Academic Libraries and Research Data Management Bibliography

Charles W. Bailey, Jr.

Houston: Digital Scholarship, 2023

The Academic Libraries and Research Data Management Bibliography includes over 350 selected English-language articles and books that are useful in understanding how academic libraries plan for, implement, provide, evaluate, and conduct studies about research data management (RDM) services. RDM deals with the full lifecycle of research data from creation to long-term preservation (or disposal). In addition to research data curation and repository support, academic libraries provide user consultation and training about a wide range of RDM issues, such as data literacy, metadata creation, open standards, funder compliance (e.g., data management plans), legal issues (e.g., copyright and privacy), data publication, and data sharing and reuse. The bibliography is available as a website and a website PDF with live links.

This bibliography does not cover conference proceedings, digital media works (such as MP3 files), editorials, e-mail messages, interviews, letters to the editor, presentation slides or transcripts, technical reports, unpublished e-prints, and/or weblog postings.

Most sources have been published from January 2012 through May 2023; however, a limited number of earlier key sources are also included. The bibliography has links to included works. Where possible, it uses Digital Object Identifier System (DOI) URLs. All links are subject to change. Should a link be dead, try entering it in the Internet Archive Wayback Machine.

Some publishers may use nontraditional citation elements and patterns, and they may omit standard bibliographic elements and substitute new ones.

Abstracts are included in this bibliography if a work is under a Creative Commons Attribution License (BY and national/international variations), a Creative Commons public domain dedication (CC0), or a Creative Commons Public Domain Mark and this is clearly indicated in the work.

Unless otherwise noted, article abstracts in this bibliography are under a Creative Commons Attribution 4.0 International License, https://creativecommons.org/licenses/by/4.0/. Abstracts are reproduced as written in the source material.

Dedication

In memory of Paul Evan Peters (1947-1996), founding Executive Director of the Coalition for Networked Information, whose visionary leadership at the dawn of the Internet era fostered the development of scholarly electronic publishing.
Researchers are increasingly being asked to ensure that all products of research activity—not just traditional publications—are preserved and made widely available for study and reuse as a precondition for publication or grant funding, or to conform to disciplinary best practices. In order to conform to these requirements, scholars need effective, easy-to-use tools and services for the long-term curation of their research data. The DataShare service, developed at the University of California, is being used by researchers to: (1) prepare for curation by reviewing best practice recommendations for the acquisition or creation of digital research data; (2) select datasets using intuitive file browsing and drag-and-drop interfaces; (3) describe their data for enhanced discoverability in terms of the DataCite metadata schema; (4) preserve their data by uploading to a public access collection in the UC3 Merritt curation repository; (5) cite their data in terms of persistent and globally-resolvable DOI identifiers; (6) expose their data through registration with well-known abstracting and indexing services and major internet search engines; (7) control the dissemination of their data through enforceable data use agreements; and (8) discover and retrieve datasets of interest through a faceted search and browse environment. Since the widespread adoption of effective data management practices is highly dependent on ease of use and integration into existing individual, institutional, and disciplinary workflows, the emphasis throughout the design and implementation of DataShare is to provide the highest level of curation service with the lowest possible technical barriers to entry by individual researchers. By enabling intuitive, self-service access to data curation functions, DataShare helps to contribute to more widespread adoption of good data curation practices that are critical to open scientific inquiry, discourse, and advancement.


Scholars at the ten campuses of the University of California system, like their academic peers elsewhere, increasingly are being asked to ensure that data resulting from their research and teaching activities are subject to effective long-term management, public discovery, and retrieval. The new academic imperative for research data management (RDM) stems from mandates from public and private funding agencies, pre-publication requirements, institutional policies, and evolving norms of scholarly discourse. In order to meet these new obligations, scholars need access to appropriate disciplinary and institutional tools, services, and guidance. When providing help in these areas, it is important that service providers recognize the disparity in scholarly familiarity with data curation concepts and practices. While the UC Curation Center (UC3) at the California Digital Library supports a growing...
roster of innovative curation services for University use, most were intended originally to meet the needs of institutional information professionals, such as librarians, archivists, and curators. In order to address the new curation concerns of individual scholars, UC3 realized that it needed to deploy new systems and services optimized for stakeholders with widely divergent experiences, expertise, and expectations. This led to the development of Dash, an online data publication service making campus data sharing easy. While Dash gives the appearance of being a full-fledged repository, in actuality it is only a lightweight overlay layer that sits on top of standards-compliant repositories, such as UC3’s existing Merritt curation repository. The Dash service offers intuitive, easy-to-use interfaces for dataset submission, description, publication, and discovery. By imposing minimal prescriptive eligibility and submission requirements; automating and hiding the mechanical details of DOI assignment, data packaging, and repository deposit; and featuring a streamlined, self-service user experience that can be integrated easily into scholarly workflows, Dash is an important new service offering with which UC scholars can meet their RDM obligations.


Academic librarians are increasingly engaging in data curation by providing infrastructure (e.g., institutional repositories) and offering services (e.g., data management plan consultations) to support the management of research data on their campuses. Efforts to develop these resources may benefit from a greater understanding of disciplinary differences in research data management needs. After conducting a survey of data management practices and perspectives at our research university, we categorized faculty members into four research domains—arts and humanities, social sciences, medical sciences, and basic sciences—and analyzed variations in their patterns of survey responses. We found statistically significant differences among the four research domains for nearly every survey item, revealing important disciplinary distinctions in data management actions, attitudes, and interest in support services. Serious consideration of both the similarities and dissimilarities among disciplines will help guide academic librarians and other data curation professionals in developing a range of data-management services that can be tailored to the unique needs of different scholarly researchers.
In addition to encouraging the deposit of research data into institutional data repositories, academic librarians can further support research data sharing by facilitating the deposit of data into external disciplinary data repositories.

In this paper, we focus on the University of Michigan Library and Dryad, a repository for scientific and medical data, as a case study to explore possible forms of partnership between academic libraries and disciplinary data repositories. We found that although few University of Michigan researchers have submitted data to Dryad, many have recently published articles in Dryad-integrated journals, suggesting significant opportunities for Dryad use on our campus. We suggest that academic libraries could promote the sharing and preservation of science and medical data by becoming Dryad members, purchasing vouchers to cover researchers' data submission costs, and hosting local curators who could directly work with campus researchers to improve the accuracy and completeness of data packages and thereby increase their potential for re-use.

By enabling the use of both institutional and disciplinary data repositories, we argue that academic librarians can achieve greater success in capturing the vast amounts of data that presently fail to depart researchers' hands and making that data visible to relevant communities of interest.

Academic research libraries are quickly developing support for research data management (RDM), including both new services and infrastructure. Here, we tell the stories of how eight different universities have developed programs of RDM support, focusing on the prominent role of the library in educating and assisting researchers with managing their data throughout the research lifecycle. Based on these stories, we construct timelines for each university depicting key steps in building support for RDM, and we discuss similarities and dissimilarities among universities in motivation to provide RDM support, collaborations among campus units, assessment of needs and services, and changes in staffing.
This article describes the novel open source tools for open data publication in open access journal workflows. This comprises a plugin for Open Journal Systems that supports a data submission, citation, review, and publication workflow; and an extension to the Dataverse system that provides a standard deposit API. We describe the function and design of these tools, provide examples of their use, and summarize their initial reception. We conclude by discussing future plans and potential impact.

This work is licensed under a Creative Commons Attribution 3.0 United States License, https://creativecommons.org/licenses/by/3.0/us/.


This article explores the types of digital information research libraries typically deal with and what factors might influence libraries' decisions to take on the work of data curation themselves, to take on the responsibility for data but market out the actual work, or to leave the responsibility to other organisations.


A range of external pressures are causing research data management (RDM) to be an increasing concern at senior level in universities and other research institutions. But as well as external pressures, there are also good reasons for establishing effective research data management services within institutions which can bring benefits to researchers, their institutions and those who publish their research. In this article some of these motivating factors, both positive and negative, are described. Ways in which libraries can play a role—or even lead—in the development of RDM services that work within the institution and as part of a national and international research data infrastructure are also set out.

Objectives: To meet the changing needs of our campuses, librarians responsible for research data services are often tasked with starting new endeavors with new populations without much support. This paper reports on a collaborative effort to build a community of practice of librarians tasked with addressing the research data needs of their campuses, describes how this effort was evaluated, and presents future opportunities.

Methods: In March of 2015, three librarians found themselves in a situation of serendipitous professional development: one was seeking to provide a new method of mentorship, and two more were working on an event, hoping to broadcast it to a wider community. From these two disparate goals, the Research Data Management (RDM) Roundtables were created. The RDM Roundtables planning committee developed a low-cost professional development day divided into two parts: a morning session that detailed an idea or solution relevant to our practice, and an afternoon roundtable discussion on practical aspects of research data services. Evaluations from these events were coded in NVivo and we report on the common themes.

Results: Participants returned sixty-one evaluations from four events. Five themes emerged from the evaluations: learning, sharing, format, networking, and empathy.

Conclusion: The events provide a valuable professional development experience for attendees, and the authors hope that by providing a description of the events' development, others will establish their own local communities of practice.


In 2017 a group of academic library and information technology staff from institutions across New England piloted a process of joining The Carpentries, an organization developed to train researchers in essential computing skills and practices for automating and improving their handling of data, as a consortium. The New England Software Carpentry Library Consortium (NESCLiC) shared a gold-level tier membership to become a Carpentries member organization. NESCLiC members attended a Software Carpentry workshop together and then participated in instructor training as a cohort, collaborating on learning the material, practicing, and beginning to host and teach workshops as a group.

This article describes both the successes and challenges of forming this new consortium, suggests good practices for those who might wish to form similar collaborations, and discusses the future of this program and other efforts to help researchers improve their computing and data handling skills.


Many libraries are keen to take on new roles in providing support for effective research data management (RDM), but lack the necessary skills and resources to do so. This article explores the approach used by the University of Sussex to engage with academic departments about their RDM practices and requirements in order to develop relevant library support services. It describes a project undertaken with three Academic Schools to inform a list of recommendations for senior management, to include areas which should be taken forward by the Library, IT and Research Office in order to create a sustainable RDM service. The article is unflinchingly honest in sharing the differing reactions to the project and the lessons learnt along the way.


Purpose: Medical libraries need to actively review their service models and explore partnerships with other campus entities to provide better-coordinated clinical research management services to faculty and researchers. TRAIL (Translational Research and Information Lab), a five-partner initiative at the University of Washington (UW), explores how best to leverage existing expertise and space to deliver clinical research data management (CRDM) services and emerging technology support to clinical researchers at UW and collaborating institutions in the Pacific Northwest.

Methods: The initiative offers 14 services and a technology-enhanced innovation lab located in the Health Sciences Library (HSL) to support the University of Washington clinical and research enterprise. Sharing of staff and resources merges library and non-library workflows, better coordinating data and innovation services to clinical researchers. Librarians have adopted new roles in CRDM, such as providing user support and training for UW's Research Electronic Data Capture (REDCap) instance.

Results: TRAIL staff are quickly adapting to changing workflows and shared services, including teaching classes on tools used to manage clinical research data. Researcher interest in TRAIL has sparked new collaborative initiatives and service offerings. Marketing and promotion will be important for raising researchers' awareness of available services.

Conclusions: Medical librarians are developing new skills by supporting and teaching CRDM. Clinical and data librarians better understand the information needs of clinical and translational researchers by being involved in the earlier stages of the research cycle and identifying technologies that can improve healthcare outcomes. At health sciences libraries, leveraging existing resources and bringing services together is central to how university medical librarians will operate in the future.


INTRODUCTION Sharing digital research data is increasingly common, propelled by funding requirements, journal publishers, local campus policies, or community-driven expectations of more collaborative and interdisciplinary research environments. However, it is not well understood how researchers are addressing these expectations and whether they are transitioning from individualized practices to more thoughtful and potentially public approaches to data sharing that will enable reuse of their data. METHODS The University of Minnesota Libraries conducted a local opt-in study of data management plans (DMPs) included in funded National Science Foundation (NSF) grant proposals from January 2011 through June 2014. In order to understand the current data management and sharing practices of campus researchers, we solicited, coded, and analyzed 182 DMPs, accounting for 41% of the total number of plans available. RESULTS DMPs from seven colleges and academic units were included. The College of Science of Engineering accounted for 70% of the plans in our review. While 96% of DMPs mentioned data sharing, we found a variety of approaches for how PIs shared their data, where data was shared, the intended audiences for sharing, and practices for ensuring long-term reuse. CONCLUSION DMPs are useful tools to investigate researchers' current plans and philosophies for how research outputs might be shared. Plans and strategies for data sharing are inconsistent across this sample, and researchers need to better understand what kind of sharing constitutes public access. More intervention is needed to ensure that researchers implement the sharing provisions in their plans to the fullest extent possible. These findings will help academic libraries develop practical, targeted data services for researchers that aim to increase the impact of institutional research.
higher education. Reviewing current best practices provides insights into the role-based responsibilities for RDM services that data services librarians perform, as well as ways to improve and create new services to meet the needs of their respective university communities.

Objectives: The objectives of this article are to provide the context of research data services through a review of past studies, explain how they informed this qualitative study, and provide the methods and results of the current study. This study provides an in-depth overview of the overall job responsibilities of data services librarians and as well as their perspectives on RDM through job analyses.

Methods: Job analysis interviews provide insight and context to the tasks employees do as described in their own words. Interviews with 10 data services librarians recruited from the top 10 public and top 10 private universities according to the 2020 Best National University Rankings in the US News and World Reports were asked 30 questions concerning their overall job tasks and perspectives on RDM. Five public and five private data services librarians were interviewed. The interviews were recorded and transcribed. The transcriptions were analyzed in NVivo using a grounded theory application of open, axial, and selective coding to generate categories and broad themes based on the responses using synonymous meanings.

Results: The results presented here provide the typical job tasks of data services librarians that include locating secondary data, reviewing data management plans (DMPs), conducting outreach, collaborating, and offering RDM training. Fewer data services librarians assisted with data curation or manage an institutional repository.

Discussion: The results indicate that there may be different types of data services librarians depending on the mix of responsibilities. Academic librarianship will benefit from further delineation of job titles using tasks while planning, advertising, hiring, and evaluating workers in this emerging area. There remain many other explorations needed to understand the challenges and opportunities for data services librarians related to RDM.

Conclusions: This article concludes with a proposed matrix of job tasks that indicates different types of data services librarians to inform further study. Future job descriptions, training, and education will all benefit from differentiating between the many associated research data services roles and with increased focus on research data greater specializations will emerge.


Comprehensive research data management is fundamental to ensuring reproducible, open scientific research. However, sufficient research data assistance is often not offered at universities, and when offered, typically only provides basic services that are viewed as optional. Integrating information specialists into research groups provides a potentially promising means of improving data management by providing personalized data management workflows. Workflows are comprehensive, executable guides that require planning, implementation, feedback, and adaptation. Comprehensive data management workflows should include a file organization scheme, the creation of data management roles for members, a data storage/sharing guide, and training and evaluation. Librarians, who regularly interact with faculty and students and are familiar with data management tools, are uniquely situated to assist with the creation and assessment of these workflows.
Data services librarians are often faced with local needs in all 12 data information literacy competencies (Carlson et al. 2011) but may not have the requisite skills in each area to offer services.

Gaining expertise in an unfamiliar competency, particularly one which has not historically been associated with the library like "data visualization and representation." can be challenging.

This paper identifies a scaffolding for librarians to gain foundational experience in data visualization and provides a case study on one library's initial service offering in this area — a "Data Visualization 101" workshop.

Dates appear regularly in research data and metadata but are a problematic data type to normalize due to a variety of potential formats. This suggests an opportunity for data librarians to assist with formatting dates, yet there are frequent examples of data librarians using diverse strategies for this purpose. Instead, data librarians should adopt the international date standard ISO 8601. This standard provides needed consistency in date formatting, allows for inclusion of several types of date-time information, and can sort dates chronologically. As regular advocates for standardization in research data, data librarians must adopt ISO 8601 and push for its use as a data management best practice.

INTRODUCTION Many research institutions have developed research data services in their libraries, often in anticipation of or in response to funder policy. However, policies at the institution level are either not well known or nonexistent. METHODS This study reviewed library data services efforts and institutional data policies of 206 American universities, drawn from the July 2014 Carnegie list of universities with "Very High" or "High" research activity designation. Twenty-four different characteristics relating to university type, library data services, policy type, and policy contents were examined. RESULTS The study has uncovered findings surrounding library data services, institutional data policies, and content within the policies. DISCUSSION Overall, there is a general trend toward the development and
implementation of data services within the university libraries. Interestingly, just under half of the universities examined had a policy of some sort that either specified or mentioned research data. Many of these were standalone data policies, while others were intellectual property policies that included research data. When data policies were discoverable, not behind a log in, they focused on the definition of research data, data ownership, data retention, and terms surrounding the separation of a researcher from the institution. CONCLUSION By becoming well versed on research data policies, librarians can provide support for researchers by navigating the policies at their institutions, facilitating the activities needed to comply with the requirements of research funders and publishers. This puts academic libraries in a unique position to provide insight and guidance in the development and revisions of institutional data policies.


Data curation is the process of managing data to make it available for reuse and preservation and to allow FAIR (findable, accessible, interoperable, reusable) uses. It is an important part of the research lifecycle as researchers are often either required by funders or generally encouraged to preserve the dataset and make it discoverable and reusable. This has been especially important as the Open Access (OA) policy is being implemented in many institutions across the nation. In facilitating research data discovery and enhancing its easier reuse, an efficient data repository and its data curation play key roles. In this article, we briefly discuss the local institutional repository at Penn State University and the general data curation practices we adopt for the deposited files and datasets, then we focus on a data analytics tool that has recently been applied to extract tabular data from PDF files. This is an enhancement to the existing data curation practices as it adds additional tabular data to deposits with PDF files where tables are often embedded and not easily reused.
Every story has a beginning, where the narrator chooses to start, though this is rarely the genesis. This story begins with the launch of the University of Virginia Library's new Research Data Services unit in October 2013. Born from the conjoining of a data management team and a data analysis team, Research Data Services expanded to encompass data discovery and acquisitions, research software support, and new expertise in the use of restricted data. Our purpose is to respond to the challenges created by the growing ubiquity and scale of data by helping researchers acquire, analyze, manage, and archive these resources. We have made serious strides toward becoming "the face of data services at U.Va." This article tells a bit of our story so far, relays some early challenges and how we've responded to them, outlines several initial successes, and summarizes a few lessons going forward.


Research data management (RDM) is a major priority for many institutions as they struggle to cope with the plethora of pronouncements including funder policies, a G8 statement, REF2020 consultations, all stressing the importance of open data in driving everything from global innovation through to more accountable governance; not to mention the more direct possibility that non-compliance could result in grant income drying up. So, at the coalface, how do we become part of this global movement?

In this article the author explains the approach being taken at the University of St Andrews, building on the research information management infrastructure (data, systems and people) that has evolved since 2006. Continuing to navigate through the rapidly evolving research policy and cultural landscape, they aim to establish services to support their research community as it moves to this "open by default" requirement of funders and governments.


Objective: To evaluate library workshops on their coverage of data management topics.

Methods: We used a modified version of Sapp Nelson's Competency Matrix for Data Management Skills, a matrix of learning goals organized by data management competency and complexity level, against which we compared our educational materials: slide decks and worksheets. We examined each of the educational materials against the 333 learning objectives in our modified version of the Matrix to determine which of the learning objectives applied.

Conclusions: We found it necessary to change certain elements of the Matrix's structure to increase its clarity and functionality: reinterpreting the "behaviors," shifting the organization from the three domains of Bloom's taxonomy to increasing
complexity solely within the cognitive domain, as well as creating a comprehensive identifier schema. We appreciated the Matrix for its specificity of learning objectives, its organizational structure, the comprehensive range of competencies included, and its ease of use. On the whole, the Matrix is a useful instrument for the assessment of data management programming.


Researchers are faced with unprecedented challenges due to the size and complexity of data, and libraries are stepping in to help by providing guidance on research data management primarily to graduate students and faculty. Currently, many universities are encouraging an undergraduate research experience where students engage in research projects in the classroom and in research labs, yet research data management is often not included as part of these opportunities. At UW-Madison, we piloted researchERS (Emerging Research Scholars), a program for undergraduates from all disciplines to learn data management skills. Focusing on core concepts as well as data ethics, reproducibility, and research workflows, the format of the program included seven evening workshops, two networking events, and one field trip. Each workshop invited campus and community speakers relevant to the workshop’s theme as a way to introduce the students to the network of available resources and data expertise and provided food for attendees. The workshops also built in customized activities to show students how to incorporate best practices into their work. Local businesses provided a tour of their facilities as well as a talk on how they leverage data. This paper will describe this program as well as the benefits and drawbacks of tailoring a research data management program toward undergraduates.


Purpose: This paper introduces the Portage Network’s Dataverse Curation Guide and the new bilingual curation framework developed to support it.

Brief Description: Canadian academic institutions and national organizations have been building infrastructure, staffing, and programming to support research data management. Amidst this work, a notable gap emerged between requirements for data curation in general repositories like Dataverse and the requisite workflows and guidance materials needed by curators to meet them. In response, Portage, a national network of data experts, organized a working group to develop a Dataverse curation guide built upon the Data Curation Network’s CURATED workflow. To create a bilingual resource, the original CURATE(D) acronym was modified to CURATION—which has the same meaning in both French and English—and steps were augmented with Dataverse-specific guidance and mapped to three conceptualized levels of curation to assist curators in prioritizing curation actions.

Methods: An environmental scan of relevant deposit and curation guidance materials from Canadian and international institutions identified the need for a comprehensive Dataverse Curation Guide, as most existing resources were either depositor-focused or contained only partial workflows. The resulting Guide synthesized these guidance materials into the CURATION steps and mapped actions to various theoretical levels of data repository services and levels of curation.

Much time and energy is now being devoted to developing the skills of researchers in the related areas of data analysis and data management. However, less attention is currently paid to developing the data skills of librarians themselves: these skills are often brought in by recruitment in niche areas rather than considered as a wider development need for the library workforce, and are not widely recognised as important to the professional career development of librarians. We believe that building computational and data science capacity within academic libraries will have direct benefits for both librarians and the users we serve.

Library Carpentry is a global effort to provide training to librarians in technical areas that have traditionally been seen as the preserve of researchers, IT support and systems librarians. Established non-profit volunteer organisations, such as Software Carpentry and Data Carpentry, offer introductory research software skills training with a focus on the needs and requirements of research scientists. Library Carpentry is a comparable introductory software skills training programme with a focus on the needs and requirements of library and information professionals. This paper describes how the material was developed and delivered, and reports on challenges faced, lessons learned and future plans.


This article reports an international study of research data management (RDM) activities, services, and capabilities in higher education libraries. It presents the results of a survey covering higher education libraries in Australia, Canada, Germany, Ireland, the Netherlands, New Zealand, and the UK. The results indicate that libraries have provided leadership in RDM, particularly in advocacy and policy development. Service development is still limited, focused especially on advisory and consultancy services (such as data management planning support and data-related training), rather than technical services (such as provision of a data catalog, and curation of active data). Data curation skills development is underway in libraries, but skills and capabilities are not consistently in place and remain a concern. Other major challenges include resourcing, working with other support services, and achieving "buy in" from researchers and senior managers. Results are compared with previous studies in order to assess trends and relative maturity levels. The range of RDM activities explored in this study are positioned on a "landscape maturity model," which reflects current and planned research data services and practice in academic libraries, representing a "snapshot" of current developments and a baseline for future research.


University libraries have played an important role in constructing an infrastructure of support for Research Data Management at an institutional level. This paper presents a comparative analysis of two international surveys of libraries about their involvement in Research Data Services conducted in 2014 and 2018. The aim was to explore how services had developed over this time period, and to explore the drivers and barriers to change. In particular, there was an interest in how far the FAIR data principles had been adopted.

Services in nearly every area were more developed in 2018 than before, but technical services remained less developed than advisory. Progress on institutional policy was also evident. However, priorities did not seem to have shifted significantly. Open ended answers suggested that funder policy, rather than researcher demand, remained the main driver of service development and that resources and skills gaps remained. While widely understood as an important reference point and standard, because of their relatively recent publication date, FAIR principles had not been widely adopted explicitly in policy.


The purpose of this paper is to explore the value to librarians of seeing research data management as a 'wicked' problem. Wicked problems are unique, complex problems which are defined differently by different stakeholders making them particularly intractable. Data from 26 semi-structured in-depth telephone interviews with librarians was analysed to see how far their perceptions of research data management aligned with the 16 features of a wicked problem identified from the literature. To a large extent research data management is perceived to be wicked, though over time good practices may emerge to help to 'tame' the problem. How interviewees thought research data management should be approached reflected this realisation. The generic value of the concept of wicked problems is considered and some first thoughts about how the curriculum for new entrants to the profession can prepare them for such problems are presented.

This work is licensed under a Creative Commons Attribution 3.0 Unported License. https://creativecommons.org/licenses/by/3.0/.


For many UK HEIs, especially research-intensive institutions, Research Data Management (RDM) is rising rapidly up the agenda. Working closely with other professional services, and with researchers themselves, libraries will probably have a key role to play in supporting RDM. This role might include signposting institutional expertise in RDM; inclusion of the topic in information literacy sessions for PhD students and other researchers; advocacy for open data sharing; or contributing to
the management of an institutional data repository. It seems that there are choices for each librarian to make, largely shaped by their existing role. For some, RDM may rapidly become a core part of their job. For others it may be something of which they simply need a greater awareness. New graduates entering the profession require a grounding in RDM-related knowledge and skills, but there is also a need for established professionals to update their competencies too.

In this context, JISC have funded the White Rose consortium of academic libraries at Leeds, Sheffield and York, working closely with the Sheffield Information School, in the RDMRose Project (link is external), to develop learning materials that will help librarians grasp the opportunity that RDM offers. The learning materials will be used in the Information School's Masters courses, and are also to be made available to other information sector training providers on a share-alike licence. A version will also be made available (from January 2013) as an Open Educational Resource for use by information professionals who want to update their competencies as part of their continuing professional development (CPD). The learning materials are being developed specifically for liaison librarians, to upskill existing professionals and to expand the knowledge base for new entrants to librarianship. It is hoped to accommodate the perspectives of any information professional, but the scope is not intended to encompass a syllabus for a data management specialist role (following the distinction made by Corrall [1]).

This article summarises current thinking developed within the project about the scope and level of such learning materials. This thinking is based on a number of sources: the literature and existing curricula and also the project vision and data collected during the project in focus groups with staff at the participating libraries.

This work is licensed under a Creative Commons Attribution 3.0 Unported License. https://creativecommons.org/licenses/by/3.0/.


INTRODUCTION Research Data Management (RDM) offers opportunities and challenges at the interface of library support and researcher needs. Libraries are in a position of balancing the capacity to provide support at the point of need while also implementing training for subject liaison librarians grounded in the practical issues and realities facing researchers and their institutions. DESCRIPTION OF PROGRAM/SERVICE The North Carolina State University (NCSU) Libraries has deployed a Data Management Plan (DMP) Review service managed by a committee of librarians with diverse experience in data management and domain expertise. By rotating librarians through membership on the committee and by inviting subject liaisons librarians to participate in the DMP Review process, our training ground model aims to develop needed competencies and support researchers through relevant services and partnerships. AUDIT OF PROGRAM/SERVICE This article presents an audit of the DMP Review service as a training ground to develop and enhance competencies as identified by the Joint Task Force on Librarians'
Competencies in Support of E-Research and Scholarly Communication. NEXT STEPS AND CONCLUSIONS The DMP Review service creates opportunities for librarians to learn valuable skills while simultaneously providing a time-sensitive service to researchers. The process of auditing competencies developed by participating in the DMP Review service highlights gaps needed to more fully support RDM and reinforces the capacity of the DMP Review service as a training ground to sustain and iterate learning opportunities for librarians engaged in research support and partnerships.


This paper charts the steps taken and possible ways forward for the University of Warwick in its approach to research data management, providing a typical example of a UK research university's approach in two strands: requirements and support. The UK government approach and funding landscape in relation to research data management provided drivers for the University of Warwick to set requirements and provide support, and examples of good practice at other institutions, support from a central national body (the UK Digital Curation Centre) and learning from other universities' experiences all proved valuable to the University of Warwick. Through interviews with researchers at Warwick, various issues and challenges are revealed: perhaps the biggest immediate challenges for Warwick going forward are overcoming scepticism amongst researchers, overcoming costs, and understanding the implications of involving third party companies in research data management. Building technical infrastructure could sit alongside and beyond those immediate steps and beyond the challenges that face one University are those that affect academia as a whole. Researchers and university administrators need to work together to address the broader challenges, such as the accessibility of data for future use and the reward for researchers who practice data management in exemplary ways, and indeed it may be that a wider, national or international but disciplinary technical infrastructure affects what an individual university needs to achieve. As we take these steps, universities and institutions are all learning from each other.


eScience related library services at Princeton University started in response to the National Science Foundation's (NSF) data management plan requirements, and grew to encompass a range of services including data management plan consultation, assistance with depositing into a disciplinary or institutional repository, and research data management instruction. These services were initially directed at science and engineering disciplines on campus, but the eScience Librarian soon realized the relevance of research data management instruction for humanities disciplines with digital approaches. Applicability to the digital humanities was initially recognized by discovery of related efforts from the history department's Information Technology (IT) manager in the form of a graduate-student workshop on file and digital-asset management concepts. Seeing the common ground these activities shared with research data management, a collaboration was formed between the history department's IT Manager and the eScience Librarian to provide a research data management overview to the entire campus community. The eScience Librarian was then invited to participate in the history department's graduate student file and digital
asset management workshop to provide an overview of other research data management concepts. Based on the success of the collaboration with the history department IT, the eScience Librarian offered to develop a workshop for the newly formed Center for Digital Humanities at Princeton. To develop the workshop, background research on digital humanities curation was performed revealing similarities and differences between digital humanities curation and research data management in the sciences. These similarities and differences, workshop results, and areas of further study are discussed.


Objectives

Many librarians are taking on new roles in research data services. However, the emerging field of data librarianship, including specific roles and competencies, has not been clearly established. This study aims to better define data librarianship by exploring the skills and knowledge that data librarians utilize and the training that they need to succeed.

Methods

Librarians who do data-related work were surveyed about their work and educational backgrounds and asked to rate the relevance of a set of data-related skills and knowledge to their work.

Results

Respondents considered a broad range of skills and knowledge important to their work, especially "soft skills" and personal characteristics, like communication skills and the ability to develop relationships with researchers. Traditional library skills like cataloging and collection development were considered less important. A cluster analysis of the responses revealed two types of data librarians: data generalists, who
tend to provide data services across a variety of fields, and subject specialists, who tend to provide more specialized services to a distinct discipline.

Discussion

The findings of this study suggest that data librarians provide a broad range of services to their users and, therefore, need a variety of skills and expertise. Libraries hiring a data librarian may wish to consider whether their communities will be best served by a data generalist or a subject specialist and write their job postings accordingly. These findings also have implications for library schools, which could consider adjusting their curricula to better prepare their students for data librarian roles.


Increasingly, users of health and biomedical libraries need assistance with challenges they face in working with their own and others' data. Librarians have a unique opportunity to provide valuable support and assistance in data science and open science but may need to add to their expertise and skill set to have the most impact. This article describes the rationale for and development of the Medical Library Association Data Services Competency, which outlines a set of five key skills for data services and provides a course of study for gaining these skills.


Academic libraries have a critical role to play as data quality hubs on campus. There is an increased need to ensure data quality within 'e-science'. Given academic libraries' curation and preservation expertise, libraries are well suited to support the
While interest in research data management (RDM) services have grown, clarifying the path between traditional library responsibilities and RDM remains a challenge. While the literature has provided ideas about services and student-researcher-focused data information literacy (DIL) competencies, nothing has yet brought these skill sets together to provide a pathway for librarians engaging in RDM. The Data Engagement Opportunities scaffold was developed to provide a strategic trajectory relating information science skills, the DIL competencies, the stages of the data life cycle, three levels of RDM engagement activities, and potential measurable outcomes. This scaffold provides direction for librarians looking to identify their current abilities and explore new opportunities.

As data as a scholarly object continues to grow in importance in the research community, librarians are undertaking increasing responsibilities regarding data management and curation. New library initiatives include assisting researchers in finding data sets for reuse; locating and hosting repositories for required archiving; consultations on workflow, data management plans, and best practices; responding to changing funder policies (Whitmire, et al. 2015) and development of department or institutional policies. Librarians looking to provide services or expand into these areas will need both foundational resources and information about engaging the network of librarians exploring data. This webliography is intended for librarians seeking to enhance their own knowledge and assist peers in improving their data management awareness.

INTRODUCTION New interest has arisen in organizing, preserving, and sharing the raw materials—the data and metadata—that undergird the published products of research. Library and information scientists have valuable expertise to bring to bear in the effort to create larger, more diverse, and more widely used data repositories. However, for libraries to be maximally successful in providing the research data management and preservation services required of a successful data repository, librarians must work closely with researchers and learn about their data management workflows. DESCRIPTION OF SERVICES Databrary is a data repository that is closely linked to the needs of a specific scholarly community—researchers who use video as a main source of data to study child development and learning. The project’s success to date is a result of its focus on community outreach and providing services for scholarly communication, engaging institutional partners, offering services for data curation with the guidance of closely involved information professionals, and the
creation of a strong technical infrastructure. NEXT STEPS Databrary plans to improve its curation tools that allow researchers to deposit their own data, enhance the user-facing feature set, increase integration with library systems, and implement strategies for long-term sustainability.


**INTRODUCTION**

Research data services have been adopted by many academic libraries. This study tracked the changes in research data management services and staffing among Association of American Universities (AAU) libraries over the past 5 years and compared them to the libraries' goals for research data management (RDM) in their strategic plan. **METHODS** This quantitative study examined libraries at the 60 U.S. AAU institutions. In order to examine longitudinal changes, portions of Briney et. al. (2015a) were used as a basis for measuring data librarian staffing and services. These trends were compared to the contemporary strategic priorities of libraries interviewed by Meier (2016), as well as against strategic plans of 2014 and 2019 available online. **RESULTS & DISCUSSION** While there have been modest increases in libraries in the sample population offering data services, most of those gains have been among the libraries that did not consider RDM a priority in 2014. Interestingly, some of the libraries that mentioned RDM as a priority in 2014 have lost data librarian positions. Over half of the libraries in this study now provide or support a data repository. Many library strategic plans that mentioned RDM as an explicit goal 5 years ago now no longer mention it. **CONCLUSION** Data librarian positions, data services, and data repositories have now become common features of large research university libraries. However, research data services are no longer as prominent in many library strategic plans at institutions where such services are more established, and libraries instead seem to be moving on to the work of rethinking the nature of the services or expanding them.

This work is licensed under a Creative Commons Attribution 3.0 Unported License. https://creativecommons.org/licenses/by/3.0/.


**Objective:** In the interest of making data findable, accessible, interoperable, and reusable (FAIR), the National Institutes of Health (NIH) will institute a new Data Management and Sharing Policy in January 2023. This policy will require researchers applying for NIH funding to submit a Data Management and Sharing Plan. As 63% of grant dollars received by University of Massachusetts Chan Medical School (UMass Chan) researchers comes from the NIH, we explored whether UMass Chan researchers are currently sharing data associated with their published research and how they shared their data.

**Methods:** PubMed was searched for articles published in 2019 with a UMass Chan researcher as either the first or last author. These articles were examined for evidence of original or reused data, the type of data, whether the article stated that data was available, and where and how to find that data.

**Results:** Of the 361 articles with original data, 26% had a data availability statement. However, most articles (71%) did not mention where data could be. The data storage
location of the estimated 1551 original datasets was similarly not mentioned for 74% of the datasets with the next largest category being available upon request (8.6%). Genomic data repositories such as the Gene Expression Omnibus were among the top repositories used by authors. Similar areas for improvement were noted for permanent identifier use (46% had a permanent identifier), using non-proprietary file formats (most popular format was Excel), and citing reused data. Authors who published open access were more likely to share their data.

Conclusions: While some researchers at UMass Chan have embraced data sharing, particularly genomic data sharing, we expect there will be more data shared in the coming years with the implementation of the new NIH Data Management and Sharing Policy.


The ASU Library is actively building relationships around and increasing its expertise in research data services. We have established a collaboration with our university's research administration in order to coordinate our distinct areas of expertise in research data services so that both entities can better support researchers all the way through the research data lifecycle. The Library embedded itself into research administration's learning management system and works with their research advancement officers to engage with researchers and staff we have not traditionally reached. Forging this new collaboration increased expectations that the Library will expand existing research data services to more investigators, so we have grown Library professionals' internal competencies by providing research data management training opportunities to meet these demands. In addition, the Library's Research Services Working Group established data competencies, workflows, and trainings so more librarians gain skills necessary to answer and assist patrons with data needs. Greater expertise throughout the Library enables us to authentically and confidently scale our research data services and form new collaborations.


There are many courses available to teach research data management to librarians and researchers. While these courses can help with technical skills, like programming or statistics, and practical knowledge of data life cycles or data sharing policies, there are "soft skills" and non-technical skills that are needed to successfully start and run data services. While there are many important characteristics of a good data librarian, reference skills, relationship building, collaboration, listening, and facilitation are some of the most important. Giving consideration to these skills will help any data librarian with their multifaceted job.

Leiden University (LU) adopted a data management regulation in 2016. The regulation embraces the Findable, Accessible, Interoperable and Reusable (FAIR) principles. To implement the regulation a programme was established. The focus of the programme was initially to raise awareness and to set up services to make data Findable and Accessible and to train researchers on data management planning. In 2019, the programme entered its second phase, with an increased focus on Interoperable and Reusable data, and on implementing the machine-actionable aspects of FAIR data. This step is non-trivial, however, because of the fast-developing FAIR research data international research field that requires fast adoption of leading practices by support professionals with the adequate skills. In this paper we describe how LU aims to close the feedback loop between international bottom-up organisations such as GOFAIR, the Research Data Alliance and CODATA on the one hand and university staff engaged in developing and implementing emerging FAIR leading practices on the other. During processes such as these, it is of crucial importance to focus primarily on the needs of researchers. We describe how LU builds up its support for FAIR data before, during and after research through its involvement in leading practices, training and consultancy and end with recommendations for other universities wanting to implement the FAIR principles.
This paper examines the intersection of legacy digital humanities projects and the ongoing development of research data management services at Vanderbilt University's Jean and Alexander Heard Library. Future directions for data management and curation protocols are explored through the lens of a case study: the (re)curation of data from an early 2000s e-edition of Raymond Poggenburg's Charles Baudelaire: Une Micro-histoire. The vagaries of applying the Library of Congress Metadata Object Description Schema (MODS) to the data and metadata of the Micro-histoire will be addressed. In addition, the balance between curating data and metadata for preservation vs. curating it for (re)use by future researchers is considered in order to suggest future avenues for holistic research data management services at Vanderbilt.


Canada's federal funding agencies are following the directions of funding agencies in the United States and United Kingdom, and will soon require a data management plan in grant applications. The University of Manitoba Libraries in Canada has started planning and implementing research data services, and education is seen as a key component. In June 2014, the New England Collaborative Data Management Curriculum (NECDMC) (Lamar Soutter Library, University of Massachusetts Medical School 2014) was piloted and used to provide data management training for a group of subject librarians at the University of Manitoba Libraries, in combination with information about data-related policies of the Canadian funding agencies and the University of Manitoba. The seven NECDMC modules were delivered in a seminar style, with emphasis on group discussions and Canadian content. The benefits of NECDMC—adaptability and flexible framework—should be weighed against the challenges experienced in the pilot, mainly the significant amount of time needed to create local content and complement the existing curriculum. Overall, the pilot showed that NECDMC is a good, thorough introduction to data management, and that it is possible to adapt NECDMC to the local and Canadian settings in an effective way.


BACKGROUND Many libraries have launched or adapted services to address the research data needs of campus faculty and students. At the University of Colorado Boulder (CU-Boulder), local demand for research data training emerged from a broader assessment of training needs for subject librarians. The findings from this assessment led to the development of a day-long workshop called DataDay!
aimed to expand and translate the skills of subject librarians into the context of research data support. DESCRIPTION OF PROGRAM The DataDay! workshop incorporated hands-on exercises with expert presentations, informal discussions, and print handouts. The workshop allowed participants to gain experience with activities like working with real data sets and developing materials for outreach about research data services. Several instruments were used to assess the workshop learning outcomes, which included changes in knowledge and comfort levels related to engaging in research data support. Assessment activities also measured how well participants applied concepts taught in the workshop to novel situations. NEXT STEPS Future research data training efforts for CU-Boulder librarians will be informed by the DataDay! workshop assessment results, and this workshop may provide a model for other institutions to use to train subject librarians to adapt to new roles in support of research data. There is also a need for the lessons learned from local training efforts like DataDay! to inform the development of resources to support the broader subject librarian community as their institutions launch and grow research data services.


Objective: To examine the effects of research data services (RDS) on the quality of data management plans (DMPs) required for a campus-level faculty grant competition, as well as to explore opportunities that the local DMP requirement presented for RDS outreach.

Methods: Nine reviewers each scored a randomly assigned portion of DMPs from 82 competition proposals. Each DMP was scored by three reviewers, and the three scores were averaged together to obtain the final score. Interrater reliability was measured using intraclass correlation. Unpaired t-tests were used to compare mean DMP scores for faculty who utilized RDS services with those who did not. Unpaired t-tests were also used to compare mean DMP scores for proposals that were funded with proposals that were not funded. One-way ANOVA was used to compare mean DMP scores among proposals from six broad disciplinary categories.

Results: Analyses showed that RDS consultations had a statistically significant effect on DMP scores. Differences between DMP scores for funded versus unfunded proposals and among disciplinary categories were not significant. The DMP requirement also provided a number of both expected and unexpected outreach opportunities for RDS services.

Conclusions: Requiring DMPs for campus grant competitions can provide important assessment and outreach opportunities for research data services. While these results might not be generalizable to DMP review processes at federal funding agencies, they do suggest the importance, at any level, of developing a shared understanding of what constitutes a high quality DMP among grant applicants, grant reviewers, and RDS providers.


Objective: This paper aims to inform on opportunities for librarians to assist faculty with research data management by examining practices and attitudes among life sciences faculty at a tier one research university.

Methods: The authors issued a survey to estimate actual and perceived research data management needs of New York University (NYU) life sciences faculty in order
to understand how the library could best contribute to the research life cycle.

Results: Survey responses indicate that over half of the respondents were aware of publisher and funder mandates, and most are willing to share their data, but many indicated they do not utilize data repositories. Respondents were largely unaware of data services available through the library, but the majority were open to considering such services. Survey results largely mimic those of similar studies, in that storing data (and the subsequent ability to share it) is the most easily recognized barrier to sound data management practices.

Conclusions: At NYU, as with other institutions, the library is not immediately recognized as a valuable partner in managing research output. This study suggests that faculty are largely unaware of, but are open to, existent library services, indicating that immediate outreach efforts should be aimed at promoting them.


Objective:

This study investigates research data management (RDM) services using a crosstab framework with the National Institutes of Health (NIH) Library as a case study to provide practical considerations for libraries seeking to improve their RDM services.

Methods:

We conducted semistructured interviews with four librarians who provide data services at the NIH Library regarding library user characteristics, RDM services provided, RDM infrastructure, and collaboration experiences. Through the analysis of interview transcripts, we identified and analyzed the NIH Library's RDM services according to Online Computer Library Center (OCLC)'s three categories of RDM services and the six stages of the data lifecycle.

Results:

The findings show that the two models' crosstab framework can provide an overview of an institution's current RDM services and identify service gaps. The NIH Library tends to take more responsibility in providing education and expertise services while relying more on information technology departments for curation services. The library provides significant support for data creation, analysis, and sharing stages to meet biomedical researchers' needs, suggesting areas for potential expansion of RDM services in the less supported stages of data description, storage, and preservation. Based on these findings, we recommend three key considerations for libraries: identify gaps in current services, identify services that can be supported via partnerships, and get regular feedback from users.

Conclusion:

These findings provide a deeper understanding of RDM support on the basis of RDM service categories and the data lifecycle and promote discussion of issues to be considered for future improvements in RDM services.

Objective: Evaluate and examine Data Literacy (DL) in the supported disciplines of four liaison librarians at a large research university.

Methods: Using a framework developed by Prado and Marzal (2013), the study analyzed 378 syllabi from a two-year period across six departments—Criminal Justice, Geography, Geology, Journalism, Political Science, and Sociology—to see which classes included DLs.

Results: The study was able to determine which classes hit on specific DLs and where those classes might need more support in other DLs. The most common DLs being taught in courses are Reading, Interpreting, and Evaluating Data, and Using Data. The least commonly taught are Understanding Data and Managing Data skills.

Conclusions: While all disciplines touched on data in some way, there is clear room for librarians to support DLs in the areas of Understanding Data and Managing Data.


The amount of research data is growing constantly, due to new technology with new potentials for collecting and analysing both digital data and research objects. This growth creates a demand for a coherent IT-infrastructure. Such an infrastructure must be able to provide facilities for storage, preservation and a more open access to data in order to fulfil the demands from the researchers themselves, the research councils and research foundations.

This paper presents the findings of a research project carried out under the auspices of DEFF (Danmarks Elektroniske Fag-og Forskningsbibliotek—Denmark's Electronic research Library) to analyse how the Danish universities store, preserve and provide access to research data. It shows that they do not have a common IT-infrastructure for research data management. This paper describes the various paths chosen by individual universities and research institutions, and the background for their strategies of research data management. Among the main reasons for the uneven practices are the lack of a national policy in this field, the different scientific traditions and cultures and the differences in the use and organization of IT-services.

This development contains several perspectives that are of particular relevance to research libraries. As they already curate digital collections and are active in establishing web archives, the research libraries become involved in research and dissemination of knowledge in new ways. This paper gives examples of how The State and University Library's services facilitate research data management with special regard to digitization of research objects, storage, preservation and sharing of research data. This paper concludes that the experience and skills of research libraries make the libraries important partners in a research data management infrastructure.


Librarians champion the value of openness in scholarship and have been powerful advocates for the sharing of research data. College and university administrators have recently joined in the push for data sharing due to funding mandates. However, the researchers who create and control the data usually determine whether and how data is shared, so it is worthwhile to look at what they are incentivized to do. The current scholarly publishing landscape plus the promotion and tenure process create a "prisoner's dilemma" for researchers as they decide whether or not to share data, consistent with the observation that researchers in general are eager for others to share data but reluctant to do so themselves. If librarians encourage researchers to share data and promote openness without simultaneously addressing the academic incentive structure, those who are intrinsically motivated to share data will be selected against via the promotion and tenure process. This will cause those who are hostile to sharing to be disproportionately recruited into the senior ranks of academia. To mitigate the risk of this unintended consequence, librarians must advocate for a change in incentives alongside the call for greater openness. Highly-cited datasets must be given similar weight to highly-cited articles in promotion and tenure decisions in order for researchers to reap the rewards of their sharing. Librarians can help by facilitating data citation to track the impact of datasets and working to persuade higher administration of the value of rewarding data sharing in tenure and promotion.


The importance of managing research data has been emphasized by the government, funding agencies, and scholarly communities. Increased access to research data increases the impact and efficiency of scientific activities and funding. Thus, many research institutions have established or plan to establish research data curation services as part of their Institutional Repositories (IRs). However, in order to design effective research data curation services in IRs, and to build active research data providers and user communities around those IRs, it is essential to study current data curation practices and provide rich descriptions of the sociotechnical factors and relationships shaping those practices. Based on 13 interviews with 15 IR staff members from 13 large research universities in the United States, this paper provides a rich, qualitative description of research data curation and use practices in IRs. In particular, the paper identifies data curation and use activities in IRs, as well as their structures, roles played, skills needed, contradictions and problems present, solutions sought, and workarounds applied. The paper can inform the development of best practice guides, infrastructure and service templates, as well as education in research data curation in Library and Information Science (LIS) schools.

Adoption of good research data management practices is increasingly important for research teams. Despite the work the research community has done to define best data management practices, these practices are still difficult to adopt for many research teams. Universities all around the world have been offering Research Data Services to help their research groups, and libraries are usually an important part of these services. A better understanding of the pressures and factors that affect research teams may help librarians serve these groups more effectively. The social interactions between the members of a research team are a key element that influences the likelihood of a research group successfully adopting best practices in data management. In this article we adapt the Unified Theory of the Acceptance and Use of Technology (UTAUT) model (Venkatesh, Morris, Davis, & Davis, 2003) to explain the variables that can influence whether new and better, data management practices will be adopted by a research group. We describe six moderating variables: size of the team, disciplinary culture, group culture and leadership, team heterogeneity, funder, and dataset decisions. We also develop three research group personas as a way of navigating the UTAUT model, and as a tool Research Data
Services practitioners can use to target interactions between librarians and research groups to make them more effective.

This work is licensed under a Creative Commons Attribution 3.0 Unported License. https://creativecommons.org/licenses/by/3.0/.

Llebot, Clara, and Steven Van Tuyl. "Peer Review of Research Data Submissions to ScholarsArchive@OSU: How Can We Improve the Curation of Research Datasets to Enhance Reusability?" *Journal of eScience Librarianship* 8, no. 2 (2019): https://doi.org/10.7191/jeslib.2019.1166

Objective: Best practices such as the FAIR Principles (Findability, Accessibility, Interoperability, Reusability) were developed to ensure that published datasets are reusable. While we employ best practices in the curation of datasets, we want to learn how domain experts view the reusability of datasets in our institutional repository, ScholarsArchive@OSU. Curation workflows are designed by data curators based on their own recommendations, but research data is extremely specialized, and such workflows are rarely evaluated by researchers. In this project we used peer-review by domain experts to evaluate the reusability of the datasets in our institutional repository, with the goal of informing our curation methods and ensure that the limited resources of our library are maximizing the reusability of research data.

Methods: We asked all researchers who have datasets submitted in Oregon State University's repository to refer us to domain experts who could review the reusability of their data sets. Two data curators who are non-experts also reviewed the same datasets. We gave both groups review guidelines based on the guidelines of several journals. Eleven domain experts and two data curators reviewed eight datasets. The review included the quality of the repository record, the quality of the documentation, and the quality of the data. We then compared the comments given by the two groups.

Results: Domain experts and non-expert data curators largely converged on similar scores for reviewed datasets, but the focus of critique by domain experts was somewhat divergent. A few broad issues common across reviews were: insufficient documentation, the use of links to journal articles in the place of documentation, and concerns about duplication of effort in creating documentation and metadata. Reviews also reflected the background and skills of the reviewer. Domain experts expressed a lack of expertise in data curation practices and data curators expressed their lack of expertise in the research domain.

Conclusions: The results of this investigation could help guide future research data curation activities and align domain expert and data curator expectations for reusability of datasets. We recommend further exploration of these common issues and additional domain expert peer-review project to further refine and align expectations for research data reusability.


Objective: The Illinois Data Bank provides Illinois researchers with the infrastructure to publish research data publicly. During a five-year review of the Research Data Service at the University of Illinois at Urbana-Champaign, it was recognized as the most useful service offering in the unit. Internal metrics are captured and used to monitor the growth, document curation workflows, and surface technical challenges faced as we assist our researchers. Here we present examples of these curation challenges and the solutions chosen to address them.
Methods: Some Illinois Data Bank metrics are collected internally by within the system, but most of the curation metrics reported here are tracked separately in a Google spreadsheet. The curator logs required information after curation is complete for each dataset. While the data is sometimes ambiguous (e.g., depending on researcher uptake of suggested actions), our curation data provide a general understanding about our data repository and have been useful in assessing our workflows and services. These metrics also help prioritize development needs for the Illinois Data Bank.

Results and Conclusions: The curatorial services polish and improve the datasets, which contributes to the spirit of data reuse. Although we continue to see challenges in our processes, curation makes a positive impact on datasets. Continued development and adaptation of the technical infrastructure allows for an ever-better experience for the curators and users. These improvements have helped our repository more effectively support the data sharing process by successfully fostering depositor engagement with curators to improve datasets and facilitating easy transfer of very large files.

https://doi.org/10.2218/ijdc.v7i1.220

http://dx.doi.org/10.1080/13614533.2016.1159969

https://digitalcommons.unl.edu/libphilprac/6760

https://doi.org/10.1016/j.acalib.2014.06.011

http://doi.org/10.18352/lq.9868

In order to align information literacy instruction with changing faculty and student needs, librarians must expand their skills and competencies beyond traditional information sources. In the sciences, this increasingly means integrating the data resources used by researchers into instruction for undergraduate students. Open access data repositories allow students to work with more primary data than ever before, but only if they know how and where to look. This paper will describe the development of two information literacy workshops designed to scaffold student learning in the biological sciences across two second-year courses, detailing the long-term collaboration between a librarian and an instructor that now serves over 500 students per semester. In each workshop, students are guided through the discovery and analysis of life sciences data from multiple sites, encouraged to integrate text and data sources, and supported in completing research assignments.


https://doi.org/10.7191/jeslib.2018.1155
Data Management Plans (DMPs) are often required for grant applications. But do strong DMPs lead to better data management and sharing practices? Several recent research projects in the Library and Information Science field have investigated data management planning and practice through DMP content analysis and data-management-related interviews. However, research hasn't yet shown how DMPs ultimately affect data management and data sharing practices during grant-funded research. The research described in this article contributes to the existing literature by examining the impact of DMPs on grant awards and on Principal Investigators' (PIs) data management and sharing practices. The results of this research suggest the following key takeaways: (1) Most PIs practice internal data management in order to prevent data loss, to facilitate sharing within the research team, and to seamlessly continue their research during personnel turnover; (2) PIs still have room to grow in understanding specialized concepts such as metadata and policies for use and reuse; (3) PIs may need guidance on practices that facilitate FAIR data, such as using metadata standards, assigning licenses to their data, and publishing in data repositories. Ultimately, the results of this research can inform academic library services and support stronger, more actionable DMPs.


Objective: This article analyzes twenty cited or downloaded datasets and the repositories that house them, in order to produce insights that can be used by academic libraries to encourage discovery and reuse of research data in institutional repositories.

Methods: Using Thomson Reuters' Data Citation Index and repository download statistics, we identified twenty cited/downloaded datasets. We documented the characteristics of the cited/downloaded datasets and their corresponding repositories in a self-designed rubric. The rubric includes six major categories: basic information; funding agency and journal information; linking and sharing; factors to encourage reuse; repository characteristics; and data description.

Results: Our small-scale study suggests that cited/downloaded datasets generally comply with basic recommendations for facilitating reuse: data are documented well; formatted for use with a variety of software; and shared in established, open access repositories. Three significant factors also appear to contribute to dataset discovery: publishing in discipline-specific repositories; indexing in more than one location on the web; and using persistent identifiers. The cited/downloaded datasets in our analysis came from a few specific disciplines, and tended to be funded by agencies with data publication mandates.

Conclusions: The results of this exploratory research provide insights that can inform academic librarians as they work to encourage discovery and reuse of institutional datasets. Our analysis also suggests areas in which academic librarians can target open data advocacy in their communities in order to begin to build open data success stories that will fuel future advocacy efforts.


This paper reports on the design, delivery and assessment of a model for internal library education around research data management (RDM). Conducted at the University of Pittsburgh Library System (ULS), the exercise and resultant instructional session employed an active learning approach, in which a group of librarians and archivists explored data issues and conventions in a discipline of their own selection and presented their findings to an audience of library colleagues. In this paper, we put forth an adaptable active learning model for internal RDM education and offer guidance for its implementation by peer libraries that are similarly building internal capacity for the design and delivery of RDM services that are responsive to disciplinary needs.


This paper is a review of the status of Research Data Management (RDM) efforts in UAE public university libraries. Approach: The investigation is through examining available literature about the topic using region-specific articles when available, librarian comments, and the information provided by UAE university library websites. Existing lessons and policy documents are sought, and plans suggested for local solutions, suggesting avenues for progress. Findings: Though not a new concept, findings indicate that local RDM activities are emerging, but knowledge of their importance of exists. Research limitations: This review is limited to public university libraries though the results and experiences could be generally relevant to more research establishments. Practical implications: Taking advantage of the existing awareness to organize tangible RDM efforts can facilitate retrieval and availability of data relevant to the region. Value: An intricate range of activities involved in the organization of RDM services is revealed.


Milewska, A., Natalia Wiśniewska, Paulina Cimoszko, and Jakub Rusakow. "A Survey of Medical Researchers Indicates Poor Awareness of Research Data Management
Research Data Management (RDM) services are increasingly becoming a subject of interest for academic and research libraries globally—this is also the case in developing countries. The interest is motivated by a need to support research activities through data sharing and collaboration both locally and internationally. Many institutions, especially in the developed countries, have implemented RDM services.
to accelerate research and innovation through e-Research but extensive RDM is not so common in developing countries. In reality many African universities and research institutions are yet to implement the most basic of data management services. We believe that the absence of political will and national government mandates on data management often hold back the development and implementation of RDM services. Similarly, research funding agencies are not yet applying sufficient pressure to ensure that Africa complies with the requirement to deposit research data in trusted repositories. While the context was acknowledged the University of Dodoma library staff realized that it is urgent to prepare for the inevitable—the time when RDM will be a requirement for research funding support.

This paper presents the results of research conducted at the University of Dodoma, Tanzania. The purpose of the research was to identify and report on relevant RDM services that need to be implemented so that researchers and university management could collaborate and make our research data accessible to the international community.

This paper presents findings on important issues for consideration when planning to develop and implement RDM services at a developing country academic institution. The paper also mentions the requirements for the sustainability of these initiatives.


Recent years have seen a growing emphasis on the need for improved management of research data. Academic libraries have begun to articulate the conceptual foundations, roles, and responsibilities involved in data management planning and implementation. This paper provides an overview of the Engineering data support pilot at the University of Michigan Library as part of developing new data services and infrastructure. Through this pilot project, a team of librarians had an opportunity to identify areas where the library can play a role in assisting researchers with data management, and has put forth proposals for immediate steps that the library can take in this regard. The paper summarizes key findings from a faculty survey and discusses lessons learned from an analysis of data management plans from accepted NSF proposals. A key feature of this Engineering pilot project was to ensure that these study results will provide a foundation for librarians to educate and assist researchers with managing their data throughout the research lifecycle.


Research Data Management (RDM) has become increasingly important for more and more academic institutions. Using the Peking University Open Research Data Repository (PKU-ORDR) project as an example, this paper will review a library-based university-wide open research data repository project and related RDM services implementation process including project kickoff, needs assessment, partnerships establishment, software investigation and selection, software customization, as well as data curation services and training. Through the review, some issues revealed during the stages of the implementation process are also discussed and addressed in the paper such as awareness of research data, demands from data providers and users, data policies and requirements from home institution, requirements from funding agencies and publishers, the collaboration between administrative units and libraries, and concerns from data providers and users. The significance of the study is that the paper shows an example of creating an Open Data repository and RDM services for other Chinese academic libraries planning to implement their RDM services for their home institutions. The authors of the paper have also observed since the PKU-ORDR and RDM services implemented in 2015, the Peking University Library (PKUL) has helped numerous researchers to support the entire research life cycle and enhanced Open Science (OS) practices on campus, as well as impacted the national OS movement in China through various national events and activities hosted by the PKUL.


This paper explores three stories, each occurring a year apart, illustrating an evolution toward a strategic vision for Library leadership in supporting research data management at the University of Sydney. The three stories describe activities undertaken throughout the Seeding the Commons project and beyond, as the establishment of ongoing roles and responsibilities transition the Library from project partner to strategic leader in the delivery of research data management support. Each story exposes key ingredients that characterise research data management support: researcher engagement; partnerships; and the complementary roles of policy and practice.


Objective and Setting: As universities and libraries grapple with data management and "big data," the need for data management solutions across disciplines is particularly relevant in clinical and translational science (CTS) research, which is
designed to traverse disciplinary and institutional boundaries. At the University of Florida Health Science Center Library, a team of librarians undertook an assessment of the research data management needs of CTS researchers, including an online assessment and follow-up one-on-one interviews.

Design and Methods: The 20-question online assessment was distributed to all investigators affiliated with UF's Clinical and Translational Science Institute (CTSI) and 59 investigators responded. Follow-up in-depth interviews were conducted with nine faculty and staff members.

Results: Results indicate that UF's CTS researchers have diverse data management needs that are often specific to their discipline or current research project and span the data lifecycle. A common theme in responses was the need for consistent data management training, particularly for graduate students; this led to localized training within the Health Science Center and CTSI, as well as campus-wide training. Another campus-wide outcome was the creation of an action-oriented Data Management/Curation Task Force, led by the libraries and with participation from Research Computing and the Office of Research.

Conclusions: Initiating conversations with affected stakeholders and campus leadership about best practices in data management and implications for institutional policy shows the library's proactive leadership and furthers our goal to provide concrete guidance to our users in this area.


Objective: The purpose of this article is to explore data visualization as a consulting service offered by a research library with particular attention to uses of visualization at various places within the research lifecycle.

Methods: Lessons learned from a year of offering data visualization as a consulting service, and two general case studies are offered.

Results: Data visualization consulting services have a few unique considerations, including setting clear expectations, considering proprietary vs open source technologies, and making sure the consulting experience is also a learning experience. In addition, we can clearly place data visualization requests, in the form of profiled case studies, in multiple parts of the research lifecycle.


Orlowska, Daria, Colleen Fallaw, Yali Feng, Livia Garza, Ashley Hetrick, Heidi Imker, and Hoa Luong. "Better Data Management, One Nudge at a Time." IASSIST Quarterly 45, no.

Objectives: In 2021, a new national center was funded by the National Library of Medicine (NLM) to build capacity for providing data services within health sciences libraries by coordinating, developing, and disseminating trainings, curricular resources, and curated pathways for learning. Research data services has been a growing service component for health sciences libraries for over a decade, and efforts have come from individuals, professional societies, task forces, and interest groups; however, there is still a great deal of unrealized potential in this area as well as growing needs driven by new requirements from funders and publishers, and increasing demand from institutions for data science skills and support for data-related research needs.

Methods: The National Center for Data Services (NCDS) was established in July 2021 and is the newest of the Network of the National Library of Medicine (NNLM)'s national offices and centers. NCDS works with the NNLM's seven regional medical libraries and national offices and centers, and draws upon extensive connections with national data and library organizations, all toward the aim of developing capacity for data services in the health information community while centering efforts toward equity, diversity, and accessibility throughout.


Objective: This paper describes a project to revise an existing research data management (RDM) course to include instruction in computer skills with robust data science tools.

Setting: A Carnegie R1 university.

Brief Description: Graduate student researchers need training in the basic concepts of RDM. However, they generally lack experience with robust data science tools to implement these concepts holistically. Two library instructors fundamentally redesigned an existing research RDM course to include instruction with such tools.
The course was divided into lecture and lab sections to facilitate the increased instructional burden. Learning objectives and assessments were designed at a higher order to allow students to demonstrate that they not only understood course concepts but could use their computer skills to implement these concepts.

Results: Twelve students completed the first iteration of the course. Feedback from these students was very positive, and they appreciated the combination of theoretical concepts, computer skills and hands-on activities. Based on student feedback, future iterations of the course will include more "flipped" content including video lectures and interactive computer tutorials to maximize active learning time in both lecture and lab.


The management of research data is now a major challenge for research organisations. Vast quantities of born-digital data are being produced in a wide variety of forms at a rapid rate in universities. This paper analyses the contribution of academic libraries to research data management (RDM) in the wider institutional context. In particular it: examines the roles and relationships involved in RDM, identifies the main components of an RDM programme, evaluates the major drivers for RDM activities, and analyses the key factors influencing the shape of RDM developments. The study is written from the perspective of library professionals, analysing data from 26 semi-structured interviews of library staff from different UK institutions. This is an early qualitative contribution to the topic complementing existing quantitative and case study approaches. Results show that although libraries are playing a significant role in RDM, there is uncertainty and variation in the relationship with other stakeholders such as IT services and research support offices. Current emphases in RDM programmes are on developments of policies and guidelines, with some early work on technology infrastructures and support services. Drivers for developments include storage, security, quality, compliance, preservation, and sharing with libraries associated most closely with the last three. The paper also highlights a 'jurisdictional' driver in which libraries are claiming a role in this space. A wide range of factors, including governance, resourcing and skills, are identified as influencing ongoing developments. From the analysis, a model is constructed designed to capture the main aspects of an institutional RDM programme. This model helps to clarify the different issues involved in RDM, identifying layers of activity, multiple stakeholders and drivers, and a large number of factors influencing the implementation of any initiative. Institutions may usefully benchmark their activities against the data and model in order to inform ongoing RDM activity.

In the last years, the scientific community and funding bodies have paid attention to collected, generated or used data throughout different research activities. The dissemination of these data becomes one of the constituent elements of Open Science. For this reason, many funders are requiring or promoting the development of Data Management Plans, and depositing open data following the FAIR principles (Findable, Accessible, Interoperable and Reusable). Libraries and research offices of Catalan universities—which coordinately work within the Open Science Area of CSUC—offer support services to research data management. The different works carried out at the Consortium level will be presented, as well the implementation of the service in each university.

Objectives: This study sought to understand the needs of biomedical researchers related to the research lifecycle and the present state of library support for biomedical research.

Methods: Qualitative interview data were collected from biomedical researchers who were asked to describe their research activities from identifying a problem to measuring the impact of their findings. Health sciences library leaders were surveyed about the services that they currently provide or plan to provide in supporting biomedical research.

Results: Library services were strongest at the beginning and end of the research lifecycle but were weaker in the conducting phase of research. Co-occurrence of codes from the qualitative data suggests that library services are on the fringe of rather than integrated into the research lifecycle.

Discussion: Findings from this study suggest that tradition-based service models of health sciences libraries are insufficient to meet the needs of biomedical researchers. Investments by libraries in services that integrate with the conducting phase of research are needed for libraries to remain relevant in their support of the research lifecycle.

Key themes in Dickens' novel, transformation and resurrection, darkness and light, and social justice are firmly connected to the work being done in data. Data librarians can make a difference in times like these: resurrecting data, transforming how students, researchers, or the public think about and use data; unearthing and bringing to light historical data that will give context and meaning to an ; and that accessible data can help address, and perhaps solve, social justice issues.
Background

Librarians and researchers alike have long identified research data management (RDM) training as a need in biomedical research. Despite the wealth of libraries offering RDM education to their communities, clinical research is an area that has not been targeted. Clinical RDM (CRDM) is seen by its community as an essential part of the research process where established guidelines exist, yet educational initiatives in this area are unknown.

Case Presentation

Leveraging my academic library's experience supporting CRDM through informationist grants and REDCap training in our medical center, I developed a 1.5 hour CRDM workshop. This workshop was designed to use established CRDM guidelines in clinical research and address common questions asked by our community through the library's existing data support program. The workshop was offered to the entire medical center 4 times between November 2017 and July 2018. This case study describes the development, implementation, and evaluation of this workshop.

Conclusions

The 4 workshops were well attended and well received by the medical center community, with 99% stating that they would recommend the class to others and 98% stating that they would use what they learned in their work. Attendees also articulated how they would implement the main competencies they learned from the workshop into their work. For the library, the effort to support CRDM has led to the coordination of a larger institutional collaborative training series to educate researchers on best practices with data, as well as the formation of institution-wide policy groups to address researcher challenges with CRDM, data transfer, and data sharing.


Background: Librarians developed a pilot program to provide training, resources, strategies, and support for medical libraries seeking to establish research data management (RDM) services. Participants were required to complete eight educational modules to provide the necessary background in RDM. Each participating institution was then required to use two of the following three elements: (1) a template and strategies for data interviews, (2) the Teaching Toolkit to teach an introductory RDM class, or (3) strategies for hosting a data class series.
Case Presentation: Six libraries participated in the pilot, with between two and eight librarians participating from each institution. Librarians from each institution completed the online training modules. Each institution conducted between six and fifteen data interviews, which helped build connections with researchers, and taught between one and five introductory RDM classes. All classes received very positive evaluations from attendees. Two libraries conducted a data series, with one bringing in instructors from outside the library.

Conclusion: The pilot program proved successful in helping participating librarians learn about and engage with their research communities, jump-start their teaching of RDM, and develop institutional partnerships around RDM services. The practical, hands-on approach of this pilot proved to be successful in helping libraries with different environments establish RDM services. The success of this pilot provides a proven path forward for libraries that are developing data services at their own institutions.


Objective: To increase data quality and ensure compliance with appropriate policies, many institutional data repositories curate data that is deposited into their systems. Here, we present our experience as an academic library implementing and managing a semi-automated, cloud-based data curation workflow for a recently launched institutional data repository. Based on our experiences we then present management observations intended for data repository managers and technical staff looking to move some or all of their curation services to the cloud.

Methods: We implemented tooling for our curation workflow in a service-oriented manner, making significant use of our data repository platform's application programming interface (API). With an eye towards sustainability, a guiding development philosophy has been to automate processes following industry best practices while avoiding solutions with high resource needs (e.g., maintenance), and minimizing the risk of becoming locked-in to specific tooling.

Results: The initial barrier for implementing a data curation workflow in the cloud was high in comparison to on-premises curation, mainly due to the need to develop in-
house cloud expertise. However, compared to the cost for on-premises servers and storage, infrastructure costs have been substantially lower. Furthermore, in our particular case, once the foundation had been established, a cloud approach resulted in increased agility allowing us to quickly automate our workflow as needed.

Conclusions: Workflow automation has put us on a path toward scaling the service and a cloud based-approach has helped with reduced initial costs. However, because cloud-based workflows and automation come with a maintenance overhead, it is important to build tooling that follows software development best practices and can be decoupled from curation workflows to avoid lock-in.


Objective: This eScience in Action article describes the collaborative development process and outputs for a qualitative data curation curriculum initiative led by a library faculty (research data specialist) at an R1 research university.

Methods: The collaborative curriculum development activities described in this article took place between 2015-2020 and included 1) a college-wide "call out" meeting with graduate methods instructors and additional one-on-one conversations, 2) a year-long training series for disciplinary faculty teaching graduate-level qualitative research methods courses, 3) guest lectures and co-curricular workshops, and 4) the development of a credit-bearing graduate-level course.

Results: This practice-based article includes a reflection on the collaborative curriculum development process and impacts, including the development of networks between the Library and qualitative researchers across campus. The article provides a proof-of-concept example for developing relevant and trustworthy library data services for humanities and qualitative social-science researchers.

Conclusions: Curriculum development activities focused predominately upon researcher-centered perspectives and identified needs. However, changes in institutional expectations for library faculty (i.e. requirement to teach credit-bearing courses) played a major role in how the curriculum was implemented, its impact and continued sustainability of outputs going forward.


Objective: This article examines a reshaped service model for research data management (RDM) founded on centralized and cohesive collaboration between multiple stakeholders at a large research university in Canada. This initiative, along with a newly formed team dedicated to RDM service provision, is a joint effort by the institution's Vice-Principal Research and Innovation (VPRI), Library, IT Services, and Research Ethics units.

Methods: This article presents a single case study methodology. The authors reflect on services such as "query the panel" sessions where researchers across all disciplines bring their questions to representatives from the Library, IT, Research Ethics, and VPRI. This case study also highlights the use of Jira's service desk software as a user management system. The authors also present descriptive statistics representing engagement with this new unit and our services.

Results: Support for RDM requires expertise from multiple domains. With a collaborative approach as a guiding principle and a focus on establishing a small, but agile team comprised of a librarian along with stakeholders from IT and VPRI, it is
possible to leverage resources and support for RDM from a broad range of units across an institution.

Conclusions: At many institutions, RDM services are siloed within the library or an adjacent campus unit. New digital technologies have profoundly transformed academic research across all disciplines, necessitating the evolution of corresponding research data-related services. The authors will conclude by outlining specific lessons learned in reshaping digital research infrastructure-related services at their institution.


Objectives: To support the library’s research data services program and provide opportunities for student employees to gain skills in supporting data science research and teaching, the author led the creation of a data fellowship program at their institution. The author describes the development of the program’s peer-to-peer support model and how that model was integrated into the library's broader research data services program.

Methods: There were two aspects to the fellowship's peer-to-peer support model. First, the data fellows provided direct support to their peers—through instruction, reference support, and outreach—who were engaged in data-related research projects and class assignments. Second, the fellows engaged in professional development and learning opportunities led by library colleagues.

Results: During the first year of the program, the data fellows provided virtual data reference support, developed two CANVAS modules, conducted an email outreach campaign, and authored several blog posts. Furthermore, the data fellows received data reference training and engaged in discussions with library experts on pertinent data-related topics. These discussions provide the fellows with knowledge and skills that can be utilized both during and after the fellowship.

Conclusions: The first year of the STEM Data Fellowship program proved to be successful in establishing the foundation for a robust peer-to-peer learning and support model for research data services at the author's institution. The author plans to further enhance and grow the program in future years by providing more opportunities for established and new data fellows to contribute to the library's ongoing RDS efforts.


This article reports on the transfer of a massive scientific dataset from a national laboratory to a university library, and from one kind of workforce to another. We use the transfer of the Sloan Digital Sky Survey (SDSS) archive to examine the emergence of a new workforce for scientific research data management. Many individuals with diverse educational backgrounds and domain experience are involved in SDSS data management: domain scientists, computer scientists, software and systems engineers, programmers, and librarians. These types of positions have been described using terms such as research technologist, data scientist, e-science professional, data curator, and more. The findings reported here are based on semi-
structured interviews, ethnographic participant observation, and archival studies from 2011-2013.

The library staff conducting the data storage and archiving of the SDSS archive faced two performance problems. The preservation specialist and the system administrator worked together closely to discover and implement solutions to the slow data transfer and verification processes. The team overcame these slow-downs by problem solving, working in a team, and writing code. The library team lacked the astronomy domain knowledge necessary to meet some of their preservation and curation goals.

The case study reveals the variety of expertise, experience, and individuals essential to the SDSS data management process. A variety of backgrounds and educational histories emerge in the data managers studied. Teamwork is necessary to bring disparate expertise together, especially between those with technical and domain education. The findings have implications for data management education, policy and relevant stakeholders.


In this paper, we discuss the various stages of the institution-wide project that lead to the adoption of the data management policy at Leiden University in 2016. We illustrate this process by highlighting how we have involved all stakeholders. Each organisational unit was represented in the project teams. Results were discussed in a sounding board with both academic and support staff. Senior researchers acted as pioneers and raised awareness and commitment among their peers. By way of example, we present pilot projects from two faculties. We then describe the comprehensive implementation programme that will create facilities and services that must allow implementing the policy as well as monitoring and evaluating it. Finally, we will present lessons learnt and steps ahead. The engagement of all stakeholders, as well as explicit commitment from the Executive Board, has been an important key factor for the success of the project and will continue to be an important condition for the steps ahead.


This paper presents the results of a research data assessment and landscape study in the institutional context of Virginia Tech to determine the data sharing and reuse practices of academic faculty researchers. Through mapping the level of user engagement in "openness of data," "openness of methodologies and workflows," and "reuse of existing data," this study contributes to the current knowledge in data sharing and open access, and supports the strategic development of institutional data stewardship. Asking faculty researchers to self-reflect sharing and reuse from both data producers' and data users' perspectives, the study reveals a significant gap between the rather limited sharing activities and the highly perceived reuse or repurpose values regarding data, indicating that potential values of data for future research are lost right after the original work is done. The localized and sporadic data management and documentation practices of researchers also contribute to the obstacles they themselves often encounter when reusing existing data.


Institutional data repositories are the acknowledged gold standard for data curation platforms in academic libraries. But not every institution can sustain a repository, and not every dataset can be archived due to legal, ethical, or authorial constraints. Data catalogs—metadata-only indices of research data that provide detailed access instructions and conditions for use—are one potential solution, and may be especially suitable for "challenging" datasets. This article presents the strengths of data catalogs for increasing the discoverability and accessibility of research data. The authors argue that data catalogs are a viable alternative or complement to data repositories,


Scholarly researchers today are increasingly required to engage in a range of data management planning activities to comply with institutional policies, or as a precondition for publication or grant funding. The latter is especially true in the U.S. in light of the recent White House Office of Science and Technology Policy (OSTP).
mandate aimed at maximizing the availability of all outputs—data as well as the publications that summarize them—resulting from federally-funded research projects.

To aid researchers in creating effective data management plans (DMPs), a group of organizations—California Digital Library, DataONE, Digital Curation Centre, Smithsonian Institution, University of Illinois Urbana-Champaign, and University of Virginia Library—collaborated on the development of the DMPTool, an online application that helps researchers create data management plans. The DMPTool provides detailed guidance, links to general and institutional resources, and walks a researcher through the process of generating a comprehensive plan tailored to specific DMP requirements. The uptake of the DMPTool has been positive: to date, it has been used by over 6,000 researchers from 800 institutions, making use of more than 20 requirements templates customized for funding bodies.

With support from the Alfred P. Sloan Foundation, project partners are now engaged in enhancing the features of the DMPTool. The second version of the tool has enhanced functionality for plan creators and institutional administrators, as well as a redesigned user interface and an open RESTful application programming interface (API).

New administrative functions provide the means for institutions to better support local research activities. New capabilities include support for plan co-ownership; workflow provisions for internal plan review; simplified maintenance and addition of DMP requirements templates; extensive capabilities for the customization of guidance and resources by local institutional administrators; options for plan visibility; and UI refinements based on user feedback and focus group testing. The technical work undertaken for the DMPTool Version 2 has been accompanied by a new governance structure and the growth of a community of engaged stakeholders who will form the basis for a sustainable path forward for the DMPTool as it continues to play an important role in research data management activities.


Research data is an essential part of the scholarly record, and management of research data is increasingly seen as an important role for academic libraries. This article presents the results of a survey of directors of the Association of European Research Libraries (LIBER) academic member libraries to discover what types of research data services (RDS) are being offered by European academic research libraries and what services are planned for the future. Overall, the survey found that library directors strongly agree on the importance of RDS. As was found in earlier studies of academic libraries in North America, more European libraries are currently offering or are planning to offer consultative or reference RDS than technical or hands-on RDS. The majority of libraries provide support for training in skills related to RDS for their staff members. Almost all libraries collaborate with other organizations inside their institutions or with outside institutions in order to offer or develop policy related to RDS. We discuss the implications of the current state of RDS in European academic research libraries, and offer directions for future research.


We asked several data librarians, archivists and educators who have had prominent and interesting careers if they would be willing to let us profile them and share some of their thoughts on the field. Six graciously agreed to be interviewed via email. Many of our respondents played key roles in developing data services and infrastructure in their respective countries, while others are involved in building the future of the field through education, advancing standards, and advocacy.

Our virtual panel includes Tuomas J. Alaterä, Finland; Ann Green and Jian Qin, United States; Guangjing Li, China; Wendy Watkins, Canada; and Lynn Woolfrey, South Africa.


Libraries have been asked to provide many new services over the past several decades. This paper aims to show how data management services were incorporated into the services that Oregon State University provides to faculty and graduate students. The lessons learned are general and applicable to any research institute that needs to manage data or help others with managing data.

This work is licensed under a Creative Commons Attribution 2.5 License. https://creativecommons.org/licenses/by/2.5/legalcode.


Objective: Given the increasing need for research data management support and education, the Spokane Academic Library at Washington State University (WSU) sought to determine the data management practices, perceptions, and needs of researchers on the WSU Spokane health sciences campus.

Methods: A 23-question online survey was distributed to WSU researchers and research support staff through the campus listserv. This online survey addressed data organization, documentation, storage & backup, security, preservation, and sharing, as well as challenges and desired support services.

Results: Survey results indicated that there was a clear need for more instruction with regard to data management planning, particularly as data management planning addresses the areas of metadata design, data sharing, data security, and data storage and backup.

Conclusions: This needs assessment will direct how RDM services are implemented on the WSU Spokane campus by the Spokane Academic Library (SAL). These services will influence both research data quality and integrity through improved data management practices.


In this paper we summarize the findings of an empirical study conducted by the EDaWaX Project. 141 economics journals were examined regarding the quality and extent of data availability policies that should support replications of published empirical results in economics. This paper suggests criteria for such policies that aim to facilitate replications. These criteria were also used for analysing the data availability policies we found in our sample and to identify best practices for data policies of scholarly journals in economics. In addition, we also evaluated the journals' data archives and checked the percentage of articles associated with research data. To conclude, an appraisal as to how scientific libraries might support the linkage of publications to underlying research data in cooperation with researchers, editors and publishers and data centres is presented.


INTRODUCTION As data-driven research becomes the norm, practical knowledge in data stewardship is critical for researchers. Despite its growing importance, formal education in research data management (RDM) is rare at the university level. Academic librarians are now playing a leadership role in developing and providing RDM training and support to faculty and graduate students. This case study describes the development and implementation of a new, credit-bearing course in RDM for graduate students from all disciplines. DESCRIPTION OF PROGRAM The purpose of the course was to enable students to acquire foundational knowledge and skills in RDM that would support long-term habits in the planning, management, preservation, and sharing of research data. The pedagogical approach for the course combined outcomes centered course design with active learning techniques. Periodic course assessment was performed through anonymous student surveys, with the objective of gauging course efficacy and quality, and to obtain suggested modifications or improvements. These assessment results indicated that the course content and scope were appropriate and that the active learning approach was effective. Assessments of student learning demonstrated that all major learning objectives were achieved. NEXT STEPS Information derived from the student surveys was used to determine how the course could be modified to improve student experience and the overall quality of the course and the instruction.


Objectives: This study follows up on previous work that began examining data deposited in an institutional repository. The work here extends the earlier study by answering the following lines of research questions: (1) What is the file composition of datasets ingested into the University of Illinois at Urbana-Champaign (UIUC) campus repository? Are datasets more likely to be single-file or multiple-file items? (2) What is the usage data associated with these datasets? Which items are most popular?

Methods: The dataset records collected in this study were identified by filtering item types categorized as "data" or "dataset" using the advanced search function in IDEALS. Returned search results were collected in an Excel spreadsheet to include data such as the Handle identifier, date ingested, file formats, composition code, and the download count from the item's statistics report. The Handle identifier represents the dataset record's persistent identifier. Composition represents codes that categorize items as single or multiple file deposits. Date available represents the date the dataset record was published in the campus repository. Download statistics were collected via a website link for each dataset record and indicates the number of times the dataset record has been downloaded. Once the data was collected, it was used to evaluate datasets deposited into IDEALS.

Results: A total of 522 datasets were identified for analysis covering the period between January 2007 and August 2016. This study revealed two influxes occurring during the period of 2008-2009 and in 2014. During the first timeframe a large number of PDFs were deposited by the Illinois Department of Agriculture. Whereas, Microsoft Excel files were deposited in 2014 by the Rare Books and Manuscript Library. Single-file datasets clearly dominate the deposits in the campus repository. The total download count for all datasets was 139,663 and the average downloads per month per file across all datasets averaged 3.2.

Conclusion: Academic librarians, repository managers, and research data services staff can use the results presented here to anticipate the nature of research data that may be deposited within institutional repositories. With increased awareness, content recruitment, and improvements, IRs can provide a viable cyberinfrastructure for researchers to deposit data, but much can be learned from the data already deposited. Awareness of trends can help librarians facilitate discussions with researchers about research data deposits as well as better tailor their services to address short-term and long-term research needs.


Objectives: This study explores data management knowledge, attitudes, and practices of bioengineering and biomedical researchers in the context of the National Institutes of Health-funded research projects. Specifically, this study seeks to answer the following questions:
What is the nature of biomedical and bioengineering research on the Illinois campus and what kinds of data are being generated?

To what degree are biomedical and bioengineering researchers aware of best practices for data management and what are the actual data management behaviors?

What aspects of data management present the greatest challenges and frustrations?

To what degree are biomedical and bioengineering researchers aware of data sharing opportunities and data repositories, and what are their attitudes towards data sharing?

To what degree are researchers aware of campus services and support for data management planning, data sharing, and data deposit, and what is the level of interest in instruction in these areas?

Methods: Librarians on the University of Illinois at Urbana Champaign campus conducted semi-structured interviews with bioengineering and biomedical researchers to explore researchers' knowledge of data management best practices, awareness of library campus services, data management behavior and challenges managing research data. The topics covered during the interviews were current research projects, data types, format, description, campus repository usage, data-sharing, awareness of library campus services, data reuse, the anticipated impact of health on public and challenges (interview questions are provided in the Appendix).

Results: This study revealed the majority of researchers explore broad research topics, various file storage solutions, generate numerous amounts of data and adhere to differing discipline-specific practices. Researchers expressed both familiarity and unfamiliarity with DMP Tool. Roughly half of the researchers interviewed reported having documented protocols for file names, file backup, and file storage. Findings also suggest that there is ambiguity about what it means to share research data and confusion about terminology such as "repository" and "data deposit". Many researchers equate publication to data sharing.

Conclusions: The interviews reveal significant data literacy gaps that present opportunities for library instruction in the areas of file organization, project workflow and documentation, metadata standards, and data deposit options. The interviews also provide invaluable insight into biomedical and bioengineering research in general and contribute to the authors' understanding of the challenges facing the researchers we strive to support.


Appendix A. Related Bibliographies


Note on the Inclusion of Abstracts

Abstracts are included in this bibliography if a work is under a Creative Commons Attribution License (BY and national/international variations), a Creative Commons public domain dedication (CC0), or a Creative Commons Public Domain Mark and this is clearly indicated in the publisher's current webpage for the article. Note that a publisher may have changed the licenses for all articles on a journal's website but not have made corresponding license changes in journal's PDF files. The license on the current webpage is deemed to be the correct one. Since publishers can change licenses in the future, the license indicated for a work in this bibliography may not be the one you find upon retrieval of the work.

Abstracts for works under the following types of Creative Commons Licenses (and their national/international variations) are not included:

- Attribution-NoDerivs
- Attribution-NonCommercial
- Attribution-NonCommercial-NoDerivs
- Attribution-NonCommercial-ShareAlike
- Attribution-ShareAlike

See the Creative Commons' Frequently Asked Questions for a discussion of how documents under different Creative Commons licenses can be combined.

Appendix B. About the Author

Charles W. Bailey, Jr. is the publisher of Digital Scholarship and a noncommercial digital artist (ORCID ID: https://orcid.org/0000-0001-8453-4402).

Bailey has over 46 years of information technology, digital publishing, and instructional technology experience, including 24 years of managerial experience in academic libraries. From 2004 to 2007, he was the Assistant Dean for Digital Library Planning and Development at the University of Houston Libraries. From 1987 to 2003, he served as Assistant Dean/Director for Systems at the University of Houston Libraries.

Previously, he served as Head, Systems and Research Services at the Health Sciences Library, The University of North Carolina at Chapel Hill; Systems Librarian at the Milton S.
Eisenhower Library, The Johns Hopkins University; User Documentation Specialist at the OCLC Online Computer Library Center; and Media Library Manager at the Learning Resources Center, SUNY College at Oswego.

Bailey has discussed his career in an interview in *Preservation, Digital Technology & Culture*. See Bailey’s vita for more details.

Bailey has been an open access publisher for over 33 years. In 1989, Bailey established PACS-L, a discussion list about public-access computers in libraries, and *The Public-Access Computer Systems Review*, the first open access journal in the field of library and information science. He served as PACS-L Moderator until November 1991 and as Editor-in-Chief of *The Public-Access Computer Systems Review* until the end of 1996.


In 1992, he founded the PACS-P mailing list for announcing the publication of selected e-serials, and he moderated this list until 2007.

In 1996, he established the *Scholarly Electronic Publishing Bibliography (SEPB)*, an open access book that was updated 80 times by 2011.

In 2001, he added the *Scholarly Electronic Publishing Weblog*, which announced relevant new publications, to SEPB. It was published through 2013.

In 2001, he was selected as a team member of *Current Cites*, and he has was a frequent contributor of reviews to this monthly e-serial until 2020.

In 2005, he published the *Open Access Bibliography: Liberating Scholarly Literature with E-prints and Open Access Journals* with the Association of Research Libraries (also a website).

In 2005, Bailey established Digital Scholarship ([http://digital-scholarship.org/](http://digital-scholarship.org/)), which provides information and commentary about digital copyright, digital curation, digital repository, open access, research data management, scholarly communication, and other digital information issues. Digital Scholarship’s digital publications are open access. Its publications are licensed under a Creative Commons Attribution 4.0 International License or a Creative Commons Attribution 3.0 Unported License. Digital Scholarship is a noncommercial publisher and it accepts no advertising.

At that time, he also established *DigitalKoans*, a weblog that covers the same topics as Digital Scholarship. From April 2005 through May 2023, DigitalKoans published over 15,000 posts.

From April 2005 through December 2022, Digital Scholarship had over 39 million page views from over 13 million visitors from 235 countries (excluding spiders and non-content files).


He also published and updated the following bibliographies, webliographies, and weblogs: the *Scholarly Electronic Publishing Bibliography* (1996-2011), the *Scholarly Electronic Publishing Weblog* (2001-2013), the *Electronic Theses and Dissertations Bibliography*
In 2011, he established the LinkedIn Digital Curation Group.

In 2010, Bailey was given a Best Content by an Individual Award by The Charleston Advisor. In 2003, he was named as one of Library Journal’s “Movers & Shakers.” In 1993, he was awarded the first LITA/Library Hi Tech Award For Outstanding Communication for Continuing Education in Library and Information Science. In 1992, Bailey received a Network Citizen Award from the Apple Library.

In 1973, Bailey won a Wallace Stevens Poetry Award. He is the author of The Cave of Hypnos: Early Poems, which includes several poems that won that award.

Bailey has written over 30 papers about artificial intelligence, digital copyright, institutional repositories, open access, scholarly communication, and other topics.

He has served on the editorial boards of Information Technology and Libraries, Library Software Review, and Reference Services Review. He was the founding Vice-Chairperson of the LITA Imagineering Interest Group.

Bailey is a digital artist, and he has made over 680 digital artworks freely available on social media sites, such as Flickr, under Creative Commons Attribution-NonCommercial licenses. A list of his artworks that includes links to high resolution JPEG images on Flickr is available.

He holds master’s degrees in information and library science and instructional media and technology.

You can contact him at: publisher at digital-scholarship.org.

You can follow Bailey at these URLs:

- Digital Artist weblog: https://charleswbaileyjr.name and RSS feed: https://charleswbaileyjr.name/feed
- Flickr: https://www.flickr.com/photos/charleswbaileyjr/

Citation


Copyright © 2023 by Charles W. Bailey, Jr.
This work is licensed under a Creative Commons Attribution 4.0 International License.