SPECIAL SECTION

Research Data Access and Preservation
by guest editors Sara Mannheimer, Brianna Marshall and Lizzy Rolando


14] Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates
by Jake Carlson, Megan Sapp Nelson, Lisa R. Johnston and Amy Koshoffer

18] Building Outreach on Assessment: Researcher Compliance with Journal Policies for Data Sharing by Kathleen Fear

22] Current Issues and Approaches to Curating Student Research Data
by Andrew Creamer

26] You’re in Good Company: Unifying Campus Research Data Services
by Cynthia R. H. Vitale, Brianna Marshall and Amy Nurnberger

29] BYOPanel – The On-the-Spot Assembled Panel on Responses to the OSTP Memo Responses by Margaret Janz

32] University Data Policies and Library Data Services: Who Owns Your Data?
by Lisa D. Zilinski, Abigail Goben and Kristin Briney

35] Research Data Services at the University of Colorado Boulder
by Shelley L. Knuth, Andrew M. Johnson and Thomas Hauser

39] DLF E-Research Network Members Attend 2015 RDAP Summit by Gail Clement and Rita Van Duinen

DEPARTMENTS

[2] Editor’s Desktop
[3] President’s Page
[5] Inside ASIS&T

COLUMN

[42] IA COLUMN
Everyone’s Job: Making Sense of the Mess by Laura Creekmore
SIS&T’s annual RDAP Summits continue to be a place where the various players in research data access and preservation can exchange ideas and experience, with the role and activities of university libraries in this arena being a major focus. This issue features nine articles related to RDAP15, held in April in Minneapolis, including summaries of panels, presentations and other related activities. Special thanks to Sara Mannheimer, Brianna Marshall and Lizzy Rolando for compiling and editing the section and to all the many contributors.

We also mark the passing of information science pioneer Claire Schultz, a president of ASIS&T when it was the American Documentation Institute. Robert Williams reviews her remarkable career and achievements in information retrieval, thesaurus construction and other areas.

The IA Column returns in this issue under the leadership of Laura Creekmore, our new associate editor for IA. Her observations on how much just a little help from an information expert could improve everyday processes dovetails nicely with the President’s Page. There, ASIS&T president Sandy Hirsh offers special thanks to and updates on the work of the Information Professionals Task Force and the ASIS&T Membership Committee, both of which are working on the important goal of bringing a broader range of information researchers and practitioners into ASIS&T or at least into collaborative relationships with us. The task force’s website will go public at the 2015 Annual Meeting. The site is in process, and you can access it at http://infoprofessionals.org/. Comments and suggestions are welcome.
We are already halfway through my term as president, and I want to share with you details of some of the important work that has been taking place this year. In addition to the strategic planning initiative that I have reported on in previous columns, the Information Professionals Task Force and the International Relations Committee have been hard at work and this column is devoted to updates on their activities.

The Information Professionals Task Force

The overarching goal of the Information Professionals Task Force (IPTF) is to broaden the Association’s scope to include practitioners in related professions, industry employers and the general public interested in information. A demonstration website is under construction with targeted content, job descriptions and links to collaborating organizations, with survey input from members in other information organizations. The International Relations Committee is tasked with promoting ASIS&T recruitment and retention of members beyond North America. Personal stories from international members, international student feedback on website development and membership dues awards will raise awareness of the benefits of Association membership for this audience.

EDITOR’S SUMMARY

In addition to ongoing work on the strategic planning initiative, an important focus for the Association has been outreach to a wider audience of information industry professionals, academics and institutions around the world. The goal of the Information Professionals Task Force is to broaden the Association’s scope to include practitioners in related professions, industry employers and the general public interested in information. A demonstration website is under construction with targeted content, job descriptions and links to collaborating organizations, with survey input from members in other information organizations. The International Relations Committee is tasked with promoting ASIS&T recruitment and retention of members beyond North America. Personal stories from international members, international student feedback on website development and membership dues awards will raise awareness of the benefits of Association membership for this audience.

KEYWORDS

Association for Information Science and Technology
strategic planning
information associations
information industry
international aspects
students
substantial new content. A key component of developing the content strategy was the creation of six user personae representing different types of potential users. The demonstration site was transferred to a WordPress platform. Working from the content strategy, existing content was reviewed, updated and revised to better address the anticipated users. We expect to have the public site ready and available for promotion before the Annual Meeting. The site in process can be viewed at http://infoprofessionals.org/ and comments and suggestions are welcome.

In parallel with these activities, the task force reached out to other information professional organizations to identify people willing to have bios on the site and representative job descriptions. Among the collaborating associations are the Coalition for Networked Information, the Association for Information Systems, the IA Summit, Information Architecture Institute, International Council on Archives and the American Medical Informatics Association. Social media and career links from the collaborating organizations will be included on the site. The site will also highlight information professional career paths and job descriptions and will provide leads to selected educational programs, jobs and collaborating professional organizations. Contacts with other organizations, including additional international ones, are ongoing. To supplement the bios obtained through this approach, task force members have also requested assistance directly from some of their professional contacts.

To identify ways to make ASIS&T more attractive to practitioners, the task force decided to conduct a survey of people in other information professional organizations, and work has begun on survey development. The task force will use members of some of the organizations contacted for the website to get survey recipients.

**International Relations Committee**

The International Relations Committee (IRC) focuses on ways ASIS&T can continue to support information science academics, students, professionals and institutions worldwide.

The IRC is co-chaired by Diane Sonnenwald and Mei-Mei Wu. Members include Fidelia Ihekwe-SanJuan, Irene Onyancha, Hans-Christoph Hobohm, Yan Zhang, Jane Greenberg, Yuelin Li, Christian Schögel, Marica Sapropficovik, Ying-Hsang Liu, and Xiaojun (Jenny) Yuan.

To assist ASIS&T headquarters and the ASIS&T social media manager in their efforts to recruit and retain members, testimonials from members in Australia, China and France were collected. These testimonials provide personal stories regarding the benefits of participating in ASIS&T and are intended for use on the ASIS&T website and in promotional materials. The IRC has also been active in obtaining feedback on the new website from students in Africa, Europe and Asia. The goal of this initiative is to ensure that the new ASIS&T website reflects the global nature of the Association, rather than reflecting a North American perspective. The Board generously allocated 40 free one-year student memberships to be awarded to students who provide feedback. Ten additional matching student memberships for students in the Asia-Pacific Chapter region were obtained from an anonymous donor. The IRC has also submitted a new membership dues proposal that takes into account all categories of membership and different income levels in countries. The proposal was endorsed by SIG/III and is currently under consideration by the Board.

I highlighted the work of these two groups in this column because their work directly aligns with the goals I outlined for the year and because information professionals (practitioners) and international members are among the categories of members that ASIS&T has identified as important for the Association’s ongoing success. As the only professional association that bridges the gap between information science practice and research, it is critical that we continue to attract and retain practitioners in our Association. We also want to make sure that our members, who span the globe, feel included and welcomed in our Association. I wish to thank these hard working committees/task forces for their significant contributions.
ASIS&T Members Elect New Board Officers and Representatives

In ASIS&T’s annual summertime election, members made their selections for president-elect and directors-at-large, while also approving an amendment to the organization’s bylaws.

The winning candidate for president-elect, who will serve a three-year term from president-elect through past president, is Lynn Silipigni Connaway. The two candidates earning spots as directors-at-large are Kathryn La Barre and Abebe Rorissa.

Lynn Silipigni Connaway, who is completing a three-year term as a director, is senior research scientist at OCLC Research where she leads the user studies activities. Her responsibilities include research projects that directly involve OCLC libraries and users, such as WorldCat data mining projects; JISC-funded investigations of digital information seekers and users in the virtual research environment; and IMLS-funded grant projects to study virtual reference services and the behavior patterns of college and university information seekers.

A member of ASIS&T since her doctoral student days, Lynn says in her position statement for the election that she believes “it is imperative in today’s globalized information environment for ASIS&T to provide value to its members.” To work toward that goal, she will use the results of the strategic planning process and work with the Board and membership on the following activities:

- Provision of career development and mentoring opportunities focused on students, early-career professionals and researchers
- Identification and implementation of new forms of member engagement
- Promotion and support of SIGs and chapters
- Recruitment of more international members

Abebe Rorissa is an associate professor in the department of information studies at University at Albany, State University of New York. Previously he worked in four countries over a 17-year period as a lecturer and systems librarian. He has also consulted for academic institutions, national governments and international organizations. His research interests include multimedia information organization and retrieval, scaling of users’ information needs/perceptions and impact of information and communication technologies.

Abebe notes in his position statement that in 13 years of membership he has served in numerous leadership positions and participated in virtually every type of activity ASIS&T sponsors. He says that “ASIS&T is on a long journey to becoming a truly international association. If elected… I will strive to broaden the geographic reach of ASIS&T with members from every corner of this global village.” To accomplish this goal, he says “ASIS&T must widen the scope of issues addressed and reach/attract a new and diverse group of academics, researchers, students and practitioners on all continents.”

Kathryn La Barre is an associate professor at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. Her research on historical and contemporary knowledge organization systems focuses on cultural heritage materials. She is co-principal investigator of the Comic Book Readership Archive (CoBRA) project to build a digital archive of materials related...
Plans Ramping Up for the 2015 Annual Meeting

With the passing of deadlines for submission of proposals for program activities at the 78th ASIS&T Annual Meeting, conference committee chairs have moved into the final stages of accepting the best-of-the-best ideas and putting form to an actual conference schedule. When attendees gather in St. Louis on November 6-10 for Information Science with Impact: Research in and for the Community, they will choose from among eight exceptional pre-conference workshops and then head into the technical program days where conference organizers have identified a keynote speaker; some 40 superb papers and 20 panel presentations.

Plenary Speaker
This year’s conference theme gives information science researchers – including academics and practitioner researchers – the opportunity to discuss the impact of their research on industry, government, local/national/global community groups, individuals, information systems and other practice contexts. The theme highlights the introduction of a new conference focus on applied research, which recognizes that basic research in information science is also inspired by and/or connected to information practice contexts. Conference chair Lisa Given, Charles Sturt University, Australia, says the plenary session at the meeting will help focus the conference theme and the dozens of concurrent sessions that will speak to it.

On Monday, November 9, Sarah Morton, co-director of the Centre for Research on Families and Relationships (CRFR), University of Edinburg,
Scotland, will discuss “Creating Impact: Issues, Challenges and Solutions.” Sarah leads her center’s knowledge exchange team, facilitating ways in which research on families and relationships can be widely used. She has been working in knowledge exchange for more than 10 years and is interested in all aspects of research use and knowledge to action, particularly social research, and issues in the co-production of research. Her research has investigated the process of assessing the impact of research on policy and practice. She has a specialty in contribution analysis and uses this approach in a variety of projects, often working with non-academic partners.

Pre-conference Workshops

As always, ASIS&T encourages its members to take advantage of pre-conference workshops, many of which are offered by ASIS&T special interest groups and offer in-depth looks into new topics or new approaches to the fields they represent. This year’s workshops will be held on Friday and Saturday, November 6 and 7. Separate registration is required for these sessions. Here’s what is planned:

**FRIDAY, NOVEMBER 6**

- So Who’s Managing All That Organizational Information Anyway? (SIG/MGT)
- Metrics 2015: Workshop on Informetric and Scientometric Research (SIG/MET)
- Knowledge Audits: A to Z (SIG/KM)
- PKM Education and Devices as Key Enablers for Growing a 21st Century Knowledge Society

**SATURDAY, NOVEMBER 7**

- 11th Annual Social Informatics Research Symposium: The Impacts of Social Informatics Research (SIG/SI)
- Conceptual Crowbars and Classification at the Crossroads: The Impact and Future of Classification Research (SIG/CR)
- Digital Liaisons: Serving and Impacting Communities Through Digital Libraries (SIG/DL)
- Information Visualization Workshop (SIG/VIS)
- 15th Annual SIG/USE Symposium (SIG/USE)

Look for descriptions of these sessions and more details in the 2015 ASIS&T Annual Meeting preliminary program coming your way later this month and on the ASIS&T website.

As ASIS&T president Sandy Hirsh notes in her President’s Page in this issue of the Bulletin, the ASIS&T Information Professionals Task Force (IPTF) is hard at work looking for new and innovative ways in which “to increase awareness and interest in the information professions.” The task force’s web page defines its charge in part as follows:

The web presence will be designed to define information professional as a generic term to refer to professionally trained workers in information intensive environments, but will also include the ASIS&T brand. The site will highlight information professional career paths and job descriptions and will provide leads to educational programs, jobs and professional organizations.

As one of many means to that end, IPTF co-chairs Diane Rasmussen Pennington and Nancy Roderer have posted a short survey that seeks individual thoughts about the roles of information professionals. The survey is available at http://svy.mk/1FyKFaA.
EDITOR’S SUMMARY
A groundbreaker and important link between library science and computer-based information science, Claire Kelly Schultz passed away in May 2015. Though Schultz’s original aim to become a doctor was derailed, her first job as a biology lab assistant and librarian launched her on a career path that ultimately provided important service to medicine, pharmacology and information science. With training in library studies, chemistry and medicine, Schultz became the librarian at the future Merck, Sharp and Dohme Research Labs and became aware of the technical challenges of literature retrieval in the field. She consulted with others at the forefront of indexing and retrieval systems to develop a successful card sorting method using Boolean logic. This success led to consulting positions and publications on information retrieval, vocabulary control, thesaurus construction and indexing. Schultz strongly advocated applying documentation practices to the developing field and to the history of the American Documentation Institute (ADI), forerunner of the Association for Information Science and Technology, and she created the first thesaurus of information science. Active in several professional societies, Schultz became the ADI’s first woman president and earned its 1980 Award of Merit.

KEYWORDS
obituaries  Association for Information Science and Technology
information science history  honors
history  biography

Claire Kelly Schultz, first woman president of the American Documentation Institute (ADI; predecessor of ASIS&T) and recipient of the 1980 Award of Merit, died May 28 in her hometown of Line Lexington, Pennsylvania. The cause of death was Alzheimer’s disease, which she had had for a number of years. Claire was an active member of ADI, the Special Libraries Association (SLA), the Medical Library Association (MLA) and the American Chemical Society (ACS). She was the first ADI member to take a serious interest in the history of ADI, publishing an historical sketch of it in 1969 followed by a series of biographical articles on early ADI pioneers [1]. Her edited volume on Hans Peter Luhn, the developer of keyword-in-context (KWIC) indexing and many other innovations in the field, was a significant contribution to the history of information science [2].

Claire was born in Etters in central Pennsylvania, and her parents were part-time farmers and merchants with limited formal education. She completed her B.S. degree in 1944, majoring in chemistry and pre-med, at Juniata College. From an early age Claire wanted to be a medical doctor and through...
various family trials and monetary tribulations managed, at age 16, to be admitted to the Woman’s Medical College of Pennsylvania. Her stay there, however, would be short: she became pregnant (she had married Wallace L. Schultz in June 1945) and was dismissed from the school for that reason. Her first job was at the Wistar Institute of Anatomy and Biology in Philadelphia as lab assistant and librarian, where she worked from 1946 to 1948. Looking for a better paying job she began taking library science courses at Drexel Institute, graduating with an M.S. in 1952. Because she was taking these library science courses and because of her chemistry background and medical training, she was hired at Sharp and Dohme Research Labs (soon to be Merck, Sharp and Dohme) in 1949 as librarian.

One of the significant problems of library services at the time was the inability of staff and users to locate needed literature, particularly in chemistry and pharmacy. Claire began to study this problem and invited Calvin Mooers to come for a visit and consultation regarding the Zator indexing and retrieval system he had developed and was attempting to sell. This solution worked fairly well for her, but the mechanics of vocabulary control and searching were too complex. She formed a partnership with Robert Ford in the pharmacology lab, who had similar problems with vocabulary control of compounds, and they discovered that the Remington Rand punched card sorter had a “logic bridge” that utilized Boolean logic for sorting and retrieval. This approach worked very well, but since it was a mechanical system it was always breaking down. This problem in turn led to exploring the use of the IBM 101 card sorter. However, it did not have logic capability so Ford had to fix up special wiring (a dial-up control board) for it to do sorting using Boolean logic. This system was very successful, and Claire made many presentations and demonstrations at her library and at SLA, ADI and other meetings. She completed her M.S. in 1952 and submitted as her thesis a paper on the system [3] and later published a paper in Casey and Perry’s 1958 book on punched cards [4].

Publicity about her system led John Mauchly, developer of the ENIAC computer, to visit for a demonstration. He was surprised at the use of Boolean logic in her system and offered her a job with his company, Sperry-Rand Univac, in Philadelphia, to do more work on her searching system and on vocabulary control. She declined this offer, fearing that if she left Merck, Sharpe and Dohme her work on the retrieval system would be lost. However, three years later, in 1958, she became a systems analyst for Sperry-Rand. Her work there mostly involved working with and consulting on various information retrieval problems relating to federal government contracts (or possible contracts), most notably, the Armed Services Technical Information Agency (ASTIA), the National Library of Medicine’s (NLM) MEDLARS project and others. The early 1950s was a time of great interest in building vocabulary control systems, and many of these contracts were related to this issue, a particular interest and specialty of Claire’s. She was also involved in the decision-making process for the best computer to use for a contract, which meant she had to know the technical details about each machine and how well it could handle the work to be done.

The late 1950s and early 1960s were a very busy time for Claire. Not only was her fifth child born during this time, but also she was attending classes at Drexel and, later, teaching there. She was also doing a lot of consulting with federal agencies with weekly meetings in Washington, DC, as part of her job with Sperry-Rand. She also somehow found the time to publish over 20 articles from 1952 to 1963, most of them relating to information retrieval, vocabulary control, thesaurus construction and indexing [5]. And, for a brief six-month period in 1957, before she began with Sperry-Rand, she helped Eugene Garfield do some of the early publicity and related work for his new company, Eugene Garfield and Associates, later the Institute for Scientific Information. She had extensive contacts with the
Claire’s work as a consultant in information retrieval and thesaurus construction brought her in contact with many of the leaders of ADI and well as SLA, MLA and other professional associations. She served on several committees of ADI before being elected president-elect in 1961 and then president in 1962. In her inaugural address as incoming president in 1962 she stressed the importance of research to members and to the society itself, began a new project, “documentation of documentation,” and urged members to collect literature of the field and build a modern indexing system for them. Despite the fact that the National Science Foundation (NSF) had just granted $16,000 to aid this project (and ADI overall) not much was done with it, much to Claire’s dismay. However, a secretariat for ADI was established during her term as president, the first executive director was hired (John Kaiser) and membership soon doubled [6]. Shortly after her presidential year she prepared the first thesaurus for the field of information science [7]. (In her oral history, Claire tells an interesting story about a mysterious John Deere, who went all about the country recruiting for ADI and then disappeared – but no one ever knew where he came from or where he went.) [8, p. 52]

During the year of her presidency of ADI Claire joined the Institute for Advancement of Medical Communications (IAMC), which had been founded by Richard Orr, well known for his studies of medical libraries, indexing and methodology for measuring the effectiveness of libraries. She continued her consulting with various federal government agencies and teaching at Drexel. One large contract that IAMC had was to collect, analyze and index the papers of the Federation of the American Societies for Experimental Biology (FASEB), the largest scientific meeting in the United States. This project resulted in the production of a “back of the book” index that was computer generated, possibly the first of its kind in the United States. Claire worked at IAMC until 1970 when it was disbanded, Dr. Orr moved to Europe to do consulting and she became a freelance consultant in many of the same areas she had worked before.

The freelance consulting work ended in 1972 when she was enticed to become the director of libraries and professor of information science at the Medical College of Pennsylvania (MCP), the same school that had kicked her out of the M.D. program. She worked there until 1982 when she retired. However, during the 10 years she was at MCP she oversaw the building and equipping of a new library, taught the medical students how to use computers and managed a number of small contracts in the areas of vocabulary control and indexing.

Claire was the ideal bridge from library science to information science. She was well prepared in the basics of library science, had a strong science background and chose to do in-depth exploration of the new trends in scientific documentation, indexing and thesaurus construction. She was also a leader in developing and implementing computer-based approaches to library problems, particularly in helping to design large information systems for the federal government. As a bridge in the larger field, she was able to deal effectively with what was becoming a split between special librarians and the new “documentalists” of ADI and, later, ASIST.

After her retirement in 1982 Claire became deeply involved in the life of being a grandmother. She and her husband, Wally, built their own catamaran and began to sail it to many places on the East Coast. After several years of doing that in the summers they gave the boat to a children’s center.

Claire was severely ill with Alzheimer’s for a number of years before her death, living in an assisted living facility, but was always well cared for by her extensive family.
Resources Mentioned in the Article

RDAP15 Summit Report: Introduction
by Sara Mannheimer

EDITOR’S SUMMARY
The 2015 ASIS&T Research Data Access & Preservation (RDAP) Summit presented an opportunity to explore the current state and challenges facing librarians providing research data services. From pre-conference meetings and the opening keynote through lightning talks, posters and roundtable discussions, Summit attendees considered data education, outreach and networking and evaluation of progress, all of which are represented in the articles of this special section. Keynote speaker Claudia Neuhauser spoke of the need for interdisciplinary collaboration in research data services, equally challenging for big data and small. Special attention was paid to curating student research data and specialized documentation. A special panel examined the mandate in the Office of Science and Technology Policy (OSTP) memo and responses to it.

KEYWORDS
meetings
data curation
library technical services
outreach services
educational activities
documentation
information policy
strategic planning

Sara Mannheimer is data management librarian at Montana State University. She can be reached at sara.mannheimer<at>montana.edu.

Editor’s Note: This special section on RDAP15 was co-edited by Sara Mannheimer, Brianna Marshall and Lizzy Rolando.

I love attending ASIS&T’s Research Data Access & Preservation Summit (RDAP). The RDAP community is a vibrant and welcoming group of researchers and practitioners, eager to share ideas and discuss the intricacies of our developing specialty. I always leave the Summit with renewed enthusiasm about the value of data management and open data and with a renewed commitment to data librarians’ role as ambassadors for the cause. This year’s Summit (RDAP15) – held April 22-23, 2015 in Minneapolis, Minnesota – did not disappoint. From the pre-conference meeting of the CLIR/DLF E-Resource Network to the closing poster session and roundtable discussions, the program was brimming with ideas to help data professionals think more clearly about data and practical strategies to help us build better research data services at our institutions.

In her opening keynote, Claudia Neuhauser, director of the University of Minnesota Informatics Institute, presented a researcher-centered approach to university data services as well as some data challenges and potential solutions. She discussed the changing research ecosystem, noting its highly collaborative, interdisciplinary nature; she discussed the new requirements and complexities surrounding research data; and she highlighted the library’s role in facilitating both collaboration and data management. One memorable moment in the keynote was when Neuhauser deftly tackled the ever-problematic term big data by suggesting this definition: Data that is too big for you. The approving audience response suggested that, although most data at our institutions may technically be “small data,” without the right resources to manage, analyze and store it, that data seems gigantic. Neuhauser’s talk
set the stage for the rest of the Summit, as we talked through research and practice surrounding data librarianship and library data services.

The remaining presentations on both days centered on a few major themes. We discussed outreach and collaboration strategies through the panels “Unifying Campus Research Data Services” and “Research Data Integration in the Purdue Libraries.” We investigated strategies for data education through the panel “Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates.” We examined assessment through the panel “The Role of Assessment in Research Data Services.” Kathleen Fear contributed her paper from that panel, “Building Outreach on Assessment: Researcher Compliance With Journal Policies for Data Sharing.”

A special panel on responses to the OSTP memo delved into how both funders and academic institutions are responding to these new data management and data sharing requirements, and the work we’re doing to educate our faculty about data archiving mandates. And we discussed data curation through the panels “Current Issues and Approaches to Curating Student Research Data” and “Beyond Metadata: Supporting Non-Standardized Documentation to Facilitate Data Reuse.” Lightning talks, posters and roundtable discussions gave us even more opportunity to explore these themes. Two lightning talks, “Research Data Services at the University of Colorado Boulder” and “University Data Policies and Library Data Services,” and an article discussing the participation of DLF E-Research Network members are also included here.

In this special section of the Bulletin, guest edited by Lizzy Rolando, Brianna Marshall and me, RDAP15 session leaders and presenters review and reflect on their experiences at the Summit and the research data issues that were explored in their respective sessions. Special thanks to this year’s program chairs, Carolyn Bishoff, University of Minnesota, and Margaret Janz, Temple University; to all of the presenters, session leaders and planning committee members; and to our sponsors, Globus and ASIS&T SIG/DL (Special Interest Group/Digital Libraries).

For additional information, please see the following resources:

- Slides from presentations and a selection of posters, available on SlideShare: www.slideshare.net/asist_org/tag/rdap15
- Shared notes from the sessions: bit.ly/rdap15notes
- Shared notes from the discussion tables: bit.ly/RDAP15_DisTable
- RDAP15 program: www.asis.org/rdap/program/

We invite you to join the conversation and stay updated on RDAP 2016 by signing up for our listserv: http://mail.asis.org/mailman/listinfo/rdap.
Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates
by Jake Carlson, Megan Sapp Nelson, Lisa R. Johnston and Amy Koshoffer

EDITORS SUMMARY
Building data information literacy among faculty, graduate students and undergrads was the focus of a 2015 RDAP Summit panel, with panelists describing programs at different institutions geared to each of these target groups. The Data Information Literacy project identified 12 key competencies for graduate students and how librarians could help build those skills. The Data Management Strategies Self-Assessment encourages junior faculty members to objectively consider their research data management practices and to prioritize issues and tasks. Identifying and addressing the data information literacy competencies of undergraduate students is challenging, with their widely diverse backgrounds and needs. Varied creative approaches, such as embedding lessons within a class, presenting workshops and developing partnerships and research mentorships, have been successful. Data Information Literacy project teams have developed educational programs, compared and integrated their experiences and offer guidelines, available online, for developing digital literacy programs at other institutions.

KEYWORDS
information literacy faculty
data curation graduate students
outreach service college students
educational activities

Researchers are under increasing pressure from federal agencies, scholarly publishers, disciplinary societies and their peers to administer their data in ways that enable them to be discoverable, understandable and used by others. However, the knowledge and skills required to fulfill these expectations are not often included as a part of higher education, leaving researchers to figure out how to manage, share and preserve their data on their own. This panel at ASIS&T’s RDAP15 explored how libraries are supporting the education of graduate students, faculty and undergraduate students with data literacy programs.

Working With Graduate Students
Jake Carlson from the University of Michigan and Lisa Johnston from the University of Minnesota presented the lessons learned from Data Information Literacy (DIL) project.

The Data Information Literacy (DIL) project was launched in 2011 with support from the Institute of Museum and Library Services. The intent of the project was to identify the competencies that graduate students should possess in working with data to be successful in their chosen disciplines and to explore roles for librarians in teaching these competencies. The project included five teams from four institutions, each of which partnered with a faculty member to design and implement an educational program informed by disciplinary cultures of practice and targeted to address specific local needs.

The DIL project methodology had each team interview graduate students and faculty advisors to dive deeply into their educational needs and gaps from the perspective of both graduate students and faculty. An outcome of this project is shown in Figure 1. It is a comparison of how faculty and students rate the importance of each of the 12 DIL competencies used in our study. In our sample we had faculty and students from a variety of scientific...
disciplines (agriculture, ecology, civil engineering, computer science and landscape architecture). One disconnect that we found was that faculty tended to highly rank more abstract concepts, such as metadata and description and data quality and documentation, whereas students focused more on the type of skill sets that they were already accustomed to using, such as data management and organization, as well as data processing and data visualization.

![Graphical comparison of faculty and student rating of importance of DIL competencies.](image)

Each of five teams developed an educational program based on what was learned in their interviews and from researching the data cultures of practice of their faculty partner’s discipline. For example, the University of Minnesota team created an online asynchronous training course with seven video modules viewable anytime. After two semesters of online only, they flipped the content to a hybrid in-person and online approach to allow more class time for hands-on activities. This change resulted in higher completion rates, even for busy graduate students [2]. Their lesson plans and syllabus for this hybrid approach is available online for anyone to download and reuse at [http://z.umn.edu/TeachDatamgmt](http://z.umn.edu/TeachDatamgmt).

After teaching their respective programs, each team compared their experiences with each other and developed a guide for academic librarians seeking to develop DIL programs of their own. The following are among our key takeaways:

- **Graduate students are key members of the research group and often on the frontline of the research process.** They collect, process, analyze and often (solely) manage the research data collected in that research.
- **Data management is often a task given to graduate students without a lot of preparation or education.** For example, students may be unfamiliar with the techniques for proper data documentation in their own disciplines at this stage in their careers. New students may not have a good understanding of storage options on their campus and may try a DIY approach with backup. They may not have a good understanding of the long-term value of the data or if the data need to be retained after they graduate.
- **Take it slow.** Don’t assume that students have learned even basic data management skills in their undergraduate programs. However, they are very skilled at managing a variety of other types of personal information, such as photos, digital documents and video on many different devices. Use these skills to help them scaffold to digital research data when introducing concepts such as organization, metadata and digital preservation file formats.
- **There may be ownership concerns for the students’ data that are not well understood.** For example, if the data were created as part of a grant or funded by a private organization, those ownership considerations impact how the data are managed and shared.
- **Students can and do ask for help.** We heard from students that they learned how to manage data by asking their peers, family members and Google or they would try to come up with their own best practices.
- **We also found some things that motivate students to participate in DIL educational programs:**
The tool developed for this outreach, the Data Management Strategies Self Assessment, can be downloaded at http://dx.doi.org/10.5703/1288284315525. This tool is intended to help junior faculty think through research data management in terms of concrete tasks and to invite reflection on practices currently in place in their laboratories and/or research groups. It also provides the faculty members a way to prioritize the most important data management tasks that should be implemented as a starting point for discussion during the workshop. The priorities of the junior faculty then guided the discussion for the workshop. No one individual who participated in the group discussion had solutions for all of the identified priority issues, but the group generally had suggestions and ideas for approaches to addressing identified problems. All identified issues were addressed during the workshop, and all junior faculty members were able to leave the session with ideas for how to begin.

This particular market, early career faculty, present many possibilities for data specialists and data librarians. In particular, finding faculty mentors who work with a number of junior faculty members may be a fruitful avenue for exploration. Additionally, approaching junior faculty members as they set up their research laboratories during their first semesters in their new positions may prove fruitful as well.

Working with Undergraduate Students

Amy Koshoffer from the University of Cincinnati (UC) is actively identifying opportunities for supporting DIL education for undergraduates.

Undergraduates have varied backgrounds in data information literacy at graduation. Those that seek out a research experience potentially learn necessary skills for the specific type of research from the research mentor. Students enrolled in a research-based course, such as biology laboratory courses, may be introduced to concepts of data collection, data analysis and data visualization, but probably not to data management, data backup and preservation. No single approach seems to cover all data information literacy competencies. Librarians involved in providing research data services, such as the three informationists at UC libraries, are looking for creative venues to deliver instruction on these skills. Some approaches include embedding in a class, offering library workshop and forming
strategic partnerships with university faculty, offices and institutes involved with undergraduate education.

At the University of Cincinnati, experiential learning is a cornerstone of the university’s approach to undergraduate education [5]. Each year over 5,000 of the 25,000 full-time undergraduates complete a semester of experiential learning, usually in the form of an industry co-op. Experiential learning greatly enhances interpersonal skills such as good communication, confidence and punctuality and professional skills such as critical thinking and collaboration. The Office of Undergraduate Research, Scholarly and Creative Endeavors (URSC) is one of several offices involved in experiential learning support. One goal of URSC is to establish research experiences as viable co-op experiences for students. The office has established a ready-for-research workshop to educate undergraduate students on how to find research opportunities. This introductory workshop focuses on career planning, the resume and basic laboratory safety training. Additionally participants in the workshop can enroll in a mentoring program known as RECON: Research, Education & Creative Opportunities Network.

Students in the RECON program are matched to more senior students already involved in a research experience, usually in the same area of interest. Partnering with this office provides UC libraries with an established and research-focused venue to introduce data information literacy skills and, with student partners in the RECON, mentors to collaborate on teaching these skills.

In collaboration with URSC, informationists in UC libraries plan to offer additional workshops on grant writing, data management, data visualization, presentations and data preservation. Workshops offered will mirror the research process from data collection through data analysis and sharing to data preservation and discoverability. One potential idea is to create a non-credit certification to award to participants who complete all the workshops. Potential research mentors may view this certificate as evidence that certificate holders will have more successful research experiences and be better contributors to the research group’s overall goals. The program will give research mentors standards that can be used to evaluate undergraduates seeking research co-op experiences, as participants in these workshops gain enhanced data information literacy skills.

Partnering with URSC may lead to additional collaborations with strategic partners for DIL instruction support, such as individual faculty interested in building data information literacy skills into their courses or other mentoring programs like UC’s Preparing Future Faculty (PFF) program for instruction collaborators.

Conclusion

More and more libraries are launching data information literacy initiatives as a component of the data services offered to their constituencies. As we make progress towards defining and delivering data information literacy programs we should also keep in mind the need to develop a community of practice of our own in this area. To this end, we have made the teaching materials generated by the DIL project teams available through the Data Information Literacy Case Study Directory (http://docs.lib.purdue.edu/dils/). We hope these materials are of use to others and inspire librarians to get involved in engaging researchers in a growing area of need.

---

**Resources Mentioned in the Article**


Building Outreach on Assessment:
Researcher Compliance with Journal Policies for Data Sharing

by Kathleen Fear

EDITOR’S SUMMARY

Discouraged by researchers’ low response to informational offerings on data sharing, librarians at the University of Rochester’s River Campus Libraries refined their outreach strategies. To understand the support that faculty members may need, they checked the data sharing policies of journals that faculty members publish in and the level of compliance. A review of journal policies and evidence of data sharing showed both requirements and compliance varied widely. The university librarians plan to offer workshops specifically on data sharing requirements for PLOS One, the journal researchers published in most in 2014. Researchers with strong data sharing practices will be encouraged to mentor colleagues. Ongoing compliance assessment will point out where further library support may be effective.

KEYWORDS

editorial policy
information sharing
data sets
evaluation
outreach services
educational activities

The University of Rochester’s River Campus Libraries have been actively developing services to meet researchers’ data management needs. Many of these initiatives have been very popular, with the marked exception of outreach around data sharing. Workshops focusing on how to share data have gone unattended. Other efforts, like a series of articles in a campus newsletter to researchers, pose a challenge to assessment: the information is out there, but there is no easy and effective way to measure its impact. This lack of uptake was both disappointing and perplexing, given the increasing emphasis on data sharing from publishers and funders.

Conversations with faculty provided some insight into their lack of interest in guidance on data sharing. Many believed that they already shared their data effectively, while others were adamant that data sharing does not apply to them. The root of the problem seemed to be that, when it came down to it, data sharing was too broad a category to be of interest to these researchers. A topic like developing a data management plan is concrete, and our outreach materials use the same language as the funders who require such plans. This specificity makes it easy for researchers to recognize that they have a need and that the library is trying to help them meet it. Data sharing, by contrast, is nebulous, and different entities talk about sharing and openness in different ways, which poses a challenge in connecting with researchers.

As a result, we changed our concentration from education and outreach on data sharing generally to trying to find opportunities for targeted outreach. At the same time, our library was making a shift: the librarians transitioned from subject librarians to outreach librarians, reflecting the libraries’ focus on building relationships with faculty, students and
departments. These two factors together prompted a collaborative project among the data librarian and the outreach librarians for computer science (Fang Wan); brain and cognitive sciences, linguistics and public health (Judi Briden); chemistry (Sue Cardinal); physics, optics and astronomy (Tyler Dzuba); and biology, math and statistics (Diane Cass). The project came together around two goals: develop effective strategies for getting data sharing information to researchers and strengthen outreach librarians’ relationships with their respective departments as well as boost their expertise in data support.

Methods

We identified journal data sharing policies and researchers’ compliance with those policies as a promising area of study. First, the outreach librarians already field questions from faculty about journals and their author guidelines, so examining their data policies represented a natural extension of work they commonly do. Additionally, assessing compliance with journal data policies is relatively straightforward. Looking at compliance with data management plans is complicated both by the long timespan between when the plan is put into place and when compliance can be assessed as well as by the breadth and diversity of those plans. Journal data policies are comparatively much simpler to assess, since typically authors must comply with the policy at the time of publication, and the same policy applies to all papers published in the journal, rather than being fine-tuned to individual projects.

We carried out the project in three stages: identifying the journals our faculty commonly publish in; reviewing those journals’ data sharing policies (or lack thereof); and evaluating whether and how well the authors met the requirements of those policies. To build the list of journals, we extracted from Web of Science all publications with at least one University of Rochester-affiliated author between January 1 and December 31, 2014. The University of Rochester has a heavy science and engineering focus, so while Web of Science does not have perfect coverage, we believed that the 2,784 articles across 1,181 journals it identified represented a reasonable sample set for this exploratory project.

From the 1,181 total journals represented in the article sample, we focused our policy review on the 109 journals in which Rochester researchers published five or more times in 2014. Of those journals, 43 required or explicitly encouraged data sharing. The compliance review narrowed the sample further, examining 161 articles across 13 journals with data sharing policies.

Good Policy, Good Sharing

In order to carry out the review we planned, we needed to define what we were looking for in both the policy review and the compliance review steps. In other words, we needed to determine, for our purposes, what a good data sharing policy looked like and what good compliance looked like.

Because the eventual goal of the project was to develop outreach strategies, we assessed journals’ data sharing policies through the lens of outreach and instruction. We were especially interested in how much assistance researchers might need to comply with the plans as written – and, in turn, how the library might provide that assistance – so our review criteria focused on the comprehensiveness of the policies. We evaluated whether the policies specified how to share the data, how to indicate in the paper that the data were available and when the data should be accessible to others. This approach is a departure from other studies of data sharing policies, which focus on how enforceable those policies are (Piwowar and Chapman, 2008, [1] and Sturges et al., 2014 [2]), particularly whether the journals require authors to submit an accession number or other identifier as proof of deposit.

It is worthwhile to consider the comprehensiveness of a policy and the quality of that policy separately. A policy can be – and indeed, many were – comprehensive, clear and straightforward to comply with, without requiring or encouraging good data sharing practices and without enabling any means of enforcement. For example, Organometallics requires that crystallographic data be submitted as supporting information and specifically states that depositing that data at the Cambridge Crystallographic Data Center, a widely used and well-respected repository in that field, does not fulfill that journal’s data sharing requirement. Others have comprehensive guidance for
have comprehensive guidance for certain data types but not others. This complicated our analysis of compliance with the policies. Because fully complying with a journal’s data sharing policy was no guarantee that the data were shared effectively (that is, openly, in a usable format or linked to the paper), we developed a three-tier rating system for reviewing articles.

Articles received the highest rating if the data were linked directly from the paper, if the data were included in the paper or supplemental information in a usable format (not just as a pdf) or if the authors provided a clear justification for why the data could not be made accessible. A second level rating flagged articles that, while technically in compliance with the appropriate journal’s policy, fell short of good data sharing practice. Examples include articles that indicated that the data were available, but with no link, accession number or other connection to where the data were held; those that promised that data were available on request only; and those that included data in the paper or supplement, but in an unusable format.

The lowest rating was reserved for papers with no sign that the data were available.

Findings and Putting Them to Use

Our findings confirmed that there is a need to better educate researchers on sharing their data. Half the articles reviewed received the lowest rating, indicating that there was no way to find or access their data. More positively, though, of the papers that did share data, 60% received the highest rating.

There did not appear to be any relationship between how comprehensive the journal’s data sharing policy was and how well authors complied with that policy. All but two journals had both compliant and non-compliant articles. The two exceptions were Monthly Notices of the Royal Astronomical Society and Earth and Planetary Sciences Letters, both of which had no non-compliant articles. Notably, those two journals also had two of the least comprehensive data policies. This reflects that journal data policies are written with an assumption of a certain amount of disciplinary knowledge. In fields like astronomy and earth science, where data sharing practices are more well-established, journals may include less guidance in their policies on the assumption that, if authors are publishing work in that field, they have already been socialized into the field’s practices around data sharing.

These findings open several avenues for outreach. Based on our findings, we can segment departments and researchers into several categories: those who are not sharing data but are supposed to; those who share, but not as well as they could; and those who are exemplary in their data sharing practices. Currently, we are working to design and test out targeted outreach strategies for each of these populations.

Reaching Recalcitrant Sharers. The journal in which researchers published the most in 2014 was PLOS One. This is also a journal with a rigorous and comprehensive data sharing policy. However, nearly 60% of the articles appearing in that journal did not share their data. Partly, this lack of sharing is a timing issue: PLOS’s data sharing policy only went into effect in April 2014, so many of the papers published last year were submitted prior to that time. But given the popularity of the journal on this campus, its interdisciplinary nature and its relatively high prestige, we plan to focus on it going forward, offering workshops on publishing in PLOS journals (with data sharing advice as an integral part) as well as directly contacting authors who published there previously with guidance and resources for sharing their data the next time they plan to publish there.

Helping Bare-Bones Sharers Do Better. Overall, the number of papers where data was shared well was higher than those that did the bare minimum to share their data, and all but three journals had at least one paper that received the highest rating in our compliance review. We plan to speak first with authors who do a good job sharing and enroll them to encourage their colleagues to do so as well. Working in partnership with those researchers is an opportunity to expose researchers to some peer pressure while at the same time offering examples of how to share data effectively and resources for doing so.

Learning from Exemplars. In our sample, the best data sharers were from earth science and astronomy, both fields whose journals had limited guidance on data sharing in their policies. The lack of guidance in their
Journal policies reflect a seemingly correct assumption that authors know what to do with their data, so our outreach to these exemplar researchers will focus on learning from them, especially on how they become acculturated into the data sharing practices in their fields. While experienced researchers may understand what they need to do with their data, it is less clear that graduate students publishing for the first time have the same knowledge. Studying how researchers gain that knowledge may open an opportunity for the library to support that process or to foster the same kind of learning in other fields. We plan to conduct interviews with researchers in these fields with an eye toward understanding their practices as well as building the libraries’ relationship with those faculty.

Future Directions

We plan to continue to monitor how well researchers are sharing their data, as well as how journals’ data sharing policies evolve. Our ongoing goal is not to get directly involved in enforcing researchers’ compliance with those policies, but rather to assess compliance as a way of identifying where support from the library might be welcome as well as determining in what form that support might be most effective. We expect that the more closely we can target our outreach efforts to the point of need (and, where necessary, make sure researchers are aware they have a need), the better uptake of our services we will see – and hopefully, the better job our researchers will do sharing their data.

Resources Mentioned in the Article


Current Issues and Approaches to Curating Student Research Data
by Andrew Creamer

EDITOR’S SUMMARY
The 2015 RDAP Summit hosted a panel discussion to address issues and challenges of curating students’ research data and digital scholarship, bringing together librarians dealing with various aspects of digital curation and management. Few institutions have a policy on handling student research data, theses and dissertations. Obstacles include insufficient know-how, time, incentive and resources. Without guidance, data files vary widely in format and quality. Yet, as one panelist pointed out, without definitive standards, data librarians must contend with a messy reality and should be reasonable in their expectations for students. Planning for improvements must start by engaging students and advisors throughout the process of data creation. Panelists agreed on the need to further explore topics such as intellectual property, ownership and security, data quality, staff skill building and outreach to stakeholders.

KEYWORDS
data curation
dissertations
institutional policy
intellectual property
data security
graduate students
college students

W
hile there is much consistency in the way that university libraries have collaborated with their undergraduate colleges and graduate schools to archive and/or publish their students’ electronic theses and dissertations (ETDs) online, there is little consistency in the way that universities are handling the archiving, curation and publication of their students’ research data and digital scholarship that underlies these ETDs. The Networked Digital Library of Theses and Dissertations (NDLTD) found that most of their member institutions had no policy on the stewardship of student data related to ETDs, and in the cases where students provided data with their ETDs, the libraries treated them as supplementary files [1]. Several years later, Collie and Witt [2] argued that libraries could seed their fledgling repositories with student research data: “Dissertation datasets represent ‘low-hanging fruit’ for universities who are developing institutional data collections.” (p.166). Yet five years later, university libraries are still figuring out the best approaches for building collections of data related to ETDs. The 2015 Research Data Access and Preservation (RDAP) Meeting’s panel “Current Issues and Approaches to Curating Student Research Data” explored these issues and challenges.

This panel was unique in that it was a hybrid made up of both invited panelists (information professionals who had conducted and presented recent research on the theme) and selected panelists (information professionals who had submitted proposals for papers that demonstrated current and innovative approaches.) The panelists included a digital curation librarian working within his organization to plan a path forward for curating students’ digital scholarship; a research data management librarian and a subject librarian working together to explore a workflow with their campus partners to inform and educate students about submitting their ETD data.
and digital scholarship; a library administrator who conducted an assessment of the supplementary data files that were submitted by students with their ETDs; and a digital repository librarian who is sorting out the complexities of ingesting, modeling and describing student research data files.

Aaron Collie, head of digital curation at Michigan State University Libraries, opened the panel with his presentation, “Building Organizational Capacity for Data Collections Using Electronic Theses and Dissertations.” He reflected on the article he published in 2011 with Michael Witt at Purdue Libraries that I referenced above. In addition to characterizing student ETD data as “low hanging fruit,” Aaron and Michael saw the curation of students’ ETD data as a scale model of the scholarly communication lifecycle, valuable collections that universities should pursue, archive and publish. So why did so many libraries fail to pursue student ETD data and use student research data to seed their fledgling data repositories? Aaron sees organizational capacity as the largest obstacle to building these ETD data collections. He described three organization-level challenges regarding digital curation that libraries need to address: people not knowing how to do the work, not enough time or incentive for people to learn and insufficient resources. According to Aaron, there is an unrealistic expectation held by academic library administrators that they can fit all digital responsibilities and expertise into just one fulltime employee (FTE), adding, “I think digital preservation is a strategic direction that is too often operationalized as an individual responsibility and skill set.” He concluded his talk by sharing the progress he is making at his library to carry out a three-year strategic plan, laying the groundwork for a collaborative approach to digital curation that will put the policies, people and technologies in place to build organizational capacity.

While Aaron’s paper highlighted the challenges that we may encounter inside the library, Dianne Dietrich, physics and astronomy librarian, and Wendy Kozlowski, data curation specialist, both from the Cornell University Library, described in their presentation, the challenges that university libraries face engaging and educating students as well as coordinating with their campus partners as they position themselves to curate and archive their students’ ETD data. Like many institutions, the Cornell University Library collaborates with Cornell’s Graduate School to ingest, archive and publish their students’ ETDs. In this model, there are several nodes of communication along a student’s path to writing his or her dissertation and graduating where students are informed about the guidelines for formatting and submitting their ETDs. For example, students commonly encounter theses and dissertation advisors, graduate program and administrative staff in their departments, and staff in the graduate school and library. Wendy and Dianne described the big challenges of making sure that the stakeholders at each of these nodes are informed about any changes to ETD submission guidelines to include submitting data with ETDs and for getting the word out to students and educating them about managing and sharing their research data. When and how many times should we engage students to prepare them to deposit their data as they move towards submitting their ETDs? In addition to communication and workflow, Dianne and Wendy also presented us with the policy challenges related to ingesting and describing data, such as often overlooked issues related to defining ownership, handling embargo situations and licensing and versioning.

Sarah Shreeves, associate dean of digital strategies at University of Miami Libraries, presented “Supplemental Files for ETDs: Diversity, Documentation and Data,” shifting the focus of the panel from preparing the library, campus partners and students for curating, archiving and publishing student ETD data to looking at lessons learned from libraries that have begun building student ETD data collections. She shared the results of her assessment of the data files that had been optionally submitted by students with their ETDs over several years at the University of Illinois at Urbana-Champaign (UIUC). Students at UIUC have the option to submit data and digital scholarship as supplementary files along with their ETDs; the files are considered appendix items and UIUC ETD submission guidelines inform students about their option to deposit and describe the process and rules for adding these digital objects to their appendices and depositing these as supplementary data files. Between 2010-2014 she found there were 6,472 ETDs submitted by UIUC students to the library. Of these students, 129 or roughly 2% submitted supplementary files along with their ETDs. Of this sample, 94 were students in a program in the sciences, 19 were in a program in the arts and humanities...
and 16 were students in a social sciences program. Eighty-eight students submitted between one and five files, 25 students submitted six to 20 files and 16 students submitted 21 or more files (she found a few instances among these 21 students with 1000+ files). The types of data files submitted included image, text, hypertext, tabular, sound and film files, among others. Roughly 64% could be characterized as data, 15% as code, 7% as data and code and the remainder as protocols and other digital material.

Sarah’s assessment offers information professionals many insights on curating student ETD data. For example, her findings help us to get a rough idea of the percentage of students a library could expect to submit supplementary data with their ETDs if the library were to add this option to its ETD submission guidelines, and it helps us to predict the major file types and average number of files. Her research also helps us to ask important questions about the best ways to curate and describe student ETD data. Should there be more oversight over the documentation quality and quantity students provide with their datasets? Should these digital objects receive their own record and metadata, and if so, what are the best ways to express the relationships among these objects and the ETD? She also highlighted the challenges that we face in preserving student data. For example, can we make the same archival/preservation commitments to supplementary data files that we do for the pdf file of the ETD?

To conclude the panel, Steve Van Tuyl, data and digital repository librarian at Oregon State University, presented “Treating Data Like Data: Unifying Data Processing Workflows for Datasets in the IR,” tempering the expectations for libraries transitioning to ingest students’ ETD data files. He presented a sobering assessment of the student data submitted with ETDs in the ScholarsArchive@OSU. Of a sample of 93 ETDs with related data in OSU’s repository, 45% were Excel files (30% of which had macros, charts and/or linked to other data), 22% were image files and 25% were document files. The remainder of the data included text, database and/or statistical software files, of which 23% were code (and 15% of these executable files), and 12% of the files were metadata. Of the 93 ETDs with data, 30% were unknown, un-operable and/or obsolete; and 3% of the ETDs were missing data files from what was listed among their manifests.

In addition to echoing Sarah’s concerns about the archival and preservation commitments library repositories are making to depositors and whether we can keep these same commitments for their data, Steve also skillfully made the point that we cannot realistically expect data management perfection from students and, that given the great diversity and uniqueness of student research, it is difficult for repositories to have one-size-fits-all, definitive standards for description and curation quality. After presenting us with the messy reality of curating, ingesting and publishing student ETD data (including sharing with us a humorous document file submitted by a student along with his ETD containing a list of songs that he had listened to while collecting field data), Steve encouraged us to walk away from the panel ready to “treat data like data,” treating the curation, ingest and publication of student research data related to ETDs in the same ways we would approach student research data not related to an ETD. He characterized this unified approach as one of iteration and encouragement, working with students to get them to try their best to meet the library’s ingest and description standards and, if necessary, tattling and getting the student’s advisor(s) involved in the process or communication chain.

The panel successfully identified several important issues and obstacles preventing libraries’ curation of students’ ETD data. Issues warranting further exploration and discussion, hopefully in future RDAP meetings, include intellectual property, ownership, copyright and licensing, data management and description quality, privacy and security, communication workflows with students and campus partners, building staff expertise and organizational capacity, embargo policies for student research data, and archival and preservation concerns. The panelists’ consensus is that student data collections are worth pursuing and have much value for the public and research enterprise, but libraries interested in harvesting this so-called low-hanging fruit need to be prepared that this fruit may be a little higher on the vine than previously thought.
Panelists Slides:

Aaron Collie

Dianne Dietrich and Wendy Kozlowski
www.slideshare.net/asist_org/cornell-etd-workflowrdap2015

Sarah Shreeves
www.slideshare.net/asist_org/rdap-15-47572605

Steve Van Tuyl

Resources Mentioned in the Article


You’re in Good Company: Unifying Campus Research Data Services
by Cynthia R. H. Vitale, Brianna Marshall and Amy Nurnberger

Editor’s Summary
With a focus that extends beyond library walls, promoting research data services must be a point of cooperation across an institution’s units and departments. An RDAP 2015 panel explored the value of advisory committees or communities of practice established at three universities to build strong research data services. Though the several committees at the institutions have different primary focus areas, each functions as a community of practice with a unique set of shared interests and goals, mutual relationships and group identification. Each is challenged by its decentralized nature across the institution and must continually communicate its activities, services and resources. These communities of practice are working toward long-term goals through a common vision and strategic planning, sharing experiences and opportunities to promote their research data services.

Keywords
- data curation
- colleges and universities
- library technical services
- collaboration
- communities of practice
- strategic planning
- academic libraries
- information sharing

Research data services are a larger-than-library issue. Experience and scholarship have shown that, given the complex nature of research data services, various units and departments across an institution must work together to provide appropriate, seamless services [1]. One method to facilitate this type of collaboration within a university setting is to form a research data services advisory committee or communities of practice around research data services.

This panel summary, from the Research Data Access and Preservation Summit 2015, focuses on the various committees and groups in place across three institutions – the University of Wisconsin-Madison (UW), Washington University in St. Louis (WUSTL) and Columbia University (CU) – specifically addressing group composition, activities and challenges and opportunities.

A useful framework for understanding knowledge management, community development and roles for research data support in a university environment is through the lens of a community or linked communities of practice (CoP). According to Wenger, McDermott & Snyder [2, p. 4], “communities of practice are groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis.”

In practice, a CoP can be made up of individuals with varying levels and areas of interest in research data. At the University of Wisconsin-Madison an interdisciplinary group of librarians, researchers, IT staff and graduate students constitute their group, called Research Data Services (RDS). Similarly, Washington University in St. Louis initiated a data translation group whose members encompass faculty, library staff, IT staff and other campus data service providers, including representatives from the Center for...
Biomedical Informatics and the Institute for Public Health Data and Training Center. Columbia University’s CoP, in contrast, consists of multiple groups, which similarly include faculty, librarians, data managers and representatives from university administrative offices such as information technology, sponsored projects, compliance and the Institutional Review Board (IRB).

In comparing the practices across these three institutions, librarians, faculty and IT staff are consistently common members when considering CoP for research data services.

The specific activities the CoP undertakes is dependent upon their goals, but Wenger-Trayner [3] indicate typical work includes:
- problem solving
- requests for information
- reusing assets
- coordination and strategy
- discussing developments.

The work of the CoP at UW, WUSTL and CU has some overlap, but also some variation, given the various purposes of the groups.

At the University of Wisconsin-Madison, RDS activities include drafting and reviewing data management plans, consultations on data management policies and best practices, training and education, and referral services to other research data resources on and off campus. For WUSTL, the goal of the data translation group is to discuss integrated approaches for moving data to information for larger collective impact. Given this goal, many of the activities center on three main categories: research, curricular and experiential learning. A number of specific outcomes have developed from these categories, including a university-wide data challenge and a new group focused on developing data-related curriculum and training for undergraduates and graduate students. Good research data management (RDM) practices and data related skills are woven throughout each of these outcomes.

Columbia University supports three groups created from members with a variety of experiences and perspectives on managing research data and materials: Columbia University Libraries/Information Services Research Data Management Interest Group (CUL/IS-RDM-IG), CU RDM community and the Research Data Advisory Committee (RDAC). The object of this variety is to develop CoP at different scales and levels, so that as they evolve there are points where their typical work and shared purposes may intersect and where they may benefit from the opportunities for interaction and support. The CUL/IS-RDM-IG focuses on knowledge sharing and skill-building. Adjacent in purpose to this group is the CU RDM community, which was founded on the basis of assembling those who are actively involved in managing data or research materials as part of a research or scholarship endeavor to find solutions for common challenges in RDM. RDAC is a newly established committee brought together to provide leadership in developing strategies, services and programs for the university to cultivate its already established investment in research data. The intent is that these groups will find compatible partners and rich sources of experience and resources in each other while engaging in fruitful correspondence and collaboration that will ultimately benefit the whole of the university community as it engages more deeply with the challenges of successfully working with research data and materials throughout the research life-cycle.

The success of a CoP is predicated on a number of factors according to Lee-Kelley, Turner and Ward [4] and Wenger [5]. They found that a successful, self-sustaining CoP includes “interaction, sharedness (of knowledge, interests and goals), ongoing participation, mutual relationships, personal identification of members with the group and the development of a continuous and conspicuous shared repertoire.” [4, p. 48] While the CoP success factors are accurate for groups and committees focused on research data services and support, they do not address the challenges found in progressing toward the attainment of these success factors.

At each of the three institutions the decentralized nature of the universities provides some challenges to creating CoP sharedness. Negotiation takes place when bringing together individuals with various backgrounds, domain expertise and technical knowledge around everything from term definitions to desired outcomes. It is only through ongoing participation that a balance of the interests and needs of committee members emerges. Strategies to overcome this challenge have focused on developing strategic plans for the CoP and committees. By developing strategic plans,
disparate group members create a shared vision of the committee, agree upon success factors, define roles, and take ownership of related strategic activities.

Another recurring challenge involves making CoP services and related activities known to the larger university community. Very often faculty members and research staff are surprised to learn of the services and resources that result from committee activities. To improve visibility of these services, university committees have initiated a cycle of assessment and re-branding or re-launching of the activities.

In comparing the composition, goals, activities and challenges of research data services committees or CoP across universities, it is evident that there are a number of common experiences that may be found. By sharing these experiences, as well as the commonly developed strategies and approaches, the authors hope that creating an ongoing conversation around sharing experiences and opportunities will assist others in shaping support and providing seamless service to faculty at their own institutions. Establishing this practice of sharing and conversation in our own community may in turn lead to broader coordination that enables stronger partnerships with faculty as research data services continue to evolve in the academic landscape.

Resources Mentioned in the Article


BYOPanel – The On-the-Spot Assembled Panel on Responses to the OSTP Memo Responses

by Margaret Janz

In February 2015, while RDAP planning was well underway, federal agencies affected by the Office of Science and Technology Policy (OSTP) memo of two years earlier began to release their responses to it. Many of the librarians who were (and continue to be) tracking the release of these documents are also RDAPers. As such, a conversation on Twitter led to the request for an #OSTPResp panel at the summit for which planning was then ongoing. The program had already been set, but the RDAP15 planning committee decided to see how and when we could fit such a session into our two-day conference. We had 30 minutes at the beginning of the second day open that we’d previously been hoping to fill with a second keynote speaker and we decided this would be our best bet.

The planning committee came up with three questions to be answered by attendees:

1. How are you educating your community about the new data policy requirements?
2. What resources have you developed about the OSTP responses?
3. What are some of the emerging issues that might be challenges for either your researchers or those who support data management at your institution?

On the first day of RDAP, we asked RDAPers standing in the hallway during the first break to answer the three questions on slips of paper. The responses were hung on poster board (on the backs of old event signs, thanks to Hyatt Minneapolis staff) and at lunch we asked everyone to put stickers next to those they wanted to hear about the most. The most voted for answers were as follows:

- University of Iowa’s creation of a Research Data Interest Group
- National Institute of Standards and Technology’s taxonomy of research data and archiving requirements
- James Madison University’s efforts to inform stakeholders on the OSTP mandate

At a short session the next morning, speakers from these institutions expanded on their first brief responses, sparking further questions and plans to carry on the conversation through social media and future meetings.
Amanda Xu from the University of Iowa:
Q: How are you educating your community about the new data policy requirements?
A: Establish Research Data Interest Group with stakeholders from across campus, e.g., libraries, ITS, Office of Sponsored Research, digital humanities

Regina Avila from the National Institute of Standards and Technology (NIST):
Q: What resources have you developed about the OSTP responses?
A: NIST has developed a taxonomy of research data (pyramid) that describes consequences or requirements of metadata/archiving at each level

Yasmeen Shorish from James Madison University:
Q: What are some of the emerging issues that might be challenges for either your researchers or those who support data management at your institution?
A: Bringing all stakeholders (OSP, faculty, admin) up to speed on all the various req’s. I’m the sole conduit for this info (currently) and I’m concerned about message penetration.

We asked each writer to speak the next morning on our DIY panel to expand on their answers. Avila and Shorish agreed and Xu asked her colleague Sara Scheib to speak about University of Iowa’s activities. And so the panel was set.

And Then What Happened?

Scheib told us that the librarians at University of Iowa had reached out to other departments on campus to create a research data interest group. They have representatives from Information Technology Services, their Office of Sponsored Research, digital humanists and others in the group. Since the OSTP responses have come out, they’ve been meeting to go through each document line by line to make sure everyone is interpreting the implications the same way.

Avila was able to give a sneak peak at the as yet unpublished taxonomy that’s been designed by NIST. The new taxonomy gives seven levels of data: working, derived, publishable, published, resource, reference and standard reference. Avila also shared a pyramid diagram (Figure 1) to describe different data categories within the NIST data taxonomy. The purpose of this taxonomy was to define a collection of terms and concepts that describe categories of scientific data that result from NIST research. These classifications provide a vocabulary that simplifies discussion at the agency when assigning requirements to various categories of data. The taxonomy was later published as an appendix to the NIST order “Managing Public Access to Results of Federally Funded Research.” In addition to the pyramid, the document suggests consequences at each level for data preservation, review and discoverability. To review the order and appendix, visit http://go.usa.gov/c7aCW.

Shorish gave an answer that spoke to a theme of the Summit: being a research data services department of one. Her greatest challenge is getting
the word out about the OSTP responses. Even with support from her Office of Sponsored Projects it is a feat to keep track of all the varied proposed policies and let all of the researchers know about the changes to their responsibilities.

The session was very short with an unfortunately small amount of time for questions. Hopefully the gracious panelists were able to spark some ideas on partnering with other institutional units and offer some comfort to attendees in the knowledge we are not alone in responding to #OSTPResp—although we might be alone in our institutions.

The RDAP15 planning committee had a final question for attendees that we wanted to discuss at the end of the panel. Regrettably, 30 minutes is not a very long time at all, and the question was thrown to Twitter and informal conversations during breaks. I’d like to ask it again here to continue the important conversation around it: How can the RDAP community help support you on OSTPResp issues?

Talk amongst yourselves or tweet to @RDAPsummit or email the past or future planning committee members. All responses to questions asked for this panel can be found here: bit.ly/rdapostp.
What is the landscape of policies covering research data and library data services at major U.S. research universities? This lightning talk discussed that question by reviewing the prevalence of university policies on data management topics such as stewardship, ownership, retention and sharing, and the existence of data services from the university library. Our research reviews the websites of 206 institutions with a Carnegie Classification on Institutions of Higher Education for research level of either “High” or “Very High” (as of July 2014). In addition to examining library data services and university data policies, we examine the content of the policies themselves. What is included in the policy? Who hosts the policy (for example, office of research, information technology, libraries)? What happens to the ownership and management of the data when a researcher leaves the institution? Are universities with data management services provided by the library more likely to have a policy on data management? At RDAP 2015, we presented our environmental scan of current library data services in the context of institutional policies and summarized the findings for future research in policy and service development.

The complete report of our research into these subjects will be available Summer 2015 in an article in the *Journal for Scholarly Communication and Librarianship*.

**Who Owns Your Data?**

This is the question we asked each other just over a year ago when we met for another symposium on research data management. It is a big question. While it may be somewhat clear for funded research, it is often not clear for unfunded research and pilot projects. Other questions then pop up, usually around policy:
What happens if a researcher leaves the institution?  
What if someone needs access to my data?  
How long do I have to keep them and how should I discard them?  
How do I set policy for my team, my lab or myself?  
But I’m collaborating: whose policy wins?  
What if there isn’t a policy?  
What help can I get?

There are many more questions, and so we sought to find answers, both for our own sanity and to help support our organizations. We determined that we wanted to understand the current landscape of university data policies as well as what library data services were available to support these policies.

A Look at Research Institutions

We decided to look deeper at universities classified by Carnegie as high and very high. Other metrics for these institutions are ARL membership, research expenditure, faculty size, student population and public/private status. We looked at all publicly accessible university data policies including but not limited to intellectual property (IP), technology transfer, copyright, information technology and other governance policies. We only analyzed policies that specifically covered research data. This analysis was a huge undertaking, and we faced uncertainty about the complete number or breadth of policies because some institutions have their policies behind an institutional log in.

Library Data Services

We also wanted to know what data services were provided by the libraries at these institutions. We again looked only for publicly accessible information, usually through the library websites. Occasionally, we found some library data services in other areas such as the office of research or information technology. We looked for services such as a data repository, data librarian and specific data services such as data management plan (DMP) consultations. This was also challenging since some institutions post on library websites, while others utilize content management systems such as LibGuides.

Currently, about 50% of the libraries surveyed offer some form of data services that go beyond simply providing links to external resources via a resource guide. Nearly 40% of the libraries have a librarian either entirely dedicated to research data management initiatives or someone for whom it is a major portion of the job description. Unsurprisingly, the latter are frequently science librarian positions. Roughly 10% have a dedicated data repository, though this number may be inflated due to consortial arrangements such as the Merritt repository, which services all of the University of California.

Data Policy Landscape

In terms of the landscape of institutional data policies, our findings are not surprising, but the process did open our eyes to the challenges that researchers, librarians and institutions face when trying to meet funder or journal requirements on public access. This is further complicated as researchers are attempting to navigate policies that many do not know exist, have not read or may not exist at all for their institution.

There were a few correlations we found in this process [1]:

- Universities that perform more research, as measured by these characteristics, offer more data services:
  - Higher Carnegie Classification  
  - ARL membership  
  - Higher research expenditure  
  - Larger faculty size

- Universities with more research, as measured by these characteristics, are more likely to have a stand-alone data policy:
  - Higher Carnegie Classification  
  - ARL membership  
  - Higher research expenditure

- Universities with data services or a data librarian are more likely to have stand-alone data policy

- Universities focus on legal repercussions of research data
  - IP policies covering data are concerned with data ownership
  - Stand-alone policies cover access, retention, separation, etc.
We also found differences in institutional data policy and funder data policy content. These variations offer an area for future analysis. Of note, this analysis found that top research institutions almost universally offer research data services. Will it soon be expected for all academic research libraries to provide data services? With the recent growth in both data services and data policy, will we see all research universities build services in this area?

These data suggest that universities’ significant concern is ownership and legal issues with regard to data, while libraries are developing programs and services aimed at the entire data lifecycle. This division offers libraries and librarians an opportunity to provide important clarity where policies exist or advice in policy development on their campus.

What’s Next?

While all of our questions were not answered in our initial query, we have a roadmap for demonstrating clear areas of further need at our institutions. This challenge is one that all research universities are presently facing, and librarians can lead the way in developing a response that considers both the researcher and the institutional point of view. Librarians can lead policy development; educate faculty and administrators about best practices; and determine how to navigate the myriad policies from funders, journals, collaborating institutions and more. We hope to see more librarians engaged in institutional research data policy development in the future.

Resource Mentioned in the Article

Research Data Services at the University of Colorado Boulder
by Shelley L. Knuth, Andrew M. Johnson and Thomas Hauser

EDITOR'S SUMMARY
The University of Colorado Boulder's Research Data Services (RDS) is a joint activity of the University Libraries and the Office of Information Technology. Started in 2011 to meet the National Science Foundation's mandate for data management plans with all grant submissions, RDS grew from helping to write the plans to indicating resources to educating researchers on their use and providing individual consultations and reviews. RDS staff also aid researchers with metadata documentation, helping to develop file documentation and later migrating to final formats. Researchers often contact RDS about long-term data storage or for assistance with data dissemination and complying with the requirements of data repositories, funding agencies or scholarly journals. Urging researchers to learn about data management, using a variety of communication channels and venues, is key. Incorporating a data management plan requirement to in-house grants and creating a data management plan competition with financial award have been most successful at promoting the RDS.

KEYWORDS
Data curation    Documentation    Metadata
Library technical services    Digital repositories

Shelley Knuth is a senior research data specialist at the University of Colorado Boulder in research computing, where she assists campus researchers with their research data needs, including data management, education and storage. She can be reached at shelley.knuth@colorado.edu.

Andrew Johnson is research data librarian at the University of Colorado Boulder Libraries where he develops and implements research data services in collaboration with campus partners and co-chairs the campus Research Data Advisory Committee. He can be reached at andrew.m.johnson@colorado.edu.

Thomas Hauser is the director of research computing at the University of Colorado Boulder, where he was instrumental in establishing centralized storage for research data and creating the Research Data Services. He can be reached at thomas.hauser@colorado.edu.

W
While research funders and journal publishers now encourage or mandate data management and sharing, researchers are often not formally trained in these practices. As a result, many universities have begun to develop programs to assist faculty, staff and students with these needs. One such effort, Research Data Services (RDS) at the University of Colorado Boulder (CU-Boulder), is a collaborative activity between research computing (RC), a division of the Office of Information Technology, and the University Libraries. Similar to other institutions, the range of data services provided includes assistance with writing data management plans (DMPs), data storage and repository advice, data processing and other needs related to research data management. In addition, RDS has experimented with a variety of novel approaches to outreach and engagement across all disciplines at CU-Boulder and with affiliated institutions in the surrounding area. The history, services, outreach and education efforts of the RDS program at CU-Boulder are described in the sections that follow.

History
RDS was developed in 2011 in response to the National Science Foundation (NSF) requirement for DMPs to be included with all grant proposals [1]. To meet the new requirements, the original RDS was developed initially to help with DMP writing by repurposing existing positions in RC and the libraries. In addition to providing DMP templates via the DMPTool [2], RDS would meet with campus personnel to help them understand the components of the DMP. As a result of a campus-wide Data Management Task Force report [3], a governance structure and additional services were added to RDS to expand offerings, while keeping the primary mission intact.
Two CU-Boulder committees provide oversight and strategic direction for RDS. RDS reports directly to the Research Data Executive Committee (RDEC). The members of this committee consist of RDS staff, the director of research computing, the senior associate dean of libraries and the associate vice chancellor for research (AVCR). This committee develops an implementation plan based on direction provided by members of the Research Data Advisory Committee (RDAC). The RDAC committee consists of members of RDEC plus researchers and support staff from a wide range of disciplines across campus. RDAC also includes data managers and curators from local and regional data centers, including the National Center for Atmospheric Research (NCAR) and the National Snow and Ice Data Center (NSIDC). The purpose of RDAC is to provide strategic direction for RDS and to provide guidance on issues most relevant in specific disciplines.

Services

The CU-Boulder RDS offers several services to its researchers, and most are free of charge. Outreach efforts, described below, have identified new future opportunities to be explored to better reach the research community at CU-Boulder.

**Data Management Consulting:** At the core of RDS are general data management services, particularly those related to assisting campus personnel with writing DMPs for grant proposals. The DMP can be the first exposure researchers have to the concept of data management, and it is important to properly lay the foundation for better data management strategies for the future. Because researchers are often the first, and for a long time only, stewards of their data, instilling good data management practices early benefits everyone. This process involves coaching on using non-proprietary data formats, understanding how the size of data can limit use, proper data collection processes and proper formatting. Often, researchers do not involve RDS until shortly before proposal submission, but this point is still early enough to produce guidance on good data management practices. CU-Boulder RDS uses these opportunities as teaching moments to encourage good practices for the future. The DMP service is primarily centered on one-on-one consulting, but requests to review draft DMPs via the DMPTool also occur.

RDS also assists researchers with metadata documentation. Generally, CU-Boulder RDS wants to encourage researchers to document their data early and often. RDS has discovered that asking researchers to learn a new language, such as XML, to document data usually results in no documentation being written. As such, RDS encourages researchers to utilize whatever means they are most comfortable with to document their data, which oftentimes is a simple text file describing their dataset. By making this process simple for the researchers, the important information about the dataset is more likely to be captured, and the text files can be converted to other formats without the need for further input from the researchers.

**Data Storage Archiving and Dissemination:** Many researchers first find RDS when looking for a place to house data for the long term. RDS staff will provide advice on matters related to data storage, archiving or curation, and will assist CU-Boulder researchers with finding a long-term solution for their data depending upon the level of storage they require. Generally, RDS encourages researchers to store data in discipline-specific locations – in essence, where they themselves would go to find data. Some researchers wish to house their data on their own managed systems; while RDS generally does not have the opportunity to connect with these researchers, we still encourage the use of proper data management techniques, no matter the researcher’s preferences or needs.

In some cases, discipline-specific storage facilities may not be beneficial to the researcher, may not exist or may not be cost effective. In these instances, we may also assist with in-house solutions. One local resource the RDS utilizes is the PetaLibrary storage infrastructure, which is an NSF subsidized service for storing research data that is managed by RC. The PetaLibrary offers a minimum of 2 terabytes (TB) of storage to any U.S.-based researcher affiliated with CU-Boulder. Each researcher pays a nominal fee, depending on the level of service, to store data on the system. The PetaLibrary has two main classes of storage available, called active and archive. Active storage keeps data on spinning disk, intended for data that is
frequently written or read. The data can be snapshotted as part of this service. A second level of service, archive, is stored on a combination of disc and tape. Under this service data not accessed after a period of time is migrated to tape. Users can also do various combinations of active and archive data storage, depending on their needs. These options include replication or off-site duplicate storage. Additional information on the PetaLibrary is available at https://rc.colorado.edu/resources/storage/petalibrary.

Another point of contact with researchers involves data dissemination. While some researchers willingly share their data, others only share data because of funding agency or journal requirements. In either case, RDS will assist researchers with finding the proper venue. For some, the process of storing data in a well-managed archive or storage facility can also be the venue for data sharing, particularly in proper data repositories. RDS encourages the use of digital object identifiers (DOIs) to ensure long-term access, and we promote data storage, archiving and preservation locations that offer DOIs. RDS also provides advice on how to prepare data for long-term storage, including proper formats and metadata as described above. Issues of data security can also be discussed with RDS, who will refer researchers to the appropriate offices on campus to handle these types of data. CU-Boulder's local resources, such as the PetaLibrary, are currently not equipped to house secure data and do not attach DOIs to data.

Education and Outreach

CU-Boulder RDS spends a large portion of time on outreach efforts to encourage researchers to learn techniques to better manage their data. Despite the increased use of digital data and data management requirements by funding agencies, researchers may not utilize university RDS effectively due to various reasons, such as lack of knowledge about campus data management services or little incentive to improve their data management practices. CU-Boulder RDS has approached these concerns with ideas to promote awareness and to entice the utilization of data management services with great success. To promote awareness, in 2014/2015 RDS set up accounts on social media, began utilizing campus and department newsletters, hosting workshops and contacting department chairs directly via email. We have also made an effort to attend departmental and institute faculty meetings on campus to introduce researchers to RDS services. This outreach has resulted in several consultation requests from faculty.

One of the more popular activities has been the introduction of data workshops across campus. Since May 2014, approximately 15 workshops taught by RDS have provided information on a wide variety of topics, including best practices in data management, how data publishing benefits careers, federal funding agency requirements around data, learning the difference between specific data formats, data transfer and storage, and how to write good data management plans. Promotion of these workshops takes place through campus newsletters, social media, email listservs and emailing the chairs of departments on campus. The most popular workshop to date has been Best Practices for Good Data Management, in which we discussed the various components of a NSF DMP and general good data management practices (slides: http://researchcomputing.github.io/meetup_fall_2014/pdfs/fall2014_meetup11_data_management.pdf).

Another activity for increasing visibility across campus is to provide financial incentives. RDS, in conjunction with RDEC, RDAC and the office of the vice chancellor for research (OVCR), developed two funding opportunities for researchers on campus to encourage good data management practices. The first was to add a DMP requirement to an existing internal grant competition – Innovative Seed Grants (ISGs). The ISGs provide seed money for innovative and collaborative research projects, in part to improve faculty competitiveness for federal grants. The addition of a DMP as part of the ISGs forced faculty to give more consideration to how they will manage their data as part of their research. Since faculty from all disciplines apply for these grants, the addition of a DMP requirement expanded outreach to faculty who normally do not apply for grants with data management requirements. In 2014, the first year of this requirement, the DMPs were only reviewed by a committee to provide feedback to the researchers. By 2015, the DMPs were both required and counted as a small part of the applicant’s score. The addition of the DMP as part of the ISGs was the most impactful action for promoting awareness of data management issues to-date. RDS received requests to review approximately 20% of
DMPs to be submitted as part of an ISG proposal. OVCR also requested RDS organize educational workshops on writing good DMPs. Approximately 20 faculty members attended these workshops.

The second funding opportunity developed by OVCR and RDS was called the Best Digital Data Management Plans and Practices Competition. This opportunity allows faculty, postdocs, research staff or graduate students to submit a DMP to the challenge, where the winners in five disciplines (arts and humanities, social sciences, life sciences, physical sciences and engineering) receive $2,000 in unrestricted university funds. The purpose of this competition is multifold – first, to promote conversations about good data management practices; second, to collect a bevy of plans that could be distributed as exemplary submissions; and third, to increase awareness of RDS. The 2014 competition drew approximately 15 submissions. A 2015 competition is already underway. Overall, tying data management

requirements to financial incentives was the most successful outreach effort organized by RDS.

Summary and Future Work

RDS at CU-Boulder, with support from the OVCR, has made great strides toward promoting data management awareness on campus since its inception in 2011. RDS has developed new funding and educational opportunities and made a strong effort toward promoting services through various outreach venues. RDS hopes to expand future services on campus by providing more tailored services to campus researchers, such as visualization and analytics services, and additional services focused on the humanities and social sciences.

Acknowledgments

The authors acknowledge support for the PetaLibrary storage resources, which were provided by NSF-MRI Grant ACI-1126839.

Resources Mentioned in the Article


DLF E-Research Network Members Attend 2015 RDAP Summit
by Gail Clement and Rita Van Duinen

To address the continued need for libraries to be engaged in developing research data management services (RDMS), the Council on Library and Information Resources (CLIR) and the Digital Library Federation (DLF) have established the DLF E-Research Network. Launched in 2014, the E-Research Network was designed to help members of academic and research libraries develop strategies for implementing e-research and research data management support services through peer-driven, shared learning experiences and through collaborative efforts across institutions. The goal of the network experience is to encourage a self-reliant, mutually supportive community.

As E-Research Network members, institutional teams are given formal and informal opportunities for networking, resource sharing and collaboration, supported by CLIR/DLF’s organizational resources, as well as access to structured curricula, webinars and personalized consultations. Through in-person meetings and shared learning activities and experiences, the DLF is building an active and growing community of practice.

Network members come from colleges and universities of varying size. To date, 13 institutions from across the United States and Canada have participated in the E-Research Network:

- California Institute of Technology
- Colgate University
- Montana State University
- Northwestern University
- Temple University
- University of Arizona
- University of Florida
- University of Illinois Urbana-Champaign

Geochemist-turned-librarian Gail Clement (http://orcid.org/0000-0001-5494-4806; @Repositorian) is head of research services at Caltech and may be reached at gclement<at>library.caltech.edu.

Rita Van Duinen is the curriculum and research strategist for the Council on Library and Information Resources (CLIR) and may be reached at rvanduinen<at>clir.org.
The California Institute of Technology (Caltech) team has a unique perspective on what it means to be a part of the E-Research network. The Caltech Library joined the 2015 E-Research Network cohort with the aim of assessing and enhancing the rich array of research services provided to their campus community (both E- and otherwise). The Caltech team is made up of Donna Wrublewski, chemistry librarian and information specialist as well as CLIR/DLF postdoctoral fellow; Gail Clement, head of research services and librarian for earth and planetary science; and astronomer George Djorgovski, director of the Center for Data Discovery, who serves as faculty advisor.

Caltech is an atypical university because the extensive and world-renowned research footprint combines with a small and intensely challenging academic program (2000 students; 300 faculty) to deliver impactful research outputs from across the entire campus population. Many investigations take place in global “Big Science” facilities such as the Seismological Laboratory, the preeminent source for southern California earthquake information; the International Observatory Network operating telescope arrays around the world; and NASA’s Jet Propulsion Laboratory, home of the Mars Science Laboratory. Other research operations on campus represent smaller science – the long tail oft discussed in the research data community. Investigations at Caltech are typically interdisciplinary and collaborative, usually extending beyond campus and geographic borders. The roles and responsibilities of the various researchers working on a given investigation therefore presents a complex and fascinating puzzle – one Caltech needs to unpack and understand in order to serve their diverse population of users effectively.

The Caltech library’s focus on user needs relies on a robust liaison program in which many of their librarians hold sci-tech degrees, have professional experience as working scientists or engineers and/or actively serve on committees and boards that advise scientific societies and publishers. The embedded science librarian model at Caltech means that each information professional offers a vertically integrated suite of services and resources for a particular disciplinary division. These services cover the gamut from collection development to open access publishing, from enhancing scholarly reputation to providing good old fashioned bibliographic reference management and online literature searches. Research data management and sharing is one important component of that vertically integrated suite. In Caltech’s view, research data support is a point on a continuum of exemplary research services they provide and demands integration into a larger context of programmatic objectives: authorship services, research effectiveness, knowledge management and born-digital research tools and infrastructure.

Engaging with the DLF E-Research Network (and the library research data movement generally) ensures that visioning, planning and implementations in support of research data are informed by the best thinking in the profession. The E-Research Network meetings and webinars have already provided Caltech opportunities to consider questions that other research libraries are asking and answering:

- What role can the library play in research data support for their campus?
- What resources and expertise do they have that complement what other campus entities already have capacity for (or are already doing)?
- What forms of needs assessment for research data support actually work in practice?
- What local infrastructure is most effective for storing research data that does not go to a disciplinary data center/subject repository (the long tail)?
- And others!

Additionally, Caltech has questions and concerns about particular aspects of research data support that cannot be addressed in isolation. Caltech is looking to the larger research data community (DLF E-Research, RDAP Summit, Research Data Alliance) to gain insight and identify solutions to the following questions:

- What mechanisms are in place to address uncertainties in research data management and sharing, such as ownership vs. contributorship, intellectual property rights, and other ethical and legal issues?
What rights statements should our diverse campus stakeholders, as data producers, providers, aggregators and publishers, attach to their outputs to enhance reuse?

Who is expected to pay the considerable costs for research data management and publication within the American scientific enterprise? Government, universities, individual principal investigators?

And others!

The resources and people encountered in the 2015 DLF E-Research cohort have already proven to be a valued source of knowledge and experience as Caltech explores the possibilities for library engagement with research data management.

At this year’s RDAP Summit, members of the 2015 E-Research Network cohort gathered in person and virtually to discuss the current state of RDMS at their institutions and to share expectations as network members. Those that attended the cohort’s kick-off meeting in person stayed on in Minneapolis to attend the Summit. This first meeting was intentionally scheduled in conjunction with RDAP, given the breadth and depth of the conference program and its relevance to the mission and goals of the network. Having network members meet early on in the cohort and then attend the conference together provided a sense of community and a shared experience as members listened and learned throughout two days of programming.

Two weeks after attending RDAP, E-Research Network members gathered virtually for a conference debriefing. During this session members identified and discussed certain conference themes that stood out as most relevant to the group. Included in the debriefing session were the following RDMS themes:

- Outreach – defining the service and message
- Teaching – What does data literacy look like? How can we build a curriculum?
- Assessment – What does RDM success look like? What types of things/files are part of the data spectrum?
- Ingest – How do we recruit and store data?
- Partnerships – Where are the efficiencies in the current RDM profession that people can share to avoid reinventing the wheel?

The debriefing session proved to be beneficial given the varied levels of existing RDMS across the member institutions. Those that are more advanced share their experiences and resources while also encouraging and advising those institutions less advanced. Institutions in the early stages of building RDMS look to the rest of the group for leadership. The community building taking place as a result of the E-Research Network is a result of many of the themes taken away from the RDAP Summit. The Summit provided network members with a wealth of learning opportunities while also giving this budding community of practice the opportunity to collectively address and explore ways in which they can work collaboratively in solving RDMS issues.

This was the first time CLIR/DLF has been represented at the RDAP Summit; by both the network’s cohort meeting and the subsequent poster session. The Summit was an ideal venue for interacting with the RDAP community and for sharing the work of the DLF E-Research Network.

DLF offers new cohorts of the E-Research Network on an annual basis. For information on how your institution can participate in future course offerings, please contact us at info<at>diglib.org.
The longer I do this work, the more I see the need. Throughout our varied work in the information fields, we spend a lot of time dealing with both the details and the big ideas. We think about the theories of categorization. We contemplate the many ways to structure metadata. We focus on the psychology of learning. These insights help us better meet the information needs of our audiences: our readers, our researchers, our colleagues, our clients.

But every day, I’m reminded again how much our work is needed by the world. How much more work there is to do. At the beginning of the summer, I took my daughter to get her driver’s license, prepared with necessary paperwork and a scheduled appointment, but no preparation could contend with the disarray, lines and lack of organization, signage and information on site. Reassurance from help desk staff was false and misleading. The visit, though ultimately successful, was frustrating, full of wasted time and a stark reminder of the prevalence of bad, poorly presented or missing information without enough information specialists to fix it all. A book by Abby Covert, How to Make Sense of Any Mess, presents useful advice for information professionals facing everyday information challenges.

But every day, I’m reminded again how much our work is needed by the world. How much more work there is to do.

At the beginning of the summer, I took my daughter to get her driver’s license. While she’d received a learner’s permit a year ago to allow her to learn to drive, now it was time for her to get the real license, to be able to drive on her own.

We received our first information related to the experience by word of mouth. A friend mentioned to me in the spring that he’d been surprised to discover there was a six-week wait to schedule a road test for new drivers, so I should visit the state’s website immediately if I hoped for my daughter to take the test and get her license as soon as she was old enough. I was at a baseball game at the time, but I was able (with some degree of backtracking and effort, but successful nonetheless) to schedule the appointment right then on my smartphone. Whew!

Each state in the United States has somewhat different rules about how you get a driver’s license, and so when it came time for the test, my daughter and I both read the state’s website carefully to ensure she had all the necessary paperwork to take with us. In retrospect, it should have been a warning sign when she and I came away with entirely different ideas about what paperwork would be required. We gathered all the documents we each thought we’d need, and we arrived 15 minutes early for her appointment.

When we got there, there was a line out the door, at least half the length of the building down the sidewalk. I asked the people in line, “Do you stand in line if you have an appointment?” No one knew. I looked at the door – no sign or information. I sent my daughter to stand in line, and I went inside to look for more info. Nothing. The long line ended at the Information Desk. There were also dozens of chairs in the room, nearly all of them occupied. I asked a few people inside, and I got both yes and no answers, and several “I have no ideas.”

So I broke in the front of the line to ask the woman staffing the information desk whether we needed to stand in line if we already had an appointment. She said, “Yes, but they’ll call your name while you’re in line.” What?!

That really made no sense to me, but nonetheless, we’d
received the official word from an employee of the driver’s license bureau, so we remained in line for an hour. We finally got to the information desk ourselves, whereupon we were informed we’d missed our appointment.

It turns out they believed they’d called my daughter’s name much earlier, while we were in line and listening for them to do so, but the officer who’d called it out so badly mangled it (my daughter’s name is even easier to pronounce than mine, though we have different last names), that neither my daughter nor I recognized it as her name. The officer was originally going to make us wait six more weeks for a new appointment, but after begging, pleading and threatening, we managed to convince them they should test her that day. (And she passed the test with flying colors. Thank goodness!) And in the end, they didn’t even look at any of the “required” paperwork we brought with us. When we left, my daughter asked me why I was so angry about the whole thing. After all, it had worked out, and she could now drive!

I told her that it frustrated me that all of the confusion could have been avoided with a little better information. It was obvious to the license bureau staff how the process worked, and based on our conversation, they clearly believed they’d explained it well. Their curse of knowledge prevented them from seeing our confusion. It would have been great to have a sign or two explaining a process that made sense. What if they’d assigned us a reservation number when we made the appointment? What if there had even been a low-tech sheet of paper where we could sign in to let them know we’d arrived? I had many solutions.

We all encounter situations like this on a regular basis. There aren’t enough information professionals in the world to fix them all.

That’s why I am so fond of Abby Covert’s recent book, *How to Make Sense of Any Mess* (http://abbytheia.com/makesense/). It ought to be part of the required reading for 101 classes in any information-related discipline. It will have the seasoned professionals among us nodding along on every page. But it’s also entirely understandable for someone whose calling in life comes nowhere near our profession. We’re all faced with information challenges every day. Most of them are pretty messy.

Covert divides her work into several sections, focusing on identifying the issue, clarifying goals, understanding the audience, getting your arms around the context and building a solid information structure for any project. Then she helps you figure out how to adapt that framework to a rapidly changing world, and most importantly, communicate and collaborate with others.

The highly digestible book treats each page as a mini-essay, with a principle and an explanation. You can dip in to grab an idea or two and get back to work or read the whole thing through in just a couple of hours. I don’t think you will, though – it’s the kind of book that makes me stop every few pages, because I’m ready to put the ideas to work on a thorny issue I’ve been dealing with.

The thing is, everyone has to communicate with other humans, every day. And most of the time, we’re blissfully unaware whether we’re actually speaking the same language. Many times we all may nod in agreement, not realizing we don’t fully understand each other.

There will always be a role for information professionals. Our work should be in greater demand every day, as the world grows more complex. But we also ought to continue to find ways to share with others the tools we’ve learned for sensemaking. Covert’s deceptively simple book is a wonderful tool for information pros and laypeople alike.