services Quarterly* 34(1): 47–59. DOI: 10.1080/02763869.2015.986783.



- Ioannidis JP, Greenland S, Hlatky MA, et al. (2014) Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* 383(9912): 166–175. DOI: 10.1016/S0140-6736(13)62227-8.
- Kim Y and Stanton JM (2016) Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology* 67(4): 776–779. DOI: 10.1002/asi.23424.
- McLure M, Level AV, Cranston, et al. (2014) Data curation: A study of researcher practices and needs. *portal: Libraries and the Academy* 14(2): 139–164.
- Mayernik MS (2015) Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology* (67)4: 973–993. DOI: 10.1002/asi.23425.
- Monastersky R (2013) The library reboot: As scientific publishing moves to embrace open data, libraries and researchers are trying to keep up. *Nature* 495(7442): 430–433.
- PARSE.Insight (2009) *Insight into Digital Preservation of Research Output in Europe: Survey Report*. Available at: [http://libereurope.eu/wp-content/uploads/PARSE-Insight\\_D3-5\\_InterimInsightReport\\_final.pdf](http://libereurope.eu/wp-content/uploads/PARSE-Insight_D3-5_InterimInsightReport_final.pdf) (accessed 9 December 2016).
- Peters C and Dryden AR (2011) Assessing the academic library's role in campus-wide research data management: A first step at the University of Houston. *Science & Technology Libraries* 30(4): 387–403.
- Piowar HA and Chapman WW (2010) Public sharing of research datasets: A pilot study of associations. *Journal of Informetrics* 4(2): 148–156.
- Poole AH (2015) How has your science data grown? Digital curation and the human factor: A critical literature review. *Archival Science* 15(2): 101–139.
- Royal Society (2012) *Final Report: Science as an Open Enterprise*. Available at: <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/> (accessed 9 December 2016).
- Salo D (2010) Retooling libraries for the data challenge. *Ariadne* (64). Available at: <http://www.ariadne.ac.uk/issue64/salo> (accessed 9 December 2016).
- Savage CJ and Vickers AJ (2009) Empirical study of data sharing by authors publishing in PLoS journals. *PloS one* 4(9): e7078.
- Scaramozzino JM, Ramírez ML and McGaughey KJ (2012) A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries* 73(4): 349–365.
- Si L, Xing W, Zhuang X, et al. (2015) Investigation and analysis of research data services in university libraries. *The Electronic Library* 33(3): 417–449. DOI: 10.1108/EL-07-2013-0130.
- Stamatoplos A, Neville T and Henry D (2016) Analyzing the data management environment in a Master's-level institution. *Journal of Academic Librarianship* 42(2): 154–160. DOI: 10.1016/j.acalib.2015.11.004.
- Steinhart G, Chen E, Arguillas F, et al. (2012) *Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements*. Available at: [https://works.bepress.com/florio\\_arguillas/1/](https://works.bepress.com/florio_arguillas/1/) (accessed 9 December 2016).
- STM (2007) *Brussels Declaration*. Available at: <http://www.stm-assoc.org/public-affairs/resources/brussels-declaration/> (accessed 9 December 2016).
- Tenopir C, Allard S, Douglass K, et al. (2011) Data sharing by scientists: Practices and perceptions. *PloS one* 6(6): p.e21101. DOI: 10.1371/journal.pone.0021101.
- Tenopir C, Sandusky RJ, Allard S, et al. (2013) Academic librarians and research data services: Preparation and attitudes. *IFLA Journal* 39(1): 70–78. DOI: 10.1177/0340035212473089.
- Toups M and Hughes M (2013) When data curation isn't: A redefinition for liberal arts universities. *Journal of Library Administration* 53(4): 223–233. DOI: 10.1080/01930826.2013.865386.
- Wellcome Trust (2016) Sharing data during Zika and other global health emergencies. In: *Wellcome Trust blog*. Available at: <https://blog.wellcome.ac.uk/2016/02/10/sharing-data-during-zika-and-other-global-health-emergencies/> (accessed 9 December 2016).
- White House Office of Science and Technology Policy (2013) *Memorandum for the Heads of Executive Departments and Agencies*. Available at: [https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf) (accessed 9 December 2016).
- Wiley C (2014) Metadata use in research data management. *Bulletin of the American Society for Information Science and Technology* 40(6): 38–40. DOI: 10.1002/bult.2014.1720400612.
- Wilson J, Jeffreys P, Patrick M, et al. (2013) *University of Oxford Research Data Management Survey 2012: The Results*. University of Oxford. Available at: <http://blogs.it.ox.ac.uk/damaro/2013/01/03/university-of-oxford-research-data-management-survey-2012-t> (accessed 9 December 2016).
- Witt M (2008) Institutional repositories and research data curation in a distributed environment. *Library Trends* 57(2): 191–201.
- Witt M (2012) Co-designing, co-developing, and co-implementing an institutional data repository service. *Journal of Library Administration* 52(2): 172–188. DOI: 10.1080/01930826.2012.655607.
- Wright S, Whitmire A, Zilinski L, et al. (2014) Collaboration and tension between institutions and units providing data management support. *Bulletin of the American Society for Information Science and Technology* 40(6): 18–21. DOI: 10.1002/bult.2014.1720400607.
- Yozwiak NL, Schaffner SF and Sabeti PC (2015) Data sharing: Make outbreak research open access. *Nature* 518(7540): 477–479. DOI: 10.1038/518477a.

## Author biographies

**Dr Manorama Tripathi** is at present working as Deputy Librarian at Jawaharlal Nehru University (JNU) New

Delhi, India. Prior to joining JNU, she had worked at Indira Gandhi National Open University, New Delhi and Banaras Hindu University, Varanasi.

**Mahesh Chand** works as Assistant Librarian in Jawaharlal Nehru University. He has professional experience of more than 20 years. He has several publications to his credit in the form of journal articles and conference publications etc. His research interest areas are digital library services and he also has expertise in digitization of Electronic Theses and Dissertations, Newspaper/Press Clippings. He has been speaker at various seminars/workshops/conferences at JNU and many other institutions. He is a life member of ILA and has been associated with various professional bodies.

**Dr Sharad Kumar Sonkar** presently works as Senior Assistant Professor in the Department of Library and Information Science, Babasaheb Bhimrao Ambedkar University Lucknow. Before joining this university he also worked in various universities and institutions like BHU, Varanasi, DESIDOC, New Delhi; Digital Library, RGUHS,

Bangalore; Central Library, Bundelkhand University, Jhansi. He has also taken a specialization course in IT Applications to Library and Information Services from NCSI, IISC, Bangalore. He has guided nine MPhil and 70 PG students. He has contributed more than 80 papers in national and international journals, books and conferences.

**Dr V. K. J. Jeevan** currently works as Deputy Librarian in the Library and Documentation Division of Indira Gandhi National Open University, New Delhi, India. Prior to this he worked as Assistant Librarian in the Central Library of Indian Institute of Technology Kharagpur. He has Master's degrees in Physics, Library & Information Science and Computer Science and a PhD in Library & Information Science and has over 23 years of professional experience. He is the recipient of 'Young Information Scientist Award' from the Society for Information Science (SIS), India and best paper awards from Raja Rammohan Roy Library Foundation (RRLF), India, Indian Library Association (ILA) and the Indian Association of Special Libraries and Information Centres (IASLIC).

## Appendix A. Coverage and content of the data repositories of India which are listed in Re3data.

Name of the repository	Parent organization	Subject/coverage	Type of data	Dataset	Size of the database	Data format
MOL TABLE <a href="http://moltable.ncl.res.in/">http://moltable.ncl.res.in/</a>	NCL, pune	Chemoinformatics, nano bio chemistry	Molecular data	44 dataset		PDF
Indiawater portal <a href="http://www.indiawaterportal.org/">http://www.indiawaterportal.org/</a>	Arghyam, public charitable foundation	Water, sanitation, agriculture, the environment, climate change, biodiversity, and other related thematic areas.	Temperature, vapour pressure and cloud, etc. Crop envapotranspiration	Numerical dataset	Year 1999–2000	XLS
Open Government Data Platform India <a href="https://data.gov.in/">https://data.gov.in/</a>	NIC, India	Government ministries and department-related data	Government-related data	Numerical dataset	2014 to till date	CSV, XLS, ODS, XML, RDF, KML, GML, RSS, ATOM
Indian Genetic Disease Database <a href="http://www.igdd.iicb.res.in/">http://www.igdd.iicb.res.in/</a>	Indian Institute of Chemical Biology Kolkata	Genetic diseases, genetic disorders	Ideogram, map text	Database covers 104 diseases with information on 8233 individuals	1993 to till date	HTML
IGDD/home.aspx Export Import Databank <a href="http://commerce.nic.in/eidb/default.asp">http://commerce.nic.in/eidb/default.asp</a>	Directorate General of Commercial Intelligence and Statistics (DGCI&S), Kolkata	Trade-related data	Import /export	Statistical	2006–2007 onwards excluded 12–13	.asp, XLS
World Data Centre for Geomagnetism <a href="http://wdci.res.in/WebUI/Home.aspx">http://wdci.res.in/WebUI/Home.aspx</a>	Indian Institute of Geomagnetism, situated in the campus of old Colaba Magnetic Observatory, Bombay	Geomagnetic	Set of analog and digital geomagnetic data	Set of analog and digital geomagnetic data as well as indices of geomagnetic activity supplied from a worldwide network of magnetic observatories.	1841 to till date	Magnetogram Image
International Molecular Exchange Consortium <a href="http://www.imexconsortium.org/">http://www.imexconsortium.org/</a>	European Molecular Biology Laboratory-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK	Life sciences	Scientific data, genome, human data	Non-redundant set of highly annotated protein-protein interaction data.	2003 to till date	XML
International Ocean Discovery Program (IODP) <a href="http://www.iodp.org/">http://www.iodp.org/</a>	Implementing organizations (USIO, CDEX and ESO) Texas A&M University	Earth sciences Geoscience	Study the composition and structure of the Earth's subseafloors	Database includes paleontological, lithostratigraphic, chemical, physical, sedimentological, and geophysical data for ocean sediments and hard rock	2003 to till date	.kml Keyhole Markup Language (KML) is an XML notation for expressing geographic annotation and visualization .htm
International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) <a href="http://www.icrisat.org/">http://www.icrisat.org/</a>	International crop Research Institute for the Semi Arid Tropics	Agriculture	Information related to tropics, crop, location and resources	Information related to tropics, crop, location and resources		
Indian National Centre for Ocean Information Services Tropflux <a href="http://www.incois.gov.in/tropflux/">http://www.incois.gov.in/tropflux/</a>	Indian National Centre for Ocean Information Services	Oceanography	Accurate air-sea heat and momentum flux data	Turbulent and longwave fluxes, and ISCCP surface radiation data for shortwave flux	1979–2013	.nc

(continued)

## Appendix A. (continued)

Name of the repository	Parent organization	Subject/coverage	Type of data	Dataset	Size of the database	Data format
BioSearch <a href="http://www.bioserach.in/microbes/">http://www.bioserach.in/microbes/</a>	Bioinformatics Centre, National Institute of Oceanography	Bioscience	Marine microbial data	Marine microbial data	1814	.php, PDF
Histome <a href="http://www.actrec.gov.in/histome/index.php">http://www.actrec.gov.in/histome/index.php</a>	Advance Centre for Treatment Research and Education in Cancer (ACTREC), Navi Mumbai & Indian Institute of Science Education and Research (IISER), Pune	Biology	Electronic Data	Human histones, their post-translational modifications and modifying enzymes		.PHP, XLS
Humanproteinpedia <a href="http://www.humanproteinpedia.org/">http://www.humanproteinpedia.org/</a>	Pandey lab & Institute of Bioinformatics		Biology (human protein)	Human protein data	Post-translational modifications, protein-protein interactions, tissue expression, expression in cell lines, subcellular localization and enzyme substrate relationships	
.txt Database on Indian Economy <a href="http://dbie.rbi.org.in/DBIE/dbie.rbi?site=home">http://dbie.rbi.org.in/DBIE/dbie.rbi?site=home</a>	Reserve Bank of India	Indian Economy	Statistical data	Real Sector, Corporate Sector, Financial Sector, Financial Market, External Sector, Public Finance, SocioEconomic Indicators. S	1955 onwa-rds	XLS
Oral Cancer Gene Database <a href="http://www.actrec.gov.in/OCD/index.htm">http://www.actrec.gov.in/OCD/index.htm</a>	Tata cancer Research Institute, Bombay	Bio information	Genes data	Genomic and proteomic profiles data	374 genes	.html
India Environmental Portal <a href="http://www.indiaenvironmentportal.org.in/">http://www.indiaenvironmentportal.org.in/</a>	Centre for Science and Environment, New Delhi	Environment	Climate change, mining, forests, water management, environmental impact assessments Statistical data	Contextual, curated, cross-tagged specialized research reports and government documents	400000	.html
India Energy Portal <a href="http://www.indiaenergyportal.org/">http://www.indiaenergyportal.org/</a>	TERI	Energy		Global energy scenarios in terms of resources, demand, supply, and installations		.htm
The Human Protein Atlas <a href="http://www.proteinatlas.org/">http://www.proteinatlas.org/</a>	Royal Institute of Technology in Stockholm, Uppsala University, Uppsala Akademiska University Hospital	Biology	Data related to spatial distribution of proteins in 44 different normal human tissues and 20 different cancer types, as well as 46 different human cell line	Normal tissue, cancer tumor, subcellular location, RNA gene and RNA isoform data	Atlas content: 718 antibo-dies and 413,568 images 2003 onwar-ds	XML, RDF & TAB format
Human Protein Reference Database <a href="http://www.hprd.org/">http://www.hprd.org/</a>	Johns Hopkins University and the Institute of Bioinformatics	Biology, bioinformatics	Protein-protein interaction data	Post-translational modifications, tissue expression, subcellular localization and protein-protein interaction data	2003 onwards	XML and FLAT files SVG format

(continued)

# Appendix A. (continued)

Name of the repository	Parent organization	Subject/coverage	Type of data	Dataset	Size of the database	Data format
North East Resources Databank <a href="http://databank.nedfi.com/">http://databank.nedfi.com/</a>	North Eastern Development Finance Corporation Ltd. (NEDFI), Dispur	Information about North East India	General, economy, agriculture, tourist, infrastructure, public finance, industry.	Information related to general, economy, agriculture, tourist, infrastructure, public finance, industry.	1980 onwards	.htm
Distract Information System for Education(DISE) <a href="http://www.dise.in/">http://www.dise.in/</a>	National University of Educational Planning and Administration, New Delhi	Education	District wise educational data	Educational statistics data	1995 onwards	.pdf
Experimental Tropical Watersheds <a href="http://bvet.obs-mip.fr/en">http://bvet.obs-mip.fr/en</a>	Indo-French Cell for Water Sciences, Bangalore Midi-Pyrénées Observatory (OMP, Toulouse)	Environment-experimental tropical watersheds/tropical ecosystems	Hydro chemical, hydrological and metrological data	Hydro chemical, hydrological and metrological data	1999 onwards	.csv
Central Food Technological Research Institute Peptide Database of CFTRI <a href="http://www.cftri.com/pepdb/">http://www.cftri.com/pepdb/</a>	CFTRI	Food and technology	Peptides data	Peptides data	865	.php
TBNET India: a national portal for Tuberculosis Initiative <a href="http://tbnetindia.ibioinformatics.org/">http://tbnetindia.ibioinformatics.org/</a>	Institute of Bioinformatics, Bangalore	Biotechnology	Clinical, epidemiological and molecular data	Clinical, epidemiological and molecular data		.htm
India Biodiversity Portal <a href="http://indiabiodiversity.org/">http://indiabiodiversity.org/</a>	Collaborative effort between five partner institutions	Biodiversity information		Biodiversity information		.htm
Chickpea Transcriptome Database CTDB <a href="http://www.nipgr.res.in/ctdb.html">http://www.nipgr.res.in/ctdb.html</a>	National Institute of Plant Genome Research	Biology	Sequence, annotation and comprehensive expression profiling data for the chickpea transcriptome.	Data to facilitate functional and applied genomics research in chickpea and other legumes.		.fasta
Ocean Data and Information System (ODIS) <a href="http://odis.incois.gov.in/">http://odis.incois.gov.in/</a>	Indian National Centre for Ocean Information Services	Earth sciences		COMAPS CTCZ GEOTRACES		.htm
Clinical Trials Registry India CTRI <a href="http://ctri.nic.in/">http://ctri.nic.in/</a> Clinicaltrials/login.php Indian Space Science Data Centre <a href="http://www.isssc.gov.in/">http://www.isssc.gov.in/</a>	National Institute of Medical Statistics Indian Space Research Organization ISRO	Health sciences Space science	Clinical trial data. Payload data and related ancillary data	Clinical trial data. Payload data and related ancillary data		.php JPEG and metadata



## Appendix B. Questionnaire used for eliciting responses from researchers.

Researchers' perception towards research data

1. Designation
2. School/Centre/Department
3. Research area
4. What do you specialize in?
5. What kind of raw data is generated while you pursue your research activities?
6. How do you organize, maintain, and preserve your data for future use?
7. Do you experience raw data storage problem?
8. Does the practice of data sharing exist in your discipline? Is there tradition to share data with others in your field?
9. Do you share your data with your colleagues or students?
10. Do you believe that raw data should be freely accessible for all to browse and use?
11. Are you aware of data repositories of your field? Do you submit your raw data to data repositories of your field?
12. Do you feel that the Central Library should offer some support in managing storing and archiving your research data for future use?
  - I. Yes, strongly feel
  - II. No, we can manage on our own
  - III. Can't say

## Appendix C. Subject areas and research interests of respondents.

Subject areas	Research interests
Physics	Nano materials Science and Research
Computer Science	Software design, algorithm writing, digital image processing, mobile and adhoc networks, distributed computing networks, cloud computing, wireless sensor Networks, Security and Resources Provisioning, NLP
Biotechnology	Molecular biology, plant virology, RNAi, molecular virology, cancer biology, molecular modelling of drug metabolism, computational quantum chemistry, human sensor, gas sensor, synthetic organic chemistry.

(continued)

## Appendix C. (continued)

Subject areas	Research interests
Applied Plant Sciences	Vegetable breeding. Plant–pathogen interaction towards identifying genes responsible for diseases or disease resistance in plants
Environmental Sciences	Renewable energy production and low cost waste water treatment technology, clean energy option, bioremediation, microbial enzyme technology, environment, water, glaciers, coastal zones
Economics	Agricultural Economics, Developmental Economics, Game Theory, Oligopoly theory, Auction theory
Sanskrit Studies	Machine translation, computational linguistics
Arts and Aesthetics	Documentary cinema, visual studies, regional and transnational cinema
Psychology	Social Psychology inter-group relations, stereotypes; discrimination, prejudices, social identity
Education	Environmental Education, Educational Technology
Library and Information science	Academic library services; e-resource management, IT applications in libraries
Language, Literature and Cultural Studies	Hindi fiction, Dalit (the marginalized) literature
International Studies	International Relations, North Africa; West Asian politics; Indian foreign policies; Palestinian refugees and United Nations

## Appendix D. Types of research data generated by respondents.

Subject areas	Type of data
Physics	Experimental data
Computer Science	Codes; textual, audio, video data, observation notes
Biotechnology	Observation notes, images, videos, sequence and culture

(continued)

**Appendix D.** (continued)

Subject areas	Type of data
Applied Plant Sciences	Pathogens in plant; gene structure, gene product; biochemical functions; DNA-protein interaction; protein-protein interaction; microscopic observations; instrument readings
Environmental Sciences	Readings, images, observation notes, experimental data
Arts and Aesthetics	Questionnaire data, notes, diaries; audio and video data, slides
Social Sciences and Humanities (Economics, Education, Library and Information Science, International Studies, Language, Literature and Cultural Studies, Sanskrit studies)	Observation notes, questionnaire data, sample data collected during field trips, interviews, videos, diaries, newspaper article, letters



International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 40–50  
© The Author(s) 2017  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216686983  
journals.sagepub.com/home/ifi



# **‘We have to make an effort with it’: Exploring the use of stages to help understand the personal information management needs of humanities and social science doctoral students managing dissertation information**

**Amber L. Cushing and Odile Dumbleton**

University College Dublin, Ireland

## **Abstract**

Fifteen doctoral students enrolled in a variety of humanities and social science programmes from a single metropolitan university in Ireland discussed their practices and habits associated with personal information management via three focus groups, in order to explore personal information management skills at different stages of the doctoral programme. Findings suggest that personal information management needs of doctoral students managing dissertation/thesis information can be categorised as three distinct stages: beginning, middle and end of the doctoral programme. In developing services to meet the needs of this population, information professionals may find it useful to classify such services via these three stages.

## **Keywords**

Academic libraries, doctoral students, instruction, outreach, personal information management

Submitted: 16 May 2016; Accepted: 10 November 2016.

## **Introduction**

According to Jones (2007: 453), personal information management (PIM):

refers to both the practice and the study of the activities a person performs in order to acquire or create, store, organise, maintain, retrieve, use and distribute the information needed to complete tasks (work related or not) and fulfill various roles and responsibilities.

As social science and humanities doctoral students are tasked with personally managing digital information over the course of the PhD programme, they often develop and/or adopt PIM practices and strategies that they may bring with them when they graduate and enter the next stage in their careers. While previous literature has addressed the PIM practices of academic researchers, these studies do not address the beginnings of where and when these researchers began to develop their skills (as doctoral students)

(Fear, 2011). Further, when literature has addressed the PIM practices of students, it has focused on undergraduates or graduate students as a whole, which ignores the fact that PhD students are distinct, often enrolled in programmes for years (Fourie, 2012; Hoffman et. al., 2008; Reed, 2015). A PhD student in their first semester is markedly different from a graduating scholar years later, in many ways. As students learn and develop into independent scholars in the process of their programmes, it is necessary to understand how their PIM needs develop in programme. Finally, while tools and advice exist to manage ‘big data’ which is more commonly used in the sciences, there is less applicable advice available

## **Corresponding author:**

Amber L. Cushing, School of Information and Communication Studies, University College Dublin, Stillorgan Road, Belfield, Dublin 4, Ireland.  
Email: [amber.cushing@ucd.ie](mailto:amber.cushing@ucd.ie)

to students in social sciences and humanities who utilise data that is less likely to be massive in scale (Reed, 2015; Sheeja, 2010). Such an understanding can be of use in designing services to meet the needs of these students.

Over the years that these scholars develop, they may turn to academic library instruction for support in utilising different tools, such as citation management systems and metrics display. There is little research about the instruction needs of doctoral students at different stages of programmes, to inform the development of services to support these budding scholars. Thus, in order to inform the development of library services for doctoral students, this exploratory research seeks to understand the ways in which humanities and social science doctoral students manage their personal information for their dissertation/PhD thesis, at different stages of their doctoral programmes: how do humanities and social science doctoral students organise their personal information related to and during the dissertation/thesis process?

## Literature review

A study that explores the PIM practices of social science and humanities doctoral students can be informed by previous PIM studies as well as library instructional strategies to meet the needs of social science and humanities doctoral students. While there is a long history of studying the information-seeking behaviour of students and post-PhD scholars, less is known about their PIM strategies (Catalano, 2013). Knowledge of PIM concepts can assist instructional librarians in their attempt to better understand the information and instructional needs of doctoral students.

### *Personal information management and library work*

According to Jones (2007: 453), personal information management:

refers to both the practice and the study of the activities a person performs in order to acquire or create, store, organise, maintain, retrieve, use and distribute the information needed to complete tasks (work related or not) and fulfil various roles and responsibilities.

Jones (2008: 46) described PIM in the context of activities that describe 'the essential activities of PIM'. These activities include: keeping activities, finding/re-finding activities, and meta-level activities; including organising; maintaining; managing privacy and the flow of information; measuring and evaluating; and making sense. In addition, Barreau

(1995: 327) defined a personal information management *system* as 'an information system developed by or created for an individual for personal use in a work environment'. As doctoral students often personally manage several types of information during the doctoral process, PIM is used in this study to refer to PIM systems adopted and/or developed by doctoral students.

Previous work has recognised the value of PIM theory and concepts as they apply to academic library work. Fear (2011), Fourie (2012) and Otopah and Dadzie (2013) all recognised the use of PIM in academic reference work: while Fourie (2012) suggested that it is up to academic reference librarians to raise awareness of PIM and its usefulness to library users, Otopah and Dadzie (2013) suggested that an understanding of PIM behaviours could be imbedded in information literacy programs for college students. Fear (2011: 74) came to the conclusion that 'there is no bright line between data management and what is more often considered personal information management'.

Many researchers believe academic librarians are in a unique position to be of help during this time of transition to digital practice (Cox et al., 2012). Lush (2014, 49) proposed that a librarian who understands the core PIM activities of their patrons 'can also equip them to better identify the PIM needs of their users, and design and provide appropriate support services to meet them', highlighting the need for librarians to both train and be trained in the PIM area. Lush (2014: 47) further argued that librarians trained in the fundamentals of PIM are 'better equipped to manage their own personal information more effectively, as well as identifying and meeting the PIM needs of library users'. Writing from the perspective of public librarians, Cushing (2016) agreed that PIM skills can be useful in librarianship, as public librarians are increasingly sought out by the general public for assistance with information management issues. Simply having access to a wealth of tools is not enough to enact changes in information-seeking and -saving behaviours: users need guidance to implement these behaviours.

### *Library services for doctoral students*

Previous research concerning the information needs and wants of doctoral students in the social sciences does not specifically focus on PIM. Most studies applied only to information produced by students of science; to humanities students below the level of doctoral student; or mixes doctoral students in a population with academic faculty (Emanuel, 2013;

Melles and Unsworth, 2015; Reed, 2015). In their needs assessment of science students, Hoffman et al. (2008: 1) found that ‘many student focus group participants said they would like subject-specific workshops’ as that would better address ‘library research skills within the context of their disciplines’. The conflict between these two distinct user groups’ information-seeking behaviour was further probed in Sheeja’s study of PhD students from several universities in India. Using students from both disciplines, her study found that ‘although similarities exist...there are significant differences as well’ (Sheeja, 2010: 529). Sallans and Lake (2014: 87) noted that among all research students: ‘the tools and practices are still lagging and generally see slow adoption rates’. Carpenter (2012: 3) found that information literacy itself ‘has not improved with wider access to technology’. Students need to be first introduced to the tools and then further supported during their continued long-term use of the tool.

### *Existing library outreach relevant to doctoral students*

In attempting to gain an understanding of doctoral student practices to inform the development of library services, it is relevant to explore how academic librarianship has reported existing outreach practices and issues related to this demographic. While several definitions of outreach exist, most outreach programmes aim to reach potential users outside of the library. In academic libraries, outreach often takes the form of library instruction, but can also include tactics such as the tailored use of librarians and tools training and support aimed at the specific disciplines. Where outreach does exist, it is typically limited to engaging students at all levels and stages and is catered to their specific processes (Reed, 2015). For example, in the medical and scientific disciplines it is common practice to construct research groups that carry out research creation and output as a team (Mackenzie, 2014). There is no such bespoke research strategy for the social science student, which appears to be isolated to undergraduate and graduate students, with no mention of PhD students.

One example of a successful humanities digital outreach project comes from a Rice University case study that focused on a Digital Scholarship Services (DSS) team created for an Institute of Museum and Library Services (IMLS) funded project, the Our Americas Archive Partnership Collection (Henry, 2014). The project was in support of faculty needs, but the project team serves as an example of how the use of ‘embedded librarians’ could be of greater help to humanities scholars. Embedded librarianship ‘takes

a librarian out of the context of the traditional library and places him or her in an “on-site” setting or situation that enables close coordination and collaboration with researchers or teaching faculty’ (Carlson and Kneale, 2011: 167). The DSS team used a data curation approach to the project, customising key processes in the workflow to further support the specific needs of the faculty such as distinct subject metadata, searchability in multiple languages, and image search. This case study embodies all of the approaches that could apply to using academic librarians in a more precise way to fully assist with the needs of social science doctoral students.

Part of outreach efforts should include better marketing of library services to PhD students in general. Further, an emphasis on ‘data management’ may be misguided, when doctoral students need assistance managing all their personal information – PIM may be a better term to use when marketing services. Mackenzie (2014: 168) states: ‘Libraries (and their librarians) are rarely identified by the research community as an important source for assistance with information’. In her survey of academic staff, Gabridge (2009: 15) found that ‘faculty do not often see librarians as being equipped to help them solve their data problems’: instead they seek the IT department for solutions to storage and access of research. She further emphasised the need for proper library outreach, stating: ‘it is in our institution’s best interest for librarians to demonstrate in compelling ways the strengths and capabilities libraries can bring to bear on these problems’.

According to Luce (2008: 44), evolving technology requires attention/training/upskill, stating: ‘there is a need for workflow tools that capture emerging communication modalities, and libraries and appropriate partners have the opportunity to fill that critical gap’. The use of reference management software and relevant usability studies is often discussed in conjunction with outreach related to information management (not PIM), but the discussion generally centred on the end-user. Tools most commonly reviewed were Endnote, Zotero, Mendeley, and RefWorks. ‘Each of these tools has its own strengths and weaknesses that will be of greater or lesser importance to different groups of users’ (Gilmour and Cobus-Kuo, 2011: 74). One thing they do have in common is how they work. In his article on the subject, Perkel (2015: 123) describes the background function of the software to work as such:

Typically, the process of dragging and dropping a PDF into an application window triggers the software to try to identify it using the DOI or title, and to retrieve relevant metadata (such as title, keyword and author names) from online servers.



This genre of software enabled users to organise, sort, search, tag, and usually store their personal library of PDFs along with their relevant metadata. The use of such applications has revolutionised how students conduct and create research, but these tools only work if researchers understand how to use them, and then include that behaviour in their research rituals. In their examination of information-management behaviour by humanities doctoral students and faculty, Melles and Unsworth (2015) studied the reference management preferences of social science graduate students and found that most would not use certain tools for managing personal information unless their peers were already using them. This work revealed a reluctance on the part of the Generation Y doctoral student to use new tools for research management: fellow students and peers were the major influence on whether or not these doctoral students decided to use a technology application and are their main source of hands-on help (Carpenter, 2012: 11).

There are various reasons for this phenomenon, including: knowing that the software exists and how to use it; continued support over the years as applications are changed and updated; as well as having access to trained staff that understands the software and can help with issues as they happen. Sustained and continuous outreach and support are the answers to this issue (Carpenter, 2012).

This research is important as we move generationally toward the formation of students raised entirely in a digital research world, but who remain lacking in properly developed social science research skills. Rowlands et al.'s (2008: 290) work proved to be a 'study that overturns the common assumption that the "Google generation" is the most web-literate'. This tells us that even the PhD students educated in an increasingly digital environment still require an introduction and proper instruction in PIM and related tools.

The research reviewed demonstrates that while some librarians have imbedded PIM concepts into their library instruction efforts and found this useful, little empirical research available has explored exactly how understanding specific student PIM strategies can be of use in understanding the needs library user populations. Little research exists on academic library services specifically designed to meet the needs of doctoral students, while outreach efforts in academic libraries often do not focus on doctoral students and their needs. An empirically developed understanding of the PIM of doctoral students could be used to develop academic libraries services and outreach for this population.

## Method

To understand the PIM of humanities and social science doctoral students managing their personal information related to completion of their dissertation/PhD thesis, we gathered data from a group of students via the use of targeted focus group sessions. Critical Incident Technique (CIT) was used to collect data on how these students engage with personal information management at different stages of the doctoral programme. This approach differs from existing literature, which typically attempts to understand students' practices in general, without parsing them by discipline, or considers doctoral students a single group, ignoring different stages within a doctoral programme. The length of time one is enrolled in a doctoral programme can differ, but can average between three and eight years, depending on the student and discipline. Such a long and varied length of time deserves greater attention in research on this demographic.

The adoption of the focus group method is a well-established qualitative technique for gathering data, whose potency only increases when used in combination with a second method (Morgan, 1996). In his study on focus groups and group interviews, Morgan (1996: 129) posits that: 'the advantages of focus groups can be maximised through careful attention to research design issues'.

Following this recommendation, we enhanced the focus group method with CIT, a methodology that provided much opportunity for careful abstraction of data from the informants (Flanagan, 1954). CIT, when applied to the development of the interview questions as a means to pointedly probe the blank spaces in the research, can reveal a new understanding of how this group manages information in the context of the higher levels of social sciences and humanities academic study.

Participants were invited to view a short presentation from two recent doctoral graduates (one in a social science discipline, the other in a humanities discipline) who discussed their PIM when completing their dissertations/PhD theses. The discussion of PIM in the presenter's PhD programme was grouped in three stages, according to the presenter: looking back, how the presenter wished to categorise the beginning, middle and end of their PhD programme. The beginning stages generally included the learning of how to select and manage personal information related to the PhD dissertation/thesis, identifying environmental influences (such as peer research behaviours), and understanding how to find support and resources such as seeking advice from peers, library services and

alternative sources, such as blogs. Middle-stage activities addressed using and understanding tools for ingest, access, and use, including considering data format types for storage and use issues. The end stage covers the creation of reliable backup actions, processes for dissemination and reuse of data, and long-term preservation planning.

These recent graduate presenters, one a new staff member in a social science department completed her PhD in 2012 and the other presenter, a postdoctoral researcher in a humanities institute, completed her PhD in 2014. Both presenters were chosen for their discipline as well as their ability to reflect on their PhD process a few years out.

These short presentations were situated as prompts to which focus group participants could respond. After the presentations, participants were divided into three focus groups. The questions posed to the focus group participants were structured using CIT: for example, groups were asked 'looking back, can you think of a time when they might have managed data differently?' As Flanagan (1954) suggested, the direction to participants to think back on an event of significance allows for researchers to effectively gather data about problems participants faced. The questions asked of the participants centred on using the three stages of the doctoral programme to elicit genuine responses and trigger memories.

The beginning stage included a question of how a student learned to manage their data, what influences had an effect on their methodology, and what resources they had that were helpful to them. Mid-stage research processes revolved around formats and the types of technology services offered by their school. Finally, the end-stage questions addressed preservation and backup planning, dissemination of data and its eventual reuse. At all stages participants were invited to cite memories of significant events that recreated their most memorable research experiences in relation to these stages and also refer to the presentations if it was helpful.

## Findings

Fifteen doctoral students from a single large metropolitan university in Ireland participated in three focus groups of four to six participants each. Participants represented all stages of the PhD programme: some were just starting out while others were close to submitting their dissertations/PhD theses. Groups included a mix of male and female participants. Focus groups were audio-recorded and then transcribed. Pseudonyms were added during the transcription process. Participant came from a variety of fields

including but not limited to architecture, English drama and film, archaeology, language and literature, philosophy, and applied social sciences.

One author used open coding 'by hand', following the memoing technique described in Corbin and Strauss (2008) and the coding technique described by Miles and Huberman (1994). Categories for analysis were developed from the interview guide as well as data that emerged from the focus groups. Once the general codes were developed, all transcripts were coded.

The authors have reported the findings by categorising characteristics associated with the beginning, middle and end stages of the personal information management needs associated with these stages of the doctoral programme. These stages do not necessarily correspond with year in programme, as all students advance at different paces. The goal of presenting findings with these three categories was to identify the characteristics that define each stage so that information professionals can better identify student needs according to stage in programme.

### *Beginning stage*

Findings suggest that the first stage of the doctoral research process is characterised by learning how to manage data, including tool selection, and an introduction to the challenges presented by pre-existing habits. Establishing good PIM habits that assist in re-finding such as file naming, storage, and executing proper backups in the early stages can pay off later when reviewing research previously collected. Even so, some students approached the doctoral process with existing work habits that needed to be enhanced, updated and sometimes relearned altogether. Thus, the beginning stage is also categorised by refining one's PIM habits to align with expectations of the research process.

For example, Jeanne described her prior role as an architect to exemplify PIM habits she formed while working in a paper environment, suggesting the transition to digital was challenging for her: 'It's my training as an architect, but I find that I physically like to place things'.

When one has previously worked with paper-based resources, changing habit to work with digital information could be a challenge. However, for those who had previously worked with digital resources, the transition to managing personal information related to a doctoral research programme was not as difficult. Peter stated: 'I've worked in an office for 15 years so I had certain file management techniques in terms of how you put the date and how you do correspondence et cetera, I'd just brought over'.

This carryover of work practice behaviour was found in the discussion of several other students, one who referenced her Masters' thesis work processes and another citing his undergraduate degree behaviours. Doctoral students often have work experience before entering their programme. This finding suggests it may be of use to library professionals to understand a doctoral student's background when making PIM recommendations, especially for those students in the beginning stages of their research programme. This can include a suggestion to evaluate previous habits developed in a working environment to assess which habits may continue to be useful in the doctoral programme and other habits that would benefit from change.

A lack of knowing how to use available technology was another detriment to those in the first stage. Many of the students spoke of a general cloud of confusion around tool use that prevented them from utilising available technologies. Donald described his fear of making a bad decision when choosing to use a new system: 'if I'm embedded in EndNote, with however many hundred references, am I really going to convert (to something else) now I'm midway?' This fear of making the wrong choice is reflected in his perceived wasting of any investment of time and money.

Other fears hold some students back from making any decisions at all toward using particular management systems. Peter summed up his frustration with the vast array of choices in technology, stating that he is a mature student who graduated in 1998, 'and the difference from then to now, in every dimension is mind-blowing'. While Donald explained his lack of formal use of any research management technology in stating that he sees 'technology as more of a barrier than anything else, than something that helps' and 'I'm just really bad with technology'. Apprehension of technology is a common theme which appears to dissipate when proper training is offered. Said Donald: 'the idea of having everything correctly formatted and correctly stored, collated, is the kind of comforting idea' that would have helped him understand the benefits of using technology to manage his data in the early stage.

While academic librarians have long engaged in instructing students how to use reference tools, it is worth considering the role of information overload doctoral students face along with the implications of long-term tool use. As some doctoral students can spend nearly a decade within a single programme, their concerns over future upkeep and use of a tool could be included in services for this demographic. Services for this beginning stage could focus on

developing criteria for selection of a tool, rather than just jumping in to training for a specific system.

In the initial stage of research, even those that had adopted use of management tools found the process of creating a file naming schema daunting. Sarah created a system for her filenames so that she could use her laptop's browser to search for files, but felt that her needs changed in her third year and had to rethink her entire research method, a time consuming and tedious task. Others found it difficult to decide how to create a filename system, perplexed by using variations of customised themes, dates, or formats. Noah was confused by the task of file structuring, wondering: 'it's like the naming of the files and whether it should be thematic, whether it should be by author, whether it should be chronological, whether – you know'.

Most participants pondered how one could possibly know what might be the best way to structure the files in the early stage, when they do not yet have the proper experience or perspective that sometimes happens in more advanced stages of the programme.

This is where instruction on PIM practices may be of use to doctoral students. As Lush (2014), Fourie (2012) and Otopah and Dadzie (2013) suggest, academic librarians with an understanding of PIM are better equipped to meet the needs of students, such as the issue of file naming. In addition to developing criteria for tool selection, assessing one's PIM may be a good place to start, in order to drive tool criteria for selection. Development of a solid criterion to select tools may alleviate much of the anxiety with making the 'wrong' decision that many of our participants discussed.

Whichever naming convention one used, it was important to build in flexibility for future stages. Several participants noted that what you might find useful in the first year can be less so in the third year. Niamh stated: 'I think in a humanities PhD the likelihood is that you're not going to know exactly' how or what to save, and that it would be 'impossible' to predict which processes would work from the start of the PhD.

The cost of using and maintaining reference management systems emerged early on as a theme, as it had an influence on some participants' selection of an initial system. EndNote was often mentioned, and the cost associated with its use was discussed by Susan when she stated: 'that when you leave [university] you lose access to EndNote, unless you continue to pay for it'.

Her university in particular was locked into a licensing scheme where the University is only allowed to recommend, use, and train students on EndNote, to the exclusion of all other proprietary or open access

systems. Alternatively, some students recommend using Zotero for reference and citation management and PDF storage. When considering switching to Zotero, Niamh asked Susan about the cost of the programme, noting that with ‘just about everything you have to pay. But [I think] it’s something like 20 USD a year? It’s not very much’.

While cost is considered to be an issue in considering tool choice, it did not appear to be an obstruction for these participants. In reviewing reference management systems, information professionals may want to address cost, as well as legacy concerns when supporting doctoral students. While Emanuel (2013) found that cost is a factor in citation management system adoption rates, studies such as Melles and Unsworth (2015) demonstrate that it is not always considered an issue considered when querying users in research design.

### *Middle stage*

The middle stage of doctoral research comprised what we believe to be a major issue in humanities and social science PhD research: the selection of tools and their cross compatibility with various data formats, as well as issues of data loss. While the first stage is characterised by more of an analysis of and transition in PIM habits and discovering information about management systems and considering options, it is typically in the middle stage of a programme that students make a decision and fully commit to use of a system. With this commitment and an increase in the development of files, come concerns over data loss. Among the three focus groups conducted, a wide selection of various types of tools were mentioned. The reference management tools EndNote, Zotero, and Mendeley were discussed the most often. For data creation formats, Microsoft Word, Excel, Google Sheets and Google Docs were the most popular. Storage application use ranged from Google Drive, Dropbox, and OneDrive. The most cited data formats included Docs, JPGs, PDFs and proprietary template files.

A priori themes for the middle stage of the PhD research process include more specific tool training (versus selection advice, as was a characteristic in the beginning stage) and PIM strategies and support. Students felt that they had randomly come across advice on a tool or technique as opposed to a formally organised introduction to tools and techniques available in their discipline from the University. Peter said that he had learned about a method called Mind Mapping from a friend he spoke with at a funeral: ‘I find the whole thing quite random. It’s quite sort of serendipitous’.

Several informants described at least one example of learning about a method from word of mouth, advice from friends or as Jennifer puts it, ‘kind of stumbled [sic] on our own’. Some had relatively good luck with advice from peers, while others did not, as disciplinary differences coloured how research was developed and maintained.

This finding suggests that while there is room for formal library instruction about tools and support for managing research, informal avenues also supply instruction. Academic librarians may wish to explore the promotion of informal data gathering as a method to inform adoption of research management tools and techniques. This is similar to the ways in which some scholars may behave – they utilise their academic network and colleagues’ suggestions when determining how to develop a project (Talia, 2002). If doctoral students are considered to be scholars in training, it would be of use to provide instruction in networking skills as a way to gather information about how to manage a research project.

Our participants came from a mixture of fields including architecture, English drama and film, archaeology, language and literature, philosophy and applied social sciences. While most of our participants spoke of managing data, the variety of disciplines represented here suggests that our students managed different types of formats for their data collection. Different formats meant that different tools were required, i.e. Excel for spreadsheets, Photoshop for images, Dropbox for audio interviews, etc. Issues of cross-compatibility were common among users of many formats especially in the data collection and reading sections of the early and middle stages. Donald reflected on personal information management and the humanities student in particular:

we have to make an effort with it, and maybe humanities more – people are less likely, particularly we don’t have to do data collection or storage or things like that. We’re just more likely to sit cross-legged in a chair, read a book, make a few notes, whenever.

An emergent theme for the middle stage was the topic of information loss, which was set in context here as computer crashes, USB stick loss, or laptop theft. Examples of hardware loss were exacerbated when there was no backup plan in place, something common to each of the stories told around this theme. Mary reported ‘learning a lesson’ when she lost a USB stick she was using to transport her data files between computers, although she was spared disaster when she discovered a backup copy. Another participant explained that she had been using an external



hard drive to mirror her work computer on her home computer, and:

when I went home from work in the evening, I could upload what I did for the day and just work away, like I had carried the laptop home. But then one day the sketch up on that computer crashed and I had all these files I didn't know what to do with.

She reported having since changed her method of transporting files to the use of cloud storage so that she would not find herself at such a loss again. The use of external hard drives, servers and various brands of cloud storage were popular among students, citing easier and better management of backups and proper use of applications that can almost effortlessly transfer files from one computer to another. This suggests that offering doctoral students basic instruction in personal preservation systems, such as creating a LOCKSS (Lots of copies keeps stuff safe) system, would be of use specifically in this stage. At the beginning stages doctoral students typically would not have accumulated enough information to be concerned with loss.

The topic of loss naturally led to recommendations for creating and preparing a viable preservation process of backups for storage and safety. Many cited the use of Dropbox, Google Drive and emailing information to oneself as methods of backup, and most had come to form some sort of system for backing up their data collection and work. Sarah noted that part of her process was to insert Word files with notes on them into her folders of PDFs, that way: 'I know these articles are meant for which purpose in future'.

Several of the participants reported working in paper notebooks that they digitised and stored with their other files, although this method was not without concern. Mary found that she was 'petrified of spilling coffee on this [notebook], or it falling out of my bag or something like that, so everything – as much as possible – is backed up'.

### End stage

In the final stage of doctoral studies, participants were most concerned about how to manage all of the information that had been accumulated and created during the research process. This included long-term preservation planning, access and reuse of data, and deciding what to keep and what to discard. Most of the students used quotes such as 'mountains of data' and 'piles of things' to describe their curated research.

In archival practice, there is a term for controlling data inventory known as 'weeding', which Pearce-Moses and Baty (2015: 25) define as 'the process of

identifying and removing unwanted materials from a larger body of materials'. In defining PIM activities, Jones (2008) refers to this activity as maintaining. Selecting what to discard was a common theme of worry among the participants, usually addressed or noticed during the transition between middle and end stages. Niamh stated that in her experience, it was almost too easy to save everything in the early stages, and that:

sometimes you have too much stuff. I have a massive library of stuff I don't need, and now I'm at the point where – I'm now at the point where I need to clear out stuff in a different folder because it's no longer relevant.

Echoing this need for a standardised personal practice, Jennifer points out that weeding was good when you 'get this huge dataset but you don't use everything, so then having to cut through it anyway'.

In the end, participants were left with data that required a long-term preservation plan, which includes issues of access for reuse, and storage and safekeeping of research. Most participants did not understand that they needed to provide continued care for their data, especially if they were to publish or reuse the datasets in any way. Only one student mentioned the use of an institutional repository, and that was in the context of belonging to a team of researchers who have their own database. When asked directly, none of the participants had considered what format their data would take or where it would exist in 20 years. This provides another point of entry for librarians to assist doctoral students in the end stage with managing data: how to manage preservation at the end of a project.

One group of participants had a discussion around the emerging trend of the sharing of data in the field of humanities. Some noted it as a funder requirement, and others believed that the trend was motivated by social media and an open access environment that encourages the sharing of data 'because of the impact you have on society'. Bob believed that the use of social media for the sharing of data got to the heart of data sharing and contributing to public ideas discovery: 'that's what we are supposed to be doing in the humanities, is [sic] about dissemination and research'.

This presents a potential opportunity to offer final-stage doctoral students instruction about how using sound practices to manage data allows for easier access and sharing of material. It can also serve as an entry point to discuss more formal PIM practices related to one's future career, if students pursue academic staff positions. Reflection on PIM



during the thesis/dissertation process could be used as reflective tool, to inform how to move forward in one's career as research professional, where demands of and responsibilities associated with PIM in research projects will continue to grow, along with other demands on one's time.

## Conclusion

In this study, we have attempted to categorise PIM needs during the social science doctoral research process via three distinct stages: beginning, middle, and end of the doctoral programme. We believe these stages are useful as they provide organisation to the needs of this demographic, that can then be more easily act as a frame of reference for information professionals. Mackenzie (2014) and Gabridge (2009) both concluded that academic librarians are not being capitalised upon in a meaningful way by doctoral students, and while we do not attempt to provide specific advice for instruction and outreach practices, we suggest a consideration of these three stages may be useful for academic librarians to consider in planning their programmes.

As concluded by Cox et al. (2012), the findings from this study suggest that academic librarians are in a unique position to provide support to social science doctoral students, and understanding their personal information needs at the different stages in programme is necessary to provide the best possible help. Time spent in doctoral programmes is typically longer than any other education programme, setting them apart from undergraduate and Master's programmes. Therefore, it may be ineffective to lump all doctoral students together when one student may have a few months' experience in programme, while another student could have five or more years of experience in programme. Findings demonstrated different needs associated with each stage.

The beginning stages of the doctoral programme were characterised by the need to consider and perhaps shift PIM habits from the workplace and or previous experiences to the research atmosphere, as well as concern over selection of tools and software to support management of resources. In this stage it may be of use for information professionals assisting this demographic to assist these students with analysis of their PIM habits in order to better understand what can be usefully applied to doctoral work and what might need to be shifted. Further, assisting these students in development of a criterion that they can utilise to assess the effectiveness of tools and software to support their information needs is critical. Specifically, development of a criterion may alleviate

concern over 'making the wrong choice' and could be addressed before offering tool instruction.

Findings demonstrated that students were ready to select a tool to support PIM practices at the middle stage of their programme, which would be the best time for librarians to offer tool instruction. It was also at this time when librarians might address concerns over loss and emphasise the usefulness of good maintaining habits, including backup of systems. Finally, helping students understand how to utilise a personal network to evaluate personal information tools and concepts may be of use and could apply to academic networking, in general.

In the end stage of the programme, participants expressed a desire to develop an understanding of weeding or appraisal (to use an archival term) of personal digital information. As weeding and appraisal were long standing strategies used by the library and information professional, it may be of use to translate the core of these concepts for use in PIM.

Overall, the use of conceptual stages may assist information professionals in better understanding the needs of doctoral students, leading to the developing of better strategies and practices to work with this population. Future research could explore the development of these strategies and practices.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References

- Barreau D (1995) Context as a factor in personal information management systems. *Journal of the American Society for Information Science* 46: 327–339.
- Carlson J and Kneale R (2011) Embedded librarianship in the research context: Navigating new waters. *College and Research Library News* 72(3): 167–170.
- Carpenter J (2012) Researchers of tomorrow: The research behaviour of Generation Y doctoral students. *Information Services and Use* 32(1/2): 3–17.
- Catalano A (2013) Patterns of graduate students' information seeking behaviour: A meta analysis of the literature. *Journal of Documentation* 69(2): 243–274.
- Cox A, Verbena E and Sen B (2012) A new role for academic librarians? Research data management. *Multimedia Information & Technology* 38(4): 29–30.
- Corbin J and Strauss A (2008) *Basics of Qualitative Research*. 3rd edn. Thousand Oaks, CA: SAGE.
- Cushing AL (2016) 'If it computes, patrons have brought it in': Personal information management and personal

- technology assistance in public libraries. *Library & Information Science Research* 38(1): 81–88.
- Emanuel J (2013) Users and citation management tools: use and support. *Reference Services Review* 41(4): 639–659.
- Fear K (2011) 'You made it, you take care of it': Data management as personal information management. *International Journal of Digital Curation* 6(2): 53–77.
- Flanagan JC (1954) The critical incident technique. *Psychological Bulletin* 51(4): 327. Available at: <https://www.apa.org/pubs/databases/psycinfo/cit-article.pdf> (accessed 12 January 2017).
- Fourie I (2012) Collaboration and personal information management (PIM). *Library Hi Tech* 30(1): 186–193. DOI: 10.1108/07378831211213292.
- Gabridge T (2009) The last mile: Liaison roles in curating science and engineering research data. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC* 265(August): 15–21.
- Gilmour R and Cobus-Kuo L (2011) Reference management software: A comparative analysis of four products. *Issues in Science and Technology Librarianship* 66(Summer): 63–75. DOI: 10.5062/F4Z60KZF.
- Henry G (2014) Data curation for the Humanities: Perspectives from Rice University. In: Ray JM (ed.) *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, IN: Purdue University Press, pp. 355–366.
- Hoffmann K, Antwi-Nsiah F, Feng V, et al. (2008) Library research skills: A needs assessment for graduate student workshops. *Issues in Science and Technology Librarianship* 53(Winter): 1. DOI: 10.5062/F48P5XFC.
- Jones W (2007) Personal information management. *Annual Review of Information Science and Technology* 41: 453–504.
- Jones W (2008) *Keeping Found Things Found: The Study and Practice of Personal Information Management*. Boston, MA: Morgan Kaufmann.
- Luce RE (2008) A new value equation challenge: The emergence of eResearch and roles for research libraries. In: CLIR (ed.) *No Brief Candle: Reconceiving Research Libraries for the 21st Century*. Washington, DC: Council on Library and Information Resources. Available at: <https://www.clir.org/pubs/reports/pub142/luce.html> (accessed 12 January 2017).
- Lush A (2014) Fundamental personal information management activities – Organisation, finding and keeping: A literature review. *Australian Library Journal* 63(1): 45–51. DOI: 10.1080/00049670.2013.875452.
- MacKenzie E (2014) Academic libraries and outreach to the sciences: Taking a closer look at research groups. *Science & Technology Libraries* 33(2): 165–175. DOI: 10.1080/0194262X.2014.914011.
- Melles A and Unsworth K (2015) Examining the reference management practices of Humanities and Social Science postgraduate students and academics. *Australian Academic & Research Libraries* 46(4): 1–26. DOI: 10.1080/00048623.2015.1104790.
- Miles M and Huberman AM (1994) *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: SAGE.
- Morgan DL (1996) Focus groups. *Annual Review of Sociology* 22: 129–152.
- Otopah FO and Dadzie P (2013) Personal information management practices of students and its implications for library services. *ASLIB Proceedings* 65: 143–160.
- Pearce-Moses R and Baty LA (2005) *A Glossary of Archival and Records Terminology*. Chicago, IL: Society of American Archivists.
- Perkel J (2015) Eight ways to clean a digital library. *Nature* 527(7576): 123–124. DOI: 10.1038/527123a.
- Reed RB (2015) Diving into data: Planning a research data management event. *Journal of eScience Librarianship* 4(1): 1–6. DOI: 10.7191/jeslib.2015.1071.
- Rowlands I, Nicholas D, Williams P, et al. (2008) The Google generation: The information behaviour of the researcher of the future. *Aslib Proceedings* 60(4): 290–310. DOI: 10.1108/00012530810887953.
- Sallans A and Lake S (2014) Data management assessment and planning tools. In: Ray J (ed.) *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, IN: Purdue University Press, pp. 87–107.
- Sheeja NK (2010) Science vs. social science: A study of information-seeking behaviour and user perceptions of academic researchers. *Library Review* 59(7): 522–531. DOI: 10.1108/02425351080001367.
- Talia S (2002) Information sharing in academic communities: Types and levels of collaboration in information seeking and use. *New Review of Information Behavior Research* 3(1): 143–159.

### Author biographies

**Amber L. Cushing** is currently a Lecturer/Assistant Professor at the School of Information and Communication Studies, University College Dublin in Ireland where she also serves as Director of the MSc in Digital Curation Programme and Director of the PhD in Information and Communication Studies Programme. She previously served as Director of the Allenstown, NH Public Library while managing her consulting business, Cushing Information Consulting. She has taught graduate courses in Digital Curation, Research Methods, Information Resources and Services, Archival Access and Technology Issues and Archival Outreach. Cushing holds a PhD in Information and Library Science from the University of North Carolina at Chapel Hill, an MLIS with a concentration in Archives Management from Simmons Graduate School of Library and Information Science, and a BA in History from Mount Holyoke College. Her main area of study concerns how individual perceptions of value in digital collections can influence institutional practices in LAMs.

**Odile Dumbleton** is a doctoral student at the School of Information and Communication Studies at University College Dublin. Her research interests lie in the study of digital cultural heritage objects and how changes in user perception might impact the viability of digital projects in a LAM context. She has held several roles including archivist, librarian and collections assistant at

a variety of cultural heritage institutions including the Chicago Field Museum, the Metropolitan Museum of Art, the MTA Museum and the Brooklyn Public Library. She holds an MSc in Digital Curation from University College Dublin, an MLS with Archives Certificate from Pratt Institute and a BA from DePaul University.



# Managing research data at an academic library in a developing country

Shamin Renwick, Marsha Winter and Michelle Gill

University of the West Indies, St Augustine, Trinidad and Tobago

International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 51–64  
© The Author(s) 2017  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216688703  
journals.sagepub.com/home/iff



## Abstract

Managing research data has become an issue for many universities. In the Caribbean, the St Augustine Campus Libraries at the University of the West Indies are keenly aware of the need to support researchers in this regard. The objectives of this study were to identify current practices in managing research data on the campus and to determine a possible role for the Campus Libraries. A pilot study of 100 researchers on the campus was conducted. Analysis of the 65 valid responses revealed that while researchers owned data sets they had little knowledge or experience in managing such. This low level of awareness is instructive and validates a role for the Campus Libraries to play in supporting researchers on campus. The Campus Libraries need to sensitize researchers about what data planning and managing research data entail as well as provide technical assistance with actual data storage.

## Keywords

Academic libraries, Caribbean, data management, research data management, research data service, University of the West Indies

Submitted: 16 May 2016; Accepted: 15 December 2016.

## Introduction

Managing research data has become more significant as universities make increasing requests for research funding in the face of shrinking national budgets in the Caribbean. In this troubled economic climate, there is a need to ensure there is value for money and that data management practices produce research data that has integrity so that it can be curated, reused, shared and further analysed. Data management should then be treated as a *sine qua non* and is aptly described by Kenneth Pimple, an academic whose work involves research ethics, as a “necessary twin to the scientific method” (Coates, 2014: 599). Research data is being viewed by funders as an asset and hence there is, no doubt, a growing requirement for universities to include data management plans in research grant proposals (Pryor, 2012: 4).

Many researchers at the University of the West Indies (UWI), St Augustine, Trinidad and Tobago, although actively engaged in research, appear to not have fully grasped the importance or benefits of managing their research data and, more so, the need to develop data management plans as part of their

research grant proposals. The UWI St Augustine Campus Libraries (referred to as “the Campus Libraries” from here on) recognize the need to implement a data management policy across the entire campus, given the need for academics to engage in research which will drive the growth and development of the Caribbean nations that contribute to the funding of the UWI. Consequently, The Campus Libraries seek to clearly identify their role in supporting researchers in the managing of their research data and, further, provide guidance on data management planning.

For the purposes of this study, the definition by Whyte and Tedds (2011) of data management is considered useful, that is: “the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results”. In developed countries, such as the United States and the

## Corresponding author:

Shamin Renwick, Alma Jordan Library, University of the West Indies, Trinidad and Tobago.  
Email: shamin.renwick@sta.uwi.edu

United Kingdom, the challenges of managing research data have led to the establishment of various initiatives that facilitate the sharing and preservation of data. In addition, on the UWI St Augustine Campus, the staff that support researchers have noted that international funding agencies ask for data plans as part of their application process for research grants. These funding agencies are conscious of the need to encourage scientific good practice and to achieve greater value for the research they sponsor, widely encourage – indeed, increasingly require – particular standards of data management and sharing to be followed. However, in many developing countries, these issues are only now being addressed as technologies become more affordable, widely available and better known.

UWI, a higher education institution located in the English-speaking Caribbean, highlights achieving excellence in research as one aspect of its mission. It envisions research as an avenue for providing creative solutions to addressing the challenges within the region, as well as helping to explore the region's potential for development, and formulating sound decision making and public policy (University of the West Indies Alma Jordan Library, 2015). Along with this assertion is the possible concern for managing not just large volumes of data, also described as “big data”, but all data, which emerges from engaging in academic research. Proper management of research data will therefore become an important part of the research cycle process. Shen and Varvel (2013: 552) advocate that the employment of data management practices in the early stages of the research process is prudent, so that data loss or discard could be prevented, hence facilitating data sets being discoverable, re-usable and easily shared.

Within the Latin America and the Caribbean, many initiatives are now emerging to ensure the proper management of research data. These include the “creation of policies in government funding agencies” and the “implementation of data repositories in universities and research institutions” (Andaur, 2016), albeit universities and research institutions in the Caribbean region have been outpaced in the establishment of many such initiatives by their counterparts in the developed nations. In addition, there is a still a great need to sensitize all researchers about these initiatives and their relevance.

The Campus Libraries have grown conscious of the need to provide data management support and envisage a role to emulate these initiatives and, with the leveraging of available information and communication technologies, hope to provide sound data management support services at the UWI.

The Campus Libraries consist of a network of libraries with one main library – the Alma Jordan Library (AJL) – and seven subsidiary or branch libraries. The AJL is described in its mission statement as the “repository and gateway for information in the support of the teaching, learning and research needs of the University community and Caribbean society” (University of the West Indies Alma Jordan Library, 2015). Being dedicated to its mission, it is continuously evaluating its potential and its position within the University so that it remains relevant to the needs of its research community. One way of ensuring relevance is to “design flexible research support services that will intersect with the research cycle at various points” (University of the West Indies Office of Research, Development and Knowledge, 2016: 23). Support for managing data is currently being explored as one type of research assistance that can be provided by the Campus Libraries to their chief stakeholders at UWI.

## Background

UWI was first established as the University College of London in 1948 and became an independent university in 1962 (University of the West Indies, 2012: vii). It comprises four campuses: the St Augustine Campus in the Republic of Trinidad and Tobago; the Mona Campus in Jamaica; the Cave Hill Campus in Barbados and the Open Campus within the various territories throughout the English-speaking Caribbean. The University is financially supported by 17 countries in the Caribbean and has a student enrolment of over 50,000.

UWI is recognized by its local and international partners, and other stakeholders, as a centre for excellence in research, especially on “matters pertaining to the Caribbean and other small island states” (University of the West Indies Alma Jordan Library, 2015). The UWI's St Augustine Campus comprises 405 full-time academic staff members in seven faculties: Engineering, Food and Agriculture, Humanities and Education, Law, Medical Sciences, Science and Technology, and Social Sciences. Each faculty has variable number of departments and affiliated research centres, the two major research centres being: the Cocoa Research Centre and the Seismic Research Centre. The UWI STA has the largest student enrollment of the four campuses: in 2014/2015, there were 18,345 inclusive of 12,405 undergraduates and 5,765 postgraduates according to current campus records. See Appendix A for a breakdown of the number of academic staff and postgraduate students at the St Augustine Campus.



### *Current research support*

Of the six strategic themes emerging from UWI's Strategic Plan 2012–2017, “research and innovation” emphasizes the need for the UWI to “create an enabling environment to support, foster and increase the output of high quality research and innovation with an emphasis on the Caribbean” and to “increase funding and strengthen research partnerships” (University of the West Indies, 2012: 6). It is these goals which have led to the establishment of various research initiatives, though quite disparate, at the UWI St Augustine Campus.

One project, the Research Information Management System (RIMS), is a tool used to identify researchers at UWI with specific knowledge and skills. RIMS allocates each researcher a profile in RIMS where they can update personal information, learn about current research activities on the campus, access internal funding sources and locate information on and apply for internal and external grants. Through RIMS, researchers can access training and assistance with the development of research proposals (University of the West Indies Office of Research, Development and Knowledge, 2016).

Another venture, initially established by the AJL in 2008 to capture and provide a centralized location and access to research generated by researchers at the UWI St Augustine Campus, is the UWI's institutional repository, UWISpace. It was formally adopted by the entire UWI system in 2012. However, unfortunately it was found that, based on the Library's experience, the repository was not the first place of choice for researchers wishing to archive their research. This could be attributed to several reasons including lack of awareness; workload issues or that they do not really understand the usefulness of the repository in providing access to their research

Yet another activity provided by UWI is the Trinidad and Tobago Research and Development Impact (RDI) Fund. This initiative is in keeping with the UWI's strategic objectives in the area of research “to support projects that address pressing development challenges and that will achieve recognizable and substantive impact in the short and medium term (3–5 years)” (University of the West Indies, 2015). This fund, provided by the Trinidad and Tobago Government but managed by the Office of the Principal of the St Augustine Campus, offers a maximum of US\$300,000 to researchers to develop projects in priority areas such as crime, violence and citizen security, public health, climate change and environmental issues, finance and entrepreneurship, technology and society, and economic

diversification and sector competitiveness. Since the establishment of the fund in 2012, 85 concept notes have been received and 31 grants totaling over US\$2,000,000 have been approved and awarded. The RDI Fund programme managers are now seeking to develop data management plans to assist researchers and are hoping that by starting a data management conversation all the personnel from RIMS, RDI Fund and the Campus Libraries would be included.

### *Overview of data management support*

The paper (Erway, 2013), “Starting the conversation: University-wide research data management policy”, succinctly outlines the key stakeholders, the issues and the questions that should be addressed when implementing data management policies and plans at an institution. The key stakeholders identified were the University, since the research is viewed as the property of the institution; the Office of Research; the ICT Department who ensures that the ICT infrastructure exists to support the data management initiative at the macro level. Erway (2013: 9) underscores that “[the] cyber infrastructure environment can offer advantages such as economies of scale, integration, and a focused approach to co-ordinating technology and expertise, computing power and the planning, acquisition, and management of storage space”.

Other key stakeholders are the researchers, academic units (faculties) and the library. Erway (2013: 10) argues that the library is poised to be a “key player in data management curation and preservation given its extensive experience with selection, metadata, collections, institutional repositories preservation, curation and access”.

Therefore, at the start of the data management conversation, the main areas to be considered by the various stakeholders are: data ownership – where clear policies must indicate who owns the collected data; funding agency requirements or criteria; kind and type of data worth keeping given cost concerns; content of data management policies – including time limits of data maintenance and retention as well as what should happen to the data when the researcher leaves the university; ethical considerations – such as security protocols involved and treatment of sensitive data; tracking data usage and its impact on promotion and tenure; access controls – that is, who should have access and whether embargos are involved. Finally, decisions on who should fund the management of the research data must also be considered.

## Literature review

### *Academic libraries and data management services*

The literature is replete with initiatives undertaken in the area of research data management by universities in developed countries. US librarians, in recent years, have taken a more active role in the lifecycle of research data as under President Barack Obama's administration, the Office for Science and Technology Policy (OSTP) issued a directive in 2013, followed by a memo and a plan that mandated data management plans and access to research data for federally funded projects. Therefore, it is not surprising that US universities are ahead of their counterparts in the UK, Australia, New Zealand and Ireland in the area of data management (Pinfield et al., 2014: 3).

Although academic libraries have been involved in storing, managing and archiving data, these institutions have now been compelled to implement research data management policies and programmes. Some research data management services have been led by the library but, generally, university technology services, research offices, data-intensive departments or groups spearhead initiatives to provide data services (Henderson and Knott, 2015). Additionally, the focus on research data opens opportunities for libraries to aid researchers in preparing and executing management plans and the libraries that become engaged with data management gain the benefit of deeper involvement with their constituents (Breeding, 2016: 17). This association implies a commitment that requires the allocation or reallocation of resources, both staff and technical infrastructure (Breeding, 2016). The identification of other data services provided on campus not only prevents duplication of services but also leads to opportunities for collaborations. The initiation of this relationship between the library and data service providers can lead to the development of a community of practice for research (Grynoch, 2016).

### *Surveys of data management needs and practices*

Surveys have been conducted to determine the data management practices and services offered at institutions (Buys and Shaw, 2015; Tenopir et al., 2015; Whitmire et al., 2015). The study by Buys and Shaw (2015) revealed that researchers at a university in Chicago stored data in a variety of ways: on departmental services/external drives – 50%; computer hard drive – 66%; flash drives (which limit data sharing and long-term preservation) – 27%; and cloud storage (such as Dropbox) – 31%. In terms of retention trends, both the humanities and the sciences preferred to

retain all types of data indefinitely. Respondents, asked about the types of training and assistance needed, disclosed that they required guidance with providing long-term access to their research data, preservation services for data storage and back-up during active projects. Tenopir et al. (2015: 17) investigated the levels of research data services (RDS) offered by universities in the US and Canada. They discovered that collaboration between the library and researchers was fruitful and provided benefits to all. And that few academic libraries have hired data librarians because it is was perceived by some institutions that there was not adequate demand for RDS services to warrant hiring a full-time data librarian.

### *Types of support offered*

As Erway (2013) alluded, there are a number of considerations and issues that must be addressed when implementing RDS. The literature provides guidance to any institution wishing to establish such an initiative.

There exist various flavours of data management services offered. Generally, the data service support provided by the library help in crafting data management plans, guidance on data management throughout the lifecycle, and data set archiving and dissemination (Swanson and Reinhart, 2016: 98). RDS include providing to the campus researchers, training or active involvement in data management planning, data management guidance during research, research documentation and metadata, research data sharing, and curation of completed projects and published data (Fearon et al., 2013). Universities have adopted several approaches when initiating their RDS. It can be implemented on a needs basis, as in the case of the Virginia Commonwealth University which initially did not develop a specific plan but sought instead to be open to the needs of their researchers, as the new Director of Data Management learned about researchers, staff, faculty, students, resources and focused on developing a solid communication plan to reach the target audience (Henderson and Knott, 2015). Conducting a pilot which provides curating services for the data that graduate students produce leading up to their electronic theses and dissertations can be a starting point for some universities (Creamer, 2015; Doty et al., 2015). University of Virginia libraries built RDS to provide data discovery, acquisitions and research software support expertise in the use of restricted data (Clairbourn, 2015). In this case, a statistician was hired to head the library's data research team and the library "experimented with the concept of blended librarianship, creating teams of scientists, social scientists, data scientists and library experts".

Developing a team approach by relying on pre-existing skillsets is also a strategy that can be adopted by universities implementing RDS. Libraries are already skilled in collecting, sharing, curating and preserving information, and can provide support in the three key competencies of research data management which are: providing access to data, advocating and supporting management, and managing data. For example, cataloguers can provide assistance with description and determining suitable metadata standards; liaison librarians can bring knowledge of discipline research practices; and reference skills may be applied in the interviews and questions with researchers prior to developing data management plans (Gynoch, 2016).

### *Challenges of implementation of research data services*

The implementation of data services is not without challenges. Tenopir et al. (2014) report in a survey of academic librarians at US and Canadian universities that while data services are developing rapidly and librarians can provide subject knowledge support to researchers, they felt more training was needed to be effective. In a study by Goldman et al. (2015: 8), they noted that librarians identified conferences, Internet-based learning and on-the-job training as primary preparation for librarians to provide data services.

Marketing and communicating to the key stakeholders on campus are critical when developing data services on a campus. The University of Montana identified three strategies for marketing data management services to researchers – partner up, be social and simplify, since obtaining buy-in from the key stakeholders is crucial (Mannheimer, 2014: 42). Mannheimer asserts that the libraries should partner with the campus stakeholders, such as the IT department and those responsible for research on the campus, and a unified front must be demonstrated to researchers. Additionally, in order to gain their trust, Mannheimer underscored the need to communicate with stakeholders which includes using social media to raise awareness at the university about data management.

As noted in the literature, the studies identified document, in general, the experiences of universities in developed countries but the literature is silent regarding a developing country perspective. In the US, the development of academic libraries' RDS has been largely driven by the government's mandate to make publicly-funded research accessible. Even though the Trinidad and Tobago Government has been funding research, this perspective has not yet taken root in the Caribbean.

The Campus Librarian, having grown conscious of the need to provide data management support to the researchers at the St Augustine Campus, convened a Data Management committee to commence examining the ways in which the Campus Libraries could provide such assistance to researchers on the campus with managing large quantities of data generated from their research. One of the first strategies identified was to conduct an exploratory survey to determine the extent to which researchers were engaged in data management and their accompanying needs as well as the role that the library could play in this respect.

### **Methodology**

In order to explore the area of interest identified, a decision was made to conduct a pilot study at the UWI St Augustine Campus Annual Research Expo, held September 2015. In the social sciences, a pilot study may be conducted as a feasibility study in preparation of a larger study or can be done to test a survey instrument. Van Teijlingen and Hundley (2001: 1) outline several reasons for conducting a pilot study which include using it as a means of assessing the full-scale study, collecting preliminary data, determining the resources needed for a planned study, designing research protocols and determining if they are effective or not. There are also limitations which include making inaccurate predications based on the pilot data. While it is not the norm to publish pilot studies, they can be useful because they provide details about lessons learned.

Results from this pilot study would inform and guide a more detailed future study, possibly including the other UWI Campuses, as well as help to facilitate decision making as it relates to the level and type of input required by the Campus Libraries in the short term.

### *Study population and sample*

The study population was the researchers at the St Augustine Campus, namely the 5,765 postgraduate students and the 405 academic staff, i.e. 6,170 researchers. The composition of academic staff and postgraduates by Faculty and Centre is outlined in Appendix A.

It was decided to target a sample 100 researchers for the purpose of this pilot study. As a result, 100 print copies of the questionnaire were available at the Campus Libraries' booth at the Campus 2-day Annual Research Expo in September 2015 and library staff manning the booth approached faculty and postgraduate students asking them to complete the questionnaire. The results of this survey were used to prepare a mini-report that was submitted to the Campus Librarian and is the foundation of this paper.

### Survey instrument: questionnaire

The survey instrument consisted of nine questions. Questions 1–4 identified researcher status on the St Augustine Campus and helped to determine whether they were involved with data collection/management or analysis within the last 10 years; the kind of research data with which they were involved; and how much data was collected. Questions 5–7 required respondents to identify their back-up and storage preferences, and what was done with the data once the research was completed. Question 8 sought to establish the researcher's need for assistance and in what specific area. The final question invited researchers to share comments or concerns about data management.

All responses were collated in Microsoft Excel. SPSS 18 software was used to produce descriptive and cross tabulation statistics. A sample of the questionnaire used in this study is shown in Appendix B.

## Results

### Response rate

Of the 100 questionnaires distributed, 65 valid responses were completed and returned: as such, there was a 65% response rate. The number of responses (65) is the equivalent of approximately 0.01% of the study population (6,170 researchers).

### Respondents

Table 1 shows the distribution of responses per faculty and for each research centre.

Respondents were from six of the seven faculties on the UWI St Augustine Campus as well as two main research centres. Each faculty has a varied number of departments and the sample also represented respondents from 27 departments and 52 subject disciplines. The respondents reflected a fair representation across the St Augustine Campus of persons engaged in research. See Appendix A for breakdown of the researchers and their Faculty/Centres.

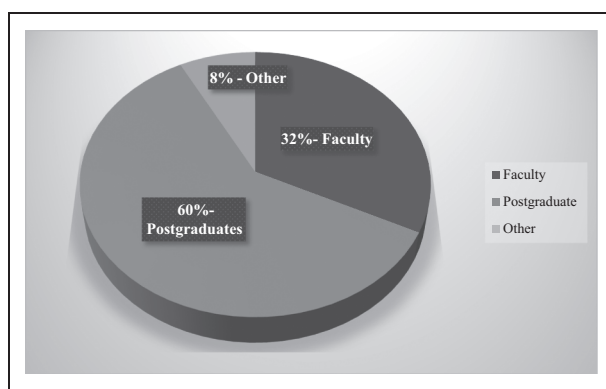
### Question 1 – Respondent status

The researcher status included faculty, postgraduate students and other persons engaging in research on the UWI St Augustine Campus. Figure 1 shows a breakdown of the respondents by status: faculty, postgraduate students or other category.

There was a ratio of 1:2 of faculty: postgraduate students. The “other” category included: a “graduated postgrad” and a “media specialist”. Academic staff was represented at each faculty except at the Faculty of Food and Agriculture. There were no postgraduates

**Table 1.** Faculties and centres participating in the survey.

Faculty	Centre	Frequency	%
Engineering		12	19
Food and Agriculture		9	14
	Cocoa Research Centre	5	8
Humanities and Education		5	8
Medical Sciences		7	11
Science and Technology		7	11
	Seismic Research Centre	1	2
Social Sciences		16	25
Total		65	100



**Figure 1.** Status of respondents.

at the two centres as postgraduates are associated with the relevant faculties. The breakdown of the status of the respondents by Faculty/Centre is shown in Appendix D Table D1.

### Question 2 – Length of time respondents involved in data management

The length of time correspondents have been involved in data management is shown in Table 2.

Approximately one-third of respondents (32%) had only recently (i.e. <1 yr) got involved in managing data and about one-quarter of respondents had been managing data for over five years. See Appendix D Table D2 for a breakdown of length of time respondents have been handling data by Faculty.

### Question 3 – Types of research data used

Table 3 shows the types of research data managed by respondents.



**Table 2.** Length of time respondents involved in managing data.

Length of time	Frequency	%
< 1 yr	21	32
2–3 yrs	14	22
3–5 yrs	13	20
>5 yrs	17	26
Total	65	100

**Table 3.** Types of research data managed.

Type of data managed	Frequency	Probability
Survey	40/65	0.62
Time Series	8/65	0.12
Experiment	27/65	0.42
Observation	30/65	0.46
Multimedia	7/65	0.11
Other	2/65	0.03

**Table 4.** Size of data managed.

Size	Frequency	%
<100GB	46	71
100–500GB	8	12
>500GB	11	17
Total	65	100

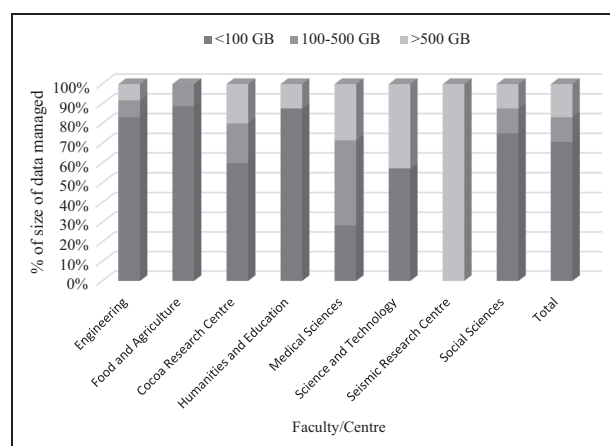
In terms of data managed, the probability of the most used (in descending order) was: survey, observation, experimental, time series and multimedia. In the category of other, the following is a list of methods of data collection identified by the researchers:

- examination
- field survey
- historical (Tobago)
- manage all data for IGDS (Institute of Gender and Development Studies)
- qualitative analysis on literature
- surveillance.

#### Question 4 – Size of data managed

Table 4 and Figure 2 illustrate the size of data managed by respondents.

It was found that few respondents (17%) were using data >500GB. Of this category, three researchers were from the Faculty of Science and Technology as well as two from Medical Sciences and two from Social Sciences. Somewhat surprisingly, it was revealed that none of the respondents in the Faculty of Engineering were using data sets > 500GB.

**Figure 2.** Bar graph of size of data managed by respondents.**Table 5.** Methods of storage.

Storage method	Frequency	Probability
Email myself	38/65	0.58
Flash drive	43/65	0.66
External hard drive	35/65	0.54
Cloud storage	26/65	0.40
Second computer	16/65	0.25

Most of the respondents (71%) were involved in data of <100GB in size. Of this category, 12 researchers were from the Faculty of Engineering; 10 from Social Sciences and eight from Food and Agriculture. See Appendix D Table D3 for a breakdown of the size of data managed by Faculty.

#### Question 5 – Methods of data storage utilized

Table 5 shows the methods of storage where respondents were asked to tick all that apply.

In order of preference, the probability of the storage methods most used was: flash drives, email oneself; external hard drives; cloud storage and, last, using a second computer. Other storage methods noted were: hardcopy, hard drive, multiple computers, Synology 8T drive.

#### Question 6 – Methods of back-up utilized

Preference for back-up methods is shown in Table 6 where respondents were asked to tick all that applied.

It was found that the descending order of the probability of using a back-up method was: external hard drives; email oneself/flash drives; cloud storage and using a second computer.



**Table 6.** Back-up methods.

Back-up method	Frequency	Probability
Email myself	32/65	0.49
Flash drive	32/65	0.49
External hard drive	39/65	0.60
Cloud storage	23/65	0.35
Second computer	14/65	0.22

**Table 7.** Data use after project.

Comment	Frequency	%
No comment	12	18
Publish	10	15
Store	29	45
Delete	2	3
Share	3	5
Further analysis	7	11
I plan to use the data collected to inform my collection of short stories	1	1
Return to lecturer (undergraduate project - collection of data, Department of Physics)	1	1

**Question 7a – What is done with data after project**

What respondents did with their data after the research project is outlined in Table 7.

Of note is that 45% stored; 15% published and 11% said further analysis was done on their data.

**Question 7b – Plans for data after project**

Table 8 illustrates the respondents' stated plans for the data after the project.

Though 55% did not respond to this question, 12% wanted to retain data for further analysis and 11% wanted to publish and store their data.

**Question 8 – Assistance required**

The assistance that respondents claimed they needed regarding their data management is shown in Table 9. Respondents were asked to tick all that applied.

The descending order of the probability that the support required will be used was: archiving for long term preservation; the associated permissions to do so; organizing for easy retrieval; storage and back-up.

**Question 9 – Concerns and comments**

Respondents were asked to identify any further concerns and comments. As this was a pilot project, this was left as an open-ended question to discover concepts which were not identified in earlier questions.

**Table 8.** Proposed plans for data after project.

Comment	Frequency	%
No comment	36	55
Publish	7	11
Store	7	11
Delete	1	1
Share	3	5
Further analysis	8	12
Generate a report to submit to superiors	2	3
Papers; background info for project funding	1	1

**Table 9.** Assistance required.

Assistance required	Frequency	Probability
Storage	16/65	0.25
Back-up	13/65	0.20
Archiving	28/65	0.43
Retrieval	19/65	0.29
Permission	20/65	0.31

**Table 10.** Concerns and comments.

No.	Categories	Number of comments and concerns in categories
1	Back-up, storage and preservation issues	13
2	Data management policy	6
3	Data analysis	3
4	Training	1
5	Other (which was a comment on the study itself)	1
6	No comment	41

There were 13 comments and 11 concerns and these have been coded and shown in Table 10.

The categories of comments/concerns in order of the number of responses were (1) back-up/storage and preservation; (2) data management policies; (3) data analysis; and (4) training. One comment indicated that the library's interest in data management was an excellent initiative. Appendix C details the concerns and comments as stated on the received questionnaires. Though 41 (63%) of respondents did not have further concerns or comments, the comments that were noted illustrate the wide-ranging areas of support required by researchers.

**Discussion**

Major findings from 65 respondents in this study, which reflect perspectives from a developing country,

were that managing data was relatively new as about one-third of respondents had only recently, i.e. less than one year, got involved in managing data and one-quarter had been managing data sets for >5 years. In terms of data collection, survey methods are the most popular with the others most used in descending order being observation, experimental, time series and multimedia. Using large data sets (i.e. >500GB) was relatively low among respondents, over 70% still dealt with data sets >100GB. This finding was a revelation for the authors who expected that large data sets would have been in greater use.

The preferred methods for storage and back-up methods were mainly flash drives, external hard drives or emailing oneself. Using cloud storage or a second computer was not as heavily used. However, probability of storage methods being used, such as, flash drive usage (0.66) and cloud storage (0.40) were greater than at a university in Chicago where these methods were used at 27% and 31%, respectively, according to Buys and Shaw (2015). Use of external hard drives was similar in both studies but emailing oneself as a method of storage was not mentioned in the US study. Of note is that after a project less than 50% stored data, only about 15% published their data and even less (11%) retained data for further analysis.

Regarding assistance identified, most wanted help with archiving of data and in order of descending probability were: help with permissions for use of the data; easy retrieval; storage and back-up of data.

Both comments and concerns identified proved useful and instructive; they identified the knowledge or concepts UWI researchers had on data management and also assisted in identifying possible roles the Campus Libraries may be able to play in assisting researchers in data management. Based on the most noted area of concern (back-up/storage and preservation) The Campus Libraries can prioritize strategies to assist in this area and the development of data management policies. Providing training in data analysis and data management is another area that can be considered. These activities should be done in collaboration with other University departments who also have an interest in the area of data management. Possibly, as well, the concept of data literacy may be included in information literacy workshops for postgraduate students and in meetings with the Faculty.

## Conclusion

This pilot instrument yielded valuable perspectives on data management on the UWI St Augustine Campus in Trinidad and Tobago, a developing country. Though the Campus Libraries had concerns that big

data may have been an issue, it was revealed that not many of the researchers in this pilot study used large data sets. However, this may increase in the near future as technology evolves, and the generation, as well as availability, of data increases. It was established that there was a need for assistance and policy from UWI researchers to manage data in general, not necessarily of only of big data sets, efficiently. There was no conclusive comment that the Campus Libraries, themselves, needed to actively provide a data management service as the question regarding assistance did not specify any specific “assistance from the Campus Libraries”. Nevertheless, it is apparent from the findings and comments that there is a role for the Campus Libraries in both raising this issue for consideration by the University’s administration as well as providing advice and technical support on the topic of data management to researchers.

## Limitations

It can be noted that whilst 65 responses may be adequate to get a general idea of developments in current data management needs and practices, the cross tab analysis showed that this number was inadequate to reflect specific Faculty considerations. (See some of the cross tabulations analysis in Appendix D tables.)

## Recommendations

Many respondents were unable to identify the types of assistance which the library specifically could provide in helping them to manage research data. This low level of awareness is instructive, and suggests the Campus Libraries need to inform researchers of the type of service they can provide with planning for data storage, protecting, archiving, storing and preserving the data as well as retention and retrieval methods to allow for further analysis later on. The Campus Libraries would also need to work with the various research departments on the campus to develop a data management policy. It is recommended that in order to do this the Campus Libraries should form two teams, one responsible for sensitization, advocacy, and policy development; and a second team to undertake the technical support of preparing data plans as well as the tasks involved of providing a data service.

## Further research

Further work identified for the Campus Libraries can be undertaken in either of two methods:

- A longer survey with more detailed questions shared widely on campus to gain an overview

of exactly who may want assistance and support from the library. This survey should try to determine the services the library can provide and whether researchers would be receptive and fully utilize these services.

- Using either purposive or snow-balling sampling methods to identify persons by who currently have data management needs and conduct either focus groups or personal interviews.

These approaches would assist the library to determine the most appropriate way forward and the optimal role that can be undertaken, accordingly. In addition, a cross-campus study would also serve to make comparisons and yield even more rich results, and the sharing of these would be of tremendous benefit to the entire University.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### References

- Andaur G (2016) Research data management in Latin America and the Caribbean: An overview. Available at: <http://learn-rdm.eu/research-data-latin-america> (accessed 12 August 2016).
- Breeding M (2016) Systems librarian. *Computers in Libraries* 36(3):15–17.
- Buyts CM and Shaw PL (2015) Data management practices across an institution: Survey and report. *Journal of Librarianship and Scholarly Communication* 3(2). DOI: 10.7710/2162-3309.1225.
- Clairbourn MP (2015) Bigger on the inside: Building research data services at the University of Virginia. *Insights* 28(2): 100–106.
- Coates H (2014) Ensuring research integrity: The role of data management in current crises. *College & Research Libraries News* 75(11): 598–601.
- Creamer A (2015) Current issues and approaches to curating student research data. *Bulletin of the Association for Information Science and Technology* 41(6): 22–25.
- Doty J, et al. (2015) Making student research data discoverable: A pilot program using Dataverse. *JLSC* 3(2): 1–25.
- Erway R (2013) *Starting the Conversation: University-Wide Research Data Management Policy*. Dublin, OH: OCLC Research.
- Fearon D, Gunia B, Pralle B, et al. (2013) SPEC Kit 334: Research Data Management Services. *Association of Research Libraries*. Available at: <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334> (accessed 15 November 2015).
- Goldman J, Kafel D and Martin ER (2015) Assessment of data management services at New England Region Resource Libraries. *Journal of eScience Librarianship* 4(1): 1–17.
- Grynoch T (2016) Implementing research data management services in a Canadian context. *Dalhousie Journal of Interdisciplinary Management* 12. DOI: 10.5931/djim.v12.i1.6458.
- Henderson ME and Knott TL (2015) Starting a research data management program based in a university library. *Medical Reference Services Quarterly* 34(1): 47–59.
- Mannheimer S (2014) Ready, engage! Outreach for library data services. *Bulletin of the Association for Information Science and Technology* 41(1):42–44.
- Pinfield S, Cox AM and Smith J (2014) Research data and libraries: Relationships, activities, drivers and influences. *PLoS ONE* (12): 1–28.
- Pryor G (2012) *Managing Research Data*. London: Facet.
- Shen Y and Varvel VE (2013) Developing data management services at the Johns Hopkins University. *Journal of Academic Librarianship* 39(6): 552–557.
- Swanson J and Reinhart AK (2016) Data in context: Using case studies to generate a common understanding of data in academic libraries. *Journal of Academic Librarianship* 42(1): 97–101.
- Tenopir C, Sandusky RJ, Allard S, et al. (2014) Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research* 36: 84–90.
- Tenopir C, Hughes D, Allard S, et al. (2015) Research data services in academic libraries: Data intensive roles for the future? *Journal of eScience Librarianship* 4(2): e1085. DOI: 10.7191/jeslib.2015.1085.
- University of the West Indies (UWI) (2012) *Strategic Plan*. Available at: [https://www.mona.uwi.edu/opair/strategic-plan/UWI+Strategic+Plan+2012-2017+\(Final\).pdf](https://www.mona.uwi.edu/opair/strategic-plan/UWI+Strategic+Plan+2012-2017+(Final).pdf) (accessed 29 April 2016).
- University of the West Indies (UWI) (2015) *RDI Fund: The UWI-Trinidad and Tobago Research and Development Impact Fund*. Available at: <https://sta.uwi.edu/rdifund> (accessed 29 April 2016).
- University of the West Indies (UWI) Alma Jordan Library (AJL) (2015) *About UWI*. Available at: <http://sta.uwi.edu/about> (accessed 4 May 2016).
- University of the West Indies (UWI) Office of Research, Development and Knowledge Transfer (ORDKT) (2016) *Research and Information Management System*. Available at: <https://sta.uwi.edu/ordkt/rims.asp> (accessed 15 May 2016).
- van Teijlingen ER and Hundley V (2001) The importance of pilot studies. *Social Research UPDATE* 35. Available at: <http://sru.soc.surrey.ac.uk/SRU35.pdf> (accessed 15 May 2016).
- Whyte A and Tedds J (2011) *Making the Case for Research Data Management*. Edinburgh: Digital Curation Centre. Available at: [www.dcc.ac.uk/resources/briefing-papers/making-case-rdm](http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm) (accessed 9 November 2015).

Whitmire AL, Boock M and Sutton SC (2015) Variability in academic research data management practices: Implications for data services development from a faculty survey. *Program: Electronic Library and Information Systems* 49(4): 382–407.

### Author biographies

**Shamin Renwick** is the Head, User Services, Alma Jordan Library, UWI, St Augustine Campus, Trinidad and Tobago. She has over 30 years' experience working in libraries and has been a school librarian; Head, Multimedia and IT Unit at a Medical Sciences library; and a Faculty Liaison Librarian for Science and Agriculture Faculties. In her current role, she is a member of the Library's senior management team and oversees all aspects of services to the academics, researchers and students including information literacy; collection development and maintenance; and outreach. Among other publications, she has co-edited a book entitled *Caribbean Libraries in the 21st Century: Changes, Challenges, and Choices*. She is a Fellow of the Chartered Institute of Library and Information Professionals (CILIP), UK and a Past President of the Association of Caribbean University, Research and Institutional Libraries (ACURIL). She is a recipient of several awards including the ACURILEANA Star 2007 for Research and Publication.

**Marsha S. Winter** is a librarian at the Alma Jordan Library, UWI, St Augustine, Trinidad and Tobago. She received her MLS from Syracuse University and has been in the field of librarianship since 2004. She has worked at the Heritage Library, National Library and Information Systems (NALIS) in Trinidad and Tobago as a reference and outreach librarian. She is currently responsible for content recruitment for the University of the West Indies institutional repository and is involved in the digitization of the AJL holdings. Her professional interests are digital libraries, institutional repositories and metadata.

**Michelle Gill** is a Faculty Liaison Librarian in Science and Agriculture at Alma Jordan Library, UWI, St Augustine Campus; in the Republic of Trinidad and Tobago. She has a BSc Major in Chemistry with Minors in Mathematics and Biochemistry. Ms Gill is a teacher by profession but entered the field of librarianship and completed her MLIS at the University of the West Indies. She has varied research interests some of which include:- emerging information communication technologies in libraries, information security of open source resources used in libraries, disaster preparedness planning in libraries, and library space redesign efforts. Her current research focuses on examining a role for academic libraries and librarians in research data management at the UWI St Augustine Campus.

## Appendix A. Number of academic staff and postgraduate students at the St Augustine Campus

Faculty	Centre	No. of academic staff		No. of postgraduate students		No. of researchers
		Sub-total	Total	Sub-total	Total	TOTAL
Engineering			72		1137	1209
Food and Agriculture		31	34		228	262
	Cocoa Research Centre	3		0		
Humanities and Education			76		958	1034
Law			10		29	39
Medical Sciences			71		531	602
Science and Technology		63	74	500	504	578
	Seismic Research Centre	11		4		
Social Science			68		2378	2446
<b>TOTAL</b>			<b>405</b>		<b>5765</b>	<b>6170</b>

Source: Internal UWI St Augustine records from the Registrar's Office and the Campus Information Technology Services (CITS), 2016.

## Appendix B – Data management planning questionnaire

The St Augustine Campus Libraries is conducting a survey to determine data management practices of UWI Faculty and postgraduate students who engage in research on the UWI St Augustine Campus.

The information gathered from this survey will ultimately be used to craft a sustainable Data Management Plan for the University of the West Indies, St Augustine Campus. Thank you for your participation

Name: \_\_\_\_\_ Faculty: \_\_\_\_\_

Dept.: \_\_\_\_\_ General Subject area: \_\_\_\_\_

1. Kindly tick your responses.

Faculty ☐ Postgraduate Student ☐ Other \_\_\_\_\_

2. Have you been engaged in or assisted with the collection/ management/analysis of data in the last ten years?

Less than a year ☐ 1–2 years ☐ 3–5 years ☐ More than 5 years ☐

3. What kind of research data did you collect/manage/analyse?

Tick all the options which apply.

- a. Survey – questionnaires, focus groups, interviews ☐
- b. Time series ☐
- c. Experimental ☐
- d. Observation ☐
- e. Multimedia ☐
- f. Other \_\_\_\_\_

4. How much data was collected?

<100GB ☐ 100–500GB ☐ >500GB ☐

5. How do you currently store your data? (Tick all those which apply)

- a. Email myself ☐
- b. Flash drive ☐
- c. External hard drive ☐
- d. In the Cloud ☐
- e. Second computer ☐
- f. Other \_\_\_\_\_

6. How do you currently back-up your data? (Tick all those which apply)

- a. Email myself ☐
- b. Flash drive ☐
- c. External hard drive ☐
- d. Cloud computing ☐
- e. Second computer ☐
- f. Other \_\_\_\_\_

7. After your research project, (a) what have you done with your data or (b) plan to do with your data?

(a) \_\_\_\_\_

(b) \_\_\_\_\_



8. With regard to managing your data, do you require assistance in any of the following areas? (**Tick all the options which apply**)

- |  |                          |
|--|--------------------------|
| a. Storage   | <input type="checkbox"/> |
| b. Back-up (current files)                               | <input type="checkbox"/> |
| c. Archiving of digital files for long-term preservation | <input type="checkbox"/> |
| d. Easy retrieval of data                                | <input type="checkbox"/> |
| e. Permission to use/gain access                         | <input type="checkbox"/> |

9. Are there any concerns or comments you would like to share regarding data management?

---



---

Thank you for completing this survey!

## Appendix C- Comments and concerns

No.	Comments (A)	Concerns (B)
1	Active data management policy must be implemented on campus	Accessibility of data to students
2	Archival databases - these should be better popularized to students who are not typically involved in research	Analysis of data
3	Excellent initiative	Confidentiality & data will not be erased
4	General info on how to manage, safeguard would be helpful	If I store my data with the university will the university have claim to all my research/data
5	I am of the view that data management must be done correctly and properly	Is there any software that allows for easy data storage, back-up and update on multiple computer systems
6	I understand the risks of using flash drives/external hard drives to back up data but I am not convinced about the security of the Cloud	Ownership; IP rights; proper handover of data
7	It is easy to lose or misplace data if it isn't labelled correctly. Ensure when saving/storing data that it is name appropriately	Privacy & protection
8	Just having someone to market the process of data mgmt. (knowledge-wise) easier and less cumbersome and challenging	Security, patent details
9	Make SPSS more simple that data management can be simplified	Storage of the volumes of data presents challenges in terms of appropriate filing and retrieval
10	Overall UWI guidelines for storage and naming of files would be helpful across the campus	Synchronizing data between hard drive and external drive
11	Should have a workshop on this; very helpful for new students - where to start	The main concern is storage and retrieval of data for public and UWI internal access for future use
12	The space provided for staff on the intranet is too limited and frequent deleting is necessary	
13	UWI should take advantage of cloud [storage] such as Moodle and Google drive. Current system cluttered and inefficient	
Total	13 Comments	11 Concerns

## Appendix D: Cross tabulation analysis by Faculty and Centres

**Table D1.** Status cross tabulation.

Faculty * Status cross tabulation		Status			Total
Count		Academic staff	PostGrad	Other	
Faculty	Cocoa Research Centre	4	0	1	5
	Engineering	5	6	1	12
	Food and Agriculture	0	9	0	9
	Humanities and Education	3	5	0	8
	Medical Sciences	5	2	0	7
	Science and Technology	1	5	1	7
	Seismic Research Centre	1	0	0	1
	Social Sciences	2	12	2	16
Total		21	39	5	65

**Table D2.** Data management time cross tabulation.

Faculty* Data management time cross tabulation		Data mgt time				Total
Count		<1 yr	2–3 yrs	3–5 yrs	>5 yrs	
Faculty	Cocoa Research Centre	0	0	1	4	5
	Engineering	3	3	1	5	12
	Food and Agriculture	4	3	2	0	9
	Humanities and Education	4	0	3	1	8
	Medical Sciences	2	1	1	3	7
	Science and Technology	3	0	3	1	7
	Seismic Research Centre	1	0	0	0	1
	Social Sciences	4	7	2	3	16
Total		21	14	13	17	65

**Table D3.** Data size cross tabulation.

Faculty * Data size cross tabulation		Data collection			Total
Count		<100GB	100–500GB	>500GB	
Faculty	Cocoa Research Centre	3	1	1	5
	Engineering	10	1	1	12
	Food and Agriculture	8	1	0	9
	Humanities and Education	7	0	1	8
	Medical Sciences	2	3	2	7
	Science and Technology	4	0	3	7
	Seismic Research Centre	0	0	1	1
	Social Sciences	12	2	2	16
Total		46	8	11	65



# Building professional development opportunities in data services for academic librarians

International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 65–80  
© The Author(s) 2016  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216678237  
journals.sagepub.com/home/iftl



**Suzanna Conrad**

California State University, USA

**Yasmeen Shorish**

James Madison University, USA

**Amanda L. Whitmire**

Stanford University, USA

**Patricia Hswe**

Andrew W. Mellon Foundation, USA

## Abstract

Research data management represents a significant professional development area for academic librarians – significant for its growing importance to the profession, since researchers are increasingly expected to comply with research data management requirements, and for the extent of competence needed by librarians to support researchers in research data management practices and plans. This article recounts how the Association of College and Research Libraries is fostering professional development opportunities in research data management. The authors describe two key endeavors: (1) the development and deployment of a needs assessment survey, which allowed insight into the types of librarians expressing the most need; and (2) planning and implementation of a pre-conference workshop for ACRL 2015, intended to prototype a future professional development offering. The article concludes by discussing additional assessment that was done following the workshop and how the pre-conference laid the foundation for proposing a “roadshow” for research data management, similar to what the Association of College and Research Libraries sponsors for scholarly communication.

## Keywords

Academic libraries, data management, data services, professional development, professional organizations

Submitted: 15 May 2016; Accepted: 6 October 2016.

## Introduction

Academic libraries have responded to changes in scholarly communication in myriad ways. Shifts in that landscape include the proliferation of data as a scholarly product and various funder requirements regarding the management of that data. Members of the Association of College and Research Libraries (ACRL) in the United States of America have responded to these changes through a series of developments, aimed at providing sustainable professional

development support to librarians in the area of research data management (RDM).

The authors are engaged in various organizations that concern RDM, such as the United Kingdom Digital Curation Centre Associates Network, Research

---

### Corresponding author:

Yasmeen Shorish, James Madison University, MSC 4601  
Harrisonburg, VA 22807, USA.  
Email: shorisyl@jmu.edu

Data Alliance, the International Association for Social Science Information Service & Technology, the Digital Library Federation, and the ACRL, have presented and published on the topic (Hswe, 2015; Hswe and Holt, 2011; Reisner et al., 2014; Shorish, 2015; Whitmire, 2013, 2015), and three of the four authors have served on the Executive Committee for the ACRL Digital Curation Interest Group. Through these broad experiences and associated perspectives, the authors formed the hypothesis that subject liaison librarians are an underserved population with respect to RDM professional development. Literature in the field gave support to this hypothesis (see Literature review). Survey assessment can direct professional development efforts and ensure that scope and purpose are derived from the best evidence available. This article details the strategies employed, including surveys, to develop a sustainable and targeted RDM professional development opportunity for academic librarians, with a focus on liaison librarians, which may have utility for other library and information science professional organizations, globally.

## Literature review

Most researchers have not been formally trained to manage their own data. In 2010 Borgman declared that the “data deluge” had arrived; she estimated that 90% of the world’s data had been created in the two years preceding her publication. Little (2012) acknowledged the struggles that academic libraries face in keeping pace with these trends, yet Scaramozzino et al. (2012) claimed that assisting faculty with their research data needs can be a growing role for university libraries. Nicholson and Bennett (2011) emphasized the library’s role in the “scholarly enterprise” as a bridge between researchers and institutional repositories to make data management part of library managed repositories’ processes and goals.

University libraries such as those at MIT, University of Wisconsin-Madison, University of Edinburgh, and University of Southampton are offering web-based guidance on best practices to researchers. Some groups or libraries, including the University of Minnesota and the California Digital Library, provide workshops on data management, individual discussions or consultations, or other outreach activities. Many libraries are using their institutional repositories as a place to house research data and fulfill grant requirements including examples from ScholarSphere at Penn State, Purdue University Research Repository (PURR), and Cornell University’s DataStaR. Librarians are currently being trained to manage data as well, i.e. University of Illinois has a Data

Curation Education Program, University of Tennessee is a partner in the Data Curation Education in Research Centers program, Syracuse University has an eScience librarian track, and the University of North Texas offers a program called Information: Curate, Archive, Manage, and Preserve.

The DuraSpace/ARL/DLF E-Science Institute is a program that was developed to help libraries create a strategic response to e-research support needs and involves a small team of individuals, including a library administrator, a data librarian, and a non-library participant (Duraspace, n.d.). Outside of an institution, regional “science boot camps” have been developed to aid science liaison librarians in keeping abreast of the changes in the field, which often includes some RDM component. The most long-running of these is the New England Science Boot Camp (e-Science Portal for New England Librarians, n.d.). DCIG offers professional development opportunities to support these new emerging roles for practicing librarians and archivists through its spring webinar series and programming at ALA Midwinter and Annual conferences (ALA Connect, n.d.).

The roles of assisting with the management and curation of research data are frequently falling under the library’s purview despite limited training opportunities and ever evolving best practice. Heidorn (2011) cited a need for librarians to pursue training in data curation and e-science, since existing skill sets leave them well positioned to assume duties in this area. Several articles and reports have been written that indicate liaison librarians are a target audience for RDM training (Cox et al. 2012; Jaguszewski and Williams, 2013; Rockenbach et al., 2015). Some programs, such as at Purdue University Libraries, have gained more traction within their campus communities (Witt, 2008, 2012), while other teaching-focused institutions may still be wondering how to tackle or even understand the greater issues. Engagement in this area is also relevant for undergraduate institutions. Shorish (2012, 2015) describes the importance for teaching-focused university libraries as well to provide education, services, or consultation on research data management and curation.

Library support for research data management is an international endeavor, as evidenced by various professional organizations across the globe. The Association of European Research Libraries (LIBER) has indicated that “Support[ing] the development of skills in RDM” is an aspect of its Scholarly Communication and Research Infrastructures Committee Strategic Priority #1: Enabling Open Science. One mechanism planned to meet this goal is the development of modules addressing policy and training (LIBER, n.d.). The

Australian National Data Service (ANDS) works with several university libraries, such as Monash University and Australian National University, to provide resources on research data management. While many of the local tools are aimed primarily at the researcher, ANDS launched “23 (research data) Things” (Australian National Data Service, n.d.) in 2016 to aid librarians and data managers in building knowledge and skills related to research data management. The Canadian Association of Research Libraries (CARL) launched the Portage initiative as an attempt to bring together RDM knowledge into an information network across Canadian research libraries (Canadian Association of Research Libraries, n.d.). Additional educational efforts captured by RDA’s Education and Training Interest Group (RDA Education and Training Interest Group, n.d.) indicate the wide range of education efforts internationally, although the list is currently populated by almost exclusively English-language activities.

## Background

In 2010 the ACRL Board of Directors approved the formation of an interest group that would, in part, “sponsor discussions or programs that share the ways in which libraries are working to meet the needs of curating a variety of content in digital form” and “to inform and educate librarians on digital curation trends and new technologies...and to collaborate with other organizations within the library profession and academe on issues concerning digital curation” (ACRL Board of Directors, 2010). The librarians who founded the Digital Curation Interest Group (DCIG), Patricia Hswe and Marisa Ramirez, purposely selected a name that would allow versatility in terms of what is submitted to methods of digital curation for the goal of meeting long-term preservation, access, use, and reuse needs. The DCIG held its first online meeting in 2011, convened in person for the first time at ALA Annual in 2012, and from that point onward has grown in membership to include more than 900 information professionals (ALA Connect, n.d.).

Not unimportantly, the same year that the DCIG was formalized, the United States National Science Foundation made official its data management plan (DMP) requirement for grant proposals to the agency. The DMP mandate and academic libraries’ responses to it became topics that the DCIG quickly saw in its purview to address, especially when no other group or committee in ACRL, at the time, seemed a plausible “home” for such issues. Management of content and data aligns logically with digital curation practices.

The robust attendance of webcasts hosted by the DCIG in 2012–2014 also evidenced librarians’ growing interest and engagement in data management. “Collaborative Data Management Services at the University of California”, “Creation of an In-House DMP tool at the University of Houston Libraries”, and “Practical Data Management” were just a few of the presentations that regularly drew almost maximum attendance.

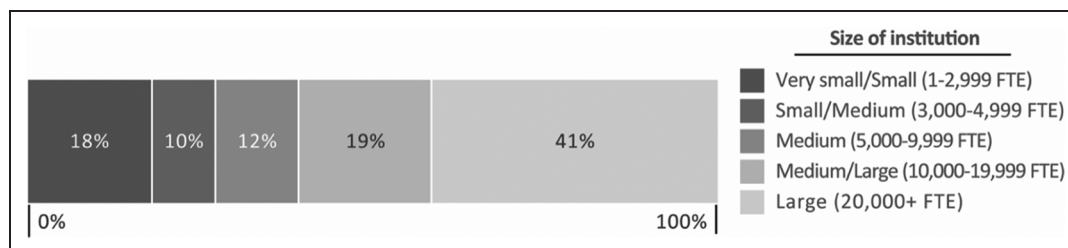
In its first few years of implementation, then, the DCIG was clearly filling a perceived gap in professional development opportunities in research data management (RDM). Moreover, when the DCIG started, although professional development opportunities dedicated to RDM had begun appearing, such as the Research Data Access and Preservation Summit (RDAP), there were still fewer offerings in RDM for subject liaison librarians, such as in the sciences. In 2013 ACRL leadership requested that its Research and the Scholarly Environment Committee (ReSEC) begin working with pertinent groups, including the DCIG, the Digital Humanities Interest Group (DHIG), and the Intersections of Scholarly Communication and Information Literacy Task Force, on determining the kind of support in RDM that ACRL should be providing its members. For instance, should there be a roadshow about RDM, similar to what the organization sponsors for scholarly communication (ACRL, n.d.)?

ReSEC devoted a portion of its 2014–2015 work plan to exploring professional development in RDM, and in early 2014, in collaboration with the other groups, the committee presented a set of recommendations to the ACRL Board. Two recommendations concerned data information literacy support, which the Board did not approve largely because of already existing efforts. Three recommendations addressed RDM support, which the Board did approve. These included the planning and coordination of a pre-conference workshop for the 2015 ACRL Conference in Portland, Oregon; the addition of information about data management to the ACRL Scholarly Communication Toolkit; and deployment of a survey to DCIG members in order to determine their needs and suggestions for data management.

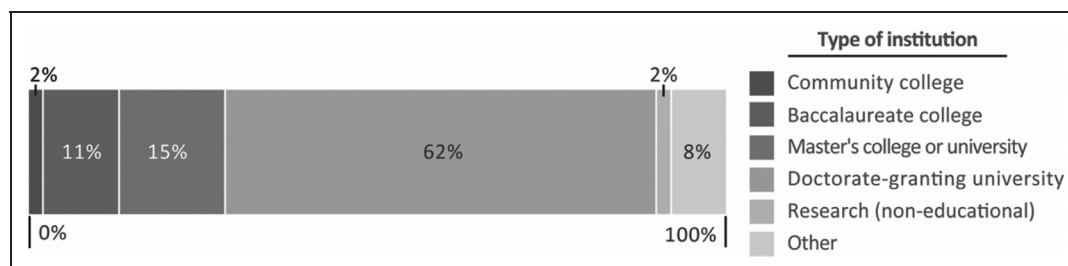
## Digital Curation Interest Group survey methodology

Acting on the recommendation to the ACRL Board, the Digital Curation Interest Group (DCIG) Executive Committee designed a professional development needs survey for its members via a 13 question survey. The survey included seven multiple choice





**Figure 1.** Institution size of survey participants.



**Figure 2.** Type of institution of survey participants.

questions, three multiple selection questions, two free-text responses, and one question asking participants to rank the importance of certain types of professional development opportunities in RDM. Since a few questions referred to the terms “research data management” and “research data curation”, these terms were defined at the onset of the survey. Assistance with survey design came from the Institutional Research Office at James Madison University, senior staff at ACRL, and the director of ALA’s Office of Research and Statistics. The survey was administered through Qualtrics and all responses were anonymous. A copy of the survey has been included in Appendix 1.

The subject group for the survey included librarians, library staff, administrators, and other personnel involved in, or interested in becoming involved in, the management and curation of research data at their institutions. The survey was emailed to 926 members of the DCIG on 8 July 2014 and closed on 31 July 2014. With a confidence level of 95% and a confidence interval of seven, 162 responses would have been needed for statistical significance. After one reminder was sent to the interest group members on 24 July 2014, a total of 195 responses were received (a response rate of 21%), which surpassed the threshold for statistical significance.

### Survey results

**Demographics.** Survey participants were asked two multiple choice demographic questions at the start of the survey to determine the size of their institutions

and the type of institution at which they were employed. Most respondents (60%) were employed by large institutions with 20,000+ FTEs (41%) or medium to large institutions with between 10,000 and 19,999 FTEs (19%). The remaining 40% were split between medium institutions with between 5000 and 9999 FTEs (12%), small or medium institutions with 3000 to 4999 FTEs (10%) and very small or small institutions with 1 to 2999 FTEs (18%). See Figure 1.

Most survey respondents (77%) were also employed by doctorate-granting universities (62%) or Master’s colleges or universities (15%). Baccalaureate colleges represented the next largest group of respondents (11%) and non-educational research institutions and community colleges were each represented by 2% of survey participants. See Figure 2.

Survey participants also had the option of entering another type of institution; 8% reported employment from a type of institution not represented in the survey options. These included a research library and museum, a law school, a non-academic research library, private institutional archives, two public libraries, someone in distance education, as well as someone from an academic library consortium.

Respondents were also queried regarding their current or anticipated roles in their institutions with the option to multi-select responses. A plurality were liaisons/subject librarians (42%). Other roles were split amongst the remaining respondents, who reported roles as administrators (22%), scholarly communication or digital repository coordinators (20%), data specialist or data services librarians (18%), and archivists and special collections librarians (16%). Of the

respondents 18% worked in other roles including librarians focusing on metadata, public services, technology, digital preservation, cataloging, special projects, digital media, digital initiatives, digital scholarship, e-resources, and collection development; interns; library technicians and clerks; specialists in digital humanities, data management, and visual resources; as well as faculty at an iSchool.

**Expertise.** The next three questions queried participants about their expertise and comfort levels with RDM topics, including how they rated their current expertise with RDM, their expertise with research data curation, and their level of preparation to engage in RDM activities at their institutions. Most participants seemed hesitant to rate themselves as experts in RDM: 65% of respondents indicated that they were completely new to the field (21%) or had limited experience in the field (44%); 35% claimed that they were intermediate experts in the field (33%) or were experts (2%). This lack of confidence increased with respect to research data curation; 78% claimed to be completely new to the field (30%) or had limited experience in the field (48%). Only 22% felt that their skills were intermediate (20%) or expert-level (2%). Overall participants seemed to rate themselves as having limited expertise in both RDM (with a mean of 2.16 out of 4) and research data curation (with a mean of 1.94 out of 4). Participants also lacked confidence about their abilities to engage in RDM activities at their institutions; on a scale of one to five, with one representing “very unprepared” and five representing “very prepared,” participants tended to rank themselves between “somewhat unprepared” and “neither prepared nor unprepared,” with a mean of 2.91 out of 5. Of the respondents 55% felt either “very unprepared” (18%), “somewhat unprepared” (27%), and “neither prepared nor unprepared” (10%); 44% felt they were more equipped to interact on the topic of research data management on their campus, while 34% felt “somewhat prepared” and 10% felt “very prepared”.

**Institutional approach.** The survey included an open-ended question asking participants to describe their institution’s approach to research data management and curation. One hundred one text responses were received. Based on some of the similarities between answers, the responses were classified by the level or stage at which the institution was in its approach to research data management and curation. Many respondents acknowledged that some efforts were being made to address services at their institutions. Of the respondents 28 indicated that the institution

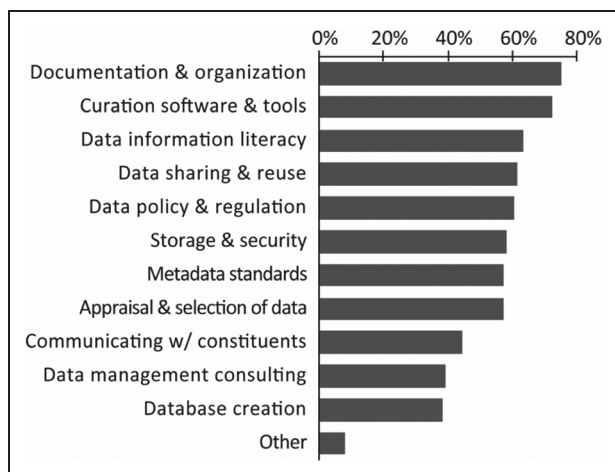
had a plan in place to offer data services of some sort. Another 30 respondents claimed that data was handled on an ad hoc basis on their campus. Twenty-three respondents were in an education, research, or conversation phase, in which they were investigating potential roles, training librarians, and/or having conversations on campus about roles for the library and other campus partners. Thirteen respondents were working on policy development either within the library or across campus. Five indicated that they were working on an institutional repository to tackle data management and curation issues and a similar five were developing outreach tactics for the campus community to either publicize current services or gauge interest in potential services. Thirteen reported that no work was in process to address data management or curation on their campuses.

Many of the survey participants also mentioned campus partners or other units that were getting involved in data management or curation. Three respondents claimed that other campus units were taking on roles of data management and curation and 18 partnered with other campus units to provide services. Nine indicated that individual departments or researchers were dealing with data management and curation issues themselves and only one mentioned that these issues are handled solely by campus IT. Fifty-two stated that the library was handling these roles without indicating any other partners.

During the review of this data, we found it interesting that many survey respondents associated data management and curation with the existence of a repository. Of the respondents 25 mentioned repositories, with a few discussing data repositories specifically. This was interesting because institutional repository software is not always conducive to housing data and data services are not strictly limited to the curation of data in a repository.

**Professional development.** The remaining five questions of the survey were geared toward determining what kinds of professional development opportunities might be the most well received for education on research data management and curation. Survey respondents were asked to select the topics they were most interested in learning more about. The most popular topics were documentation and organization best practices; curation software and tools; data information literacy; data sharing and reuse; and data policy and regulation. All topics that were of interest to respondents are detailed in Figure 3.

Other topics that survey respondents wrote in included digital preservation, digital archiving, funding data management and curation, community



**Figure 3.** Interest in research data management and curation topics.

practices, email archiving, building buy-in from faculty and campus administration, data manipulation, etc.

Respondents were also asked what types of activities would be most useful for them based on their professional needs. Most (81%) indicated that reviewing best practices and practicing with tools and software would be the most helpful. Many (65%) were interested in reading/listening and discussing case studies while 30% were interested in learning about theory; 5% wrote in responses including data-centric conferences such as RDAP, “listening to and engaging others”, “ideas on how to move forward”, “conducting research studies”, and getting involved in best practice authoring.

**Delivery methods.** We were also interested in determining what the best method of delivery would be for professional development such as webcasts, asynchronous online classes, one-day workshops such as a roadshow, multi-day data management institutes, ACRL conference sessions, or other methods. Participants were asked to rank these five options and were provided the opportunity also to rank and write in “other” options. Webcasts ranked the highest, with a mean of 2.12 (with one as most important and six as least important). One-day workshops and asynchronous online classes were ranked at about the same level with workshops at a mean of 2.74 and online classes at 2.84. ACRL conference sessions were ranked at a mean of 3.60 and multi-day data management institutes at 3.83. Perhaps it can be inferred that lower time commitments with more flexibility are preferable for many of the survey participants since webcasts, one-day workshops, and online classes ranked highest. A few respondents indicated other opportunities they would be interested in including

downloadable articles, resources on best practices for self-study, MLS curricula, educational opportunities for those not associated with the library, ACRL sessions at ALA Conferences, free online information, and collaborative projects with “learn-by-doing” participation.

ACRL Participants were also asked what more they would like to see from ACRL in terms of professional development offerings on research data management and curation. The need for some sort of resource collection of best practices including curricula, training materials, case studies, and toolkits was mentioned by 13 respondents. One respondent referred to this as a “clearinghouse” for research data management and curation best practices. Ten were concerned about cost of opportunities and suggested more free resources. Nine were interested in more online opportunities. Six respondents mentioned more discussion of outreach and collaboration with their constituents. Five asked for higher level training opportunities for the more advanced practitioners, mentioning that more tiered opportunities would be helpful. Three were interested in seeing more training on subject specific metadata standards. Two wanted to see opportunities for non-research institutions and/or opportunities for undergraduates specifically.

**Existing opportunities.** In the final question, we asked respondents what professional development opportunities they had already participated in, regarding the topics of research data management and curation. Many respondents (63%) had attended a webinar; sponsors of these webinars included the Digital Curation Interest Group, bepress, National Information Standards Organization (NISO), Society of American Archivists (SAA), the Association for Information Science & Technology (ASIS&T), Special Libraries Association (SLA), Association of College & Research Libraries (ACRL), American Library Association (ALA), Library Journal, the Association of Research Libraries (ARL)/Digital Library Federation (DLF) E-Science Institute, Preservation and Archiving Special Interest Group (PASIG), LYRASIS, ARL, and state and regional institutes. Less than half (41%) had attended a research data-focused conference such as the Research Data Access and Preservation (RDAP) Summit or the International Digital Curation Conference (IDCC); 33% had participated in an online class such as those available on Coursera and only 15% had attended an e-science bootcamp while 34% indicated that they attended other types of professional development events. Some of these included the Data Curation Profiles Toolkit Workshop; an ACRL data management online course; the

Interuniversity Consortium for Political and Social Science Research (ICPSR) Data Curation Summer Institute; DataONE involvement; the e-Science Institute; various other national, state and local workshops; on-campus events; self-study and review of literature, etc.

**Summary.** Responses from this survey were helpful in informing the development of an ACRL pre-conference, especially with establishing an approach that would reach the right audience. We were surprised by many of the survey responses; for instance, the interest in professional development opportunities on metadata standards for data management and curation was unexpected and something that was incorporated both in the pre-conference and used for a later DCIG webinar. Additionally, the expertise levels and levels of comfort with research data management and curation topics were lower than expected. Many professionals were clearly seeking better professional development opportunities to help with preparation and confidence so that they would be better able to implement services at their own institutions.

### ACRL 2015 Pre-conference

In some ways, the pre-conference was designed as a pilot program for what we would propose for future stand-alone professional development offerings. We had the DCIG survey responses to help focus the scope of the pre-conference. Respondents indicated that they were looking for more practical knowledge, so rather than focus on the theory of data management and curation, we took a practical strategies approach. Three learning outcomes were defined:

1. Define data management as it relates to data information literacy in order to build upon existing information literacy pedagogy.
2. Develop a framework for determining the most appropriate scale for data management services based on institutional circumstances.
3. Develop strategies for engaging faculty and students on data management issues in order to advance data information literacy.

To further refine the content, we sent a survey to all registrants for the pre-conference to gauge experience level and collect demographic information (see Appendix 2). Approximately 60 participants attended the pre-conference. About half of the respondents indicated that they were subject librarians, while the rest were distributed across administrator and data services roles.

The session was seven hours long, including a 90-minute break for lunch and two 15-minute breaks. We chunked the information into “modules”, units with a single theme that we delivered in a sequential manner and that all participants would move through together. While a focus on the practical was the main direction of the pre-conference, we felt that a strong foundation that related data to the research and scholarly life cycle was a critical starting point.

#### *Module 1. Introduction: Data and scholarly communication*

The first, 50-minute module was built to introduce data as another information type to be incorporated into library support for the research and scholarly life cycles. Data can be considered as both a contributor to, and product of, scholarship. Libraries exist as stewards of information and we have continued to diversify and evolve the ways with which we engage with that information. Weaving information literacy (IL) into our role as stewards is not a stretch: understanding information and how to find and use it effectively is the reason for the stewardship – there is not much point to stewarding information if no one can find and use it. This relationship with IL is analogous to the role that scholarly communication plays with IL. The ACRL white paper on the intersections of scholarly communication and IL reinforces this perspective and even makes note of the importance of treating data as an aspect of IL (ACRL Working Group on Intersections of Scholarly Communication and Information Literacy, 2013). Ending the first module with a focus on IL instruction and scholarly communication allowed for a smooth transition to the second module, which focused effective data IL instruction strategies.

#### *Module 2. Data management: Developing an instruction strategy*

Providing introductory RDM best practices instruction is one of the most common areas of engagement for libraries who are starting out in building a data services program (ACRL Research Planning and Review Committee, 2012; Cox et al., 2012; Tenopir et al., 2012). It is an opportunity to provide a needed service that has a relatively low barrier for entry, and can be a mechanism for getting a tangible research data service (RDS) off the ground. We started this 70-minute module by reviewing the current range of data management instruction offerings being provided by academic libraries. It is important to recognize that RDM instruction can be designed and conducted in



many ways that go beyond the 50-minute one-shot. The instruction should both meet the needs of the intended audience, and reflect the preferences of the librarian offering it.

**Module 2a. Introduction to data information literacy instruction.** In this module of the workshop, we first reviewed a range of current data information literacy (DIL) instruction offerings happening in academic libraries, and then provided suggestions regarding things to keep in mind while developing instruction within your institution. Instructional offerings that we reviewed included credit-bearing graduate-level courses (Borgman, 2015; Creamer, n.d.; Whitmire, 2014; Wright, n.d.), a non-credit flipped classroom, discipline-specific course (Johnston and Jeffries, 2015), workshops series on RDM basics (Coates, 2013; Muilenburg et al., 2015), workshop series on applied topics (e.g. how to write a data management plan, how to create metadata, how to keep a lab notebook, etc.), and workshop one-shots (UK Digital Curation Centre, n.d.). After reviewing these examples, we had workshop participants engage in a five-minute pair-and-share exercise to discuss DIL instruction. It focused on whether or not there were examples of instruction that we did not mention. Following the report-out from the activity, we then reviewed a suggested pedagogical approach to teaching DIL. The approach is explained in Whitmire (2015). Briefly, the approach is based on the following observation: the tenets of RDM best practices are discipline agnostic, but the application of best practices is very discipline- and situation-specific. Given that, if you really want your pupils to absorb the information and use it, they need to see how the information translates to their workflow, or better yet – practice it during class activities. The approach to developing course content involves outcomes-centered course design, with active-learning components to promote engagement and internalization of information (Whitmire, 2015).

**Module 2b. Instruction activity.** In order to engage workshop participants with the topic of DIL instruction, we conducted an activity where they could work on designing their own hypothetical piece of instruction. In this case, they were asked to design a 50-minute instruction session on metadata. We encouraged them to consider the following aspects in designing their workshop:

1. Audience – who is the workshop for?
2. Staffing – who will do the teaching? Do they need professional development?

3. Learning outcomes – what do you want attendees to learn?
4. Content – what major topics will you cover?
5. Active learning – how will you get students engaged with the material?

The purpose of this activity was to give librarians a chance to: (1) think about the large variety of different educational opportunities there are regarding data management; and (2) to strategize lesson plans for potentially challenging environments where they may have a multitude of different user needs in one session. This exercise was intended to introduce librarians to the most important considerations of designing DIL instruction, and give them a feel for the process and effectiveness of active learning.

### **Module 3. Engagement with the campus community**

**Module 3a. Reflective writing and discussion.** At the beginning of the 80-minute third module, participants were instructed to reflect upon their environments and note instructional efforts that might address certain challenges, using large-easel pad paper to write down their responses as a group. A PowerPoint slide provided a few examples, including lack of, or untrained, personnel; campus infrastructure; library stakeholders; and administrative stakeholders. Eight groups were formed. They discussed and recorded challenges for five minutes and were given ten minutes to report back as a group. Some challenges reported by one group included IT dismissal of unfunded efforts; lack of understanding of storage vs. preservation; how to market services; the challenge with finding champions; convincing researchers of the value; and issues with timing being either preventative or rescue. Another group used the examples from the PowerPoint slide as a starting point and addressed those topics. For instance, they found that lack of trained personnel was a problem because many might be hesitant to learn new tasks and their current skills were ill-defined. The same group expressed challenges with library stakeholders in committing to new duties and identifying what duties they might need to discontinue. A third group discussed issues with staff training, lack of staffing, and fear of failure, but identified small-scale solutions such as brown-bag lunches, online training, and partnerships with other university departments to kick start initiatives. The remaining groups reported similar challenges to the three mentioned above. After the groups compiled a list of challenges together, a moderator led the discussion where these challenges were shared.



**Module 3b. Panel discussion with framing questions.** After participants completed the reflective writing and discussion portion of the module, they had a chance to hear, via a moderated panel discussion, how colleagues in the field have addressed similar or other challenges in RDM services. The panel consisted of three librarians from a variety of environments: a small liberal arts college, a large research university, and a university that serves as the academic health center of its state. The following questions framed their discussion and remarks:

- Whom have you partnered with at your campuses? How did you decide on whom to work with?
- How have you framed this topic as you've done outreach at your various institutions with a range of stakeholders?
- What are the top three to five points to convey to various audiences when making a case for institution-wide data management services?
- What have you learned from these collaborations and relationships? How has this knowledge informed service models and program development at your institution?

In their responses panelists recounted partnerships with faculty and other librarians that enabled new approaches to teaching about data management, including use of actual research data sets, and encouraged more collaboration and relationship building with campus entities, such as the Office of the Vice-President for Research and the Office of Sponsored Programs.

**Module 3c. Role play activity.** As a final activity for the third module, participants were encouraged to pair up for a role play activity. Each pair received two cards; each card had a role defined. We felt it was important to acknowledge that conversations about data services are not limited to subject liaisons and faculty members; often librarians and library administration must correspond with administrators on campus to promote or establish services. Roles that participants could play included a librarian and faculty member; a librarian and college administrator or dean; and a library administrator and an administrator from an Office of Research or Sponsored Programs. The faculty member role also had three separate versions to reflect the challenges for discussing data needs across disciplines. Faculty member roles included social science researchers, an engineering professor, and a biologist. All cards included questions and prompts for both the library and the campus representative. Participants

were given ten minutes to role play and then ten minutes to report back to the group about their experiences.

#### **Module 4. Data: Taking it home**

This 60-minute module was designed to give attendees the opportunity to develop a tangible plan for how they could apply what they learned at the workshop back at their home institution. We provided an "Action Plan" worksheet (see Appendix 3) with three sections. In the first section, attendees were prompted to list internal and external stakeholders, and potential partners for research data services at their campus. Next, we asked them to list "drivers", things that would help achieve their vision for RDS, and "barriers", or things that might keep them from achieving their vision. A third table asked them to consider the following: one thing they could do immediately after returning home, one thing that would take some time, and one thing that they had no idea how to start. Finally, there was a blank area labeled "Action Plan" where they could bring all of the previous information to bear in drafting or brainstorming a plan of action for developing or expanding RDS. Participants that were comfortable sharing their drivers and barriers could do so, but this exercise was mainly focused on giving them a structured plan to refer back to when they returned from the conference.

#### **Module 5. Looking to the future**

The pre-conference concluded with some thoughts regarding the ongoing training and education needs, with respect to RDM. Many librarians may find that data has become a new medium that they need to build fluency with, but they lack the time or ability to seek out continuous education for it. Moreover, in the event that they are able to attend workshops or training sessions, the pipeline from administration that would allow for application of these new skills is often absent. How can we build that pipeline that asks, "How will your library benefit by having this person participate?" "How will this person apply these learned skills?" This last, 45-minute module focused the discussion on these issues, prompting nearly all the attendees to realize that they had not had any kind of conversation on these matters within their organization. Discussion around sustainability also came up, with the concept of "team librarianship" garnering some attention. In this model, not every librarian is expected to be an expert across domains, but there is an expectation of foundational knowledge and the practice of referral to the content expert. This can alleviate the pressure to master the

many emerging areas of support by an academic library, such as RDM, copyright, digital humanities, and so on.

The end of this module promoted a follow-up webinar in an attempt continue the conversation and assess how much of the activity worksheet was able to be carried back to the institution. Lastly, a post-assessment survey (see Appendix 4) was distributed on paper to all attendees, to compare to the pre-assessment survey data.

### Pre- & post-assessment of ACRL 2015 pre-conference

In order to best determine how to propose a professional development program that would be most effective, it was critical that the pre-conference undergo some form of assessment. While the pre-conference was not intended to be identical to an association-wide program, there were portions of it that could serve as a foundation for a program. Moreover, we wanted to assess the areas of need from the audience and how they viewed the purpose/engagement of the association in this area.

Three “comfort-level” Likert scale questions were included on both the pre- and post-surveys for the March pre-conference. These questions were intended to evaluate, in a generalized manner, the comfort of participants with RDM activities before and after the pre-conference. These questions were:

1. How prepared do you feel you are to teach research data management (in a classroom setting) at your institution?
2. How prepared do you feel you are to offer research data management support (consultation) to researchers at your institution?
3. How prepared do you feel you are to engage external stakeholders (e.g. faculty, campus IT, university administrators) in conversations about research data management at your institution?

The scale was a five-point range from “1: very unprepared” to “5: very prepared”.

In short, there was an increase in perceived comfort-level across all three questions, with preparedness to offer RDM consultations exhibiting the least gain (Table 1). Some caveats to this data analysis must be stated. The data are not from matched pairs. That is, we do not know that the same person who took the pre-survey took the post-survey and what the differences in that individual’s responses were. We had 31 pre-survey responses and 38 post-survey responses to these questions, so while we know that

**Table 1.** Results from the pre- and post-surveys regarding ACRL 2015 ‘Brass Tacks’ pre-conference participants’ views on their ability to provide RDM services.

	Pre-survey	Post-survey	Change
How prepared do you feel you are to teach research data management (in a classroom setting) at your institution?	2.03	3.47	+1.44
How prepared do you feel you are to offer research data management support (consultation) to researchers at your institution?	2.29	3.26	+0.97
How prepared do you feel you are to engage external stakeholders (e.g. Faculty, campus it, university administrators) in conversations about research data management at your institution?	2.36	3.76	+1.38

Numbers shown are means.  $N = 31$  for the pre-survey;  $N = 38$  for the post-survey.

the populations overlap, we cannot link the data back to individuals. Despite the fact that we cannot state that “Person A” demonstrated comfort-level gains, these aggregated responses do allow us to look at the trend across the cohort.

We also provided opportunities for open-ended response to several questions on the post-survey to assess what skills participants got from the pre-conference, what they felt was missing from the content, if they liked the delivery format of the session, and what role they thought ACRL should play in providing continuing education in this area. Of the 38 responses to the question of what role ACRL should play, not a single response indicated that ACRL should not engage in this area (an unsurprising finding, considering the venue). Of those responses 15 indicated that ACRL had a “strategic”, “important”, “major”, “imperative”, or “clear” role in providing professional development for RDM, especially to subject liaison librarians. Other responses suggested types of engagement, such as workshops, education materials, training sessions, “more than webinars”, and online courses. Several mentioned that RDM was

important to incorporate into IL and be marketed as a subject liaison responsibility.

As a last effort to assess the success of the pre-conference, we also organized a follow-up webinar for attendees to share what they had taken back to their institutions, what thoughts they had about the topics discussed at the pre-conference, what they might have been able to implement, and what they felt might have been missing from the pre-conference. We also shared results from post assessment surveys with the group. This webinar occurred a little under two months after the pre-conference. Four of the pre-conference participants attended, which was less than 10% of the pre-conference attendees.

## Discussion

Our survey results indicated that many were interested in one-time, short events such as webcasts, so it may seem counterintuitive that the resulting event was a full one-day pre-conference at an ACRL conference. However, introducing approaches to RDM are not easy to do in short webcasts unless they are part of a series; in this regard, a one-day event was more conducive to providing a robust professional development experience. Additionally, the survey participants' interest in shorter events can be correlated to many of the challenges that these participants face in their institutions, which emerged during the pre-conference discussions. Many are understaffed, underfunded, and do not have dedicated resources for RDM. It is our hope that additional professional development opportunities may help establish the importance of RDM in academic libraries among library and campus administrators, which may result in increased support for those activities at the home institution.

We learned that conducting a follow-up webinar to a professional development offering was unlikely to be successful. In an attempt to build community and continue the conversation, we conducted a webinar one month after the pre-conference. We had very few attendees and it was challenging to implement due to technological issues. In retrospect, there might have been a better mechanism to promote discussion between pre-conference participants than a one-time webinar that many were unable to attend, such as an asynchronous discussion. It would be interesting to investigate what the impact of the community groups and MeetUp sessions facilitated by the ANDS 23 (research data) Things (Australian National Data Service, n.d.) has been and if these efforts have resulted in positive outcomes.

Given the DCIG survey data, the pre-conference assessment survey data, and the expertise of the

authors, we presented a series of recommendations to the ACRL Board in June 2015. We provided three professional development models: a RDM roadshow, in the same vein as the ACRL Scholarly Communication Roadshow; moving DCIG from an interest group to a section, which within the organizational structure of ACRL would allow for more structure, consistency, and committee support for outreach and engagement; and lastly, an additional module or track focusing on data information literacy to the ACRL Immersion Program (ACRL, n.d.) After some discussion, the ACRL Board approved the establishment of a RDM roadshow and encouraged DCIG leadership to begin the process of petitioning a move to section. The RDM roadshow planning process transitioned from the authors to the ACRL Research and Scholarly Environments Committee, which has oversight of the Scholarly Communication Roadshow.

## Conclusion

The work presented in this article only begins to touch on progress towards creating sustainable and targeted RDM professional development opportunities for liaison librarians; future initiatives that this work has influenced or encouraged are still in development. The results of the survey provide some detail about comfort levels of ACRL DCIG members faced with responsibilities to offer RDM services in their institutions as well as preferences for delivery of professional development on the topic. The results of the survey informed much of the ACRL pre-conference discussed in this article and these results may continue to offer insight into what kinds of RDM opportunities subject liaisons need so that they can address shifting job expectations.

The pre-conference served as a test ground for future ACRL professional development opportunities on RDM. Multiple modules were presented with topics such as data and scholarly communication, data management instruction strategies, engaging the campus community, creating individual action plans, and discussing future roles for subject liaisons in regards to RDM services. When comparing the pre- and post-assessment of the pre-conference, participants indicated that their levels of preparation increased because of the contents of the pre-conference; participants felt more prepared to teach research data management, to engage external stakeholders on campus, and to offer research data management support. The establishment of a RDM roadshow, and its oversight from the Research and Scholarly Environments Committee of ACRL, indicate that the establishment of professional development opportunities in RDM for

academic librarians is important and should be ongoing.

## Acknowledgments

The authors would like to acknowledge the following individuals from ACRL for supporting the development of this research data management educational opportunity: Mary Ellen Davis, Kara Malenfant, Margot Conahan, and the members of the 2015 ACRL Board. Additional thanks to Molly Keener for serving as advisor to the pre-conference planning group.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References

- ACRL Board of Directors (2010) *Creation of ACRL Digital Curation Interest Group*. Chicago, IL: ACRL Board of Directors Action Form.
- ACRL Research Planning and Review Committee (2012) 2012 top ten trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News* 73: 311–320.
- ACRL Working Group on Intersections of Scholarly Communication and Information Literacy (2013) *Intersections of Scholarly Communication and Information Literacy: Creating Strategic Collaborations for a Changing Academic Environment*. Chicago, IL: Association of College and Research Libraries.
- ALA Connect (n.d.) *ACRL Digital Curation Interest Group*. Available at: <http://connect.ala.org/node/132171> (accessed 17 August 2016).
- Association of College and Research Libraries (n.d.) *Immersion Program*. Available at: <http://www.ala.org/acrl/immersion> (accessed 13 May 2016).
- Australian National Data Service (n.d.) *23 (Research Data) Things*. Australian National Data Service. Available at: <http://www.ands.org.au/partners-and-communities/23-research-data-things> (accessed 13 May 2016).
- Australian National Data Service (n.d.) *23 (Research Data) Things: Community Groups*. Australian National Data Service. Available at: <http://www.ands.org.au/partners-and-communities/ands-communities/23-research-data-things/community-groups> (accessed 13 May 2016).
- Borgman CL (2015) *Syllabus for Data Management and Practice, Part I, Winter 2015*. Available at: <https://works.bepress.com/borgman/358/download/> (accessed 16 November 2016).
- Canadian Association of Research Libraries (n.d.) *Portage Home*. Available at: <https://portagenetwork.ca/> (accessed 13 May 2016).
- Coates HL (2013) Developing a data management lab: Teaching effective methods for health and social sciences research. Poster presented at the *Data Information Literacy Symposium*, West Lafayette, IN. Available at: <http://hdl.handle.net/1805/3569> (accessed 16 November 2016).
- Cox A, Verbaan E and Sen B (2012) Upskilling liaison librarians for research data management. *Ariadne*. Available at: <http://www.ariadne.ac.uk/issue70/cox-et-al> (accessed 16 November 2016).
- Creamer A (n.d.) *Scientific Data Management LIS 532G-01 Fall*. Available at: <https://sites.google.com/a/vt.edu/srdm/home> (accessed 22 April 2016).
- Duraspace (n.d.) *E-Science Institute Course Description*. Available at: <http://duraspace.org/esi-course-description> (accessed 17 August 2016).
- e-Science Portal for New England Librarians (n.d.) *Science Boot Camp*. Available at: <http://esciencelibrary.umassmed.edu/professional-educ/science-boot-camp> (accessed 17 August 2016).
- Heidorn PB (2011) The emerging role of libraries in data curation and e-science. *Journal of Library Administration* 51: 662–672. DOI:10.1080/01930826.2011.601269.
- Hswe P (2015) Peering outward: Data curation services in academic libraries and scientific data publishing. In: Bonn M and Furlough M (eds.) *Getting the Word Out: Academic Libraries as Scholarly Publishers*. Chicago, IL: Association of College and Research Libraries, pp. 221–248.
- Hswe P and Holt A (2011) Joining in the enterprise of response in the wake of the NSF data management planning requirement. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC* 274: 11–17.
- Jaguszewski J and Williams K (2013) *New Roles for New Times: Transforming Liaison Roles in Research Libraries*. Washington, DC: Association of Research Libraries. Available at: <http://www.arl.org/component/content/article/6/2893> (accessed 16 November 2016).
- Johnston L and Jeffries J (2015) *Teaching Civil Engineering Data Information Literacy Skills: An e-Learning Approach*. *Data Information Literacy Case Study Directory* 3. DOI: 10.5703/1288284315479.
- LIBER (n.d.) *Enabling Open Science: Scholarly Communication & Research Infrastructures*. Available at: <http://libereurope.eu/strategy/strategic-direction-1-enable-open-science/> (accessed 17 August 2016).
- Little G (2012) Managing the data deluge. *Journal of Academic Librarianship* 38(5): 263–264. DOI:10.1016/j.acalib.2012.07.005.
- Muilenburg J, Lebow M and Rich J (2015) Lessons learned from a research data management pilot course at an academic library. *Journal of eScience Librarianship* 3. DOI:10.7191/jeslib.2014.1058.
- Nicholson SW and Bennett TB (2011) Data sharing: Academic libraries and the scholarly enterprise. *portal: Libraries and the Academy* 11(1): 505–516.



- RDA Education and Training Interest Group (n.d.) *RDM Educational Efforts - V2*. Available at: [https://docs.google.com/spreadsheets/u/1/d/10RTW-nZk0x\\_mpQw2VAltcc656MV9EeCaDe2IM4umb4/edit](https://docs.google.com/spreadsheets/u/1/d/10RTW-nZk0x_mpQw2VAltcc656MV9EeCaDe2IM4umb4/edit) (accessed 13 May 2016).
- Reisner BA, Vaughan KTL and Shorish Y (2014) Making data management accessible in the undergraduate Chemistry curriculum. *Journal of Chemical Education* 91(11): 1943–1946.
- Rockenbach B, Ruttenberg J, Tancheva K, et al. (2015) *Association of Research Libraries/Columbia University/Cornell University/University of Toronto Pilot Library Liaison Institute. Report*. Washington, DC: Association of Research Libraries.
- Scaramozzino JM, Ramírez ML and McGaughey KJ (2012) A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries* 73(4): 349–365.
- Shorish Y (2012) Data curation is for everyone! The case for Master's and Baccalaureate institutional engagement with data curation. *Journal of Web Librarianship* 6: 263–273. DOI:10.1080/19322909.2012.729394.
- Shorish Y (2015) Data information literacy and undergraduates: A critical competency. *College & Undergraduate Libraries* 22(1): 97–106. DOI: 10.1080/10691316.2015.1001246.
- Tenopir C, Birch B and Allard S (2012) *Academic Libraries and Research Data Services: Current Practices and Plans for the Future*. White Paper. Washington, DC: Association of College and Research Libraries.
- UK Digital Curation Centre (n.d.) *An Introduction to Research Data Management*. Available at: <http://www.dcc.ac.uk/events/workshops> (accessed 22 April 2016).
- Whitmire AL (2013) Thoughts on 'eResearch': A scientist's perspective. *Journal of eScience Librarianship* 2: 68–72. DOI:10.7191/jeslib.2013.1045.
- Whitmire A (2014) *GRAD 521 Research Data Management Syllabus and Lesson Plans*. Available at: [https://figshare.com/articles/GRAD\\_521\\_Research\\_Data\\_Management\\_Syllabus\\_and\\_Lesson\\_Plans/1003834](https://figshare.com/articles/GRAD_521_Research_Data_Management_Syllabus_and_Lesson_Plans/1003834) (accessed 16 November 2016).
- Whitmire AL (2015) Implementing a graduate-level research data management course: Approach, outcomes, and lessons learned. *Journal of Librarianship and Scholarly Communication* 3(2). DOI:10.7710/2162-3309.1246.
- Witt M (2012) Co-designing, co-developing, and co-implementing an institutional data repository service. *Journal of Library Administration* 52(2): 172–188. DOI:10.1080/01930826.2012.655607.
- Witt M (2008) Institutional repositories and research data curation in a distributed environment. *Library Trends* 57(2): 191–201. DOI:10.1353/lib.0.0029.
- Wright S (n.d.) *LibGuides: NTRES 6600: Research Data Management Seminar: Home*. Available at: <http://guides.library.cornell.edu/ntres6600> (accessed 22 April 2016).

## Author biographies

**Suzanna Conrad** is the Head of Library Information Systems at California State University, Sacramento. Suzanna obtained her MSLIS from the University of Illinois, Urbana-Champaign in 2011 and her MBA from the University of East London in 2006. Suzanna was the 2015–2016 Convener for the ACRL Digital Curation Interest Group.

**Yasmeen Shorish** is the Data Services Coordinator at James Madison University. She received her MSLIS from the University of Illinois, Urbana-Champaign. Her research focuses on changes in scholarly communication, data information literacy, and issues related to representation and social justice in librarianship.

**Amanda L. Whitmire** is Head Librarian and Bibliographer at the Harold A. Miller Library, and Assistant to the Director, at Hopkins Marine Station of Stanford University. She received her PhD in Oceanography from Oregon State University.

**Patricia Hswe** is Program Officer in Scholarly Communications at The Andrew W. Mellon Foundation. Previously, she was Co-Department Head, Publishing and Curation Services, at The Pennsylvania State University Libraries. She holds an MLIS from the University of Illinois, Urbana-Champaign, and a PhD in Slavic Languages and Literatures from Yale University.

## Appendices

### Appendix I: DCIG survey

In an attempt to better inform the organization of membership needs around research data management and curation, the Digital Curation Interest Group is soliciting your feedback. This survey is completely voluntary and results will be kept confidential and anonymous. While digital curation encompasses a wider scope than just research data, this survey is focusing on just that facet. The survey should take no more than 10 minutes and will help ACRL better serve you.

In this survey, “research data management” refers to the activities necessary for documenting, maintaining, and making accessible the data collected in the course of research. The data are discipline agnostic and may be in any format. “Research data curation” refers to the ongoing stewardship of research data over its useful life-cycle. One way to frame this relationship is that



one can manage research data without necessarily providing ongoing curation.

- 1) What is the size of the institution at which you are currently working? (In this survey, FTE refers to full-time equivalent enrollment)
  - Very small/small (1–2999 FTE)
  - Small/medium (3000–4999 FTE)
  - Medium (5000–9999 FTE)
  - Medium/Large (10,000–19,999 FTE)
  - Large (20,000+ FTE)
- 2) What is the type of institution at which you are currently working?
  - Community college
  - Baccalaureate college
  - Master's college or university
  - Doctorate-granting university
  - Research institution (non-educational)
  - Government institution (federal or state)
  - Other \_\_\_\_\_
- 3) What is your current (or anticipated) role at your institution?
  - Administrator
  - Liaison/Subject Librarian
  - Archivist/Special Collections Librarian
  - Data Specialist/Data Services Librarian
  - Scholarly Communication/Digital Repository Coordinator
  - Other \_\_\_\_\_
- 4) Which of the following user groups do you serve? (Check all that apply)
  - Faculty
  - Staff
  - Undergraduate students
  - Graduate students
  - Post-Docs
  - Public
  - Librarians
  - Independent researchers
  - Other: \_\_\_\_\_
- 5) How would you rate your current expertise with research data management?
  - Completely new to the field
  - Limited experience in the field
  - Intermediate expert in the field
  - Expert in the field
- 6) How would you rate your current expertise with research data curation?
  - Completely new to the field
  - Limited experience in the field
  - Intermediate expert in the field
  - Expert in the field
- 7) How prepared do you feel you are to engage in research data management activities at your institution?
  - Very unprepared
  - Somewhat unprepared
  - Neither prepared nor unprepared
  - Somewhat prepared
  - Very prepared
- 8) Please describe your institution's approach to research data management and curation.
- 9) What research data management and curation topics are you most interested in learning more about? (Check all that apply)
  - Metadata standards
  - Curation software and tools
  - Communicating with constituents
  - Data management consulting
  - Data information literacy
  - Database creation
  - Appraisal and selection of data
  - Data sharing and reuse
  - Data policy and regulation
  - Documentation and organization best practices
  - Storage and security best practices
  - Other \_\_\_\_\_
- 10) Based on your professional needs, what types of activities would be most useful to you? (Check all that apply)
  - Reviewing best practices
  - Learning about theory
  - Reading/listening and discussing case studies
  - Practicing with tools and software
  - Other \_\_\_\_\_
- 11) Which of the following continuing education opportunities would you be most interested in ACRL offering? (Rank in order of importance)
  - Webcasts
  - Asynchronous online class
  - One-day workshop (regional/"roadshow")
  - Multi-day data management institute
  - ACRL Conference sessions
  - Other \_\_\_\_\_

- 12) What would you like to see ACRL doing more of in terms of professional development offerings in the area of research data management and curation?
- 13) Please list what other professional development opportunities you have engaged with on the topic of research data management and curation. (Check all that apply)
  - Attended a research data-focused conference, e.g. RDA Summit, RDAP, DCC\
  - Participated in an e-science bootcamp
  - Participated in an online class, e.g. Coursera
  - Attended a webinar
  - If so, who sponsored it? \_\_\_\_\_
  - Other \_\_\_\_\_
- 5) How prepared do you feel you are to teach research data management (in a classroom setting) at your institution?
  - Very unprepared
  - Somewhat unprepared
  - Neither prepared nor unprepared
  - Somewhat prepared
  - Very prepared
- 6) How prepared do you feel you are to offer research data management support (consultation) to researchers at your institution?
  - Very unprepared
  - Somewhat unprepared
  - Neither prepared nor unprepared
  - Somewhat prepared
  - Very prepared

## Appendix 2: Pre-conference pre-assessment

- 1) What is the size of the institution at which you are currently working? (In this survey, FTE refers to full-time equivalent enrollment)
  - Very small/small (1–2999 FTE)
  - Medium (3000–9999 FTE)
  - Medium/Large (10,000–19,999 FTE)
  - Large (20,000+ FTE)
- 2) What type of institution do you work at?
  - Community College
  - Baccalaureate College
  - Master's College or University
  - Doctorate-granting University
  - Research Institution (non-educational)
  - Government Institution (federal or state)
  - Other \_\_\_\_\_
- 3) What is your current (or anticipated) role at your institution?
  - Administrator
  - Liaison/Subject Librarian
  - Archivist/Special Collections Librarian
  - Data Specialist/Data Services Librarian
  - Scholarly Communication/Digital Repository Coordinator
  - Other \_\_\_\_\_
- 4) How would you rate your current expertise with research data management?
  - Completely new to the field
  - Limited experience in the field
  - Intermediate expert in the field
  - Expert in the field
- 7) How prepared do you feel you are to engage external stakeholders (e.g. faculty, campus IT, university administrators) in conversations about research data management at your institution?
  - Very unprepared
  - Somewhat unprepared
  - Neither prepared nor unprepared
  - Somewhat prepared
  - Very prepared
- 8) In your opinion, how prepared is your library/unit to offer research data management support?
  - Very unprepared
  - Somewhat unprepared
  - Neither prepared nor unprepared
  - Somewhat prepared
  - Very prepared
- 9) What skills or information do you hope to get out of this pre-conference?
- 10) Have you taken advantage of other professional development opportunities regarding research data management?
  - YES
  - NO

## Appendix 3: Action Plan Worksheet

Action Plan Worksheet – fill out each column for your own local environment

Internal Stakeholders	External Stakeholders	Potential Partners

Drivers (helps to achieve your vision)	Barriers (keeps you from achieving vision)

One thing I can do when I get back	One thing that will take some time	One thing that I have no idea how to start

Action Plan:

#### Appendix 4: Pre-conference post-assessment

1. How prepared do you feel you are to teach research data management (in a classroom setting) at your institution?

- Very unprepared
- Somewhat unprepared
- Neither prepared nor unprepared
- Somewhat prepared
- Very prepared

- 2) How prepared do you feel you are to offer research data management support (consultation) to researchers at your institution?

- Very unprepared
- Somewhat unprepared
- Neither prepared nor unprepared
- Somewhat prepared
- Very prepared

- 3) How prepared do you feel you are to engage external stakeholders (e.g. faculty, campus IT, university administrators) in conversations about research data management at your institution?

- Very unprepared
- Somewhat unprepared
- Neither prepared nor unprepared
- Somewhat prepared
- Very prepared

- 4) What skills or information did you get out of this pre-conference?

- 5) What information was MISSING that you would have liked us to cover?

- 6) Did you feel that this workshop structure was an effective way to learn about this topic? Why or why not?

- 7) What role should ACRL play in providing continuing education programming for research data management?



# Open access and open data on natural disasters

International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 81–88  
© The Author(s) 2016  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216678236  
journals.sagepub.com/home/iff



**Marisa R. De Giusti**

Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina; Universidad Nacional de La Plata, Argentina

**Gonzalo L. Villarreal**

Universidad Nacional de La Plata, Argentina

**Carlos J. Nusch**

Universidad Nacional de La Plata, Argentina

**Analía V. Pinto**

Universidad Nacional de La Plata, Argentina

**Ariel J. Lira**

Universidad Nacional de La Plata, Argentina

## Abstract

After a flood which devastated the city of La Plata (Argentina) in April 2013, a survey of the water situation in the region revealed that relevant studies and projects related with this area were scattered and lacked visibility. This prompted academic, scientific, technologic, and governmental institutions to take action and work together to compile, organize, and disseminate available data in an unprecedented response to an environmental disaster in the region. For this purpose, a dedicated collection was created in the SEDICI, the institutional repository for the Universidad Nacional de La Plata, and subsequently, the Observatorio Medioambiental La Plata (OMLP), a repository for data on environmental research, was launched. This document describes these collaboration initiatives which aim to anticipate and reduce the impact of natural disasters, their main characteristics, their resources, and their current progress. Projects driven by the OMLP are also described, along with a description of their technical and infrastructure deployment as per the open access policies.

## Keywords

Digital repositories, environmental observatory, floods, open access

Submitted: 9 May 2016; Accepted: 28 September 2016.

## Introduction

On 2 April 2013, the city of La Plata, capital of Buenos Aires province, was devastated by the worst flood in its history. A storm of unprecedented intensity in the region resulted in a death toll of 89 human lives, thousands of evacuated residents, a city that was paralyzed for weeks, and economic losses adding up to millions (Cipponeri et al., 2014). This catastrophe had a strong impact on public opinion and prompted to action thousands of volunteers from across the

country, who helped the affected population return to their homes, and later rebuild or recover their lost property. The whole population was shocked by

---

### Corresponding author:

Gonzalo L. Villarreal, Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata, Calle 49 y 115 SN, Departamento de Ciencias Básicas, Facultad de Ingeniería, Universidad Nacional de La Plata, Argentina.  
Email: gonzalo@prebi.unlp.edu.ar

witnessing how a modern city, architecturally planned in the 19th century to be a governmental hub for the main province in the country, was devastated in a matter of hours by intense rainfall.

In the aftermath of the disaster, there was a need to take time to analyze and understand what had happened, and how it had escalated to such a catastrophe. It was concluded that the storm had a volume beyond anything on existing records (Benítez and Scarpati, 2005; Romanazzi, 2015) and that the water infrastructure and crisis management capacity of the city was not enough for a storm of such magnitude, not even for strong storms of a much lesser intensity (Facultad de Ingeniería, 2013). It was also noted that every piece of information, data, studies, reports, and all research related to the water and environmental situation of La Plata was scattered, as most of it was out of reach for the general public, and documents which could be accessed showed a level of disorganization and lack of proper display that made it practically useless for conducting further studies and projects.

One of the most immediate results of this analysis was the creation of a large amount of infrastructure works undertaken by the La Plata municipality, Buenos Aires province, and the national government, with the aim of improving and strengthening the city's ability to withstand these weather-related catastrophes (Romanazzi and Urbitzondo, 2014). As an alternative to the work done by government organizations, a series of initiatives came forth from local academic and research institutions looking to gather, simplify, optimize, and ensure open access to all the information available at that point.

This work describes actions and initiatives created by local academic and scientific institutions, in which SEDICI, the institutional repository of the UNLP, engages as a stakeholder in charge of managing and providing access to knowledge related to water infrastructure, risk management and analysis, crisis response, and environmental studies for the region. It also describes the subsequent creation of an environmental observatory capable of hosting research data gathered from projects related to these topics, with the collaboration of the National University of La Plata and other relevant institutions on a regional and national level.

### **Water emergency: A dedicated collection in the institutional repository**

In the previously mentioned conditions, the National University of La Plata took as a first measure to task their central repository, the Intellectual Creation Diffusion Service (SEDICI), with the creation of a

dedicated space capable of hosting and disseminating works related to local water and environmental infrastructure. For this, faculty members and researchers were asked to compile and store their research material in this new space, designed as a community inside the repository, and called Emergencia Hídrica [Water Emergency] (Servicio de Difusión de la Creación Intelectual, 2013a).

### **Goals and objectives of the repository**

SEDICI is the institutional repository for the National University of La Plata, created to host, preserve, and offer visibility to the production of its own academic units and additional branches. Some of its main goals are: (a) to manage, preserve, and disseminate the intellectual production of the university, together with the socialization of the knowledge resulting from the findings of its researches and studies; (b) foster the visibility and use of said academic and scientific production; (c) analyze and evaluate the impact and excellence of research financed with public funds; and (d) promote the adequate use of the property right of authors over their own work and the rights of the institution over its production (Villareal et al., 2008).

### **Organization of the repository**

Given the large amount of works and the typological complexity of the different resources, SEDICI organizes its archive in different communities, sub-communities, and collections in order to provide users with fast, organized, and intuitive access to the diverse records. The distribution in communities and collections in the repository reflects the structure of academic units in the university and also certain text genres (such as books or journals). Another set of collections is based on specific preservation or digitalization goals to provide access to works from institutions related to the UNLP, e.g. the Academia Nacional de Agronomía y Veterinaria (ANAV) or the Red de Universidades con Carreras en Informática (RedUNCI). Most of these special collections contain distinct works that are prominently non-textual, as is the case of the Open Educational Resources (learning objects), the UNLP Museum Network (in which physical objects are stored, represented by images and sound clips), the Vigo Experimental Art Center, Nanotechnology, and so on.

From a technological standpoint, after a migration in 2012 from a custom system, SEDICI uses the DSpace software as its main digital resource, while offering a series of services including document digitalization, online dissemination, digital preservation, and support for teaching and interoperability with



other systems, among other things (De Giusti et al., 2013; Texier, et al., 2012).

### Open access policy

SEDICI adheres to the open access guidelines for the definition of usage licenses of hosted works, and plays an active role in open access training, awareness, and promotion activities both at a local and international level through workshops, postgraduate courses, an online blog, social networks, conferences, and scientific meetings. The repository promotes the creation and dissemination of works under Creative Commons licenses, giving rise to many initiatives, including the UNLP Journals Portal, the UNLP Conferences Portal, and the UNLP Books Portal.

With the new collection called Emergencia Hídrica, SEDICI opens its self-archiving circuit to accommodate works related to water emergency matters, modifying its online portal to highlight this new collection inside the repository (Figure 1), with a view to assimilating existing works jointly with faculty members and researchers to coordinate efforts and immediately disseminate all the results from previous projects and researches, as well as from those created since the flood.

### Self-archival procedure

The self-archiving circuit is a procedure that allows every professor, researcher, or head of a project to upload works deemed relevant to this collection or another, through the use of a simplified web assistant. This assistant prompts the user in charge of the self-archiving process to input data for the document (authors, title, source organization, abstract, keywords), depending on the type of work, which can be an article, book, conference object, physical object, etc. After this step, the system allows the user to choose a Creative Commons license by answering two questions that will automatically determine which of the six different usage licenses will be assigned to the work.

Once the self-archiving process is completed, repository administrators receive the information and verify that all data is correct, normalize some of the metadata, incorporate additional metadata needed for cataloging activities to ensure resource preservation, while using the uploaded files to generate versions of the work that are best suited for its preservation (i.e. PDF/A documents). Once this verification is complete, the work is finally published into the repository. The Emergencia Hídrica collection currently contains around 130 works including journal articles,



**Figure 1.** List of featured collections in the SEDICI home page, including the Emergencia Hídrica collection. Servicio de Difusión de la Creación Intelectual (2016).

books and book chapters, reports, dissertations, and conference papers.

### Community training

On 7 May 2013, an intensive one-day workshop took place in the UNLP, in which researchers and specialists from different universities gathered to join their efforts and knowledge towards understanding the recent events and preventing similar situations from happening in the future. Part of the staff of the repository attended this workshop to give a series of short presentations, which included a video self-archiving, made available in the repository's YouTube channel (Servicio de Difusión de la Creación Intelectual, 2013b), to show how the repository worked, its collections, and the convenience of self-archiving as a way for researchers to quickly upload their works and later find them in a single location for future research and reference.

After the one-day workshop, the collection gained a large number of works that were incorporated to the existing material in the repository, which, while keeping its place in its source collections (for example, as dissertations) were immediately linked to the Emergencia Hídrica collection without the need of additional uploads. Currently, more and more works are being submitted to the collection, and further works are linked to it whenever an administrator deems their topic to be relevant to the matter. For example, the article by Canevari (2015), published in *Actas de Periodismo y Comunicación*, was immediately added to the collection

### Observatorio Medioambiental La Plata

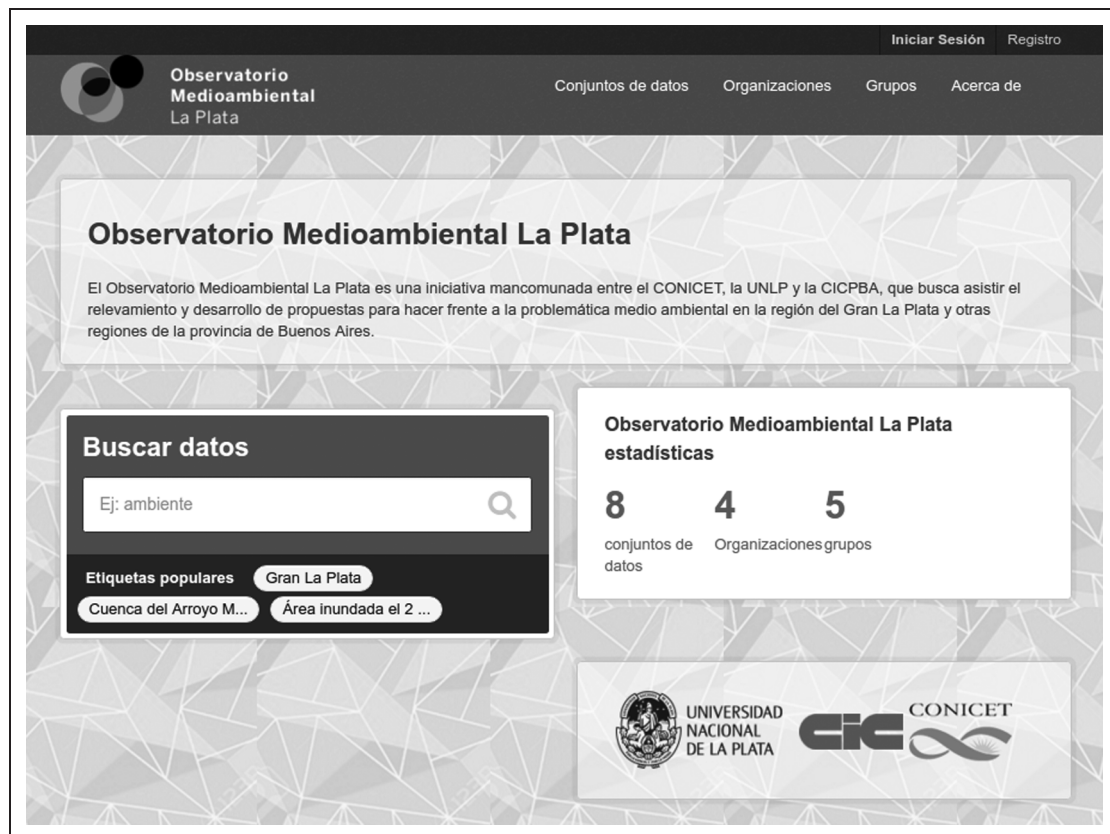
Launched in 2015, the Observatorio Medioambiental La Plata (OMLP) aims at disseminating research data and surveys carried out in these areas, driven by the increasing relevance of environmental issues and climate change challenges. This initiative is coordinated by the National University of La Plata and two of the most prestigious research institutions with the largest resource infrastructure in the country: the National Scientific and Technical Research Council (CONICET) and the Commission for Scientific Research of Buenos Aires (CICPBA). The observatory is a shared services center with a Steering Committee in charge of developing the center's activities, including a coordinator and an assistant coordinator as well as a Technical Advisory Commission, for the purposes of influencing decision-making on environmental management at national, provincial and municipal levels, and an Advisory Committee for Beneficiary Institutions. This observatory coordinates the efforts and available human and technological resources of multiple national, provincial and municipal entities, universities, scientific and technological institutions, and various private organizations which share an interest in environmental issues. Additionally, regarding the collaboration with government institutions, the observatory enables the production of specific reports. The first specific action taken by the OMLP was gathering the results, surveys and raw data obtained by the Proyectos de Investigación Orientada (PIO), a set of projects related to these issues, as well as offering a platform for their classification, cataloging and dissemination, providing access to the general public.

#### About the Proyectos de Investigación Orientada

The PIOs are the result of a special joint call by the Universidad Nacional de La Plata and the CONICET, launched at the end of the first semester 2013 with the

aim of financing research projects to solve water challenges in the region. Since then, multidisciplinary work is being done to approach future emergencies considering the relevant technical, urban, and social factors. Below is a list of on-going PIOs to be included in the OMLP at this initial stage:

1. Construcción de un Sistema Integrado de Gestión del Riesgo Hídrico en la Región del Gran La Plata: This project aims at increasing resilience of the whole environmental, social, and territorial system and its capability to face and recover from a threat by generating an integrated risk management system.
2. Estrategias para la gestión integral del territorio. Vulnerabilidades y Procesos de Intervención y Transformación con Inteligencia Territorial. Métodos y técnicas científicas ambientales, sociales y espaciales: This project covers two scenarios in the area called Gran La Plata, referring to problems and solutions in two extremely vulnerable areas in terms of environmental and social risk – the most critical areas in the basin of Arroyo Maldonado (La Plata) and the area which surrounds the refinery Refinería La Plata YPF (Ensenada and Berisso).
3. Evaluación y análisis de Riesgo Ambiental en el Área Gran La Plata: This project's goal is to obtain knowledge so as to correctly assess environmental variables, as well as the risks and consequences caused by the floods by promoting information gathering, monitoring and multidisciplinary studies with a view to improving decision-making and ultimately preserve the environment and people's life quality.
4. Las inundaciones en La Plata, Berisso y Ensenada: Análisis de riesgos y estrategias de intervención. Hacia la construcción de un observatorio ambiental: This project analyzes and explains the characteristics of floods in La Plata, Berisso and Ensenada, which can pose a territorial and environmental threat aiming to build a flood risk assessment matrix. This project's goal is to evaluate and propose action and intervention strategies to prepare for and prevent this type of threats in the short, medium and long term. Another of its goals is to lay the foundations for implementing an open data environmental observatory to provide input for public policies on social and environmental sustainability.
5. Comunicación y territorio: construcción de mapas territoriales de comunicación: This



**Figure 2.** Preview of OMLP's website homepage. Observatorio Medioambiental La Plata (2016a).

project aims to creating a georeferencing system from the point of view of complex systems and the research-action-engagement approach (investigación-acción-participación, IAP) to build territorial network maps in parallel with communication network maps, while showing their correlation with relations and practices among communities, organizations, and public policies. Its ultimate goal is to systematize information in an online, interactive, and intuitive digital platform for governmental, academic, and media organizations and institutions.

From the brief descriptions of these projects, their multidisciplinary nature becomes evident, including issues as diverse as mapping communication networks, risk analysis, evaluation of environmental parameters, and territorial management. The selection of IT tools that enable the operation of the OMLP is based on this diversity, prioritizing those that provide flexibility and extension capabilities to add new formats, metadata schemas or even customized visualization tools for different data sets.

Selected to support the OMLP, with aesthetic adjustments and functional additions, is an open

source software (Free Software Foundation, 2007) that enables the creation of open data websites. This tool uses the Python programming language and is being continuously improved both by the development team and the community of programmers who propose improvements and changes to the software, providing a plugin-based architecture that simplifies and promotes tool extension with new elements and features. Beyond its current features and tools, CKAN offers tremendous potential in terms of technical design and user community, making it a future-proof solution. (Winn, 2013). These features, coupled with clear and complete documentation available online (CKAN Docs, 2013) and the adoption of cataloging, security, and interoperability standards, are the key elements assessed when selecting this platform for the OMLP.

Aesthetically, the adjustments simplify browsing and provide a minimalist interface. The project's interinstitutional nature (CONICET-UNLP-CIC) is also highlighted (Figure 2). Much of the built-in functionality at this initial stage is based on plugins to simplify visualization of data provided by the different projects: *Text view*, for XML, JSON formats and other highlighted-syntax text formats; *DataExplorer*, for easy data access, filtering, charting, and mapping



**Figure 3.** Visualization of geographical data in KML format. Observatorio Medioambiental La Plata (2016b).

in CSV and MS Excel files; *GeoJSON*, to draw points on a map based on JSON files (Figure 3); *Image View*, for PNG, JPEG or GIF files; and *PDF Viewer*, to visualize files within each resource pages.

In the near future, more plugins will be added to enable multiple services and operations in the repository, both for external users and administrators, such as a plugin to enable different workflow configurations – implementation of this plugin is currently under analysis. Furthermore, while CKAN supports OAI-PMH by default, its interoperability will be implemented in subsequent stages.

The metadata schema offered by CKAN is used for data cataloging and implementation of the DCAT standard (Data Catalog Vocabulary, 2014) through a CKAN plugin which is currently being considered. This schema consists of a RDF vocabulary which enables management of an organization's open data to describe and disseminate them; DCAT standard is the main global standard for a vocabulary designed to simplify data search and interoperability of online data catalogs.

The site for the Observatorio Medioambiental La Plata (2016a, 2016b) data repository will initially provide the following services: dataset search and download and data visualization depending on their format. The contents submitted during this initial stage

include GIS layers, high-resolution images, text documents, and spreadsheets with survey answers. Initial submittals are being used to determine the most convenient file formats for each scenario and to assess whether they can efficiently be integrated in the web portal – e.g. for map or image visualization. It is worth stressing that the OMLP web portal is not yet accessible to the public; it will be launched once the final stage is completed, including tests, format assessments, and specification of a relevant workflow.

## Conclusions

This document describes the multiple actions taken at the local, provincial, and national level in response to a natural disaster. Especially noteworthy is the swift reaction of regional and national academic and research institutions, in a joint effort to prevent the occurrence of events such as the flood which devastated the city of La Plata. The scientific and academic community uses this knowledge via the SEDICI repository, to gather its production and works in related areas in a single collection accessible to the whole community, with a view to socializing knowledge for early alerts; simultaneously, appointed experts collaborate in multidisciplinary projects funded to give new answers in the area affected by



the disaster. Five teams, each including more than 50 multidisciplinary experts, collect, characterize, and systematize a huge amount of information, including primary data; the institutions involved then create the OMLP to meet the special requirements of the authorities and the community concerning the development of statistics and models based on gathered data. With the aim of simplifying visualization tasks and data exploitation, a new technology tool is created – specifically, a data repository where professionals classify, conduct studies, preserve, and disseminate the knowledge gained in the area with the ultimate goal of aiding their communities.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### References

- Benítez MI and Scarpati OE (2005) Las inundaciones en la ciudad de La Plata Su análisis en relación con las precipitaciones durante las últimas décadas del siglo XX. *Geograficando* 1: 111–121. Available at: <http://sedici.unlp.edu.ar/handle/10915/13922> (accessed 27 April 2016).
- Canevari T (2015) El campo de la comunicación desde un paradigma científico emergente en contextos de vulnerabilidad socio-ambiental y riesgo hídrico: Caso La Plata, Berisso y Ensenada. *Actas de Periodismo y Comunicación* 1(1). Available at: <http://sedici.unlp.edu.ar/handle/10915/51962> (accessed 25 April 2016).
- Cipponeri M, Salvioli ML, Larrivey G, et al. (2014) Vulnerabilidad de la población de la ciudad de La Plata (Argentina) a precipitaciones extraordinarias. *Aqua-Lac* 6(2): 11–20. Available at: [http://www.unesco.org/uy/phi/aqualac/fileadmin/phi/aqualac/Numero\\_6\\_vol\\_2/02\\_Vulnerabilidad\\_de\\_la\\_poblacion\\_de\\_la\\_ciudad\\_de\\_Plata.pdf](http://www.unesco.org/uy/phi/aqualac/fileadmin/phi/aqualac/Numero_6_vol_2/02_Vulnerabilidad_de_la_poblacion_de_la_ciudad_de_Plata.pdf) (27 accessed April 2016).
- CKAN Docs (2013) *Overview CKAN 2.6.0a Documentation*. Available at: <http://docs.ckan.org/> (accessed 20 April 2016).
- Data Catalog Vocabulary (DCAT) (2014) *W3C Recommendation 16 January 2014*. Available at: <https://www.w3.org/TR/vocab-dcat/> (accessed 20 April 2016).
- De Giusti MR, Villarreal GL, Terruzzi FA, et al. (2013) Interoperabilidad entre el Repositorio Institucional y servicios en línea en la Universidad Nacional de La Plata. In: *PKP international scholarly publishing conferences*, Ciudad de México, México. Available at: <http://sedici.unlp.edu.ar/handle/10915/27406> (accessed 30 April 2016).
- Facultad de Ingeniería, Universidad Nacional de La Plata (2013) *Estudio sobre la inundación ocurrida los días 2 y 3 de abril de 2013 en las ciudades de La Plata, Berisso y Ensenada. Report*. Argentina, Mayo: Universidad Nacional de La Plata. Available at: <http://sedici.unlp.edu.ar/handle/10915/27334> (accessed 13 April 2016).
- Free Software Foundation (2007) *GNU Affero General Public License*. Available at: <https://www.gnu.org/licenses/agpl-3.0.html> (accessed 8 September 2016).
- Observatorio Medioambiental La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de La Plata and Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (2016a) *Bienvenida – Observatorio Medioambiental La Plata*. Available at: <http://omlp.sedici.unlp.edu.ar/> (accessed 25 November 2016).
- Observatorio Medioambiental La Plata, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de La Plata and Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (2016b) *Lugares: Método Stlocus – Fases originales de stlocus – Observatorio Medioambiental La Plata*. Available at: <http://omlp.sedici.unlp.edu.ar/dataset/fases-originales-de-stlocus/resource/bc0aaeb1-632c-4a3b-a75e-f85acce49bc2> (accessed 25 November 2016).
- Romanazzi PG (2015) Aproximación a la estimación estadística de la precipitación máxima probable (PMP) para La Plata, Buenos Aires, Argentina. In: *III jornadas de investigación, transferencia y extensión de la Facultad de Ingeniería*. La Plata, Argentina, pp. 452–458. Available at: <http://sedici.unlp.edu.ar/handle/10915/47844> (accessed 27 April 2016).
- Romanazzi PG and Urbiztondo AJ (2014) Evaluación del desagüe existente y proyectado con un modelo dual: cuenca arroyo del Gato, La Plata, Buenos Aires, Argentina. In: *II Congreso internacional de hidrología de llanuras*, Santa Fe, Argentina, 22–26 September 2014, pp. 1–10. Available at: <http://sedici.unlp.edu.ar/handle/10915/40968> (accessed 13 April 2016).
- Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata (2013a) *Emergencia Hídrica*. Available at: <http://sedici.unlp.edu.ar/handle/10915/26450> (accessed 13 April 2016).
- Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata (2013b) *Tutorial de Auto-archivo de Emergencia Hídrica*. Available at: <https://youtu.be/umLtdb5i2LE> (accessed 13 April 2016).
- Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata (2016) *Repositorio Institucional de la Universidad Nacional de La Plata*. Available at: <http://sedici.unlp.edu.ar> (accessed 13 October 2016).
- Texier J, De Giusti MR, Oviedo NF, et al. (2012) El Uso de Repositorios y su Importancia para la Educación en Ingeniería. In: *World engineering education forum (WEEF)*, Buenos Aires, Argentina. Available at: <http://sedici.unlp.edu.ar/handle/10915/22943> (accessed 19 April 2016).



- Villarreal GL, De Giusti MR, Sobrado A, et al. (2008) SeDiCI (Servicio de Difusión de la Creación Intelectual). *D-Lib Magazine* 14(9/10): 1–11. Available at: <http://www.dlib.org/dlib/september08/villarreal/09villarreal.html>
- Winn J (2013) Open data and the academy: An evaluation of CKAN for research data management. In: *IASSIST 2013*, Cologne, Germany, 28–31 May 2013.

### Author biographies

**Marisa R. De Giusti** received her PhD in Computer Science from La Plata National University, Argentina. She is also a telecommunications engineer and literature professor. She has been the Director of SEDICI repository since 2003, of CIC-Digital repository since 2014, and of the Information Management Services Center (CESGI) since 2016. She is also the coordinator of OMLP. She is a professor in the UNLP Computer Science School, and her research interests include the fields of digital repositories, digital preservation, and open access.

**Gonzalo L. Villarreal** received his PhD in Computer Science from La Plata National University, Argentina. He also holds a degree in Computer Systems and in Computing Analysis. He is a professor in the UNLP Computer Science School, the sub-director of the Information Management Services Center (CESGI), and has published several articles on digital repositories, academic output visibility and impact, and systems interoperability.

**Ariel J. Lira** holds a degree in Computer Science from La Plata National University, Argentina, and has worked

in research and development at SEDICI repository since 2005. He is a member of the experts committee in the National Council of Digital Repositories in Argentina (SNRD) and a member of the advisory board in the Information Management Services Center (CESGI). He is sub-coordinator of OMLP where he has collaborated from its beginning.

**Analia V. Pinto** is currently working on her thesis to obtain the Editing Specialist degree at the Journalism and Social Communication School of La Plata National University. She has worked as a repository administrator since 2008 in SEDICI, is a free-lance editor and has been teaching Creative Writing workshops since 2010. She has an upcoming poetry book and several awards in that field, and is working on other poetry books and a novel. Her research interests include editing and digital editing, digital preservation, and book proof-reading and design.

**Carlos J. Nusch** is a literature professor at La Plata National University, Argentina, and has worked at SEDICI repository and Library Linkage Project (PREBI) since 2001. He is a member of the advisory board in the Information Management Services Center (CESGI) and part of the administrative staff of OMLP, where he has collaborated from its beginning. He also works in Classical Studies and Medieval Literature at the Latin Studies Center (CEL) and Social Sciences and Humanities Research Center (IdIHCS). He has published several papers in collaborative academic work, digital repositories, political discourse analysis and literature.



# Building a Research Data Management Service at the University of California, Berkeley: A tale of collaboration

International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 89–97  
© The Author(s) 2017  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216686982  
journals.sagepub.com/home/ifi



**Jamie Wittenberg**

Indiana University, Bloomington, USA

**Mary Elings**

University of California, Berkeley, USA

## Abstract

University of California, Berkeley's Library and the central Research Information Technologies unit have collaborated to develop a research data management program that leverages each organization's expertise and resources to create a unified service. The service offers a range of workshops, consultation, and an online resource. Because of this collaboration, service areas that are often fully embedded in IT, like backup and secure storage, as well as services in the Library domain, like resource discovery and instruction, are integrated into a single research data management program. This case study discusses the establishment of the program, the obstacles in implementing it, and outcomes of the collaborative model.

## Keywords

Academic libraries, data services, LIS as a profession

Submitted: 13 May 2016; Accepted: 28 September 2016.

## Context: Establishing the partnership

The University of California at Berkeley is one of the top research universities in the country, receiving over \$730 million in research funding last year and supporting over 100 research centers (Best College Reviews, 2016). In addition, UC Berkeley supports 170 academic departments and programs that are home to over 10,000 graduate students, 27,000 undergraduates, and 1600 full-time faculty. This community is dispersed across over 1230 acres in hundreds of buildings working in countless organized research units, centers, institutes, laboratories, facilities, and groups (UC Berkeley, n.d.).

With such an active and highly distributed research environment, the university has a significant task in providing research support to its campus community. An area of particular focus in the last year has been the adoption of open access (OA) policies that aim to make UC research outputs widely accessible. Despite the adoption of the UC Open Access Policy by the Academic Senate in 2013 and issuance of an expanded OA policy in 2015, these policies did not cover data specifically.<sup>1</sup>

As research changes and evolves, the services and needs of the community are evolving, especially with current data-driven research activities that rely on access to diverse data resources, data-intensive methods, and distributed computing tools and platforms in addition to meeting unfolding new federal requirements for data re-use and data security (Ferguson et al., 2014). In response to this evolving environment and its needs, the research community is seeking support in managing, storing, sharing, and preserving the data they produce in order to maintain the viability, reproducibility, and re-use of research data.

Tenopir et al. (2014) demonstrate in "Research data management services in academic research libraries and perceptions of librarians" that technical (hands-on) research data services are less common than informational (consulting) services. This lack

## Corresponding author:

Jamie Wittenberg, Indiana University, 107 S Indiana Ave,  
Bloomington, IN 47405, USA.  
Email: jamie.wittenberg@gmail.com

of technical data services in libraries may be addressed through a Library and IT partnership, and the UC Berkeley program attempted to address both technical and informational research data needs through such a partnership.

In 2015, the UC Berkeley Library and Research Information Technologies (Research IT) joined forces to develop a research data management (RDM) program to support this need for its large and active research community. The Library and Research IT partnership brought together two key organizations participating in the research process. Research IT is a unit situated in UC Berkeley's Office of the Chief Information Officer. Research IT provides research computing technologies, consulting, and community for the Berkeley campus. Research IT works in close partnership with the Office of the Vice Chancellor for Research and other campus technology services units, including the Library. The UC Berkeley Library connects students and scholars to information and services in support of research across campus. The Library seeks to select and create, organize and protect, provide and teach access to resources that are relevant to our campus programs. Together, these two organizations support the depth and breadth of campus research needs, which are increasingly digital in nature.

The goal of this collaborative partnership is to develop a program that will bring together the campus-wide systems and technical knowledge of Research IT with the research support and preservation expertise of the Library. This collaboration is a change for these two organizations and represents a new way of working together where each group is contributing to the process and sharing the costs. It is part of a push from several campus leaders, including leadership in Research IT and the Library, to build meaningful service collaborations between groups charged with providing campus wide services. It serves as a useful model of two large and diverse organizations taking joint ownership of a campus need, and working together to meet that need.

The collaboration of the Library and Research IT around the topic of RDM grew out of earlier work on the Research & Academic Engagement (RAE) Benchmarking project (2013), which was an effort by UC Berkeley's Research IT group and Educational Technology Services, with involvement from the Library.<sup>2</sup> The benchmarking project looked at existing and planned technology services and compared them with a set of peer institutions to help Berkeley develop a strategy for improving research, teaching, and learning technology support. One of the areas

RAE looked at was RDM, and the Library and Research IT recognized a shared interest in this area, as well as shared expertise, that could be brought together to advance the topic and provide support services.

A natural partnership flowed from the success of that project. By bringing the combined expertise of the Library and Research IT to bear on the emerging needs around RDM, we could advance use of services supported by Research IT and expand adoption of RDM as part of the public facing mission of the Library.

With offerings like high performance computing (HPC), virtual computing environments, and infrastructure services available through Research IT, and the Library's focus on research support and data management, the collaborative partnership covered many of the bases in a RDM portfolio. The more consultative role of the Library and the service-oriented role of Research IT completed the picture in terms of a RDM program, and thus the partnership was formed.

### **RDM program and goals: Improving campus support for research**

The stated objective of this effort was to establish a program for RDM services at the UC Berkeley campus level, through a joint partnership between Research IT and the Library. The goals in year one (January 2015–December 2015) were to design and deliver workshops, develop an RDM service guide, and develop an RDM consulting service.

The programmatic goal of the RDM initiative is to improve campus support for research output across all domains and subject areas, offering services around research data to help researchers steward, protect, and disseminate their data. Research data includes tabular and numeric data, text, images, audiovisual content, code, or any other actionable information generated during the research process. This typically excludes administrative data like financial and student records, as well as technical data, like the operations information generated by servers and laboratory equipment. RDM supports research data across domains and organizations, particularly in the areas of planning, organization, active data management, and sharing.

Research is highly distributed at Berkeley, and so are the services that support research. The efforts around creating a centralized RDM program can also be viewed as an attempt to knit together and coordinate a range of specialized and somewhat siloed services funded by departments, organized research units, and external "soft" money. The RDM program

aims to establish workflows and policies related to activities surrounding research data at Berkeley in addition to developing consulting, active data management, and training offerings. Digital Humanities has been managing the bulk of data management and curation requests focused on humanities data.

### **Contributions: Library and IT roles**

The RDM initiative at Berkeley is led by a core group consisting of leadership from both organizations, each committing one administrator to the team. The effort is managed by a team made up of a project manager, the research data management analyst, and an IT project manager. The core group working under this direction includes librarians and technical staff in the library, Research IT staff, a staff member from the California Digital Library, and a staff member from the UC Berkeley Campus Shared Services-Information Technology group. The core group meets bi-weekly and activities and deliverables are kept on a master calendar managed by the IT project manager. Meetings are led by the program manager, who also prepares the meeting agendas and keeps meeting notes.

By providing for equal staffing and equal participation, the program is expected to promote equal engagement in this effort on both sides. While there are no plans to establish a separate RDM unit within either organization, the work will continue to be coordinated among library and Research IT staff going forward. RDM will become part of what these groups provide, and that work will be shared among the participants.

As part of this effort, the Library and Research IT agreed to share support for a full-time RDM analyst who would split time between each organization. This position reports to both entities and has a physical space in both offices. The Library has provided a space for bi-weekly meetings and workshops, which to date have largely been focused on creating a cohort among librarians.

The role of the Library group in the RDM program has been to bring expertise in supporting the research process. The inclusion of librarians in the sciences, social sciences, and humanities brought a broad perspective to the core group. These participants are also part of a larger consulting network of departmental liaisons and subject specialists who are involved in research support on a day to day basis. These librarians offer support for and provide access to several data services including DASH, EZID, and the DMPTool, all hosted by the California Digital Library, another key partner in the collaboration.

The role of the Research IT group in the RDM program is to provide direction in the areas of active data management and data security, bringing expertise in data transfer, storage, and security. Research IT encompasses two groups that work very closely with RDM: Berkeley Research Computing (BRC) and Digital Humanities at Berkeley. Both of these groups are actively involved in projects that support the goals of RDM. BRC offers consultation and builds services related to high performance computing support and infrastructure. They are involved in experimental work on virtual workstations that are piloting solutions for RDM use cases – for example, developing an analytics environment for textual humanities data.

The partnership between the Library and IT is critical to the success of the RDM program, as is partnership with other organizations on campus like Educational Technology Services, the Berkeley Institute for Data Science, and the D-Lab. Support and participation by CDL is also central in this effort and will be increasingly important as the program moves forward.<sup>3</sup>

### **Professional culture: Navigating library and IT culture**

The cultural differences between the Library and Research IT organizations posed some challenges during the development of a joint program. It is important to note that the Research IT group has been involved in long-term work in museum informatics through the development of a collection management platform (CollectionSpace) and, consequently, Research IT has been deeply engaged in the libraries, archives, and museums space on multiple community source projects. This is highly unusual for a research computing group, and has been important in forging relationships between organizations. Despite this, fundamental cultural differences between the organizations emerged.

As detailed by Verbaan and Cox (2014: 211) in their discussion of occupational sub-cultures in RDM collaborations, librarians and IT staff have different and occasionally competing perspectives on RDM, wherein “Broadly speaking, IT Services focused on short term data storage; Research Office on compliance and research quality; librarians on preservation and advocacy”. This description of focus and scope aligns with the experience of the RDM program at UC Berkeley working with central IT more broadly.

For example, Library positions, being academic, are more flexible than IT staff positions, and it is not the norm for librarians to have a percentage of their position assigned to projects. In IT, it is typical to



have a 10% appointment or 50% appointment to a project where time spent on the project is tracked and assessed. A senior librarian provided feedback that the project had more IT-focused elements than library-focused elements. Perhaps one reason for this is that the time commitment of librarians is not as explicitly defined as the time commitment of IT staff; there were occasional misunderstandings related to workload, role, and commitment. As a result, some work related to the RDM program skewed more in the IT interest (active data management, storage), than the librarian interest (scholarly communication, preservation, research).

One significant example of an area where cultural difference between the Library and Research IT emerged was in approaches to researcher privacy. As established by the American Library Association (2002):

Protecting user privacy and confidentiality has long been an integral part of the mission of libraries. The ALA has affirmed a right to privacy since 1939... In keeping with this principle, the collection of personally identifiable information should only be a matter of routine or policy when necessary for the fulfillment of the mission of the library.

Leadership in Research IT preferred that identifying information like research names and departments be collected and shared among other consulting groups in order to provide a higher level of coordinated service. However, the Library has a more conservative stance towards information-sharing and does not systematically collect this kind of patron data. The resolution has been an endeavor to jointly draft a privacy policy.

Because RDM is an emerging field, it helps to have people working on the project that have a professional development mindset. Outreach and partnership with other organizations working in RDM is crucial to providing services that are relevant. Some examples of this are attending method and tool-based workshops related to scholarly communication, digital scholarship, and transparent research at UC Berkeley. We found that planning in these activities was an important part of the project.

### **Implementation: Consulting, resources, and training**

Developing the RDM Guide<sup>4</sup> was the first step in preparing to launch the program. The Guide is designed to serve as a resource for both service providers (consultants and librarians) as well as researchers. Content for the Guide was written collaboratively

by members of the team, based on area of expertise. It was developed in Drupal and is hosted by Pantheon, a web-hosting platform. The public-facing Guide contains content organized loosely by stages in the research lifecycle. Content consists of best practices, service offerings at UC Berkeley, useful tools, and case studies. There is also a back-end to the Guide, called the *Knowledge Base*, which is accessible to core team members only. The *Knowledge Base* serves as a tracking and record-keeping system that consultants use to document details of their consultations. This system is used primarily for program assessment.

Building the RDM Guide was an important part of the program because it offered the first opportunity for Research IT and the Library to collaborate on an enduring and publicly available RDM resource. Librarians and IT staff researched and wrote content together, defining the scope of the project and sharing knowledge.

As the RDM Guide took shape, development began on the consulting service. The RDM Consulting service is supported by three “triage” staff members who respond to requests and reach out to the broader consulting network to refer questions they are unable to answer. This network includes domain specialists, data scientists, qualitative data experts, librarians, and IT staff. There are many existing consulting services on UC Berkeley’s campus, including in Digital Humanities, Berkeley Research Computing, the Data Lab, and the Berkeley Institute for Data Science. It was important that the RDM consulting service integrated well with these existing services, and this allowed the team to borrow protocols and practices from partner organizations.

Building the consulting network was, in large part, an outreach and engagement objective. There were several individuals and groups that were already stakeholders in the RDM program who could serve as consultants, but one of the drivers for the development of the RDM program was bringing together distributed pockets of data management expertise. The consulting network was an opportunity to leverage knowledge in a range of domains, like cloud storage or metadata standards, for a research application.

The first goal of the RDM program was to train the staff that would make up the consulting network. Staff training for the RDM initiative has focused on three major groups: partner organizations, IT support, and librarians. Partner organizations did not receive formal training, but were engaged through a series of meetings and presentations. Following the September 2015 soft launch, RDM developed a training model targeting IT support staff and librarians. This model



proposed to create a cohort of early adopters that would participate in RDM training and serve as a point person for their unit or division. Cohort models have proved successful in training librarians, as demonstrated by Nardine and Moyo (2013) and Witteveen (2015). This group of early adopters made up Cohort 1.

Central IT (CSS-IT) support staff responsibilities are location-based, and a staff member is designated to a campus zone. That staff member will then respond to service requests within that zone. These staff are on the front lines in terms of responding to IT problems, some of which are related to research data. Because IT staff operate independently in this way, each zone representative was recruited for participation in Cohort 1. A total of seven members of the central IT group participated, including two supervisors.

UC Berkeley librarians typically operate within a division structure that partitions librarians and library staff based on domain. Library divisions include: Arts & Humanities, Engineering & Physical Sciences, Instructional Services, Social Sciences, and Life & Health Sciences. Because the University Library system at Berkeley comprises 32 constituent and affiliated libraries, these divisions can contain multiple libraries. Thus, the RDM team made the decision to recruit two representatives from each division that could serve as members of Cohort 1. Representatives were selected by the RDM team and division heads, based on expressed interest in RDM activities and the data-intensive nature of the librarian's role. A total of 11 librarians participated, including four division heads.

Cohort 1 members committed to a semester-long program that consisted of an orientation, a workshop, and an evaluation. The goal of the orientations was to introduce members to the need for and principles of RDM, to demonstrate the use of our online documentation, and to provide them with contacts for referral in the event that they or a colleague are asked an RDM question. Four orientations were offered during fall 2015: two for librarians, and two for IT. It was important to provide training for these groups separately in order to target existing workflows and tap in to referral processes within these organizations. Following these orientations, the RDM group presented RDM developments at a library-wide meeting and encouraged librarians and staff to seek out cohort members for more information, or with questions.

The fall 2015 workshop brought all of Cohort 1 together to introduce cohort members to different aspects of RDM and to some tools that they might find useful when engaging with researchers. The

workshop began with a keynote by John Chodacki, director of the California Curation Center (UC3) at the CDL. Several of the data management tools that Berkeley uses are developed and supported by the CDL, so this also provided an opportunity for relationship-building between these organizations. Following the keynote, three RDM team members gave lightening talks highlighting RDM use cases. One focused on data security, one focused on writing codebooks, and one focused on data confidentiality. Participants then split into small groups made up of both librarians and IT staff.

These groups completed an exercise that involved responding to various scenarios with RDM components. One sample scenario asked participants:

I am a researcher in agricultural economics and I have been publishing my data on my department's password-protected server, but my department is no longer going to maintain a server. What should I do to make sure that people can still find my data?

Participants then collaborated to answer the following questions:

- Who in the data management consulting network could help you answer this?
- What services exist at Berkeley that might provide support?
- Are there data privacy or security considerations?
- Are there policy, copyright, or intellectual property considerations?
- Where in the RDM Guide would you look for an answer?

This group exercise offered an opportunity for participants to practice working through some of the issues researchers face when interacting with data, as well as to work with their fellow cohort members.

The final element of the workshop was delivering two demonstrations of tools, both developed and supported by the CDL. The first, the DMPTool, is widely used at research institutions across the United States. It offers step-by-step guidance to researchers who are completing a Data Management Plan to fulfill the requirements of a funding organization – usually as part of a grant application. Data Management Plan review can serve as an effective basis for librarian training in RDM (Davis and Cross, 2015). The second demonstration was of DASH, an interface for data deposit into the Merritt data repository. Because UC Berkeley does not have an institutional data repository, DASH serves that function. Currently, the service is free to researchers and subsidized by

the university library, which makes it an attractive option for researchers who are interested in depositing their data and an important tool for Cohort 1 to be familiar with. Cohort 1 members were given access to test sites for each tool and encouraged to experiment with them.

In response to feedback from Cohort 1 members, the RDM Library Training group, made up of librarians and IT staff, shifted direction in 2016. Librarians requested training that was more nuanced, more concrete, and more directly relevant to their everyday activities. Several analyses have identified liaison librarians as critical to the success of an RDM program, and liaison librarian training was thus prioritized (Cox and Pinfield, 2014; Soehner et. al., 2010). The training team developed a 12-month, domain-based proposal for a training program for librarians.<sup>5</sup> The program divided the year into two-month training cycles. Each two-month training cycle targets a single domain, based on the existing library division structure. Library divisions will partner with the RDM team to create specialized content relevant to their domain. During a division's training cycle, the RDM team and division representative(s) collaboratively build workshop curricula and deliver two workshops. A monthly "Topics in Research Data Services" series, tailored to the domains of the training cycle, will support librarians and library staff as they develop a broad understanding of the challenges researchers face and gain confidence discussing various aspects of data management and stewardship. The first training cycle focuses on the Social Sciences Division. The curriculum was approved by the head of the Social Sciences division and developed in partnership with the Anthropology and Qualitative Data Librarian.

### **Outcomes: Resolving consultations, raising awareness, and training**

The RDM program has been successful in several areas: raising awareness of the program among UC Berkeley researchers, resolving RDM requests, training service providers in IT and the Library, and meeting project milestones on schedule.

In the 12 weeks between the service launch and the end of the semester, the program hosted or participated in 28 events, ranging from invited talks, to town hall presentations, to workshops, to demonstrations of the Guide. These events targeted both service providers and researchers. The program received 28 consulting requests from 19 departments and organized research units, 25 of which were resolved by the end of the quarter. The majority of researchers requesting

consultations were faculty and staff, closely followed by graduate students. Requests from undergraduates were rare. The Guide received 556 unique visitors who viewed approximately 2700 pages. The most frequently visited pages, after the home page, were: Data Management Best Practices, Consulting, and Data Management Planning.

Consultants were able to resolve many RDM questions, but several areas emerged as areas of greater need, with less support. In particular, active data management and securing research data need greater attention. Two working groups have convened to address these areas and develop recommendations.

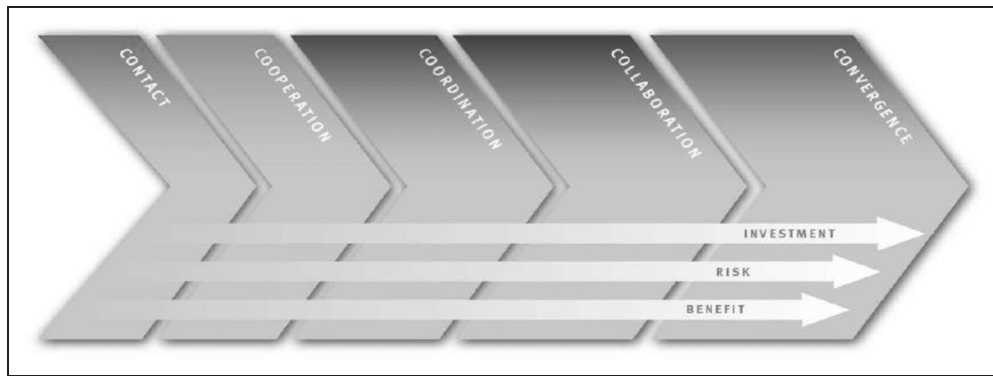
Domain-based training has proved to be very successful, with high levels of participation and engagement from librarians. This training is more successful than the generalized RDM training that attempted to target service providers from all domains and organizations.

Project management has been a very helpful part of the program development. As we have ramped up the work, having a project manager who kept the group on target and focusing on achieving goals across these two groups was very successful. Sticking to a firm calendar has helped the project manager to keep the deliverables on track. With only a handful of staff with time committed in real hours (FTE), other staff and librarians have had to make an effort to remain involved and committed given other priorities around their regular work.

The program now serves as an organized mechanism to help us better understand future researcher and support staff needs. It will help us determine where to focus our time and resources in an essential support area that is evolving fairly rapidly. In addition, we are building the foundation for future work, which will include a broader campus launch of RDM services and the development of additional services. The RDM consulting network helps to share important information with other campus service areas, such as computation (BRC) and learning analytics (ETS). All of this work taken together is helping us build a broad, meaningful service collaboration between groups charged with providing campus wide services.

### **Reflections on collaboration**

A collaboration of this type is not a simple undertaking for two large, complex, and disparate organizations like the Library and Research IT, but the shared interest in RDM support provided a common goal. A collaboration of this type can vary widely in terms of extent and outcomes, falling along a continuum ranging from a simple interaction over a common goal to



**Figure 1.** The Collaboration Continuum.

Source: Reprinted from 'Beyond the silos of the LAMs: Collaboration among libraries, archives, and museums' (Zorich et al., 2008) OCLC Research, p. 11. Copyright 2008 by the OCLC Online Computer Library Center, Inc. Reprinted with permission.

highly interdependent activities that involve shared risks and benefits. There is a model that is useful in discussing the trajectory of such partnerships called "The Collaboration Continuum" (Zorich et al., 2008). In that model (see Figure 1), partnerships move from basic contact through increasingly deepening relationships between the parties involved to a point of actual convergence. When a partnership reaches convergence, the collaboration is so ingrained that the parties no longer see it as a collaboration, but rather as a shared infrastructure that both parties have come to rely upon. Because the Library/Research IT partnership is a complex one, it might be instructive to look at how it has moved across this continuum.

In the case of the RDM service model, the process began with contact between two administrative leaders of the organizations. This started with an initial meeting to explore the idea of launching an RDM service. Research IT has a stake in the research process from the research cyberinfrastructure (RCI) side – tools, services, and community – and the Library has a similar position in supporting the research process through instruction, research design, access to resources, and publishing expertise. When they decided to work together on developing the program, the two parties moved from the contact stage of the partnership to cooperation, which made no commitment of time, money, or space, and had nothing in writing, but was simply an informal agreement to move the partnership forward.

As the partnership progressed, the parties agreed to coordinate on writing a job description for the shared RDM analyst and putting together a working group, which required a time commitment on both sides. Next they coordinated efforts to establish a calendar of work and deliverables which was managed by the assigned project manager and the leadership group.

Because this stage required a written agreement of how the analyst position would be shared and paid for, a commitment of some FTE of a project manager to the effort, as well as a commitment of time on the part of the two parties to meet regularly, this moved the project further down the continuum toward coordination.

As we see the partnership now, where we have the financial commitment of a shared position, a contribution of space where the analyst can work in each office, and written commitments of FTE to the project, we have reached the higher level of collaboration. At this point there is more investment from each party and a higher level of risk than in previous stages of the partnership. Should one of the partners back out of the collaboration, there would be a financial loss in staffing and time to untangle resources and dissolve written agreements. Having shared communication and program management responsibilities in the project has been a key method by which UC Berkeley has mitigated this risk.

On the plus side, each party has gained through sharing the work towards a common goal. The library has formed relationships and gained knowledge from interactions and sharing information with the Research IT staff. The Research IT group has gained a greater understanding of the research support process and the work done by librarians in this space. This has broadened the network of consultants across the campus that both groups can reach out to for support of research needs, so the campus community will also benefit. We have learned from each other and are better at what we do as a result.

Where the work to this point was largely additive, as we have moved toward true collaboration, the work is becoming more transformative as we begin to share work and reduce duplication of effort. This stage

suggests a level of trust between the partners, where they share risks and responsibilities, as well as the rewards.

The RDM program at Berkeley has not yet reached a point of convergence where each partner has become completely dependent on the other. This was not part of the program's stated goals, even though it is the next logical step in a collaboration. For this to happen, the Research IT group and the Library would need to, for example, commit resources to permanently support the shared position, or dedicate a shared space for this work, supported by shared funding. We would need to serve each other's missions in a way that dedicates resources across the partnership, or establish a formal partnership that forms a new organization to support this work. As the collaboration moves forward, these goals may become desirable but, for now, the close partnership will continue to work toward the shared RDM goals and continue to build an extended network of partners across campus as we move down the continuum.

### **Next steps: Formalizing and scaling the program**

The Research IT and Library partnership has come a long way in terms of their collaboration in a relatively short time. The collaboration has evolved into a successful venture to date and will continue to evolve as the RDM program establishes additional trainings and workshops, continues to develop its guide to services, and continues to share expertise across the two partners, as well as the extended network of partners.

As the program explores possibilities for additional services related to secure and active RDM, collaboration with other campus organizations is becoming increasingly important. As Wilson and Jefferies (2013: 245) discuss in "Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford", researchers prefer data management guidance that is specific to their discipline and methodology. This drive towards the provision of RDM services on a domain-specific basis necessitates domain expertise. This expertise does exist at UC Berkeley, but it is distributed among departments, research units, administration, and support teams. Partnering with these organizations is necessary to provide the support that researchers are looking for.

Because of the success of programs like this and a driving need for holistic solutions to research computing problems, Research IT is becoming increasingly involved in forging new collaborations with organizations at Berkeley and with other UC campuses. This includes a pilot project for managing OCR data

between Research IT, the Library, the D-Lab, and Digital Humanities. This project uses new analytics environments developed by Berkeley Research Computing to make licensed OCR software available to the entire campus community. Furthermore, Research IT is spearheading a consulting project that centers around a bi-annual consulting summit. This summit brings together consulting groups from IT, Educational Technology, the Library, the Berkeley Institute for Data Science, and the Geospatial Innovation Facility. In addition, the Library is partnering with UC San Diego to deliver a Data Carpentry workshop for librarians. These efforts aim to promote collaboration across these groups, and improve the impact and quality of research support services.

The library role has changed in terms of being better prepared to address data management and preservation needs as part of the broader research process. We are seeing this reflected in new library positions that include digital methods and data support as part of their portfolio. This situates these skills within the library and indicates that an RDM community is beginning to be built within the library, one which could extend to include other UC libraries. In addition, the library has gained an understanding of other services offered across campus and identified experts that can serve as partners, consultants, or referrals.

In 2016, the Library and Research IT hope to be able to substantially support to the RDM needs of campus. The RDM collaboration will continue to build relationships between IT and Library groups, pilot new services in active data storage, strengthen partnerships with the California Digital Library and researchers, and broker access to secure computing environments. By 2017, the program will be focused on formalizing RDM efforts within the institutional structure.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **Notes**

1. University of California, *Open Access Policy for the Academic Senate of the University of California*, <http://osc.universityofcalifornia.edu/open-access-policy> (24 July 2013). The scope of the policy only applies to scholarly articles.



2. Research and Academic Engagement (RAE) Benchmarking Project: <https://www.ets.berkeley.edu/projects/rae-services-peer-benchmarking>
3. These organizations represent the program's primary partners in the UC system: Research IT: <http://research-it.berkeley.edu/>; Berkeley Research Computing: <http://research-it.berkeley.edu/programs/berkeley-research-computing>; Digital Humanities at Berkeley: <http://digitalhumanities.berkeley.edu/>; California Digital Library: <http://www.cdlib.org/>; Berkeley Institute for Data Science: <https://bids.berkeley.edu/>; Educational Technology Services: <https://www.ets.berkeley.edu/>; D-Lab: <http://dlab.berkeley.edu/>
4. [researchdata.berkeley.edu](http://researchdata.berkeley.edu)
5. Educational resources associated with the librarian training program may be found at: <http://n2t.net/ark:/b6078/d1v88t>

## References

- American Library Association (2002) *An Interpretation of the Library Bill of Rights*. American Library Association. Available at: <http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy> (accessed 13 May 2016).
- Best College Reviews (n.d.) *The 50 Top Research Universities*. Available at: <http://www.bestcollegereviews.org/top-research-universities/> (accessed 13 May 2016).
- Cox A and Pinfield S (2014) Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science* 46(4): 299–316. DOI: 10.1177/0961000613492542.
- Davis HM and Cross WM (2015) Using a data management plan review service as a training ground for librarians. *Journal of Librarianship & Scholarly Communication* 3(2): 1–20. Available at: <http://jlsc-pub.org/articles/abstract/10.7710/2162-3309.1243/> (accessed 12 January 2017).
- Ferguson AR, Nielson JL, Cragin MH, et al. (2014) Big data from small data: Data-sharing in the 'long tail' of neuroscience. *Nature Neuroscience* 17(11): 1442. DOI: 10.1038/nn.3838.
- Nardine J and Moyo L (2013) *Learning Community as a Model for Cultivating Teaching Proficiencies among Library Instructors: A Case Study*. Available at: <http://library.ifla.org/106/1/100-nardine-en.pdf> (accessed 1 September 2016).
- Soehner C, Steeves C, Ward J, et al. (2010) E-science and data support services: A study of ARL Member Institutions. *Association of Research Libraries*. Available at: <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf> (accessed 21 September 2016).
- Tenopir C, Sandusky RJ, Allard S, et al. (2014) Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research* 36(2): 84–90. DOI: <http://doi.org/10.1016/j.lisr.2013.11.003>.
- University of California, Berkeley (n.d.) *By the Numbers*. Available at: <http://www.berkeley.edu/about/bythenumbers> (accessed 13 May 2016).
- Verbaan E and Cox AM (2014) Occupational sub-cultures, jurisdictional struggle and third space: Theorising professional service responses to research data management. *Journal of Academic Librarianship* 40(3–4): 211–219. DOI: <http://dx.doi.org/10.1016/j.acalib.2014.02.008>.
- Wilson JAJ and Jefferies P (2013) Towards a unified university infrastructure: The data management roll-out at the University of Oxford. *International Journal of Digital Curation* 8(2): 235–246. DOI: 10.2218/ijdc.v8i2.287.
- Witteveen A (2015) Better together: The cohort model of professional development. *Library Journal*, 3 December 2015. Available from: <http://lj.libraryjournal.com/2015/12/careers/better-together-the-cohort-model-of-professional-development/> (accessed 14 April 2016).
- Zorich D, Waibel G and Erway R (2008) Beyond the silos of the LAMs: Collaboration among libraries, archives, and museums. *OCLC Research*. Available from: <http://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf> (accessed 21 September 2016).

## Author biographies

**Mary W. Elings** is the Interim Head of Technical Services and Principal Archivist for Digital Collections at the Bancroft Library at the University of California, Berkeley. Her current research and work concentrates on issues surrounding born-digital materials, high-capacity digitization, digital humanities and social sciences, and research data management. Ms. Elings co-authored the seminal article "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums", and taught for over 10 years in the MLIS programs of Syracuse University and Catholic University, and is a regular technology guest-lecturer in the John F. Kennedy University Museum Studies program.

**Jamie Wittenberg** is the Research Data Management Librarian and Head of Scholarly Communication at the Indiana University, Bloomington Libraries. Previously, she served as the Research Data Management Service Analyst at the University of California, Berkeley. Her research interests include long-term preservation of born-digital and digitized scholarship, brokering meaningful access to research outputs, and facilitating the stewardship and reuse of research data.



International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 98–104  
© The Author(s) 2017  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216682041  
journals.sagepub.com/home/ifi



# Quality evaluation of data management plans at a research university

James E. Van Loon, Katherine G. Akers, Cole Hudson  
and Alexandra Sarkozy

Wayne State University, USA

## Abstract

With the emergence of the National Science Foundation requirement for data management plans, academic librarians have increasingly aided researchers in developing these plans and disseminating research data. To determine the overall quality of data management plans at Wayne State University, the Library System's Research Data Services team evaluated the content of 119 plans from National Science Foundation grant proposals submitted between 2012 and 2014. The results of our content analysis indicate that, while most researchers understand the need to share data, many data management plans fail to adequately describe the data generated by the project, how data will be managed during the project, or how data will be preserved and shared after the completion of the project. Our results also show that data management plan deficiencies vary across academic units, suggesting the need for differentiated outreach services to improve the strength of data management plans in future National Science Foundation grant proposals.

## Keywords

Data management, data sharing, evaluation, National Science Foundation, quality, research data

Submitted: 13 May 2016; Accepted: 7 November 2016.

## Introduction

Researchers are increasingly asked to provide access to their research data. Two key pieces of policy have set the tone concerning research data sharing in recent years: the National Science Foundation (NSF) 2011 requirement for the inclusion of data management plans (DMPs) in all grant proposals, and the 2013 memo by the Office of Science and Technology Policy requiring all major federal funding agencies to facilitate access to the publications and data resulting from federally funded research. As such, other federal funding agencies, including the National Institutes of Health, the Department of Energy, and the National Aeronautics and Space Administration now require or will soon require DMPs.

According to NSF guidelines, a DMP is a supplementary document of no more than two pages that describes how the proposal will conform to the funding agency's policy on providing access to research data (National Science Foundation, 2014). The DMP is reviewed as part of the intellectual merit or broader impact of each NSF proposal. Although the content

requirements for DMPs vary slightly across different NSF directorates, DMP elements expected for all directorates include data types and formats, methods of data sharing, and policies for data reuse and redistribution.

Wayne State University is a “doctoral university: highest research activity” according to the Carnegie Classification of Institutions of Higher Education (Indiana University Center for Postsecondary Research, 2015). Wayne State University has an annual research expenditure of over \$245m (University Research Corridor, 2016) and received almost \$12m in NSF research grants in fiscal year 2015 (National Science Foundation, 2016). In 2013, a team of librarians and specialists established Research Data Services (RDS) to provide outreach, consultation, and training on research data management and sharing to

## Corresponding author:

James E. Van Loon, University Libraries, Wayne State University,  
5265 Cass Avenue, Detroit, MI 48202, USA.  
Email: jevanloon@wayne.edu

Wayne State University faculty and research support staff. To further understand faculty research data management practices and to direct the future efforts of the RDS team, we analyzed the content of DMPs submitted by Wayne State researchers, focusing solely on NSF proposals due to the volume of NSF funding at our institution and the relative maturity of DMP requirements for this agency. The objectives of our study were to: (1) evaluate their overall quality and adherence to NSF guidance, and (2) determine whether academic units differ in their adherence to NSF guidelines.

### **Literature review: Content analysis of DMPs**

The overall quality of NSF DMPs has been evaluated in previous studies. Curty et al. (2013) used an online survey to assess attitudes and practices around data management planning among 966 NSF awardees from across the country and then analyzed the content of 68 DMPs volunteered by a subset of these researchers. They found several weaknesses in the DMPs, including dependence on informal or personal methods of sharing data (e.g. emailing upon request) and failure to address metadata standards and policies for data reuse/redistribution. As part of a pilot project to provide data management services to NSF applicants at the University of Michigan, Nicholls et al. (2014) acquired 104 DMPs from successful proposals from Engineering faculty and analyzed how well the DMPs conformed to NSF guidance. They concluded that although most DMPs were of acceptable quality, many lacked required elements, such as identification of the individuals responsible for data management and specification of the period of data retention. Bishoff and Johnston (2015) analyzed the content of 182 DMPs solicited from researchers at the University of Minnesota and found significant variation across DMPs in data sharing methods, the intended audience for sharing, and data preservation strategies.

Other studies have focused on evaluating NSF DMPs to specifically assess researchers' methods of data preservation and sharing, including the use of an institutional repository (IR) to provide access to data. Parham and Doty (2012) reviewed the content of 181 DMPs at the Georgia Institute of Technology, focusing on whether researchers indicated that they would use the IR to share their research data. They often found outdated or inaccurate references to the IR, presumably due to researchers' practice of sharing "boilerplate" DMP language across academic departments, suggesting the need to develop consistent language about repository services for research data and

to target IR awareness efforts to specific departments. Also, Mischo et al. (2014) examined 1260 DMPs at the University of Illinois and found no significant association between data storage methods and proposal funding success, although they discovered an increasing reliance on their IR as a venue for research data preservation over time.

Recently, the Data Management Plans as a Research Tool (DART) project, led by Rolando et al. (2015), developed and tested an evaluation rubric for NSF DMPs to create a robust and standardized assessment tool for DMPs to enable cross-institutional comparisons. An early version of the DART rubric was used by Samuel et al. (2015) to assess 29 DMPs from Engineering faculty at the University of Michigan. They found that the overall quality of DMPs varied greatly and identified elements that were often missing from DMPs, including clear roles and responsibilities for data management, metadata standards for describing research data, and policies for protecting intellectual property rights.

### **Motivation for the present study**

Although other researchers have evaluated the overall quality or specific elements of DMPs, we evaluated the quality of NSF DMPs at Wayne State University to: (1) characterize the content of DMPs created by researchers at our institution, and (2) identify significant variations in DMP content between academic units. Another potential outcome of this study was knowledge of specific and chronic deficiencies in DMPs that might help our team in developing tailored outreach and education for WSU faculty, administrators, research support staff, and other librarians.

### **Methodology**

We approached Wayne State University's Sponsored Program Administration (SPA) office with a proposal to study NSF DMP quality in 2014. SPA was receptive to our proposal and provided read-only access to the pre-award administrative system and support for compiling the DMP sample. Our study fell within the scope of program evaluation/quality improvement activities as defined by Wayne State's Institutional Review Board (IRB) and thus did not require IRB approval.

We compiled all funded NSF proposals between 2012 and 2014 and a roughly equal number of unfunded NSF proposals. After omitting proposals containing no DMP or for which the DMP content was minimal (e.g. conference or travel proposals), our final sample consisted of 119 DMPs from five WSU academic units as summarized in Table 1. To

**Table 1.** Final sample of DMPs.

Academic unit	Number of DMPs
College of Education <sup>a</sup>	2
College of Engineering <sup>b</sup>	61
College of Liberal Arts and Sciences <sup>c</sup>	50
Law School	1
School of Medicine <sup>d</sup>	5
Total	119

<sup>a</sup>Departments: Teacher Education (1), Theoretical & Behavioral Foundations (1).

<sup>b</sup>Departments: Biomedical Engineering (2), Chemical Engineering (14), Civil Engineering (6), Computer Science (13), Electrical & Computer Engineering (11), Engineering Technology (3), Industrial & Systems Engineering (5), Mechanical Engineering (7).

<sup>c</sup>Departments: Biological Sciences (4), Chemistry (22), Geology (5), Mathematics (8), Physics (11).

<sup>d</sup>Departments: Anatomy (1), Pediatrics (1), Pharmacology (1), Physiology (2).

maintain confidentiality of NSF proposal content, the DMPs were secured on a password-protected, internal library server for the duration of the study.

The DMPs were evaluated using a modified version of a rubric previously used by researchers at the University of Michigan (Nicholls et al., 2014; Samuel et al., 2015). Our rubric (Appendix 1) consisted of 15 items addressing the inclusion of information requested by the NSF (National Science Foundation, 2014) and other common pieces of information often found in DMPs. Two evaluators independently applied the rubric to each DMP, and any inconsistencies between evaluators were discussed and ultimately reconciled. Descriptive statistics for each rubric item were calculated for the full sample and separately for two major subgroups: the College of Engineering and the College of Liberal Arts and Sciences. Furthermore, we examined statistically significant differences in DMPs between the College of Engineering and the College of Liberal Arts and Sciences using Chi-square tests, with statistical significance set at  $p < 0.05$ .

## Results

### Overall quality of DMPs

Table 2 summarizes the proportion of DMPs containing each recommended element for the full sample and the two major subgroups: the College of Engineering and the College of Liberal Arts and Sciences. For the full sample, nearly half of the DMPs (49%) specified the individual(s) responsible for data management/sharing. A minority of DMPs (8%) specified the total amount of data that would be generated or the rate of data generation. Most DMPs (81%)

characterized data in terms of either its type (e.g. mass spectrometry data, scanning electron microscope images) or format (e.g. file extensions, name(s) of software used to collect the data). Less than half of the DMPs (38%) mentioned specific metadata standards or methods of data description (e.g. codebook, readme file). More than half of the DMPs (60%) discussed data back-up during the active project period. A vast majority of DMPs (92%) expressed an intention to share at least some data after completion of the project, but less than half (43%) specified the duration of data preservation.

We further addressed the specific methods by which researchers intended to share their data. For the full sample, the most frequently specified method of data sharing was posting data on personal websites/databases (51%; Table 2). The second most common methods were providing data upon request (e.g. by email; 24%) or depositing data in a dedicated data repository (24%). Among the DMPs, 13% mentioned sharing data through supplemental materials submitted alongside journal articles. Interestingly, a substantial proportion of DMPs (20%) stated that research data would be shared via journal articles (not as supplemental material) or conference presentations, indicating that some researchers do not distinguish between their results (i.e. summary data in tables and graphs) and the underlying data that support their results (i.e. individual-level or “raw” data in various file formats).

We also evaluated DMP content related to policies for data sharing. For the full sample, less than half of the DMPs (42%) mentioned policies for intellectual property, and only about one in five DMPs included statements about policies for data reuse/redistribution or protecting sensitive information.

Table 3 shows a breakdown of how researchers characterized the data generated by their project. For the full sample, less than half of the DMPs (42%) included both general (i.e. data types, such as mass spectrometry data or scanning electron microscope images) and specific (i.e. data format, such as file extensions or the name(s) of software used to collect the data) descriptions of the expected data, and smaller proportions of DMPs included either general or specific descriptions of data (32% and 7%, respectively) but not both. A substantial proportion of DMPs (19%) completely lacked a description of the data to be generated.

### Differences in DMP content between Engineering and Liberal Arts and Sciences

The full sample of DMPs ( $n = 119$ ) contained two major subgroups: the College of Engineering ( $n = 61$ )



**Table 2.** Elements contained in DMPs.

DMP element	Full sample	College of Engineering	College of Liberal Arts and Sciences
Basic elements			
1. Responsible individual	49%	44%	50%
2. Amount of data	8%	11%	6%
3. Expected types/formats	81%	87%	72%
4. Description/metadata	38%	36%	36%
5. Data backup	60%	59%	60%
6. Intention to share data	92%	93%	90%
7. Duration of data preservation	43%	56%	26%
Method of data sharing			
8. Email on request	24%	18%	30%
9. Personal website or database	51%	57%	50%
10. Journal articles or conferences	20%	28%	14%
11. Supplemental material	13%	2%	28%
12. Data repository	24%	18%	28%
Data sharing policies			
13. Reuse or redistribution	19%	30%	8%
14. Sensitive information	20%	26%	6%
15. Intellectual property	42%	64%	18%

**Table 3.** Characterization of data types/formats in DMPs.

Characterization of expected data types/formats	Full sample	College of Engineering	College of Liberal Arts and Sciences
Absent or unclear	19%	13%	28%
General (i.e. type)	32%	29%	34%
Specific (i.e. format)	7%	7%	8%
Both general and specific	42%	51%	30%

and the College of Liberal Arts and Sciences ( $n = 50$ ). Therefore, we analyzed differences in DMP content between these two academic units. Whereas 28% of DMPs from Liberal Arts and Sciences expressed the intention to share data via journal supplemental materials, only 2% of DMPs from Engineering expressed this intention ( $\chi^2(1, n = 111) = 16.3, p < 0.001$ ; Table 2). Slightly over half of Engineering DMPs (56%) specified the duration of data preservation, but only 26% of Liberal Arts and Sciences DMPs contained this element ( $\chi^2(1, n = 111) = 9.9, p = 0.002$ ). Furthermore, DMPs from Engineering were more likely to describe policies for data reuse or redistribution (30%;  $\chi^2(1, n = 111) = 7.9, p = 0.005$ ) and safeguarding sensitive information (26%;  $\chi^2(1, n = 111) = 7.9, p = 0.005$ ) or intellectual property (64%;  $\chi^2(1, n = 111) = 23.6, p < 0.001$ ) than DMPs from Liberal Arts and Sciences (reuse/redistribution: 8%, sensitive information: 6%, intellectual property: 18%). No other differences in DMP

elements between the two major subgroups were statistically significant.

## Discussion

We found substantial variation in the quality of individual NSF DMPs from Wayne State University researchers. Of the DMPs, 92% indicated that at least some data would be shared with others after the completion of the projects, which demonstrates that Wayne State researchers largely understand that the NSF expects broad data sharing. However, similar to previous studies (Bishoff and Johnston, 2015; Curty et al., 2013; Nicholls et al., 2014), we found that many DMPs failed to adequately describe the data that would be generated by the project, how data would be managed during the project, or how data would be preserved and shared with others after the completion of the project. In particular, we found that 51% of DMPs did not identify the individual(s) responsible for data management, which may be problematic for proposals involving multiple principal investigators or cross-institutional collaboration or for labs with high turnover rates for graduate students and research staff. Most DMPs (92%) did not provide an estimate of the total amount or expected rate of data generation, which is important for choosing the most appropriate data storage and preservation methods. Of the DMPs, 57% did not specify the duration that data would be preserved after the project or policies governing how other researchers might reuse or

redistribute their data, suggesting that researchers often do not carefully think about the lifespan of the data beyond the active period of the project. Furthermore, a majority of DMPs (62%) did not mention specific metadata standards or methods of data description methods, indicating that the data might not be easily discoverable by or understandable to other researchers in the long term.

In terms of data sharing methods, we found that researchers often rely on informal methods of providing access to data, such as sharing data through email upon request (24%) or through personal or project-specific websites or databases (51%). Only 24% of researchers stated that they would deposit data into a dedicated data repository. Informal data sharing methods, particularly sharing via email upon request, have been found to be less reliable for long-term data access than the use of a dedicated repository. Vines et al. (2014) found that the odds of successfully receiving data in response to an email request fell at the rate of 17% per year and that the chances of locating working email addresses for authors also dropped by 7% per year. Thessen et al. (2016) found that more than one-third of email requests for datasets received no response and that the overall success rate for email requests was 40%. They also found that sharing upon request was inefficient, requiring an average of 7.8 emails between the requester and data holder to negotiate a successful data transfer. Furthermore, Savage and Vickers (2009) found that only 10% of datasets requested by email were successfully received. Therefore, our RDS team will work to make Wayne State University researchers aware of the disadvantages of informal data sharing methods and encourage them to use more reliable and persistent methods of data sharing.

Interestingly, similar to previous findings by Bishoff and Johnston (2015), we discovered that a substantial proportion of DMPs (20%) stated that data would be shared via journal articles or conference presentations. In these cases, it was clear that researchers were not referring to sharing data through supplemental files accompanying journal articles; rather, they considered the publication of journal articles themselves as a way to share data. Although it is certainly expected that the results of research (i.e. interpreted, summary data in graphs and tables) would be shared through journal articles and conference presentations, these are not valid avenues of sharing the actual data underlying those results (i.e. uninterpreted, individual-level data in a variety of file formats). We believe that this may stem from a tendency for researchers to use the terms “data” and “results” interchangeably, which

suggests that researchers could benefit from greater awareness of the NSF and Office of Management and Budget definitions of “research data”.

Most NSF proposals in our sample originated from two academic units (the College of Engineering and the College of Liberal Arts and Sciences), allowing us to examine differences in DMP content between Engineering and basic science researchers. DMPs from Engineering researchers were less likely to mention data sharing through supplemental materials accompanying journals articles compared with DMPs from Liberal Arts and Sciences faculty. This finding suggests the need to improve awareness among Engineering researchers of the possibility of sharing research data via this method, which we will incorporate into future outreach efforts. Also, DMPs from Liberal Arts and Sciences researchers were less likely to specify the duration of data preservation and to describe policies for reuse/redistribution and protecting sensitive information and intellectual property rights compared with DMPs from Engineering faculty. These findings indicate a need to inform Liberal Arts and Sciences faculty about the importance of thinking about the lifespan of their data beyond the period of the project and considering whether steps should be taken to safeguard aspects of their data while also allowing the broadest access possible.

## Conclusion

By employing content analysis, we have characterized the level of quality and the variation between different academic units in NSF DMPs written by Wayne State researchers. We find that many DMPs provide an incomplete or ambiguous description of how research data will be managed and shared with others, suggesting that there is substantial room for improvement in DMP quality at our institution. Furthermore, we found several differences in DMP content between proposals from Engineering versus Liberal Arts and Sciences. These results indicate a need for the library to provide greater outreach, education, and consultation on developing strong DMPs and best practices in research data management and dissemination, and suggest that these efforts should be tailored to the needs and practices of particular groups of researchers.

Finally, we note that performing a DMP quality evaluation at our university has been a valuable experience for our RDS team, providing an opportunity to increase our knowledge of the grant application and data management planning process, to foster relationships between our team and university administrators

and other research support staff, and to create a DMP-related workshop for other librarians.

## Acknowledgements

The authors wish to thank Gail Ryan and Tim Foley of the Wayne State University Office of Sponsored Programs Administration for their support and guidance in this work.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References

- Bishoff C and Johnston L (2015) Approaches to data sharing: An analysis of NSF Data Management Plans from a large research university. *Journal of Librarianship and Scholarly Communication* 3(2): eP1231. DOI: 10.7710/2162-3309.1231.
- Curry R, Kim Y and Qin J (2013) What have scientists planned for data sharing and reuse? A content analysis of NSF awardees' data management plans. In: *Research data access & preservation summit 2013*, Baltimore, MD, USA. Available at: <http://surface.syr.edu/ischoolstudents/2> (accessed 2 May 2016).
- Indiana University Center for Postsecondary Research (2015) *Carnegie Classification of Institutions of Higher Education, 2015 Edition*. Available at: <http://carnegieclassifications.iu.edu/> (accessed 2 May 2016).
- Mischo WH, Schlembach MC and O'Donnell MN (2014) An analysis of data management plans in University of Illinois National Science Foundation grant proposals. *Journal of eScience Librarianship* 3(1): 31–43. DOI: 10.7191/jeslib.2014.1060.
- National Science Foundation (2014) *Chapter II - Proposal Preparation Instructions*. Available at: [http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg\\_2.jsp#dmp](http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#dmp) (accessed 2 May 2016).
- National Science Foundation (2016) *Award Summary: by Top Institutions*. Available at: <http://dellweb.bfa.nsf.gov/Top50Inst2/default.asp> (accessed 2 May 2016).
- Nicholls NH, Samuel SM, Lalwani LN, et al. (2014) Resources to support faculty writing data management plans: Lessons learned from an engineering pilot. *International Journal of Digital Curation* 9(1): 242–252. DOI: 10.2218/ijdc.v9i1.315.
- Parham SW and Doty C (2012) NSF DMP content analysis: What are researchers saying? *Bulletin of the American Society for Information Science and Technology* 39(1): 37–38. DOI: 10.1002/bult.2012.1720390113.
- Rolando L, Carlson J, Hswe P, et al. (2015) Data management plans as a research tool. *Bulletin of the American Society for Information Science and Technology* 41(5): 43–45. DOI: 10.1002/bult.2015.1720410510.
- Samuel SM, Grochowski PF, Lalwani LN, et al. (2015) Analyzing data management plans: Where librarians can make a difference. In: *122nd ASEE annual conference & exposition*, Seattle, WA, USA. Available at: <https://www.asee.org/public/conferences/56/papers/12072/view> (accessed 2 May 2016).
- Savage CJ and Vickers AJ (2009) Empirical study of data sharing by authors publishing in PLoS journals. *PLoS one* 4(9): e7078. DOI: 10.1371/journal.pone.0007078.
- Thessen AE, McGinnis S and North EW (2016) Lessons learned while building the Deepwater Horizon Database: Toward improved data sharing in coastal science. *Computers and Geosciences* 87: 84–90. DOI: 10.1016/j.cageo.2015.12.001.
- University Research Corridor (2016) *About the University Research Corridor*. Available at: <http://urcmich.org/about/> (accessed 2 May 2016).
- Vines TH, Albert AYK, Andrew RL, et al. (2014) The availability of research data declines rapidly with article age. *Current Biology* 24(1): 94–97. DOI: 10.1016/j.cub.2013.11.014.

## Author biographies

**James E. Van Loon** is a Science Librarian at Wayne State University in Detroit, MI. He holds an MS in Mechanical Engineering and an MLIS in Library and Information Science from Wayne State University. He provides research and instructional support to the College of Engineering, and research data services as a member of the library's RDS team.

**Katherine G. Akers** PhD is a Biomedical Research and Data Specialist in the Shiffman Medical Library at Wayne State University. She provides support for research data management and sharing, scholarly publishing, and meeting funding agency requirements for clinicians and researchers at the Wayne State University School of Medicine and across campus. Katherine was originally trained as a researcher in psychology and neuroscience and was a CLIR Postdoctoral Fellow in Academic Libraries.

**Cole Hudson** is a Digital Publishing Librarian and Adjunct Professor at Wayne State University in Detroit, MI. His research interests include open access repositories and open source software development. He holds an MS in Library and Information Science and MA in English Literature from Rutgers University and the University of Arkansas, respectively.

**Alexandra Sarkozy** MIS is a Learning and Research Support Librarian at Wayne State University. She supports

students and researchers in the College of Nursing and across other science disciplines with expert searching, library instruction, and scholarly communications advising. As part of the Research Data Services Working Group, Alexandra also provides training and recommendations to researchers creating data management plans.

## Appendix I

### Wayne State University DMP evaluation rubric

#### Basic DMP elements

1. Are the individual(s) responsible for data management specifically named (or referred to as “the PI”)?

- 1 = yes
- 0 = no/not clear

2. Is the total amount of expected data and/or expected rate of data generation specified?

- 1 = yes
- 0 = no/not clear

3. Are the file formats of expected data specified (e.g. file extensions, name of data collection software)?

- 0 = no/not clear
- 1 = general description (e.g. mass spectrometry data)
- 2 = specific description (e.g. file extensions, software used)
- 3 = both general and specific description

4. Will specific metadata standards and/or other description methods (e.g. readme files, codebooks, and lab notebooks) be used?

- 1 = yes
- 0 = no/not clear

5. Is a method of data backup (e.g. RAID, remote backup, external hard drive) specified?

- 1 = yes
- 0 = no/not clear

6. Will any data and/or code be made accessible after the study?

- 1 = yes
- 0 = no/not clear

7. Is the duration of data/code preservation specified?

- 1 = yes
- 0 = no/not clear
- Method of data sharing

8. Will data/code be provided (e.g. emailed) upon request?

- 1 = yes
- 0 = no/not clear

9. Will data/code be posted on personal or project-specific website or database?

- 1 = yes
- 0 = no/not clear

10. Will data be shared via journal articles or conference presentations?

- 1 = yes
- 0 = no/not clear

11. Will data/code be submitted to journals as supplemental material?

- 1 = yes
- 0 = no/not clear

12. Will data be deposited in a dedicated data repository/archive?

- 1 = yes
- 0 = no/not clear
- Data sharing policies

13. Are policies for data re-use or redistribution specified?

- 1 = yes
- 0 = no/not clear

14. Do policies for data access and sharing specify protections against disclosure of sensitive information?

- 1 = yes
- 0 = no/not clear

15. Do policies for data access and sharing specify protections for safeguarding intellectual property rights?

- 1 = yes
- 0 = no/not clear





# Team-based data management instruction at small liberal arts colleges

**Ryan Clement**  
Middlebury College, USA

**Amy Blau**  
Whitman College, USA

**Parvaneh Abbaspour**  
Lewis & Clark College, USA

**Eli Gandour-Rood**  
University of Puget Sound, USA

International Federation of  
Library Associations and Institutions  
2017, Vol. 43(1) 105–118  
© The Author(s) 2016  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/0340035216678239  
journals.sagepub.com/home/iff



## Abstract

This paper describes a collaborative approach taken by librarians at five small, regional liberal arts colleges to developing/enhancing research data management services on their campuses. The five colleges collectively belong to a consortium known as the Northwest Five Consortium. Over 10 months, librarians from the five schools collaborated to plan a data management and curation workshop with the goals of developing relationships with researchers working with data, developing their own research data management skills and services, and building a model for future training and outreach around institutional research data management services. This workshop brought together research teams including faculty, students, and librarians, and incorporated active learning modules as well as in-depth pre-workshop discussion. This article will discuss the context and background for this workshop, the model itself, and the outcomes and possibilities for future developments.

## Keywords

Communities of practice, data services, information literacy and instruction, preservation and conservation, research data management

Submitted: 17 May 2016; Accepted: 19 September 2016.

## Introduction

Over the past decade academic libraries at institutions of all sizes have been developing services to support research data management and curation (Choudhury, 2008; Goldstein and Oekler, 2011). Spurred by mandates passed down from federal funding agencies that are creating new imperatives for data management and sharing among the research community, academic libraries have seized this opportunity to address this evolving information need on campus (Heidorn, 2011; Hswe and Holt, 2011; Walton, 2010). Extending their training in organizing and managing information, as well as digital preservation and records

management, librarians are disposed both by their skill-set and their roles in scholarly communications support to address these needs (Brandt, 2007: 365). Moreover, the provision of research data services is an opportunity for libraries to demonstrate their relevance to the campus community at a time when conceptions of scholarly communication are evolving. On campuses where libraries compete with other

## Corresponding author:

Ryan Clement, Middlebury College, 110 Storrs Ave, Middlebury, VT Vermont 05753, USA.  
Email: rclement@middlebury.edu

departments for resources, the provision of support for research data management (RDM) can be a fresh approach for libraries to build their value on campus in both the scholarly communications and instructional domains.

### *Common challenges in developing research data management services*

**Training and time investment.** There are some common challenges encountered by librarians working to establish RDM services. Although data management is a ready match for librarians' skill-sets, they may lack specific training in these areas or be unfamiliar with the domain-specific knowledge needed to apply more general curation skills effectively. Furthermore, given that the vast majority of research data is produced digitally, managing this information requires a perpetual investment of time to stay up to date with new technologies. RDM is also a relatively new field experiencing rapid development across a variety of disciplines, requiring further time to stay abreast of domain-specific best practices.

**Building campus relationships.** Beyond requirements for professional development and training, the establishment of RDM services and workflows requires developing stakeholder relationships across the institution. Providing effective RDM frequently requires that the library develop new relationships and build existing partnerships across campus with units such as Information Technology (IT) and Sponsored Research. As familiar and accepted as conversations about RDM have become in the academic library community, other units may be entirely unfamiliar with the library's new (and often self-defined) mandate to support data management and may, for a variety of reasons, be reticent to engage with the library to address these needs. This initial caution may extend to the very stakeholders these services aim to support – researchers.

**Outreach to faculty partners.** Beyond these initial questions about new services and roles, librarians seeking to establish RDM services may encounter other challenges in working with faculty researchers. Among these is the deceptively simple task of informing researchers that these new services exist, and retraining faculty to reach out to the library when RDM needs arise. This retraining process involves two further challenges – convincing the researchers that the library is capable of providing these services, and defending the novel concepts of data management plans in particular and data sharing in general.

Researchers may fiercely defend disciplinary traditions around sharing research, and see these new mandates and requests for their output as an encroachment, even as these requirements have often been crafted with the utmost deference to establish disciplinary practices.

**Background on the Northwest Five Consortium institutions.** This paper describes a collaborative approach taken by librarians at five small, regional liberal arts colleges to developing/enhancing research data management services on their campuses. These colleges are Lewis & Clark College, Reed College, Whitman College, Willamette University, and University of Puget Sound. The five colleges collectively belong to a consortium known as the Northwest Five Consortium (NW5C), which has the mission of promoting cross-institutional collaboration and is supported by a grant from the Andrew W. Mellon Foundation. Over 10 months, librarians from the five schools collaborated to plan a data management and curation workshop with the goals of developing relationships with researchers working with data, developing their own RDM skills and services, and building a model for future training and outreach around institutional RDM services. The workshop was supported with a mini-grant from the NW5C Fund for Collaborative Inquiry.<sup>1</sup> This article will discuss the context and background for this workshop, the model itself, and the outcomes and possibilities for future developments.

All five institutions are small, private liberal arts colleges, with student enrollments ranging between 1400 and 2600 full-time equivalent (FTE). They share several metrics by which the Carnegie Classification of Institutions of Higher Education (Indiana University Center for Postsecondary Research, n.d.) classifies colleges and universities in the United States. They are four-year Baccalaureate Colleges with an Arts & Sciences Focus – the majority of degrees awarded at each college are non-professional undergraduate degrees (Bachelor of Arts or Bachelor of Sciences in the arts, humanities, social sciences, or natural sciences, rather than in disciplines such as business, nursing, or engineering). All five institutions are primarily residential (at least half of the undergraduate students live on campus and at least 80% attend full-time) and all rank among the most selective (80th to 100th percentile) of baccalaureate institutions. The concentration on undergraduate education in a residential setting, the relatively small number of students and the generally low student to faculty ratio mean that there is a strong expectation that students will know their professors and fellow

students well, and that the personal attention that students receive from faculty and staff will translate into “high-impact” educational experiences for students (Kuh et al., 2008). The faculty expectations at selective liberal arts colleges generally follow a “teacher-scholar” model; while excellent teaching is a fundamental expectation, that teaching is rooted in disciplinary scholarship. Faculty at these institutions therefore have active research agendas, which often rely upon research assistance from their undergraduate students.

While many of the challenges libraries face in establishing RDM services are common across a wide variety of institution types, some of these challenges are magnified in unique ways at small liberal arts colleges. The intimate, student-centered environment that is characteristic of liberal arts colleges poses additional challenges to developing RDM services that must be addressed by librarians in these contexts. Faculty tend to be engaged in a particularly time-intensive style of teaching and mentoring, leaving limited time to spend on lab management. In this context, data management can fall by the wayside. Research assistants are often undergraduate students with little exposure to, and instruction in, managing research data. Furthermore, there tends to be a high turnover rate among these research assistants; many will spend only 2–3 years working in a lab before graduation, sometimes with interruptions for time off or study abroad.

These immediate, situational challenges are often compounded by the larger institution-level privileging of curricular needs over research in the allocation of resources. By virtue of being in a smaller institution there is often less infrastructure (both human and technological) in place to support research and fewer resources that can be devoted to these needs than in large research-intensive universities. This holds true in the library where data management responsibilities are often added to a traditional liaison position in addition to other instructional, technical, and functional work roles.

At the same time, the opportunities for data management and data information literacy at liberal arts colleges and primarily undergraduate institutions (PUIs) are also significant. The close-knit environment lends itself to collaboration between librarians and faculty, and data management services can act to reinforce or grow liaison relationships, especially with departments less reliant on the library for more traditional services. The combination of liaison responsibilities with RDM services allows for the flexibility to be able to consult with faculty and students *in situ*, in a laboratory setting, if that works best for the research group. There are many possibilities

for developing innovative approaches to data management and curation services and for providing data information literacy training for undergraduate students.

## Literature review

### Library data services

While most of the initiators of library data services have been large research universities (Antell et al., 2014; Cox and Pinfield, 2014; Heidorn, 2011; Soehner et al., 2010; Tenopir et al., 2012, 2013, 2014), there is a growing body of literature about data services at Master’s universities and PUIs (Goldstein and Oelker, 2011; Scaramozzino et al., 2012; Shorish, 2012; Stamatoplos et al., 2016; Toups and Hughes, 2013), and about how outreach efforts related to data literacy and data management may be implemented with undergraduate students at a range of institution types (Ball and Medeiros, 2012; MacMillan, 2010; Mooney et al., 2014; Piorun et al., 2012; Qin and D’Ignazio, 2010; Reisner et al., 2014; Shorish, 2015; Stephenson and Caravello, 2007; Strasser and Hampton, 2012; Zilinski et al., 2014). Librarians at PUIs have argued that while resources may be limited at smaller institutions, the same imperatives to act as data stewards and to transmit best practices apply (Shorish, 2012). The key factor for smaller institutions is the adaptation of emerging best practices for their environments:

Large research institutions may have more resources and staff, and their need for data curation may be greater. But we at smaller institutions are poised to learn from their pioneering work, borrow accordingly, and tailor data support services to the local needs of our patrons. (Toups and Hughes, 2013: 232)

Library data support services for undergraduate students draw upon sets of competencies, or literacies, related to data production or use. Calzada Prado and Marzal (2013: 126) define *data literacy* as “the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle and ethically use data”. Qin and D’Ignazio (2010: 2) define *science data literacy* as a praxis-based skillset with emphasis on “functional ability in data collection, processing, management, evaluation, and use”. Carlson et al. (2011: 634) distinguish *data information literacy* from data literacy, statistical literacy, and information literacy with respect to the production of information in addition to its consumption, bringing the various literacies together as it “merges the concepts of researcher-as-producer and researcher-as-consumer of data products”. They situate data

information literacy competencies within the ACRL Information Literacy Competency Standards in order to propose the essential components of a data information literacy program administered by librarians (Carlson et al., 2011: 652). Considering data information literacy in light of the ACRL *Framework for Information Literacy for Higher Education*, Shorish (2015: 100) argues that data information literacy “should be treated as any of the other literacy competencies and incorporated into the workflow of outreach librarians”.

Many examples of librarian-led data information literacy instruction for undergraduates are in classroom settings. However, it can be difficult to find space in a tight undergraduate science curriculum for a stand-alone data literacy or data management course (Qin and D’Ignazio, 2010: Lessons Learned 3). This is one reason to look to the undergraduate research experience as a place to integrate data literacy or data management skills into undergraduate research experiences; another reason is the ability to immediately relate skills learned to their authentic contexts. In their needs assessment of data information literacy for faculty and students, Carlson et al. (2011: 648) find that while there were common areas of need, faculty needs were related to data they created themselves, while the students surveyed were using data from external sources as part of a course. However, Stamatoplos (2009: 240) points out that the differences between information needs for students and faculty are reduced when students are undertaking authentic research projects.

### *Undergraduate research*

The Council on Undergraduate Research defines undergraduate research as “an inquiry or investigation conducted by an undergraduate student that makes an original intellectual or creative contribution to the discipline” (Rowlett et al., 2012: 2). Undergraduate research differs from most coursework because of its open investigative nature, usually managed through the mentorship of a faculty researcher (Stamatoplos, 2009: 237). Involvement with undergraduate research has been identified as a “high-impact” practice which contributes to student engagement and success (Kuh et al., 2008: 14). While the Boyer Report called for making inquiry-based learning the standard at research universities in order to improve the quality of undergraduate experience at such institutions (Boyer Commission on Educating Undergraduates in the Research University, 1998: 15), the emphasis on active, collaborative learning and faculty mentorship has been standard at liberal

arts colleges, where undergraduate students rather than graduate students work as research assistants for faculty research groups (Shorish, 2015: 101).

By taking part in undergraduate research experiences (UREs), students not only help to create new knowledge, but also learn about the norms of scholarly work in their discipline. Thus, undergraduate research experiences allow students to be integrated into communities of practice (Hunter et al., 2007: 37). However, formal instruction in data information literacy is not necessarily the norm in these communities of practice. Hunter et al. (2007: 46) identified eight categories of gains for students who participated in science UREs, including intellectual gains of “thinking and working like a scientist”, learning professional norms of “becoming a scientist”, and acquiring skills in various information literacy-related categories such as “work organization”, “computer”, and “information retrieval”. Hunter et al. do not consider data management competencies explicitly; they might be included in any of the above-mentioned categories, although the praxis-based elements probably are best categorized with skills.

The undergraduate research experiences described by Hunter et al. do not include instruction or mentorship from librarians. Stamatoplos (2009: 239) calls for formal involvement of librarians with undergraduate research programs, as students conducting independent research have even greater need for advanced information and library skills. Hensley (2015: 722) has also shown ways in which library information literacy (IL) initiatives can intersect productively with UREs, as they share emphases on developing “critical thinking and problem-solving skills”. Respondents to her survey provided IL instruction for undergraduate researchers in a variety of formats, including one-on-one with students or faculty/mentors, workshops with lab groups, and involvement as team-teacher or as instructor of record in a credit-bearing course (Hensley, 2015: 730). The examples of IL topics for undergraduate research, however, focused primarily on database searching and citation management; explicitly data-related aspects of IL accounted for just over 2% of the topics taught (numeric and spatial data 1%, data visualization .7%, developing a data management plan .5%) (Hensley, 2015: 735).

An instance of librarians working directly with undergraduate research groups is the social-work pilot project undertaken by Mooney et al. They describe their work embedding librarians with undergraduate research teams over the course of two semester-long research projects in social work, with biweekly meetings (Mooney et al., 2014: 374). Among the advantages to this approach are just-in-time lessons that fit



the development and workflows of the projects. Mooney et al. identify the potential for partnership between librarians and research teams at a place of knowledge production. The involvement of librarians with research groups helps to integrate a range of information literacy skills into the undergraduate research experience, something that Mooney et al. find lacking overall:

It is our assessment that data management and the broader scope of data information literacy, indeed even basic library research skills, are not widely perceived as explicit goals of participation in undergraduate research despite an overarching goal for the advancement of real world research experience. (Mooney et al., 2014: 371)

Because the approach taken to embed librarians in these projects will not scale, Mooney et al. (2014: 383) foresee offering librarian-run workshops for faculty and undergraduate researchers, complemented by librarian attendance at research team meetings.

### *Librarian training*

The growth in interest in data services in libraries also has significant impact on librarian training in data services. While library and information schools are adding data services and data curation tracks for new graduates (Creamer et al., 2012; Keralis, 2012), additional training of librarians, especially in departmental liaison or institutional repository roles, may be needed for them to evaluate and develop services in their libraries (Bresnahan and Johnson, 2013), especially if data curation roles were not part of their original career plans (Pryor and Donnelly, 2009: 164). Tenopir et al. (2013: 72) find that while over 75% of librarians who support research data services as an integral part of their job responsibilities feel prepared to do so, less than 50% of those who occasionally support research data services feel prepared.

Both online and in-person training opportunities in RDM exist for librarians, with a broad range of instructional models, fee structures, and time commitments (ANDS, n.d.; EDINA and Data Library, University of Edinburgh, n.d.; Guy, 2013; Research Data Netherlands, 2016; Rice, 2014; Verbakel et al., 2013). A few models of librarian training that explicitly incorporate face-to-face training or interaction with researchers are of interest for the incorporation of elements important to the workshop model discussed in this article. The Data Information Literacy Project ([datainfo.org](http://datainfo.org)) presents a model in which librarians familiarize themselves with types of research in specific disciplinary areas, interview researchers using the Data Curation Profile, and then work with faculty

to develop data information literacy programs for graduate or undergraduate students (Carlson et al., 2015). The immersive Informatics training program at the Library at the University of Melbourne, Australia, embedded library and IT support staff in a research context with data producers for a ten-module series of topics presented over 16 weeks (Shadbolt et al., 2014). In order to provide required training for a library representative for Statistics Canada data at participating institutions across Canada, the Data Liberation Initiative (DLI) instituted a common curricular model built on regional workshops. The in-person regional trainings created a “sense of community” among participants and trainers that continued beyond the workshop proper (Watkins et al., 2004: 19).

### **Northwest Five Consortium workshop model in theory**

The planning process for the workshop took place over 10 months and was conducted primarily via videoconferencing meetings by the five librarian co-facilitators. The group was convened by the principal investigators in September of 2014 to respond to an open invitation from the NW5C for grant applications. Funding was available to support thematic workshops: cross-disciplinary initiatives to enhance teaching and learning at all member institutions. The grant application was submitted in October, and notification of funding approval was received in December. The librarian co-facilitators met more frequently beginning in January 2015 to plan the workshop. In these meetings the planners formulated a workshop model to help address the common challenges faced by librarians developing data management services at the NW5C member colleges. The model proposed bringing together teams from each school for a one-and-a-half-day workshop led by an outside facilitator. Each detail of the model from team composition to the breakdown of the curriculum was designed to maximize the impact of the workshop on the larger endeavor of establishing RDM services at each of the institutions.

### *Team model*

The team model proposed bringing together a team from each school composed of a faculty researcher, one or two student researchers working with that faculty member, a librarian, and an educational technologist or IT support staff. Each team united those involved in addressing the data management challenges in a particular lab. This team model also served to model the type of librarian-researcher collaboration

the planners sought to develop on their campuses. To this end, once a team was identified from each institution, the librarian would work with the team to undertake a preliminary assessment of current RDM practices.

### *Facilitation*

Due to the relative inexperience of the majority of the planning team with RDM services, an outside facilitator who was skilled at leading a diverse audience through an introductory data management curriculum was hired. Selecting an outside facilitator brought two advantages: first, it allowed the librarians to fully engage as members of their teams, working side-by-side through each activity; second, it provided an opportunity for the librarians to receive the RDM training themselves, learning both the content as well as the approach to teaching RDM.

### *Curriculum*

The workshop curriculum was crafted by the planners, in collaboration with the facilitator, with the aim of developing participant awareness and understanding of data management issues. The curriculum for the first day was adapted from the New England Collaborative Data Management Curriculum (Lamar Soutter Library, University of Massachusetts Medical School, n.d.) as well as the DataONE modules (DataONE, 2015) and involved lecture and discussion sections interspersed with group activities to create an active, engaging experience. The curriculum was calibrated based on the participating teams' interests or stated RDM challenges (this question was posed in the recruitment process). The following half-day of the workshop was set aside as work time for teams to apply the lessons learned the previous day while the librarians gathered separately to talk about their work to establish RDM services at their respective institutions and further develop their community of practice.

### **Workshop model in practice**

The workshop model evolved over the course of the planning process. Recruitment challenges affected the makeup of the teams and their disciplinary backgrounds, changes were made to the pre-workshop preparation process to accommodate the needs of participants, and the curriculum itself was modified prior to the delivery of the workshop to provide more participant-centered outcomes. The foundational elements of the model remained: a flexible, segmented curriculum centered around active-learning modules;

a team of researchers from each institution; participation by faculty researchers who generate data in their teaching lab groups; and involvement of undergraduate research assistants.

The planned makeup of the institutional teams, each consisting of a faculty principal investigator, one or two student researchers, a librarian, and an educational technologist or IT support person, was a deliberate attempt to bring together stakeholders with diverse perspectives and complementary skillsets. In fact, none of the teams ended up in the planned configuration of one faculty, one student, and one IT support, although all retained the librarian and faculty participants. All of the librarians reported approaching multiple faculty members at their institutions before each was successful in identifying a faculty member who fit all of the necessary criteria to attend: an interest in RDM, an active research program with undergraduate students involved in generating and working with data, and, crucially, availability on the date of the workshop. Although the original intent had been to recruit faculty from both social science and science disciplines, the planners found it challenging to recruit from the social sciences, since several of the librarians themselves were primarily liaisons to natural science departments. The final participating teams consisted of two chemists, one developmental biologist, and two environmental scientists. The fifth team, with a sociologist, started but did not complete the first day of the workshop.

In addition to the narrowing of disciplinary scope, there was not a consistent configuration of team participants. One institution sent two faculty researchers in environmental science but lacked an undergraduate student researcher, while the chemists, coming from separate institutions, each brought two undergraduate students. None of the teams brought a dedicated technology specialist. Another team brought in place of an IT support person a second faculty participant who, while not a member of the primary research group, was the primary technician for the instruments which were used by the members of the research group in their data collection. This proved to be a helpful team configuration, as the second faculty participant, while not primarily a teaching faculty member, had substantial experience both with the learning needs of undergraduate students, and also with the nature of the data sets being generated by undergraduates within their home department. During workshop activities requiring teams to make plans for future data management workflows within their lab, the second faculty participant enabled the group to delineate very specific procedures based on her expertise and familiarity with the instruments most commonly used by the lab.

Further modifications to the model were made during the pre-workshop preparatory period, when the librarians engaged with their faculty members and completed a modified Data Curation Profile (DCP) interview (Carlson, 2010). Several faculty members expressed some resistance to the time-intensive nature of completing an entire DCP. During the course of the oral interviews, the librarians were able to take stock of each faculty participant's comfort level and experience with data management principles, as well as their immediate needs. In general, faculty had little to no experience explicitly engaging with RDM as a skillset, and their priorities were focused on learning how to increase efficiency in their lab settings. These were unsurprising concerns for faculty at teaching-focused institutions, but did cause the librarians to suggest prioritizing the creation of practical documents such as lab protocols as an outcome of the workshop.

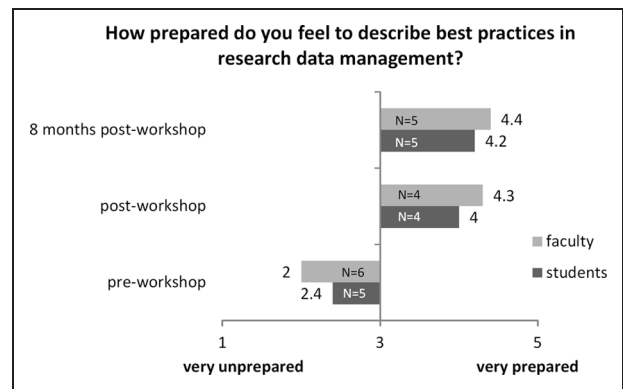
While incorporating these changes, the facilitator did retain as a central component of the curriculum a series of engaging activities to allow the participants to begin to immediately begin putting into practice skills related to the best practices being covered. This emphasis on active learning was a key element of the workshop's success, and was mentioned by the majority of participants in the feedback collected during the post-workshop assessment.

## Outcomes and assessment

### Overall outcomes

Overall, the workshop was successful in meeting several high-level goals of the planners. The team model developed new learning opportunities for students while also increasing faculty and staff communication in a collaborative cross-institutional environment. Within the workshop itself, participants increased their awareness and understanding of data management topics. The NW5C workshop piloted a model for librarian-researcher collaboration across campuses, and the incorporation of the individual DCP interviews allowed librarians, together with researchers, to create a preliminary assessment of RDM practices. The process of applying for the grant and planning the workshop itself allowed librarians to better understand the RDM landscape at their own institutions and at peer institutions. These findings are based on three assessments carried out before, immediately after, and eight months after the workshop.

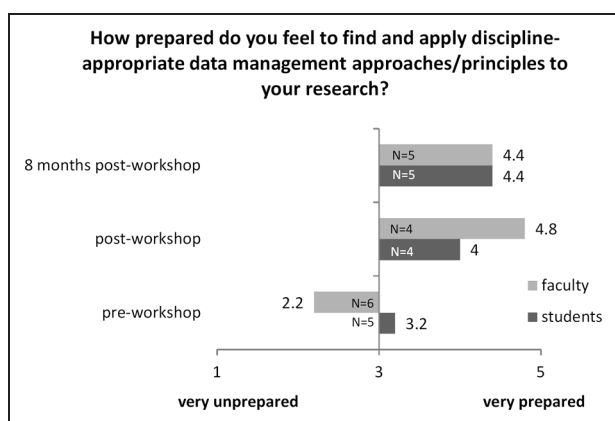
A pre-workshop survey was distributed to all participants to assess base levels of understanding of and familiarity with the material to be covered. A post-workshop survey was distributed immediately



**Figure 1.** Results from pre- and post-workshop surveys of faculty and student workshop participants on describing best practices in research data management.

following the workshop. Separate assessment instruments were used for the faculty/student research teams and the librarian participants. A follow-up survey was distributed to faculty/student research teams ( $n = 12$ , due to unavailability of one team) eight months after the workshop to assess their continued awareness of RDM and curation best practices and their longer-term evaluations of the workshop's effectiveness. Questions from all instruments are available in Appendix A. Total registered attendees of the workshop included 14 faculty/student researchers and six librarians; pre-workshop completion was 79% for the faculty/student research teams ( $n = 11$ ) and 100% for the librarians ( $n = 6$ ) while post-workshop completion was 57% for the faculty/student teams ( $n = 8$ ) and 67% for the librarians ( $n = 4$ ). The lower post-workshop completion rate was primarily due to a faculty/student team from one institution leaving the workshop. The follow-up survey had a response rate comparable to the two earlier assessments, with an overall response rate of 83% from students and faculty ( $n = 10$ ). While the sample sizes are too small to yield statistically significant inferences, the trends in both the numerical data and the qualitative responses support the conclusion that the workshop successfully met its objectives. All respondents who completed the post-workshop survey agreed with the statement that the workshop “was an effective way to learn about Research Data Management”.

The quantitative ratings from students, faculty, and librarians show that in all cases, participants' confidence in their knowledge of RDM practices increased as a result of the workshop (see Figures 1 and 2). The answers to qualitative questions provide more insight into what participants found most useful. Both students and faculty noted specific skills (e.g. file naming conventions, file types, readme files) as well as general concepts (e.g. data curation, metadata) as



**Figure 2.** Results from pre- and post-workshop surveys of faculty and student workshop participants on finding and applying data management principles appropriate for their research.

important takeaways from the workshop. After the workshop, participants expanded their definitions of “data management” to emphasize the importance of planning, an aspect that had been highlighted throughout the workshop presentations, and was absent from pre-workshop answers.

### Faculty outcomes

All responses to the post-workshop survey from faculty members referenced the importance of working in teams and fostering collaborations. Faculty supported extending the workshop’s mission to other groups on their home campuses. One faculty member wrote, “Working in teams was KEY. We need a follow up either at our own institution or another NW5C workshop!” Following the workshop, faculty members defining the “research data life cycle” were more likely to provide more detailed explanations, specifically emphasizing data storage and preservation issues. Several faculty members commented that the preliminary DCP work was helpful in preparing for the workshop and for further RDM planning. Multiple faculty commented on their intent to establish protocols and training procedures for their labs which were previously non-existent. In general, faculty feedback focused on appreciation for newly learned skills which could be immediately implemented in their labs for data backups, efficient data storage, and consistency in recording and formatting data.

Following the workshop, faculty members informally reported to librarians that their research groups had either instituted or begun to develop new file-naming plans and research data workflows to facilitate more consistent documentation and backup. The answers to qualitative questions in the follow-up survey indicate that the researchers have pursued these

preliminary plans. One group has “a protocol for file naming and creating/storing metadata on every set of experimental data we produce”. Another has “used the file naming conventions routinely and [has] converted old, ‘pre-workshop’ files to the convention”. Research groups report that they still face challenges in areas such as storage, backing up data and dealing with older files, not to mention the consistent implementation of file-naming and metadata conventions. In one case, a faculty member noted, “I believe some of these were addressed in the workshop, but the ‘doing’ is much harder” and suggested further communication and support among research groups and librarians to encourage both brainstorming and follow-through.

The quantitative data indicate that after eight months, faculty (and students) felt at least as well prepared to describe best practices and apply discipline-appropriate data management to their research as they had immediately after the workshop, and their answers to the open-ended question of how they would describe RDM were consistent with that self-assessment. One faculty answer in particular summed up key components of data management:

The data we collect in our lab is the foundation of our work and as such is extremely valuable so we must take good care of it. This means we must keep our data useful (describe it thoroughly with detailed and consistent metadata), keep our data organized (use consistent file naming conventions and store it in the correct folders), and keep our data safe (back it up consistently and provide accurate links in our notebooks).

### Student outcomes

Student evaluations immediately following the workshop included the recognition of the value of the skills and concepts covered. A student wrote:

I found it very helpful to be introduced to the concepts behind data management . . . Being aware of the advantages of data management, I can progress in my career and find various applications for [the concepts]. This seems far more advantageous than simply being told by a PI that the lab follows XYZ protocols, which I may or may not continue to apply in the future.

One of the student researchers voiced a desire that future workshops provide more space for interaction and discussion with student workshop participants from other institutions. While not all teams were able to bring students along, there was a sense that the teams who did so found it valuable. Students were able to disseminate the lessons they learned on



returning to their campuses both informally, in talking with other student research assistants in their labs, and, in at least one case, through a formal presentation to the research teams in their department.

### *Librarian outcomes*

In their post-workshop assessments, all librarian participants reported a significant increase in comfort levels with the subject matter and in feeling prepared to do outreach to faculty or colleagues. All librarians mentioned networking or collaboration with NW5C peers when asked about the most important outcome of the workshop. Librarians also raised some issues that were not within the explicit scope of the workshop, such as building a shared data repository, and posed questions about how data management plans might be applied on an individual level to undergraduate research.

The lessons learned for the librarian organizers of the workshop encompassed planning issues on many levels. The librarians had varying levels of training in RDM before the workshop, and some would have appreciated additional “train the trainer” instruction before the workshop itself. Cross-campus collaborations were possible in large part due to technological affordances of shared online documents and Google Hangouts. Having a standing online meeting was very important for planning purposes. Post-workshop use of some of these digital connections could have been improved; having a listserv or other digital meeting space set up before the end of the workshop would have facilitated its ongoing impact.

### **Conclusion**

The NW5C Data Curation Workshop model is a successful, and sustainable, model for developing RDM services and building communities of practice for small liberal arts colleges. While the costs for the first iteration of this workshop model were grant-funded, the planners believe that costs for future iterations of the workshop would be greatly reduced and potentially feasible without grant funding. The primary reduction in costs comes from the fact that the planning librarians gained valuable RDM skills themselves, as well as being exposed to how the facilitator actually taught the material, thus providing valuable professional development training while simultaneously providing outreach services to faculty constituents. Aside from this facilitation cost, the most expensive part of the implementation was reimbursement of travel costs for participants – a situation which was exacerbated by the remote nature of at least one participating institution. Taking further

advantage of digital tools and distance-learning technologies for long-distance collaboration could lower these costs even further.

Faculty outcomes in the area of RDM skills were positive, but the most important long-term outcomes for faculty participants were their development as RDM evangelists for their campuses. In seeking to create a campus culture that values solid RDM practices, librarians need faculty partners who will speak to their importance. Another important outcome was the opportunity to extend the classic liberal arts teacher-scholar model to the domain of RDM skill development. While undergraduate student research assistants are more transitory than graduate research assistants, students attend liberal arts institutions because of the belief in education through close contact with faculty both inside and outside of the classroom. RDM training at the undergraduate level can also help to prepare upcoming graduate students to recognize and promote the importance of data management. This workshop demonstrated the feasibility of faculty, undergraduates, and librarians learning and developing their RDM skills in a collaborative environment.

The development of a community of practice was one of the most powerful outcomes of this experience for the planning librarians. The planning librarians have stayed in regular contact with each other concerning RDM developments on their campuses, and have continued working together on conference presentations as well as articles about their model. These activities have continued despite one of the planners moving to a new position at an institution on the opposite coast. An often unspoken challenge of starting an RDM program at a liberal arts college library is that, in contrast to larger institutions, the librarian tasked with developing these services is typically working alone. With many of the models for these services coming from larger institutions, the new data librarian may have few colleagues to turn to when wondering what will or will not work on their campus, and how to scale these services appropriately. Though just beginning, the development of a community for discussion, planning, and sharing of best practices particular to a smaller institution is invaluable.

A second iteration of the model, with reduced costs, was planned for the summer of 2016, and there is interest in replicating this instructional model for liberal arts institutions outside of the Pacific Northwest. The modules and tools used to develop this workshop are all openly available, and the planners are enthusiastic about sharing their experiences, including the successes and challenges. Whether they choose to proceed independently, or as part of the

growing community of liberal arts colleges developing RDM services, institutions that use a research-group-based approach to teaching RDM will not only improve practices on their own campus, but also help to strengthen relationships and change campus cultures.

### Acknowledgements

The authors gratefully acknowledge those who dedicated their time and effort to this project: Mini-grant Co-PIs: Mark Dahl, Dena Hutto; Other Workshop Planners: Gary Klein, Michael Spalti; Workshop Facilitator: Amanda Whitmire, PhD; and all workshop participants.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### Note

1. The co-principal investigators on this mini-grant were Dena Hutto, Norman F. Carrigg College Librarian, Reed College and Mark Dahl, Director of the Aubrey R. Watzak Library, Lewis & Clark College.

### References

- ANDS (n.d.) *23 (Research Data) Things*. Australian National Data Service. Available at: <http://ands.org.au/partners-and-communities/23-research-data-things> (accessed 18 April 2016).
- Antell K, Foote JB, Turner J, et al. (2014) Dealing with data: Science librarians' participation in data management at Association of Research Libraries institutions. *College & Research Libraries* 75(4): 557–574.
- Ball R and Medeiros N (2012) Teaching integrity in empirical research: A protocol for documenting data management and analysis. *Journal of Economic Education* 43(2): 182–189.
- Boyer Commission on Educating Undergraduates in the Research University (1998) *Reinventing Undergraduate Education: A Blueprint for America's Research Universities*. Stony Brook, NY: SUNY Stony Brook; Carnegie Foundation for the Advancement of Teaching.
- Brandt DS (2007) Librarians as partners in e-research: Purdue University Libraries promote collaboration. *College & Research Libraries News* 68(6): 365–396.
- Bresnahan MM and Johnson AM (2013) Assessing scholarly communication and research data training needs. *Reference Services Review* 41(3): 413–433.
- Calzada Prado J and Marzal MÁ (2013) Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri: International Journal of Libraries & Information Services* 63(2): 123–134.
- Carlson J (2010) *Data Curation Profiles Toolkit*. Available at: <http://docs.lib.purdue.edu/dcptoolkit/> (accessed 21 November 2016).
- Carlson J, Fosmire M, Miller CC, et al. (2011) Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy* 11(2): 629–657.
- Carlson J, Sapp Nelson M, Bracke M, et al. (2015) *The Data Information Literacy Toolkit*. Purdue University. Available at: <http://docs.lib.purdue.edu/dilsymposium/2013/interviewinstruments/1/> (accessed 30 March 2016).
- Choudhury GS (2008) Case study in data curation at Johns Hopkins University. *Library Trends* 57(2): 211–220.
- Cox AM and Pinfield S (2014) Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science* 46(4): 299–316.
- Creamer AT, Morales ME, Kafel D, et al. (2012) A sample of research data curation and management courses. *Journal of eScience Librarianship* 1(2): 4.
- DataONE (2015) *DataONE Education Modules*. Available at: <https://www.dataone.org/education-modules> (accessed 30 March 2016).
- EDINA and Data Library, University of Edinburgh (n.d.) *DIY Research Data MANTRA Training Kit for Librarians*. Available at: <http://datalib.edina.ac.uk/mantra/libtraining.html> (accessed 20 March 2016).
- Goldstein S and Oelker S (2011) Planning for data curation in the small liberal arts college environment. *Sci-Tech News* 65(3). Available at: <http://jdc.jefferson.edu/scitechnews/vol65/iss3/4> (accessed 20 March 2016).
- Guy M (2013) *RDM Training for Librarians*. DCC RDM services case studies, Digital Curation Centre. Available at: <http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/Training-for-librarians-case-study.pdf> (accessed 20 March 2016).
- Heidorn PB (2011) The emerging role of libraries in data curation and e-science. *Journal of Library Administration* 51(7/8): 662–672.
- Hensley MK (2015) A survey of instructional support for undergraduate research programs. *portal: Libraries and the Academy* 15(4): 719–762.
- Hunter A-B, Laursen SL and Seymour E (2007) Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education* 91(1): 36–74.
- Hwse P and Holt A (2011) Joining in the enterprise of response in the wake of the NSF data management planning requirement. *Research Library Issues* 274: 11–17.
- Indiana University Center for Postsecondary Research (n.d.) Definitions. In: *The Carnegie Classification of Institutions of Higher Education*. 2015 edn. Available at: [http://carnegieclassifications.iu.edu/classification\\_descriptions/basic.php](http://carnegieclassifications.iu.edu/classification_descriptions/basic.php).
- Keralis SDC (2012) Data curation education: A snapshot. In: *The Problem of Data*. CLIR Publication No. 154. Washington, DC: Council on Library and Information

- Resources, pp. 32–43. Available at: <https://www.clir.org/pubs/reports/pub154/pub154.pdf> (accessed 6 April 2016).
- Kuh GD, Schneider CG and Association of American Colleges and Universities (2008) *High-impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter*. Report, Washington, DC: Association of American Colleges and Universities.
- Lamar Soutter Library, University of Massachusetts Medical School (ed.) (n.d.) *New England Collaborative Data Management Curriculum*. Available at: <http://library.umassmed.edu/necdmc/index> (accessed 30 March 2016).
- MacMillan D (2010) Sequencing genetics information: Integrating data into information literacy for undergraduate biology students. *Issues in Science & Technology Librarianship*. Available at: <http://www.isltl.org/10-spring/refereed3.html> (accessed 11 May 2015).
- Mooney H, Collie WA, Nicholson S, et al. (2014) Collaborative approaches to undergraduate research training: Information literacy and data management. *Advances in Social Work* 15(2): 368–389.
- Piorun M, Kafel D, Leger-Hornby T, et al. (2012) Teaching research data management: An undergraduate/graduate curriculum. *Journal of eScience Librarianship* 1(1). Available at: <http://escholarship.umassmed.edu/jeslib/vol1/iss1/8> (accessed 11 May 2015).
- Pryor G and Donnelly M (2009) Skilling up to do data: Whose role, whose responsibility, whose career? *International Journal of Digital Curation* 4(2): 158–170.
- Qin J and D'Ignazio J (2010) Lessons learned from a two-year experience in science data literacy education. In: *International Association of Scientific and Technological University Libraries, 31st annual conference*. Available at: <http://docs.lib.purdue.edu/iatul2010/conf/day2/5> (accessed 30 March 2016).
- Reisner BA, Vaughan KTL and Shorish YL (2014) Making data management accessible in the undergraduate chemistry curriculum. *Journal of Chemical Education* 91(11): 1943–1946.
- Research Data Netherlands (2016) *Essentials 4 Data Support*. Available at: <http://datasupport.researchdata.nl/en/> (accessed 6 April 2016).
- Rice R (2014) Research data MANTRA: A labour of love. *Journal of eScience Librarianship* 3(1). Available at: <http://escholarship.umassmed.edu/jeslib/vol3/iss1/4> (accessed 17 April 2016).
- Rowlett RS, Blockus L and Larson S (2012) Characteristics of excellence in undergraduate research (COEUR). In: Hensley N (ed.) *Characteristics of Excellence in Undergraduate Research (COEUR)*. Washington, DC: Council on Undergraduate Research, pp. 2–19. Available at: [http://www.cur.org/assets/1/23/COEUR\\_final.pdf](http://www.cur.org/assets/1/23/COEUR_final.pdf) (accessed 31 March 2016).
- Scaramozzino JM, Ramírez ML and McGaughey KJ (2012) A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries* 73(4): 349–365.
- Shadbolt A, Konstantelos L, Lyon L, et al. (2014) Delivering innovative RDM training: The immersive informatics pilot programme. *International Journal of Digital Curation* 9(1): 313–323.
- Shorish Y (2012) Data curation is for everyone! The case for Master's and Baccalaureate institutional engagement with data curation. *Journal of Web Librarianship* 6(4): 263–273.
- Shorish Y (2015) Data information literacy and undergraduates: A critical competency. *College & Undergraduate Libraries* 22(1): 97–106.
- Soehner C, Steeves C and Ward J (2010) *E-Science and Data Support Services: A Study of ARL Member Institutions. Report*. Washington, DC: Association of Research Libraries. Available at: <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf> (accessed 31 March 2016).
- Stamatoplos A (2009) The role of academic libraries in mentored undergraduate research: A model of engagement in the academic community. *College & Research Libraries* 70(3): 235–249.
- Stamatoplos A, Neville T and Henry D (2016) Analyzing the data management environment in a Master's-level institution. *Journal of Academic Librarianship* 42(2): 154–160.
- Stephenson E and Caravello PS (2007) Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review* 35(4): 525–540.
- Strasser CA and Hampton SE (2012) The fractured lab notebook: Undergraduates and ecological data management training in the United States. *Ecosphere* 3(12): 1–18.
- Tenopir C, Birch B and Allard S (2012) *Academic Libraries and Research Data Services: Current Practices and Plans for the Future. Report*. Association of College & Research Libraries, June. Available at: [http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir\\_Birch\\_Allard.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf) (accessed 31 March 2016).
- Tenopir C, Sandusky RJ, Allard S, et al. (2013) Academic librarians and research data services: Preparation and attitudes. *IFLA Journal* 39(1): 70–78.
- Tenopir C, Sandusky RJ, Allard S, et al. (2014) Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research* 36(2): 84–90.
- Toups M and Hughes M (2013) When data curation isn't: A redefinition for liberal arts universities. *Journal of Library Administration* 53(4): 223–233.
- Verbakel E, Noordegraaf M, de Smaele M, et al. (2013) Data-intelligence training for library staff. In: *Framing the digital curation curriculum, International conference*, Florence, Italy, 6–7 May 2013, CEUR Workshop Proceedings, paper 13. Available at: <http://ceur-ws.org/Vol-1016/paper13.pdf> (accessed 6 April 2016).
- Walton G (2010) Data curation and the academic library. *New Review of Academic Librarianship* 16(1): 1–3.



Watkins W, Hamilton E, Boyko E, et al. (2004) Creating a national peer-to-peer training program for data librarians in Canada. *IASSIST Quarterly* 28(2/3). Available at: [http://www.iassistdata.org/sites/default/files/iqvol282\\_3watkins.pdf](http://www.iassistdata.org/sites/default/files/iqvol282_3watkins.pdf) (accessed 17 April 2016).

Zilinski LD, Nelson MS and Van Epps AS (2014) Developing professional skills in STEM students: Data information literacy. *Issues in Science and Technology Librarianship* 77. Available at: <http://dx.doi.org/10.5062/F42V2D2Z> (accessed 30 March 2016).

### Author biographies

**Ryan Clement** has been the Data Services Librarian at Middlebury College, in Middlebury, VT, since 2015. Prior to this, he was the Data Services Librarian at Reed College in Portland, OR. He holds an MSc in Information from the University of Michigan School of Information, as well as an MA in Philosophy from Temple University. He currently serves on the American Psychological Association Library Advisory Board. He has published work in *Internet Reference Services Quarterly* and the recent book from the Association of College and Research Libraries (ACRL), *Databrarianship: The Academic Data Librarian in Theory and Practice*. He has presented on data services and instruction at liberal arts colleges at the ACRL Annual Conference, the Digital Library Federation Forum, the Research Data Access & Preservation Summit, and the Oregon Library Association Annual Conference. His research interests include data services planning and outreach, data/information literacy, library assessment, and user experience/service design.

**Amy Blau** has been Instructional and Data Services Librarian at Whitman College in Walla Walla, Washington since 2014. She holds an MSc in Information from the University of Michigan School of Information and a PhD in Comparative Literature from the University of Illinois, Urbana-Champaign. Her professional and research interests include data services planning and outreach, information literacy, digital preservation, and the digital humanities.

**Eli Gandour-Rood** is the Science Liaison Librarian at the University of Puget Sound, a private liberal arts college in Tacoma, WA. He received his MLIS from the University of Washington Information School in 2013. Currently, he serves as the Vice-President/President-Elect of the Washington State Chapter of the Association of College and Research Libraries.

**Parvaneh Abbaspour** is the Science & Data Services Librarian at Lewis & Clark College in Portland, Oregon. She received her MLIS from McGill University in 2013.

## Appendix A: Assessment questions

### Pre-workshop survey, librarians

1. Prior to signing up for this workshop, how familiar were you with the data researchers at your institution were working on? (1 = Very Unfamiliar, 5 = Very Familiar)
2. How prepared do you feel to talk to faculty about their research data? (1 = Very Unprepared, 5 = Very Prepared)
3. How prepared do you feel to describe best practices in research data management to faculty or students? (1 = Very Unprepared, 5 = Very Prepared)
4. Please briefly describe or list what actions constitute effective data management.
5. How prepared do you feel to explain to other librarians what research data is? (1 = Very Unprepared, 5 = Very Prepared)
6. How do you define research data when talking to other librarians?
7. How prepared do you feel to support researchers in finding and applying discipline-appropriate data management approaches/principles to their research? (1 = Very Unprepared, 5 = Very Prepared)
8. How would you rate the current importance of good data management at your institution? (1 = Very Unimportant, 5 = Very Important)
9. How prepared do you feel to explain the research data lifecycle at your institution? (1 = Very Unprepared, 5 = Very Prepared)
10. Please briefly describe the research data lifecycle at your institution. What areas does your library currently support?
11. What areas of data management are you prepared to support for your researchers? What other departments on campus might you reach out to to assist in data management?
12. What is your institution? (coded as alpha character)

### Pre-workshop survey, faculty/students/staff

1. How prepared do you feel to describe best practices in research data management? (1 = Very Unprepared, 5 = Very Prepared)
2. Please briefly describe or list what actions constitute effective data management.
3. How prepared do you feel to describe the research data lifecycle? (1 = Very Unprepared, 5 = Very Prepared)
4. Please briefly describe the research data lifecycle.



5. How prepared do you feel to find and apply discipline-appropriate data management approaches/principles to your research project? (1 = Very Unprepared, 5 = Very Prepared)
6. What do you most hope to learn from this workshop?
7. What motivated you to attend this workshop this summer?
8. What is your institution? (alpha-coded)
9. What is your role?

#### *Post-workshop survey, librarians*

1. Prior to signing up for this workshop, how familiar were you with the data researchers at your institution? (1 = Very Unfamiliar, 5 = Very Familiar)
2. How prepared do you feel to explain to other librarians what research data is? (1 = Very Unprepared, 5 = Very Prepared)
3. How do you define research data when talking to other librarians?
4. How prepared do you feel to talk to faculty about their research data? (1 = Very Unprepared, 5 = Very Prepared)
5. How prepared do you feel to support researchers in finding and applying discipline-appropriate data management approaches/principles to their research? (1 = Very Unprepared, 5 = Very Prepared)
6. How prepared do you feel to explain the research data lifecycle at your institution? (1 = Very Unprepared, 5 = Very Prepared)
7. Please briefly describe the research data lifecycle at your institution. What areas does your library currently support?
8. What areas of data management are you prepared to support for your researchers? What other departments on campus might you reach out to to assist in data management?
9. How would you rate the current importance of good data management at your institution? (1 = Very Unimportant, 5 = Very Important)
10. Please briefly describe or list what actions constitute effective data management.
11. What stands out as the most important skills or pieces of information you gained from this workshop?
12. What information or activity was missing that you would have liked to cover?
13. Do you feel that this workshop structure was an effective way to learn about this topic? Why or why not?

14. Would you recommend a workshop on this topic to your colleagues? Why or why not?
15. Do you have other suggestions for how NW5C colleges can (individually or collaboratively) develop and improve their support of research data management?
16. What is your institution? (coded as alpha character)

#### *Post-workshop survey, faculty/students/staff*

1. How prepared do you feel to describe best practices in research data management? (1 = Very Unprepared, 5 = Very Prepared)
2. Please briefly describe or list what actions constitute effective data management.
3. How prepared do you feel to describe the research data lifecycle? (1 = Very Unprepared, 5 = Very Prepared)
4. Please briefly describe the research data lifecycle.
5. How prepared do you feel to find and apply discipline-appropriate data management approaches/principles to your research project? (1 = Very Unprepared, 5 = Very Prepared)
6. What stands out as the most important skills or pieces of information you gained from this workshop?
7. What information or activity was missing that you would have liked to cover?
8. Do you feel that this workshop structure was an effective way to learn about this topic? Why or why not?
9. What, if anything, from this workshop would you most like to see shared with peers and colleagues at your institution?
10. Do you have other suggestions for how NW5C colleges can (individually or collaboratively) develop and improve their support of research data management?
11. Other comments
12. What is your institution? (alpha coded)
13. What is your role?

#### *Follow-up survey, faculty/students/staff*

1. What is your institution? (alpha coded)
2. What is your status?
3. How prepared do you feel to find and apply discipline-appropriate data management approaches/principles to your research project?
4. How prepared do you feel to describe best practices in research data management?
5. How would you describe effective research data management to a new research assistant?

who joined your laboratory/research group this year? (2-3 sentences)

6. Describe how you used something you learned in the workshop in the last 8 months. (1-3 sentences)
7. Describe any issues or challenges you have encountered in the past 8 months with respect to the storage, organization, or sharing of your research data that you wish had been covered in the workshop.
8. How do you think research data management skills/practices/concepts might be incorporated into regular classes or labs in the curriculum?

## Abstracts

### قُطُفَات

#### Research data management in Switzerland: National efforts to guarantee sustainability of research outputs

إدارة البيانات البحثية في سويسرا: الجهود الوطنية؛ لضمان استدامة  
النتائج البحثية:

Pierre-Yves Burgi, Eliane Blumer, Basma  
Makhlouf-Shabou

العدد رقم 43،1 من مجلة الإفلا المتخصصة:

المُلخَص:

يتناول مؤلفو البحث في هذا المقال، عملية إدارة دورة حياة البيانات في أحد المشروعات السويسرية، ويؤكدون حفظ البيانات على المدى الطويل، وبناءً على التحليل المُستفيض للوثائق والمقابلات شبه المُظمة، يهدف المشروع إلى تقديم خدمات وطنية تلبي احتياجات الباحثين في إدارة دورة حياة البيانات، وتشمل:

طرق الحفظ المُختلفة، خطة إدارة البيانات، حلول إدارة البيانات النشطة، الطرق المُختلفة للحفظ طويل المدى، التدريب، ووجود منفذ موحد للمعلومات ومكان يُمكن التواصل معه؛ للحصول على الدعم، بالإضافة إلى عرض مُختلف محاور عمل المشروع، يصف الباحثون أسلوب إدارة استراتيجي ونموذجًا مرئيًا لتطوير نماذج العمل الجديدة وهو مفتاح بناء خدمات ناجحة.

#### Brief Assessment of Researchers' Perceptions towards Research Data in India:

تقييم مُختصر لمفاهيم الباحثين عن البيانات البحثية في الهند:

Manorama Tripathi, Mahesh Chand, Sharad  
Sonkar, Jagjeevan Jeevan

العدد رقم 43،1 من مجلة الإفلا المتخصصة:

المُلخَص:

يُناقش البحث أهمية البيانات الأولية في التنمية والبحث، وتشمل الدراسة مُقابلة مع أكثر من 40 من الباحثين وأعضاء هيئة التدريس؛ لمعرفة مفهومهم عن البيانات الأولية، وأشارت النتائج أن المكتبات يُمكن أن تلعب دورًا بالغ الأهمية في دعم الباحثين في تنظيم وحفظ

البيانات الأولية؛ لاستخدامها في المُستقبل، ويُمكن للمكتبات أن تضع نظامًا في الجامعة يُشجع الباحثين وأعضاء هيئة التدريس على وضع المعلومات الأولية الموجودة لديهم في المستودعات المؤسسية التي أسستها مُعظم المكتبات الجامعية.

#### Exploring the use of stages to help understand the PIM:

Amber Cushing, Odile Dumbleton

العدد رقم 43،1 من مجلة الإفلا المتخصصة:

المُلخَص:

ناقش 15 من طلبة الدكتوراه من جامعة متروبوليتان بأيرلندا المسجلين في مُختلف برامج العلوم الإنسانية والاجتماعية مُمارساتهم وعاداتهم في إدارة البيانات الشخصية من خلال ثلاثة فرق تركيز؛ لاستكشاف مهارات إدارة المعلومات الشخصية في مُختلف مراحل رسالة الدكتوراه، وتُشير النتائج إلى أننا يُمكن أن نُصنف احتياجات طلبة الدكتوراه العاملين على الأطروحات والرسائل في إدارة المعلومات الشخصية إلى ثلاثة مراحل: البداية، المُنتصف، النهاية، ويُمكن لأخصائيي المكتبات الاستفادة من تصنيف الخدمات التي يطورونها لتلبية احتياجات هذه الفئة وفقًا لهذه المراحل الثلاثة.

#### Managing research data at an academic library in a developing country

إدارة البيانات البحثية في المكتبات الأكاديمية بالدول النامية:

Shamin Renwick, Marsha Winter, Michelle Gill

العدد رقم 43،1 من مجلة الإفلا المتخصصة:

المُلخَص:

لقد أصبحت إدارة البيانات البحثية محط نظر العديد من الجامعات، ففي منطقة الكاريبي تتمتع مكتبات سانت أغسطين بجامعة وست إنديز بوعي كبير بالحاجة إلى دعم الباحثين في هذا الصدد، تهدف هذه الدراسة إلى تحديد الممارسات الحالية في إدارة البيانات البحثية ومعرفة الدور المُمكن أن تلعبه مكتبات الجامعة، وتم إجراء دراسة تجريبية من مئة باحث، أشار تحليل 65 من الردود المُتاحة أن الباحثين لا يملكون إلا القليل من المعرفة والخبرة في إدارة البيانات بالرغم من

أن لديهم مجموعات من البيانات، تُعد قلة الوعي في هذه الحالة مُفيدًا للمكتبات لأنه يسمح لها بلعب دور كبير في دعم باحثي الجامعة، يجب على مكتبات الجامعة أن توعي الباحثين بتخطيط البيانات وتفاصيل إدارة البيانات وتقديم المساعدة التقنية في تخزين البيانات.

### **Building professional development opportunities in data services for academic librarians**

إتاحة فرص التنمية المهنية للمكتبيين الأكاديميين العاملين على خدمات البيانات:

**Suzanna Conrad, Yasmeen Shorish, Amanda Whitmire, Patricia Hswe**

العدد رقم 43،1 من مجلة الإفلا المُتخصصة:

المُلخص:

تمثل إدارة البيانات البحثية حقلاً هاماً من حقول التنمية المهنية للمكتبيين المهنيين نظراً لأهميته للمهنة، حيث يتزايد توقع أن يكونوا على مستوى مُتطلبات إدارة البيانات البحثية، والكفاءة اللازمة؛ لدعم الباحثين في ممارسة إدارة البيانات البحثية والتخطيط لها، ويُعيد هذا المقال النظر في كيفية تشجيع جمعية المكتبات الجامعية والبحثية (ACRL) إتاحة فرص التنمية المهنية في مجال إدارة البيانات البحثية، ويتناول الكتاب محاولتين أساسيتين: الأولى: وضع استبيان لتقييم الاحتياجات وتوظيفه، مما سمح بمعرفة أنواع المكتبيين والتعبير عن أكثر الاحتياجات ضرورة، الثانية: التخطيط لورشة عمل وتنفيذها قبل مؤتمر ACRL عام 2015؛ لوضع تصور للمستقبل المهني، وتُختتم المقالة بمناقشة ما أُجري بعد ذلك من تقييمات بعد ورشة العمل وكيف شكلت خارطة طريق لإدارة البيانات البحثية.

### **Open access and open data on natural disasters collections**

التداول الحر وحرية إتاحة بيانات المجموعات التي تتناول الكوارث الطبيعية:

**Marisa Raquel De Giusti, Gonzalo Luján Villarreal, Carlos Javier Nusch, Analía Pinto, Ariel Jorge Lira**

العدد رقم 43،1 من مجلة الإفلا المُتخصصة:

المُلخص:

بعد الفيضان الذي دمر مدينة لابلاتا عام 2013، كشف أحد الاستبيانات التي أجريت عن وضع المياه في الإقليم أن الدراسات والمشاريع المتعلقة بهذه المنطقة مُبعثرة وتفتقر إلى التواجد، وهو ما دفع المؤسسات الأكاديمية، والعلمية، والحكومية إلى العمل معاً؛ لجمع، وتنظيم، ونشر البيانات المُتاحة في رد فعلٍ غير مسبوق للكارثة البيئية التي حلت بالإقليم، ولهذا الغرض خصيصاً تم تأسيس مجموعة في مستودع جامعة لابلاتا الوطنية، وبالتالي، عمل مستودع OMLP لأبحاث البيئة، تتناول هذه الوثيقة مبادرات التعاون التي تهدف إلى الحد

من آثار الكوارث الطبيعية وما تتميز به هذه المبادرات من سمات أساسية، ومصادرها، وما تُحرزه حالياً من تقدم.

### **Building a Research Data Management Service at UC Berkeley:**

بناء خدمات لإدارة البيانات في جامعة كاليفورنيا في بركلي:

**Jamie Wittenberg, Mary Elings**

العدد رقم 43،1 من مجلة الإفلا المُتخصصة:

المُلخص:

لقد تعاونت الوحدة المركزية لتكنولوجيا المعلومات البحثية بمكتبة جامعة كاليفورنيا في بركلي، لعمل برنامج لإدارة البيانات البحثية يرتقي بخبرات المؤسسات ومصادرها؛ من أجل تقديم خدمة موحدة توفر مُتسماً من ورش العمل، والخدمات الاستشارية، والمصادر الإلكترونية، وبفضل هذا التعاون تم إضافة خدمات أخرى إلى خدمات تكنولوجيا المعلومات، مثل: التخزين المأمون والاحتياطي، بالإضافة إلى خدمات في مجال المكتبات، مثل: استكشاف المصادر والاستفادة منها، ودمج كل ما سبق في برنامج واحد، تُناقش دراسة الحالة هذه تأسيس هذا البرنامج والعقبات التي واجهت تطبيقه، ونتائج هذا النموذج من التعاون.

### **Quality evaluation of data management plans at a research university**

تقييم خطط إدارة البيانات في إحدى الجامعات البحثية كفيلاً:

**James E. Van Loon, Katherine G. Akers, Cole Hudson, Alexandra Sarkozy**

العدد رقم 43،1 من مجلة الإفلا المُتخصصة:

المُلخص:

مع ظهور حاجة مؤسسة العلوم الوطنية إلى خطط إدارة البيانات، تزايدت مُساعدة المكتبيين للباحثين في وضع هذه الخطط ونشر البيانات البحثية؛ ولتقييم الجودة العامة لخطط إدارة البيانات في جامعة وين ستيت، قام فريق خدمات البيانات البحثية بتقييم محتوى 119 خطة لإدارة البيانات من مقترحات المنح المُقدمة إلى مؤسسة العلوم الوطنية بين عام 2012 و2014، وتُشير نتائج هذا التقييم أنه بالرغم من فهم مُعظم الباحثين للحاجة إلى مُشاركة البيانات تفشل العديد من خطط إدارة البيانات في إدارة البيانات التي يولدها البرنامج، وكيفية إدارتها خلال البرنامج، كما توضح النتائج مواطن القصور في خطط إدارة البيانات والتي تختلف من وحدة أكاديمية لأخرى، وتوضح الحاجة إلى خدمات توعوية مُتباعدة؛ لتقوية الخطط المُستقبلية الخاصة بعروض منح مؤسسة العلوم الوطنية.

### **Team-based data management instruction at small liberal arts colleges**

محاضرات في إدارة البيانات بالكليات الصغيرة للفنون الليبرالية:



**Ryan Clement, Amy Blau, Parvaneh Abbaspour, Eli Gandour-Rood**

العدد رقم 43، 1 من مجلة الإفلا المُخصصة:

الملخص:

يصف هذا البحث المنهج التعاوني الذي يتبناه المكتبيون في خمسة من كليات الفنون الحرة بالأقاليم في تطوير خدمات إدارة البيانات البحثية، تنتمي الكليات الخمسة إلى اتحاد يُعرف بـ[اسم الاتحاد]، وتعاون مكتبيو

هذه الكليات على مدار عشرة شهور؛ لعمل ورشة عمل لإدارة وتنظيم البيانات تهدف إلى تنمية العلاقات مع الباحثين الذين يتعاملون مع البيانات وتطوير مهاراتهم في تقديم خدمات البيانات البحثية، وبناء نموذجًا للتدريب والتوعية مُستقبلاً، جمعت هذه الورشة فرقًا بحثية تضم الجامعة، والطلاب، والمكتبيين واستُخدم فيها موادًا تعليمية وخبرات عميقة من المناقشات التي سبقتها، سُنِّق هذا المقال محتوى هذه الورشة وخلفياتها، والنموذج نفسه، ونتائجها، والفرص المُتاحة للتطوير مُستقبلاً.

## 摘要

**Research data management in Switzerland:  
National efforts to guarantee sustainability of  
research outputs**

瑞士的研究数据管理：国家为保证研究成果可持续性的努力

皮埃尔-伊夫·布尔吉，艾莲娜·布鲁默，巴斯玛·马克鲁夫-夏布

国际图联杂志，43-1，5-21

摘要：

在本文中，作者报告了一个正在瑞士进行的数据生命周期管理(Data Life-Cycle Management, DLCM)国家项目，主要关注数据的长期保存。该项目以广泛的文档分析和半结构化访谈为基础，旨在为满足最相关的研究人员的DLCM需求提供国家服务，其中包括：建立数据管理计划的准则、主动数据管理解决方案、长期保存存储方案、相关培训，以及单点访问和支持服务。除了介绍项目的工作重心之外，作者还描述了一个用于开发新的商业模式的战略管理和精益启动模板，这是建立可行的服务的关键。

**Brief Assessment of Researchers' Perceptions  
towards Research Data in India**

研究人员对印度研究数据的看法的简短评估

马诺拉玛·特里帕西，马亥什·钱德，夏拉德·宋卡尔，贾吉文·吉文

国际图联杂志，43-1，22-39

摘要：

本文阐述了原始数据对科学和研究发展的重要性。该研究采访了40名研究人员和教师的，以了解他们对原始数据的看法。文章认为，在向研究人员提供扩展性支持服务时，图书馆可以发挥关键作用，从而对原始数据进行组织、归档和保存，以供将来使用。图书馆可在大学层次上建立一个系统，鼓励研究人员和教师将他们的原始研究数据存储在大数大学图书馆都开发有的机构知识库中。

**Exploring the use of stages to help understand the  
PIM**

探索对阶段的使用，从而帮助理解PIM

安柏·库欣，奥戴尔·顿波顿

国际图联杂志，43-1，40-50

摘要：

来自爱尔兰一所大都市大学的15名博士生参加了各种人文和社会科学课程，并在三个重点小组中讨论了他们与个人信息管理相关的做法和习惯，从而在博士课程的不同阶段探索个人信息管理技能。研究结果表明，博士生管理论文/论文信息的个人信息管理需求可以分为三个不同阶段：博士课程的开始阶段、中间阶段和结束阶段。在提供服务以满足这些人的需要时，信息专业人员可能发现通过这三个阶段对这些服务进行分类是有益的。

**Managing research data at an academic library in  
a developing country**

在发展中国家的学术图书馆管理研究数据

夏敏·任威客, 玛莎·温特, 米歇尔·吉尔

国际图联杂志, 43-1, 51-64

摘要:

管理研究数据已成为许多大学的一个问题。在加勒比地区, 西印度群岛大学的圣奥古斯丁校区图书馆非常清楚在这方面需要支持研究人员。本研究的目的是确定在校园内管理研究数据的当前做法, 并找出校园图书馆可能发挥的作用。对校园里的100名研究人员进行了试点研究。65个有效答复的分析表明, 虽然研究人员拥有数据集, 他们在管理他们的研究数据方面具有很少的知识或经验。这种低水平的意识具有启发性, 并证明了校园图书馆在校园支持研究人员方面所起的作用。校园图书馆需要使研究人员了解数据规划和管理研究数据所需的内容, 并为实际数据存储提供技术援助。

#### **Building professional development opportunities in data services for academic librarians**

为学术图书馆员建立数据服务方面的专业发展机会

苏珊娜·康拉德, 雅思敏·薛丽什, 阿曼达·怀特米尔, 帕特里西亚·赫思维

国际图联杂志, 43-1, 65-80

摘要:

研究数据管理(Research data management, RDM)因其对专业与日俱增的重要性而成为学术图书馆员的一个重要专业发展领域。研究人员越来越期望遵守RDM要求, 以及图书馆员在RDM中支持研究人员所需的能力实践和计划。本文重述了大学和研究图书馆协会 (Association of College and Research Libraries, ACRL) 如何在RDM中培养专业发展机会。作者描述了两个关键行动: 1)开发和部署需求评估调查, 从而深入了解表达最需要的图书馆员的类型; 以及2)为2015年ACRL举行会前研讨会的规划和实施, 旨在为未来的专业发展提供原型。文章最后讨论了在研讨会之后进行的额外评估, 以及会前如何为RDM

提出“路演”提供基础, 这类似于ACRL赞助学术交流。

#### **Open access and open data on natural disasters collections**

开放获取和关于自然灾害收集的开放数据

玛丽莎·拉凯尔·德·吉尤斯蒂, 冈扎罗·路汗·维拉利尔, 卡洛斯·哈维尔·努什, 安娜莉亚·品脱, 阿里尔·荷热·里拉

国际图联杂志, 43-1, 81-88

摘要:

2013年洪水摧毁了拉普拉塔城, 该地区的水情况调查表明, 与该地区相关的研究和项目分散, 缺乏可见性。这促使学术界、科学界和政府机构合作, 以前所未有的方式编制、组织和传播现有数据, 从而应对该地区的环境灾难。为此, 在拉丁美洲国家大学机构存储库SEDICI中创建了一个专门的馆藏, 随后建立了Observatorio Medioambiental La Plata (OMLP), 这是一个关于环境研究数据存储库。本文描述了旨在预测和减少自然灾害的合作举措, 及其其主要特点, 以及所利用的资源和目前进展的影响。此外, 文章还描述了由OMLP成立的项目, 以及基于开放访问策略的技术和基础设施部署。

#### **Building a Research Data Management Service at UC Berkeley**

在加州大学伯克利分校建立研究数据管理服务

杰米·惠滕伯格, 玛丽·艾琳斯

国际图联杂志, 43-1, 89-97

摘要:

加州大学伯克利分校图书馆和中央研究信息技术部门合作开发了一个研究数据管理计划, 利用每个组织的专业知识和资源创建一个统一的服务。该服务提供各种研讨会、咨询和在线资源。这一协作将通常完全嵌入IT中的服务领域(如备份和安全存储)以及图书馆域中的服务(如资源发现和指

导)都集成到一个单一的研究数据管理程序中。本案例研究讨论了计划的建立、实施计划的障碍以及合作模式的成果。

### **Quality evaluation of data management plans at a research university**

研究型大学的数据管理计划的质量评估

詹姆斯·E. 范·龙, 卡特琳·G. 阿克斯, 科尔·胡森, 亚历桑德拉·萨科齐

国际图联杂志, 43-1, 98-104

摘要:

随着国家科学基金会(National Science Foundation, NSF)对数据管理计划(DMP)提出需求, 学术图书馆员越来越多地协助研究人员开发DMP和传播研究数据。为了确定韦恩州立大学的DMP的整体质量, 图书馆系统的研究数据服务(RDS)团队评估了2012年至2014年期间提交的NSF拨款建议中的119个DMP的内容。我们的内容分析结果表明, 虽然大多数研究者了解需要共享数据, 但许多DMP无法充分描述项目生成的数据、如何在项目期间管理数据, 或如何在项目完成后保留和共享数据。我们的研究结果还表明, DMP缺陷在不同学术单位之间有所不同, 因此需要差异化的扩

展服务, 以提高未来NSF拨款建议中DMP的实力。

### **Team-based data management instruction at small liberal arts colleges**

基于团队的小型文科学院的数据管理教学

莱恩·克莱门, 艾米·布劳, 帕尔凡内·阿巴斯普, 艾力·甘杜尔-鲁德

国际图联杂志, 43-1, 105-118

摘要:

本文介绍了五家小型地方文理学院图书馆员在其校园开发/加强研究数据管理服务的合作方法。五个学院共同属于一个称为[协会名称]的财团。10个多月以来, 来自五所学校的图书馆员合作计划一个数据管理和策划研讨会, 目的是与研究数据的研究人员建立关系, 开发他们自己的RDM技能和服务, 以及为以后围绕机构RDM服务的培训和外展活动建立一个模型。该研讨会汇集了包括教师、学生和图书馆员在内的研究团队, 并包括主动学习模块以及深入的研讨会前讨论。本文将讨论本次研讨会的背景、模型本身以及未来发展的结果和可能性。

## **Sommaires**

### **Research data management in Switzerland: National efforts to guarantee sustainability of research outputs**

#### **Gestion des données de recherche en Suisse: efforts nationaux pour garantir la durabilité des résultats de la recherche.**

Pierre-Yves Burgi, Eliane Blumer, Basma Makhoulf-Shabou

IFLA Journal, 43 -1, 5-21

Résumé:

Dans cet article, les auteurs dressent un rapport sur un projet Suisse traitant de la problématique récurrente du Data Life-Cycle Management (DLCM / Gestion

du Cycle de Vie des Données), avec une attention particulière portée à la préservation à long terme. Basée sur une analyse approfondie des documents ainsi que sur des interviews semi structurés, le projet a pour objectif de fournir des services nationaux pour répondre au plus près aux besoins des chercheurs en DLCM, à savoir: des lignes directrices pour établir un plan de gestion des données, des solutions dynamiques de gestion des données, des options de stockage pour la préservation à long terme, de la formation ainsi qu'un guichet unique d'accès et de contact pour l'assistance. Les auteurs ne présentent pas seulement les différents axes de travail du projet, ils décrivent également la gestion stratégique adossée à un modèle de démarrage pour permettre le développement de nouveaux modèles de gestion, ce qui constitue la clé pour la mise en place de services fiables.

### **Brief Assessment of Researchers' Perceptions towards Research Data in India**

#### **Evaluation succincte de la Perception des Données de Recherche qu'ont les chercheurs en Inde**

Manorama Tripathi, Mahesh Chand, Sharad Sonkar, Jagjeevan Jeevan

IFLA Journal, 43 -1, 22-39

#### **Résumé:**

Ce document décrit longuement l'importance des données brutes pour le développement de la science et de la recherche. L'étude comprend des interviews de 40 chercheurs et universitaires pour comprendre leur perception des données brutes. Le rapport suggère que les bibliothèques pourraient jouer un rôle crucial en intensifiant leur soutien aux chercheurs par l'organisation, l'archivage et la préservation des données brutes en vue d'une utilisation future. Les bibliothèques pourraient élaborer un système au niveau universitaire au sein duquel les chercheurs et les universitaires seraient encouragés à déposer leurs données brutes dans les entrepôts de données institutionnels qui ont déjà été mises en place dans la plupart des bibliothèques universitaires.

### **Exploring the use of stages to help understand the PIM**

#### **Etude sur la prise en compte de différents stades pour aider à la compréhension du PIM (Personal Information Management / Gestion des Renseignements Personnels)**

Amber Cushing, Odile Dumbleton

IFLA Journal, 43 -1, 40-50

#### **Résumé:**

Quinze étudiants en doctorat inscrits dans différents cursus en sciences humaines et sociales au sein d'une même Université Métropolitaine en Irlande ont débattu de leurs pratiques de leurs habitudes vis-à-vis de la PIM (gestion des renseignements personnels). Ils ont créé trois groupes de discussion pour analyser les compétences en gestion des renseignements personnels mis en œuvre aux différents stades de leurs cursus. Leurs conclusions suggèrent que les besoins en gestion des renseignements personnels des étudiants en doctorat qui gèrent des informations pour leurs dissertations et thèses peuvent être classés en trois stades distincts: le début, le milieu et la fin du cursus doctoral. En développant des services pour répondre aux besoins de cette population, les professionnels de l'information pourraient trouver utile de s'en inspirer pour classer les dits services selon les trois stades définies.

### **Managing research data at an academic library in a developing country**

#### **La gestion des données de recherche dans une bibliothèque universitaire d'un pays en voie de développement**

Shamin Renwick, Marsha Winter, Michelle Gill

IFLA Journal, 43 -1, 51-64

#### **Résumé:**

La RDM (Research Managing Data / Gestion des Données de Recherche) est devenu un enjeu pour de nombreuses universités. Dans les Caraïbes, les bibliothèques « St. Augustine Campus Libraries » du « University of the West Indies » (Université des Antilles) sont très conscientes de la nécessité de soutenir les chercheurs dans ce domaine. Les objets de cette étude furent d'identifier les pratiques actuelles dans la gestion des données de recherche au niveau du campus et d'en déduire une action possible de la part des bibliothèques universitaires. Une étude pilote a été conduite auprès de 100 chercheurs du campus. L'analyse des 65 réponses validées a révélé que, bien que les chercheurs disposaient de jeux de données, ils n'avaient que peu de connaissances ou d'expérience au niveau de la gestion de leurs données de recherche. Ce faible niveau de prise de conscience est instructif et valide le besoin d'un soutien aux chercheurs du campus, aide qui peut être proposée par les bibliothèques universitaires. Ces dernières doivent sensibiliser les chercheurs sur les implications de la planification et de la gestion des données de recherche, et leur proposer une assistance technique pour un stockage effectif des données.

### **Building professional development opportunities in data services for academic librarians**

#### **Elaborer des opportunités de perfectionnement professionnel en matière de service des données pour les bibliothécaires universitaires**

Suzanna Conrad, Yasmeen Shorish, Amanda Whitmire, Patricia Hswe

IFLA Journal, 43 -1, 65-80

#### **Résumé:**

La RDM (Research Data Management / Gestion des Données de Recherche) représente un élément de perfectionnement professionnel significatif pour les bibliothécaires universitaires – elle est significative car elle devient une composante de plus en plus importante de la profession, dans la mesure où les chercheurs doivent se conformer de plus en plus aux exigences de



la RDM, et également parce que cela représente de nouvelles compétences à acquérir pour les bibliothécaires pour leur permettre de soutenir les chercheurs dans la planification et la mise en pratique de la RDM. Cet article retrace la façon dont l'ACRL (Association of College and Research Libraries / Association des Bibliothèques Universitaires et de Recherche) favorise les possibilités de perfectionnement professionnel autour de la RDM. Les auteurs décrivent deux axes clés: 1) le développement et l'extension d'une enquête d'évaluation des besoins, qui a permis de savoir quels étaient les bibliothécaires qui exprimaient le plus de besoins; et 2) la planification et la mise en place d'un atelier pré-colloque pour l'ACRL 2015, pour établir le prototype d'une proposition de formation professionnelle future. Dans sa conclusion, l'article commente une évaluation complémentaire réalisée à la suite de l'atelier ainsi que la proposition issue de ce pré-colloque concernant une tournée de présentation de la RDM similaire à l'action soutenue par l'ACRL au niveau de la communication universitaire.

#### **Open access and open data on natural disasters collections**

##### **Collecte de données relatives aux catastrophes naturelles: données et accès libres**

Marisa Raquel De Giusti, Gonzalo Luján Villarreal, Carlos Javier Nusch, Analía Pinto, Ariel Jorge Lira

IFLA Journal, 43 -1, 81-88

##### **Résumé:**

Suite à une inondation qui a dévasté la ville de La Plata en 2013, une étude sur la situation hydraulique de la région révéla que les recherches et projets liés à cette zone avaient été dispersés et manquaient de visibilité. Cela incita les universitaires, les scientifiques et les instances gouvernementales à travailler ensemble pour compiler, organiser et diffuser toutes les données disponibles dans un effort sans précédent, afin d'être prêts à réagir face à toute catastrophe écologique dans la région. A cet effet, une collecte de données spécifique a été mise en place au sein de la SEDICI, le dépôt institutionnel de la Universidad Nacional de La Plata. Dans un second temps l'Observatorio Medioambiental La Plata (OMLP) fut créé; un dépôt pour les données sur la recherche environnementale. Ce document décrit ces initiatives collaboratives ayant pour objectif d'anticiper et de réduire l'impact des catastrophes naturelles, leurs principales caractéristiques, leurs ressources et les progrès en cours. On y décrit également les projets menés par l'OMLP, avec leur déploiement

technique et celui des infrastructures conformément à leur Politique de Libre Accès.

#### **Building a Research Data Management Service at UC Berkeley**

##### **Mise en place d'un Service de Gestion des Données de Recherche à UC Berkeley**

Jamie Wittenberg, Mary Elings

IFLA Journal, 43 -1, 89-97

##### **Résumé:**

La bibliothèque de l'UC Berkeley (Université de Berkeley) et la Central Research Information Technologies Unit (unité central de la Recherche et des Technologies de l'Information) ont collaboré pour mettre en place un programme de gestion des données de recherche qui puisse avoir un effet de levier sur toutes les expertises et les ressources de chaque organisation dans le but de créer un service unifié. Le service propose une gamme d'ateliers, de consultations et de ressources en ligne. Du fait de cette collaboration, des services qui sont souvent complètement incorporés dans les Technologies de l'Information, comme la sauvegarde et le stockage, ainsi que des services du domaine des bibliothèques, telles que la divulgation et l'enseignement des ressources, sont intégrées dans un seul programme de gestion des données de recherche. Cette étude de cas traite de la mise en place du programme, des obstacles rencontrés pour sa mise en œuvre, et des résultats du modèle collaboratif.

#### **Quality evaluation of data management plans at a research university**

##### **Evaluation qualitative des plans de gestion de données dans une université de recherche**

James E. Van Loon, Katherine G. Akers, Cole Hudson, Alexandra Sarkozy

IFLA Journal, 43 -1, 98-104

##### **Résumé:**

Avec l'émergence des exigences définies par la National Science Foundation (NSF, la Fondation Nationale pour la Science) en ce qui concerne les Data Management Plans (DMP / Plans de Gestion des Données), les bibliothécaires universitaires ont de plus en plus aidé les chercheurs dans le développement des DMP et dans la diffusion des données de recherche. Afin de déterminer la qualité générale des DMP à la Wayne State

University (Université de Wayne State), l'équipe de la Library System's Research Data Services (RDS / Service de Recherche de Données) a évaluée le contenu de 119 DMP parmi les propositions de subvention de la NFS déposées entre 2012 et 2014. Les résultats de notre analyse de contenu démontrent que, bien que la plupart des chercheurs comprennent la nécessité du partage des données, de nombreux DMP ne décrivent pas de manière adéquate les données générées par le projet, ni la manière dont les données seront gérées pendant le projet, ni comment les données seront préservées et partagées à l'issue du projet. Nos résultats montrent également que les lacunes dans les DMP varient selon les secteurs universitaires, ce qui tend à montrer la nécessité de mettre en place des services différenciés pour permettre la consolidation des DMP des futures propositions de subvention de la NFS.

### **Team-based data management instruction at small liberal arts colleges**

#### **Formation à la gestion de données en équipe dans de petites facultés des arts**

Ryan Clement, Amy Blau, Parvaneh Abbaspour, Eli Gandour-Rood

IFLA Journal, 43 -1, 105-118

#### **Résumé:**

Ce document décrit une approche collaborative menée par les bibliothécaires de cinq petites facultés régionales des arts afin de développer et d'améliorer les services de gestion des données de recherche dans leurs universités. Ces cinq facultés font toutes parties d'un consortium connu sous le nom de (CONSORTIUM NAME). Pendant 10 mois, les bibliothécaires des cinq facultés ont collaboré pour planifier un atelier de gestion et de conservation des données avec pour objectif le développement des relations avec les chercheurs qui travaillent avec des données, le développement de leurs propres compétences et services en Research Data Management (RDM / Gestion des Données de Recherche), et la construction d'un modèle pour la formation et la diffusion des services institutionnels de RDM dans l'avenir. Cet atelier a réuni des équipes de chercheurs, incluant des membres du corps professoral, des étudiants et des bibliothécaires, et comprenait des modules de formation actifs ainsi que des discussions approfondies. Cet article analysera le contexte et les circonstances de la mise en œuvre de cet atelier, le modèle de l'atelier en lui-même, ainsi que ses résultats et les opportunités de développement futurs.

## **Zusammenfassungen**

### **Research data management in Switzerland: National efforts to guarantee sustainability of research outputs**

#### **Forschungsdatenmanagement in der Schweiz: Nationale Bestrebungen zur Gewährleistung der Nachhaltigkeit von Forschungsergebnissen**

Pierre-Yves Burgi, Eliane Blumer, Basma Makhoul-Shabou

IFLA-Journal, 43 -1, 5-21

#### **Zusammenfassung:**

In diesem Artikel berichten die Autoren über ein zurzeit in der Schweiz laufendes Datenmanagement-Projekt über den gesamten Lebenszyklus (Data Life-Cycle Management, DLCM) von Forschungsdaten mit Schwerpunkt bei der langfristigen Aufbewahrung. Anhand einer umfassenden Dokumentenanalyse sowie semi-strukturierten Interviews sollen mit diesem Projekt den schweizerischen Diensten Informationen über die relevantesten DLCM-Bedürfnisse der Wissenschaftler an die Hand gegeben werden, unter anderem

Richtlinien für die Erstellung eines Datenmanagementplans, aktive Datenmanagementlösungen, langfristige Speicheroptionen, Schulungen und die Einrichtung eines einzigen Zugangs und Ansprechpartners für Unterstützung. Neben der Präsentation der einzelnen Arbeitsrichtungen des Projekts beschreiben die Autoren eine Vorlage für das strategische Management und den ersten Ansatz für die Entwicklung neuer Geschäftsmodelle, die der Schlüssel für die Einrichtung praktikabler Dienste sind.

### **Brief Assessment of Researchers' Perceptions towards Research Data in India**

#### **Kurze Einschätzung der Wahrnehmung von Wissenschaftlern zu Forschungsdaten in Indien**

Manorama Tripathi, Mahesh Chand, Sharad Sonkar, Jagjeevan Jeevan

IFLA-Journal, 43 -1, 22-39

#### **Zusammenfassung:**

Diese Arbeit befasst sich mit der Bedeutung von Ausgangsdaten für die Entwicklung von Wissenschaft und

Forschung. In der Studie ist ein Interview mit 40 Wissenschaftlern und Fakultätsangehörigen enthalten, das deren Auffassung zu Ausgangsdaten darstellt. Dem Artikel zufolge können Bibliotheken eine Schlüsselrolle spielen, indem sie die Wissenschaftler bei der Organisation, Archivierung und sicheren Aufbewahrung von Ausgangsdaten für die künftige Nutzung stärker unterstützen. Danach können Bibliotheken auf akademischer Ebene ein System entwickeln, in dem Wissenschaftlern und Fakultätsangehörigen die Möglichkeit geboten wird, ihre Ausgangs-Forschungsdaten in die institutionellen Repositorien einzustellen, die die meisten Universitätsbibliotheken inzwischen entwickelt haben.

### **Exploring the use of stages to help understand the PIM**

#### **Erforschung der Verwendung von Phasen zum Verständnis von PIM**

Amber Cushing, Odile Dumbleton

IFLA-Journal, 43 -1, 40-50

#### **Zusammenfassung:**

Fünfzehn Doktoranden unterschiedlicher Programme im Bereich der Geistes- und Sozialwissenschaften einer Universität in Irland haben ihre Vorgehensweisen und Gewohnheiten im Zusammenhang mit dem Management personenbezogener Angaben (PIM) in drei Fokusgruppen diskutiert, um die PIM-Fähigkeiten in verschiedenen Phasen ihres Doktorandenprogramms zu erkunden. Demzufolge lassen sich die PIM-Bedürfnisse von Doktoranden für das Management von Daten der Dissertations-/Abschlussarbeit in drei unterschiedliche Phasen einteilen, und zwar in die Anfangs- die mittlere und die Abschlussphase des Doktorandenprogramms. Bei der Entwicklung von Diensten zur Unterstützung dieser Gruppen ist die Einteilung solcher Dienste in die oben genannten drei Phasen durch Informationsmanager sinnvoll.

### **Managing research data at an academic library in a developing country**

#### **Management von Forschungsdaten in Universitätsbibliotheken von Entwicklungsländern**

Shamin Renwick, Marsha Winter, Michelle Gill

IFLA-Journal, 43 -1, 51-64

#### **Zusammenfassung:**

Das Management von Forschungsdaten ist für viele Universitäten ein Thema. Im karibischen Raum sind sich die St. Augustine Campus Libraries der University of the West Indies der Wichtigkeit, Wissenschaftler in diesem Bereich zu unterstützen, besonders bewusst. Mithilfe dieser Studie sollten aktuelle Vorgehensweisen beim Management von Forschungsdaten auf dem Campus aufgezeigt und die etwaige Rolle der Campus-Bibliotheken ermittelt werden. Dazu wurde eine Pilotstudie mit 100 Wissenschaftlern auf dem Campus durchgeführt. Die Analyse der 65 validen Antworten ergab, dass die Forscher, während sie ihre Daten selbst aufbewahrten, über sehr wenig Wissen oder Erfahrung beim Management dieser Forschungsdaten verfügten. Dieses geringe Bewusstsein in dieser Frage ist sehr illustrativ und zeigt, welche wichtige Rolle die Campus-Bibliotheken bei der Unterstützung der Forscher des Campus spielen können. Es ist nun Aufgabe der Bibliotheken, das Bewusstsein der Forscher dahingehend zu schärfen, was Datenplanung und das Management von Forschungsdaten beinhalten und welche technische Unterstützung sie bei der Datenspeicherung spielen können.

### **Building professional development opportunities in data services for academic librarians**

#### **Schaffung professioneller Entwicklungsmöglichkeiten von Daten-Diensten für Bibliothekare an Uni-Bibliotheken**

Suzanna Conrad, Yasmeen Shorish, Amanda Whitmire, Patricia Hswe

IFLA-Journal, 43 -1, 65-80

#### **Zusammenfassung:**

Das Forschungsdatenmanagement (Research Data Management, RDM) ist für Bibliothekare an Universitätsbibliotheken aufgrund seiner zunehmenden Bedeutung für diesen Berufszweig ein wichtiges berufliches Fortbildungsthema, da von Wissenschaftlern verstärkt erwartet wird, dass sie die RDM-Anforderungen erfüllen, und Bibliothekare deshalb zunehmend die Kompetenzen besitzen müssen, um Wissenschaftler bei RDM-Praktiken und -Plänen zu unterstützen. Dieser Artikel beschreibt die Maßnahmen der Association of College and Research Libraries (ACRL) zur Förderung der Fortbildungsmöglichkeiten im Bereich RDM. Die Autoren beschreiben darin die zwei größten Herausforderungen: 1) die Entwicklung und Einrichtung einer Umfrage nach dem aktuellen Bedarf, die Einblicke in die beiden Bibliothekarszweige verschaffte, bei denen der größte Bedarf besteht; und 2) Planung

und Implementierung eines Workshops im Vorfeld der ACRL-Konferenz 2015, der einen ersten Ansatz für künftige Bildungsangebote liefern sollte. Der Artikel schließt mit der Diskussion eines zusätzlichen Assessments, der im Anschluss an den Workshop durchgeführt wurde, und wie die Veranstaltung im Vorfeld der Konferenz die Grundlagen für einen Vorschlag einer „RDM-Roadshow“ lieferte, ähnlich dem, was die ACRL im Bereich der wissenschaftlichen Kommunikation fördert.

### **Open access and open data on natural disasters collections**

#### **Open Acces und Open Data für Datensammlungen von Naturkatastrophen**

Marisa Raquel De Giusti, Gonzalo Luján Villarreal, Carlos Javier Nusch, Analía Pinto, Ariel Jorge Lira

IFLA-Journal, 43 -1, 81-88

Zusammenfassung:

Nach einer Überschwemmung im Jahr 2013, die weite Teile von La Plata verwüstete, kam eine Studie zur Wassersituation der Region zu dem Ergebnis, dass relevante Studien und Projekte für dieses Gebiet nur sehr vereinzelt vorlagen und schlecht auffindbar waren. Dies war der Anstoß zu einer Zusammenarbeit akademischer, wissenschaftlicher und behördlicher Einrichtungen mit dem Ziel, verfügbare Daten in einer beispiellosen Reaktion auf eine Naturkatastrophe in der Region zu sammeln, zu organisieren und zu verbreiten. Dazu wurde eine spezielle Kollektion in SEDICI zusammengestellt, dem institutionellen Repositorium der Universidad Nacional de La Plata, anschließend wurde das Observatorio Medioambiental La Plata (OMLP), ein Repositorium für Umweltforschungsdaten eingerichtet. Diese Arbeit beschreibt die Initiativen dieser Zusammenarbeit, mit denen das Ausmaß von Naturkatastrophen, ihre wesentlichen Merkmale, Ressourcen und der aktuelle Fortschritt vorhergesehen und reduziert werden sollen. Außerdem werden vom OMLP betriebene Projekte dargestellt, zusammen mit einer Beschreibung ihrer technischen und infrastrukturellen Einrichtung im Sinne der Open-Access-Richtlinien.

### **Building a Research Data Management Service at UC Berkeley**

#### **Aufbau eines Managementdienstes für Forschungsdaten an der Universität von Kalifornien, Berkeley**

Jamie Wittenberg, Mary Elings

IFLA-Journal, 43 -1, 89-97

Zusammenfassung:

Die Bibliothek und die Abteilung der zentralen Forschungsdatentechnologien an der Universität von Kalifornien, Berkeley, haben bei der Entwicklung eines Forschungsdaten-Managementprogramms zusammengearbeitet, bei dem beide Organisationen ihre Expertise und Ressourcen eingebracht haben, um einen gemeinschaftlichen Dienst einzurichten. Der Dienst bietet eine Vielzahl an Workshops, Beratungsangeboten und eine Online-Plattform. Dank dieser Zusammenarbeit konnten Servicebereiche, die häufig vollständig in IT integriert sind, wie Backup und sichere Speicherung, sowie Dienste im Bibliothekenbereich, wie die Suche nach Hilfsmitteln und Anleitungen, in einem einzigen Forschungsdaten-Managementprogramm zusammengefasst werden. Diese Fallstudie bespricht die Einrichtung des Programms, die Hürden, die bei der Implementierung zu bewältigen waren und die Ergebnisse des Kooperationsmodells.

### **Quality evaluation of data management plans at a research university**

#### **Qualitätsbeurteilung der Datenmanagementpläne an einer Forschungsuniversität**

James E. Van Loon, Katherine G. Akers, Cole Hudson, Alexandra Sarkozy

IFLA-Journal, 43 -1, 98-104

Zusammenfassung:

Seit dem Vorliegen der Forderung der National Science Foundation (NSF) nach Datenmanagementplänen (DMP) haben Bibliothekare an Uni-Bibliotheken Wissenschaftler zunehmend bei der Entwicklung von DMPs und der Veröffentlichung von Forschungsdaten unterstützt. Zur Feststellung der allgemeinen Qualität der DMPs an der Wayne State University hat das Research Data Services (RDS)-Team der Bibliothek den Inhalt von 119 DMP von Förderanträgen, die zwischen 2012 und 2014 bei der NSF eingereicht wurden, geprüft. Die Ergebnisse unserer Inhaltsanalyse weisen darauf hin, dass, auch wenn sich die meisten Wissenschaftler der Notwendigkeit der Veröffentlichung ihrer Daten bewusst sind, viele DMPs die im Rahmen des Projekts generierten Daten, die Verwaltung der Daten während des Projekts oder die Speicherung und Verbreitung der Daten nach Abschluss des Projekts nicht angemessen beschreiben. Außerdem zeigen unsere Ergebnisse, dass die Mängel der DMPs je nach



Факultät unterschiedlich sind, was nahelegt, dass jeweils unterschiedliche Unterstützungsdienste notwendig sind, um die Qualität von DMPs für künftige bei der NSF eingereichte Förderanträge zu verbessern.

### **Team-based data management instruction at small liberal arts colleges**

#### **Team-basierte Datenmanagement-Fortbildung an kleinen geisteswissenschaftlichen Hochschulen**

Ryan Clement, Amy Blau, Parvaneh Abbaspour, Eli Gandour-Rood

IFLA-Journal, 43 -1, 105-118

#### **Зusammenfassung:**

Diese Arbeit beschreibt einen gemeinsamen Ansatz von Bibliothekaren an fünf kleinen, regionalen geisteswissenschaftlichen Hochschulen in Bezug auf die

Entwicklung/Verbesserung der Forschungsdaten-Managementdienste auf ihrem jeweiligen Campus. Die fünf Hochschulen gehören gemeinsam zu einem Konsortium mit dem Namen [CONSORTIUM NAME]. 10 Monate haben die Bibliothekare an der Planung eines Workshops über Datenmanagement und Kuration zusammengearbeitet. Ziel war die Entwicklung von Beziehungen zu den Wissenschaftlern, die mit Daten arbeiten, die Entwicklung ihrer eigenen RDM-Kenntnisse und -dienste sowie der Aufbau eines Modells künftiger Fortbildungsprogramme und Unterstützung im Bereich institutioneller RDM-Dienste. Der Workshop, an dem Forschungsteams einschließlich Fakultäten, Studenten und Bibliothekaren teilnahmen, bot aktive Lernmodule und sowie eingehende Diskussionen im Vorfeld des Workshops. Die Autoren beschreiben in diesem Artikel den Kontext und die Hintergründe dieses Workshops, das Modell selbst sowie die Ergebnisse und die Möglichkeiten für künftige Entwicklungen.

## **Рефераты статей**

### **Research data management in Switzerland: National efforts to guarantee sustainability of research outputs**

#### **Управление научными данными в Швейцарии: Национальные усилия по обеспечению перспективного использования результатов исследований**

Пьер-Ив Бурги, Элейн Блюмер, Басма Макхлуф-Шабу

IFLA Journal, 43 -1, 5-21

#### **Аннотация:**

В настоящей статье авторы рассказывают о национальном проекте Управления жизненным циклом данных (DLCM), который сейчас осуществляется в Швейцарии и главной целью которого является долгосрочное хранение данных. В основе проекта лежит обширный анализ документальной базы, а также полу структурированные интервью, а его цель заключается в предоставлении на национальном уровне услуг, соответствующих наиболее актуальным запросам исследователей в сфере DLCM, которые включают в себя: руководства по разработке плана управления данными, действенные решения в сфере управления данными, способы сохранения и долгосрочного хранения данных, обучение, а также единую точку доступа

и контакта для получения содействия. Наряду с описанием различных направлений реализации проекта авторы также приводят модель стратегического управления и экономического старта для развития новых бизнес-моделей, что является ключевым фактором для обеспечения конкурентоспособности предоставляемых услуг.

### **Brief Assessment of Researchers' Perceptions towards Research Data in India**

#### **Сжатая оценка отношения исследователей к исследовательским данным в Индии**

Манорама Трипатхи, Махеш Чанд, Шарад Сонкар, Джагдживан Дживан

IFLA Journal, 43 -1, 22-39

#### **Аннотация:**

В настоящей работе рассматривается вопрос важности первичной информации с точки зрения развития научных исследований. Данная работа включает в себя опросы 40 лиц, как непосредственно занимающихся исследованиями, так и членов преподавательского состава; цель опросов - понять, как их участники воспринимают первичные, необработанные данные. В работе высказывается предположение, что библиотеки могут играть решающую роль в оказании исследователям помощи

в систематизации, архивировании и сохранении первичных данных для последующего использования. Библиотеки на уровне университетов могли бы разработать систему, которая способствовала бы заинтересованности исследователей и преподавательского состава в размещении своих первичных исследовательских данных в хранилищах данных учебных заведений, которые созданы большинством университетских библиотек.

### **Exploring the use of stages to help understand the PIM**

#### **Использование разделения на этапы как средство лучшего понимания управления личной информацией**

Эмбер Кушинг, Олди Дамблтон

IFLA Journal, 43 -1, 40-50

Аннотация:

Пятнадцать докторантов, участвующих в различных гуманитарных и обществоведческих программах одного столичного университета Ирландии, обсудили собственные методы и приемы, связанные с управлением личной информацией, при помощи трех фокус-групп, с целью установления уровня управления личной информацией на различных этапах докторантуры. Согласно полученным результатам, потребности в управлении личной информацией докторантов, занятых обработкой информации для написания диссертации/выпускной работы, можно разделить на три четких этапа: начало, середина и окончание докторантуры. При подготовке услуг, удовлетворяющих потребностям указанной социальной группы, работники сферы информации, вероятно найдут полезным использовать классификацию таких услуг в соответствии с указанными тремя этапами.

### **Managing research data at an academic library in a developing country**

#### **Управление научными данными в библиотеке учебного заведения развивающейся страны**

Шамин Ренвик, Марша Уинтер, Мишель Джил

IFLA Journal, 43 -1, 51-64

Аннотация:

Управление научными данными стало актуальной задачей для многих университетов. На островах

Карибского моря, в Библиотеках кампуса Сент-Августин Университета Вест-Индии очень остро осознают необходимость оказания исследователям поддержки в данной сфере. Задачей настоящей работы является выявление существующих в текущий момент методов управления научными данными в пределах кампуса, а также определение роли, которую могут играть Библиотеки кампуса в данном процессе. В рамках пробного исследования были опрошены 100 изыскателей в пределах кампуса. В результате анализа 65 достоверных ответов выяснилось, что несмотря на то, что изыскатели располагают определенным объемом данных, они обладают ограниченными знаниями или опытом в части управления своими научными данными. Сведения о невысоком уровне осведомленности изыскателей очень полезны, они указывают на востребованность одной из функций, которую библиотеки кампуса могут выполнять для поддержки работающих в рамках кампуса исследователей. Библиотекам кампуса следует сформировать у изыскателей понимание результативности планирования данных и управления научными данными, а также обеспечить техническое содействие непосредственно в части хранения данных.

### **Building professional development opportunities in data services for academic librarians**

#### **Формирование возможностей для профессионального развития в области информационного обслуживания у библиотекарей учебных заведений**

Сюзанна Конрад, Ясмин Шориш, Аманда Уитмайр, Патрисия Хсве

IFLA Journal, 43 -1, 65-80

Аннотация:

Управление научными данными представляет собой сферу, в которой библиотекари учебных заведений могут достичь значительного развития своих профессиональных навыков, и значительность эта объясняется как ростом важности такого развития для данной профессии, что связано с растущими ожиданиями в части соответствия исследователей требованиям в сфере управления научными данными, так и уровнем профессиональной подготовки, требуемым от библиотекарей для обеспечения исследователей в части методов и планов управления научными данными. В настоящей статье рассказывается, как Ассоциация библиотек колледжей и научных библиотек (ACRL) расширяет

возможности профессионального развития в сфере управления научными данными. Авторы описывают два ключевых направления деятельности: 1) разработку и использование опросов для оценки потребности, которые позволяют определить, какие именно библиотекари выражают наибольшую заинтересованность; и 2) планирование и проведение предварительного семинара перед конференцией ACRL 2015 г., целью которого являлась выработка модели предлагаемого будущего профессионального развития. В заключение статьи обсуждается дополнительная оценка, которая была проведена по окончании семинара, а также то, как семинар заложил основу для предложения “презентации” управления научными данными, подобной той, которую предложил ACRL для научной коммуникации.

### **Open access and open data on natural disasters collections**

#### **Свободный доступ и открытые данные об информационных материалах, касающихся стихийных бедствий**

Мариса Ракель Де Хьюсти, Гонсало Лухан Вильярреаль, Карлос Хавьер Нуск, Аналия Пинто, Ариэль Хорхе Лира

IFLA Journal, 43 -1, 81-88

Аннотация:

После наводнения, опустошившего в 2013 году город Ла-Плата, обзор водной обстановки в регионе показал, что актуальные исследования и проекты в указанной сфере разрознены и недостаточно наглядны. Это побудило учебные, научные и правительственные организации к совместной работе в области сбора, систематизации и распространения имеющихся данных в рамках беспрецедентной реакции на экологическую катастрофу в регионе. Для этой цели в вузовском электронном хранилище данных SEDICI, принадлежащем Национальному университету Ла-Платы, была создана тематическая подборка, а затем было создано электронное хранилище под названием Экологическая обсерватория Ла-Платы (OMLP), где размещаются материалы в сфере исследований, связанных с охраной окружающей среды. В настоящей работе описаны указанные выше проекты в рамках сотрудничества, направленного на прогнозирование и смягчение последствий стихийных бедствий, перечислены их основные характеристики, их ресурсы и их текущее состояние. Также описаны

проекты, реализуемые в рамках OMLP, включая описание их внедрения с технической точки зрения, а также с точки зрения инфраструктуры согласно политике Свободного доступа.

### **Building a Research Data Management Service at UC Berkeley**

#### **Создание Службы управления научными данными в Калифорнийском университете в Беркли**

Джейми Уиттенберг, Мери Илингс

IFLA Journal, 43 -1, 89-97

Аннотация:

Библиотека Калифорнийского университета в Беркли и головное подразделение отдела Научных информационных технологий совместно разработали программу управления научными данными, которая позволяет задействовать специальные знания, опыт и ресурсы каждой организации с целью создания единой службы. Эта служба предлагает серию семинаров, консультирование, а также онлайн-ресурс. Благодаря этому сотрудничеству те области, которые часто полностью находятся в ведении ИТ, такие как архивирование и безопасное хранение, а также услуги, включенные в сферу ответственности Библиотеки, такие как поиск ресурсов и информационное обеспечение, включены в единую программу управления научными данными. В рамках данного исследования конкретного примера обсуждаются процесс создания программы, препятствия на пути ее внедрения, а также конечный результат внедрения указанной модели сотрудничества.

### **Quality evaluation of data management plans at a research university**

#### **Качественная оценка планов в области управления данными одного исследовательского университета**

Джеймс Е. Ван Лоон, Кэтрин Д. Эйкерс, Коул Хадсон, Александра Саркози

IFLA Journal, 43 -1, 98-104

Аннотация:

С появлением требования Национального научного фонда (NSF) относительно планов управления данными вырос масштаб помощи, оказываемой исследователям библиотеками учебных заведений в

части разработки планов управления данными и распространения научных данных. С целью определения общего качества планов управления данными в Университете Уэйна команда Службы системных научных данных библиотеки провела оценку содержания 119 планов управления данными из заявок на гранты NSF, поданных между 2012 и 2014 гг. Результаты проведенного нами анализа содержания свидетельствуют, что в то время как большинство исследователей понимают необходимость обеспечения коллективного использования данных, многим планам управления данными не хватает грамотного описания самих данных, формируемых в рамках проекта, того, как будет осуществляться управление данными в ходе реализации проекта или того, как будет организовано хранение данных и их коллективное использование после завершения проекта. Также результаты показывают, что недостатки планов управления данными различны у разных академических структурных единиц, и это дает основания говорить о наличии потребности в дифференцированных информационных мероприятиях, направленных на усиление планов управления данными в рамках будущих заявок на гранты NSF.

#### **Team-based data management instruction at small liberal arts colleges**

**Обучение управлению данными на командной основе в небольших колледжах свободных искусств.**

Райан Клемент, Эйми Блау, Парванех Аббаспур, Эли Грандур-Роод

IFLA Journal, 43 -1, 105-118

Аннотация:

В данной работе описана стратегия сотрудничества, использованная библиотекарями пяти маленьких региональных колледжей свободных искусств в рамках разработки/расширения услуг по управлению научными данными в своих кампусах. Данные пять колледжей вместе входят в состав объединения, известного как [НАЗВАНИЕ ОБЪЕДИНЕНИЯ]. В течение более чем 10 месяцев библиотекари из указанных пяти школ сотрудничали с целью разработки плана семинара по управлению данными и курированию данных, задачами которого являются формирование взаимоотношений с исследователями, которые работают с данными, развитие их собственных навыков и услуг в области управления научными данными, а также создание модели для проведения в будущем обучающих и информационных мероприятий по теме оказываемых этими учебными заведениями услуг в сфере управления научными данными. Данный семинар свел вместе группы исследователей, включая преподавательский состав, учащихся и библиотекарей, и объединил в себе как элементы активного обучения, так и углубленное обсуждение до начала семинара. В настоящей статье обсуждается контекст и подоплека данного семинара, непосредственно сама модель, а также его результаты и перспективы будущего развития.

## **Resúmenes**

**Research data management in Switzerland: National efforts to guarantee sustainability of research outputs**

**Gestión de datos de investigación en Suiza: iniciativas nacionales para garantizar la sostenibilidad de los resultados de investigación**

Pierre-Yves Burgi, Eliane Blumer,  
Basma Makhlouf-Shabou

IFLA Journal, 43 -1, 5-21

Resumen:

En este artículo, los autores explican un proyecto nacional de gestión del ciclo de vida útil de los datos

(DLCM, por sus siglas en inglés) que se está llevando a cabo en Suiza, centrándose básicamente en la preservación a largo plazo. Este proyecto, basado en un extenso análisis de documentos y entrevistas semi-estructuradas, pretende crear servicios nacionales para responder a las necesidades de DLCM más urgentes de los investigadores, entre las que se incluyen: directrices para establecer un plan de gestión de datos, soluciones de gestión de datos activos, opciones de almacenamiento y preservación a largo plazo, formación y un punto único de acceso y contacto para la obtención de asistencia. Además de presentar las distintas vertientes de trabajo del proyecto, los autores describen una plantilla de *lean startup* y gestión estratégica para desarrollar nuevos modelos de negocio, algo fundamental para la creación de servicios viables.



**Brief Assessment of Researchers' Perceptions towards Research Data in India****Breve evaluación de las percepciones de los investigadores sobre los datos de investigación en India**

Manorama Tripathi, Mahesh Chand, Sharad Sonkar, Jagjeevan Jeevan

IFLA Journal, 43 -1, 22-39

**Resumen:**

Este artículo analiza la importancia de los datos en bruto para el avance de la ciencia y la investigación. El estudio incluye una encuesta a 40 investigadores y docentes para conocer su percepción sobre los datos en bruto. Se propone que las bibliotecas pueden desempeñar un papel crucial en la ampliación del apoyo a los investigadores para organizar, archivar y preservar datos en bruto para su uso futuro. Las bibliotecas pueden desarrollar un sistema que anime a los investigadores y los docentes universitarios a depositar sus datos en bruto en los repositorios institucionales que la mayoría de las bibliotecas universitarias tienen.

**Exploring the use of stages to help understand the PIM****Análisis del uso de etapas para comprender la GIP**

Amber Cushing, Odile Dumbleton

IFLA Journal, 43 -1, 40-50

**Resumen:**

Quince estudiantes de doctorado matriculados en programas de humanidades y ciencias sociales de una universidad metropolitana irlandesa comentaron sus prácticas y hábitos en relación con la gestión de información personal (GIP) a través de tres grupos de interés, al objeto de analizar destrezas de gestión de información personal en diferentes etapas del programa de doctorado. Los resultados indican que las necesidades de gestión de información personal de los estudiantes de doctorado que manejan información sobre tesis y disertaciones se puede clasificar en tres etapas: inicio, centro y fin del programa de doctorado. A la hora de desarrollar servicios para satisfacer las necesidades de esta población, a los profesionales de la información les podría resultar útil clasificar dichos servicios en función de estas tres etapas.

**Managing research data at an academic library in a developing country****Gestión de datos de investigación en una biblioteca universitaria de un país en vías de desarrollo**

Shamin Renwick, Marsha Winter, Michelle Gill

IFLA Journal, 43 -1, 51-64

**Resumen:**

La gestión de los datos de investigación se ha convertido en un problema para muchas universidades. En el Caribe, las bibliotecas del Campus St. Augustine de la Universidad de las Indias Occidentales son muy conscientes de la necesidad de apoyar a los investigadores en este sentido. Los objetivos de este estudio eran identificar las prácticas actuales en relación con la gestión de datos de investigación en el campus y determinar el posible papel de sus bibliotecas. Se realizó un estudio piloto con 100 investigadores del campus. Un análisis de las 65 respuestas válidas reveló que, aunque los investigadores disponían de conjuntos de datos, poseían pocos conocimientos y experiencia para gestionarlos. Esta escasez de conocimientos resulta esclarecedora y justifica el papel de las bibliotecas del campus como apoyo para los investigadores. Las bibliotecas del campus deben concienciar a los investigadores sobre lo que entraña la planificación y la gestión de datos, además de prestarles asistencia técnica con el almacenamiento de los datos.

**Building professional development opportunities in data services for academic librarians****Creación de oportunidades de desarrollo profesional en servicios de datos para bibliotecarios universitarios**

Suzanna Conrad, Yasmeeen Shorish, Amanda Whitmire, Patricia Hswe

IFLA Journal, 43 -1, 65-80

**Resumen:**

La gestión de datos de investigación (GDI) constituye un área de desarrollo profesional importante para bibliotecarios universitarios: importante por su creciente importancia para la profesión, puesto que se espera que los investigadores cumplan determinados requisitos de GDI, e importante por el grado de competencia que los bibliotecarios necesitan para apoyar a los investigadores en las prácticas y los planes de GDI. Este artículo describe cómo la Association of College and Research Libraries (ACRL) promueve las oportunidades de desarrollo profesional en GDI. Los autores describen dos iniciativas clave: 1) el desarrollo y la puesta en marcha de una encuesta de evaluación de necesidades, que

proporcionó información sobre los tipos de bibliotecarios con más necesidades; y 2) la planificación e implementación de un taller previo al congreso para ACRL 2015, destinado a diseñar una oferta de desarrollo profesional. El artículo concluye debatiendo la evaluación complementaria realizada después del taller y cómo el taller previo al congreso sentó las bases para proponer un «roadshow» para la GDI, similar a la que la ACRL promueve para la comunicación científica.

### **Open access and open data on natural disasters collections**

#### **Acceso abierto y datos abiertos en colecciones sobre catástrofes naturales**

Marisa Raquel De Giusti, Gonzalo Luján Villarreal, Carlos Javier Nusch, Analía Pinto, Ariel Jorge Lira

IFLA Journal, 43 -1, 81-88

#### **Resumen:**

Después de una inundación que devastó la ciudad de La Plata en 2013, un estudio sobre la situación del agua en la región reveló la enorme dispersión y falta de visibilidad de los estudios y proyectos realizados en este campo. Ello incitó a las instituciones académicas, científicas y gubernamentales a colaborar para compilar, organizar y divulgar los datos disponibles, como una respuesta sin precedentes ante una catástrofe medioambiental en la región. A tal fin, se creó una colección especializada en SEDICI, el repositorio institucional de la Universidad Nacional de La Plata y, posteriormente se puso en marcha el Observatorio Medioambiental La Plata (OMLP), un repositorio de datos sobre investigación medioambiental. Este artículo describe estas iniciativas de colaboración encaminadas a prever y reducir el impacto de las catástrofes naturales, sus principales características, sus recursos y su progreso actual. También se explican los proyectos impulsados por el OMLP, junto con una descripción de su despliegue técnico y de infraestructuras regulado por políticas de acceso abierto.

### **Building a Research Data Management Service at UC Berkeley**

#### **Creación de un servicio de gestión de datos de investigación en UC Berkeley**

Jamie Wittenberg, Mary Elings

IFLA Journal, 43 -1, 89-97

#### **Resumen:**

La biblioteca y el departamento de tecnologías de la información sobre investigación de UC Berkeley han colaborado para desarrollar un programa de gestión de datos de investigación basado en la experiencia y los recursos de cada organización al objeto de crear un servicio unificado. El servicio ofrece una serie de talleres, consultas y un recurso online. Gracias a esta colaboración, áreas de servicios que suelen estar totalmente integradas en la TI, como la copia de seguridad y el almacenamiento de datos, y servicios del ámbito bibliotecario, como la búsqueda de recursos y la formación, se integran en un único programa de gestión de datos de investigación. Este caso práctico debate la implantación del programa, los obstáculos encontrados y los resultados del modelo colaborativo.

### **Quality evaluation of data management plans at a research university**

#### **Evaluación de la calidad de los planes de gestión de datos en una universidad dedicada a la investigación**

James E. Van Loon, Katherine G. Akers, Cole Hudson, Alexandra Sarkozy

IFLA Journal, 43 -1, 98-104

#### **Resumen:**

Cuando la Fundación Nacional para la Ciencia (NSF) comenzó a exigir planes de gestión de datos (PGD), los bibliotecarios universitarios empezaron a ayudar a los investigadores a desarrollar estos planes y divulgar datos de investigación. Para determinar la calidad general de los PGD en la Wayne State University, el equipo de servicios de datos de investigación (SDI) del sistema bibliotecario evaluó el contenido de 119 PGD incluidos en solicitudes de subvención de la NSF presentadas entre 2012 y 2014. Los resultados del análisis de contenidos indican que, aunque muchos investigadores conocen la necesidad de compartir datos, muchos PGD no describen correctamente los datos generados por el proyecto, cómo se gestionarán los datos durante el mismo ni cómo se preservarán y compartirán los datos tras la conclusión del proyecto. Nuestros resultados demuestran que las deficiencias de los PGD varían de unas unidades académicas a otras, lo que indica la necesidad de servicios de extensión diferenciados para mejorar la eficacia de los PGD en las solicitudes de subvención de la NSF.

**Team-based data management instruction at small liberal arts colleges****Formación sobre gestión de datos en equipo en escuelas de arte pequeñas**

Ryan Clement, Amy Blau, Parvaneh Abbaspour,  
Eli Gandour-Rood

IFLA Journal, 43 -1, 105-118

**Resumen:**

Este artículo describe un método colaborativo adoptado por bibliotecarios de cinco pequeñas escuelas de arte regionales para el desarrollo/la mejora de los servicios de gestión de datos de sus respectivos campus. Las cinco escuelas pertenecen a un consorcio

denominado [NOMBRE DEL CONSORCIO]. Los bibliotecarios de cinco escuelas colaboraron durante 10 meses para planificar un taller de gestión y custodia de datos con los objetivos de desarrollar relaciones con los investigadores que trabajaban con los datos, desarrollar sus propias destrezas y servicios de GDI y crear un modelo para la formación y la extensión futuras en torno a los servicios de GDI institucionales. Este taller reunió a grupos de investigación formados por docentes, estudiantes y bibliotecarios, e incorporó módulos de aprendizaje activos, así como un profundo debate previo al taller. Este artículo describe el contexto y los antecedentes del taller, el modelo en sí, los resultados y las posibilidades de avances futuros.

