# Research Data Access and Preservation Summit

**2019-05-15 – 2019-05-17**

**Coral Gables, FL**

---

## Day 1 | May 15 2019

## Shalala Student Center
## 1330 Miller Drive
## Coral Gables, FL.

---

| | |
|---|---|
| 9:00 – 9:30 | Registration, coffee, and light morning food |

---

| | |
|---|---|
| 9:30 – 10:45 | **Opening** |
| | *Tisha Mentnech, RDAP Summit co-chair (welcome)* |
| | *Kristin Briney, University of Wisconsin-Milwaukee* |
| | Community-building icebreaker |

---

| | |
|---|---|
| 10:50 a.m. – 11:50 a.m. | **Repositories and curation panel** |
| | Thinking outside the repository: Supporting partial curation |
| | *Greg Janée, University of California, Santa Barbara* |
| | |
| | 15 min |
| | |
| | The library community has coalesced around the repository as the principal embodiment of data curation. Data is curated if it is in – and curated by placing it in – a repository. The repository's role in the context of curation has benefited from work including specifications (OAIS), open |

source toolkits (Fedora), turnkey solutions (DSpace), communities (Samvera), and certifications (CoreTrustSeal). This is all good but the danger of equating curation with storage in a repository is that the data types and use cases that don't very well fit the repository model are excluded. These include datasets that continuously evolve and grow over years, even decades, data for which only partial curation is desired, and most importantly, data that is produced by researchers and organizations with limited means and resources. Often, such data is left entirely uncurated. We seek to define curation as a set of characteristics independent of the repository. The FAIR principles are four characteristics that define a modern notion of data reuse. We add six attributes that, in aggregate, capture additional key aspects of curation: Inventoried, Owned, Controlled, Isolated, Structured, and Robustly stored. These concepts are not new. Repository systems implement these and certifications mandate them. The purpose of enumerating these outside the context of repository systems is to allow for their implementation via new, novel, and possibly independent means, and to define new, intermediate levels of support in which data is only partially curated. In this way we hope to see fewer datasets that are entirely uncurated, particularly those originating from researchers and organizations with limited resources.

## Using information architecture to combat biases in information systems
*Megan O'Donnell, Iowa State University Library*

15 min

This presentation will explore how information architecture choices can be used to reduce the harm of biases found in information systems. The presenter's experience designing the organizational infrastructure of a new data repository will be used as an example of how seemingly simple design decisions can have unexpected ramifications when examined through an intersectional feminist lens. This talk will share how problems were discovered, what decisions were made, and how the process exposed a growing need for flexibility within information systems in order to counter various system, organizational, and culture biases.

## Developing metadata curation processes for data that can't be shared openly
*Rebecca Grant, Springer Nature (with Graham Smith and Iain Hrynaszkiewicz, Springer Nature)*

15 min

This presentation will describe the development of a metadata curation process for data which cannot be shared openly and provide examples of how the application of this process has increased the accessibility of these sensitive datasets. We welcome community discussion and review of this approach, which might be useful to many research data and repository managers, and to journal editors and publishers.

Since 2017 the publisher Springer Nature has provided an optional Research Data Support service to researchers who wish to share their data in a repository but require assistance to do so. In our experience of delivering this service, and evidenced by researcher surveys, we have found that many researchers don't have the time or expertise needed to safely share their sensitive data. Some do not make their data available at all, while others choose to make them available on request after publishing their article or deposit in a controlled access repository. Through statements of data availability which now accompany many published articles, readers may be aware that these datasets could be available them, but simple "available on request" statements don't provide adequate metadata - which might enable other researchers to assess whether the data may be reusable, for example for meta-analyses of clinical research studies.

To support researchers with sensitive data, we developed a curation process which can be applied to all research data, even if raw data aren't shared. By gathering contextual information regarding the author's study and the format and content of their datasets, rich metadata records can be constructed to support the discoverability and accessibility of the data.

Questions and discussion

15 min

---

12:05 p.m. – 1:15 p.m.        Lunch and topic tables

---

1:20 p.m. – 2:20 p.m.        **Data services panel**

Advocating for Change: Understanding the Landscape of Academic Data Professionals
*Joanna Thielen, Oakland University (with Amy Neeser, University of California-Berkeley)*

15 min

The institutions in which we work as data professionals are the very power structures that define which communities are included and excluded. From access to public services such as education and technology to which voices are represented in our digital archives, these structures favor and perpetuate dominant groups and ideologies. The pipelines into these institutions are often murky, and without diverse voices and perspectives these systems will continue to perpetuate themselves. In order to better understand the pipelines into these institutions, we examined academic data professional job postings from January 2013 to July 2018 collected from professional listservs and job boards. This analysis of job postings for data professionals will determine trends in responsibilities and qualifications. We examine factors including job title, salary information, diversity, equity, and inclusion statements, degrees required, rank, as well as areas and levels of experience. We are also looking at these factors through the lens of different Carnegie Classifications of universities, from doctoral granting to special focus schools.

By understanding trends in responsibilities and qualifications, data professionals can better negotiate for fair pay and benefits. As a community, we can also advocate for more equitable hiring processes and a stronger commitment to diversity, equity and inclusion. By understanding the pipelines into the institutions in which we work, the better we can advocate for change.

### Bridging the Divides – A Mixed Method Investigation of Multidisciplinary Research Data Management Practices
*Shahira Khair, University of Victoria (with Jacqueline Quinless, University of Victoria)*

15 min

Emerging requirements for more responsible and open research mandated by granting agencies and scholarly publishers are driving the need for better tools and services to support researchers in managing their data. In Canada, recent national policies governing funded research will shape how both Canadian researchers and their home institutions manage digital research data in the coming years. Research libraries across Canada are exploring ways to work collaboratively with research communities to support their development of robust RDM practices. To understand the diversity of this digital landscape, the University of Victoria Libraries conducted a campus-wide, multidisciplinary investigation involving all ranks of faculty members, post-doctoral fellows, and graduate students. The results have helped us to assess the current level of preparedness for adopting RDM practices in different research communities, and to identify challenges and opportunities for the library and other university stakeholder communities. The mixed method approach taken by this study included a campus-wide survey, in-depth interviews with researchers, and focus group sessions with librarians,

resulting in a dynamic source of quantitative and qualitative data we use to explore differences in RDM practices across disciplines, barriers research communities face in managing and sharing their data, and gaps in current infrastructure and supports on campus. With this information, a series of recommendations is provided informing how the University of Victoria and its Libraries can cohesively bridge research services, infrastructure, and disciplinary groups across campus to improve the support provided to researchers in managing their research data.

## Value in those dusty file cabinets? Identifying analog data from our campus community
*Shannon Farrell, University of Minnesota (with Julie Kelly and Kristen Mastel, University of Minnesota)*

15 min

Research data is a growing focus in academic libraries; however, little attention is given to older data in print. Before research data was produced and preserved in electronic format, lab notebooks, field books, and data sheets were common data formats. Despite its format, this data is as important as its digital surrogates and could be significant for data sharing and reuse. We suspect that analog data resides in many places across academic institutions. In informal conversations with researchers, we discovered that several researchers were holding on to old, paper data spanning many decades. However, to quantitatively investigate our hypothesis, we sent out a survey to faculty and researchers in the agricultural and biological sciences, disciplines that we thought may have a high proportion of historical, analog data.

In the survey, we asked faculty and researchers if they possessed analog data, how they are currently storing the materials, topics and dates the data covers, and their plans for preservation. We also wanted to know what their opinions and experiences were around reusing and sharing the analog data, and if they have or are interested in making the data machine-readable or electronic. Finally, we asked them if documentation for the data exists.

In this presentation, we will discuss the survey results, our plans to conduct follow-up interviews on the same topics, and our hope to help researchers to better organize and preserve their data as well as increase its visibility and prepare it for reuse. We will also cover how we plan to engage with other universities working on similar analog data-related projects.

2:25 p.m. – 3:25 p.m.          **Lightning talks**

## Building a Carpentries Community
*Julie Goldman, Harvard Library*

5 min

The New England Software Carpentry Library Consortium, or NESCLiC for short, has brought together library staff from nine academic libraries: Yale, Harvard, Tufts, Dartmouth, UMass Amherst, UMass Medical, UConn, Brown, and Mt. Holyoke. The group's goal was to create a network of Carpentries instructors in New England, to share instruction, and ultimately, to develop new lesson material. Sharing the membership between the nine libraries allowed NESCLiC members to pilot the Carpentries approach together, rather than separately, and to lower the costs for each institution. Learn how the instructors are staying involved in this consortium, building their own carpentry skills, giving back to their communities, and collaborating around New England (and beyond)!

## A Case Study in Collaborative Open Research Training
*Matthew Harp, Arizona State University Library*

5 min

The ASU Library is working to support the inclusive mission of Arizona State University with new models of research services and data management training in partnership with ASU's Office of Knowledge Enterprise and Development (KED). We are cementing this partnership through thematic programming we integrate into KED's learning management ecosystem and by collaborating with their instructional designers to create a hybrid of online and in-person instruction experiences. To avoid siloed systems, liaison librarians work directly with research advancement officers who are already trusted members of research project teams. The liaisons are also informed and supported by the library's Research Services Working Group, a cross functional body who ensure engagement activities are properly resourced and in line with library strategic priorities. This lightning talk demonstrates how our partnership programmatically targets research advancement officers, faculty, and students for training on responsible and open research methods. The talk also addresses the work of building competencies and confidence among library professionals to provide enhanced, transdisciplinary data management support in a university with over 3,500 faculty and 7,000 graduate students.

## Building a research data management community
*Megan Hardeman, Figshare*

5 min

Sharing knowledge and best practice across communities, regardless of location and experience in the field of research data management, is hugely important to us at Figshare. In this talk, we'll showcase examples of how community members - primarily research data managers - exchange knowledge across regions and across levels of experience in the space. We'll discuss the systems and tools they use to collaborate, the methods they use to share their successes and challenges, and how they facilitate an open and accepting community.

## Investigating netCDF: Building file format knowledge to inform library services through conversation, reviewing documentation, and tool experimentation
*Samuel Sciolla, University of Michigan School of Information*

5 min

For research data librarians, providing services often requires familiarity with a range of file formats, some highly specialized to specific domains. While they cannot be expected to have an intimate knowledge of all formats they encounter, librarians can learn strategies for rapidly developing a working knowledge of one format, as well as documenting that knowledge for the benefit of their colleagues. In Summer 2018 I worked with a mentor from the University of Michigan Library to build up internal knowledge for the Research Data Services department on netCDF. A format for storing multi-dimensional arrays, netCDF is prevalent in the climate and atmospheric sciences and numerous Michigan researchers have deposited netCDF files in the library's research data repository, Deep Blue Data. In order to assist researchers in ensuring the quality of their data and its usefulness to others, we sought to identify the structure and affordances of the format, available tools and methods for working with the files, and relevant metadata standards. We also sought to create a repeatable process and a template for presenting our findings. Over two months, we met with three researchers to discuss their research and how they work with netCDF; reviewed documentation and other literature on the format; and experimented with tools for manipulation and visualization of the data netCDF files contain. The final product was a Data Curation Format Profile, which identifies informational resources, poses questions for curators to consider, and provides basic instructions on using netCDF-related tools. This presentation will explore our methods and findings and highlight the value for data curators of speaking with researchers about their relationship with file formats.

## Local Research Data Management Course for Medical Students Ventures Online
*Tess Grynoch, University of Massachusetts Medical School*

5 min

The Lamar Soutter Library at UMass Medical School developed a week-long course on Research Data Management Fundamentals in 2015 as one of the Flexible Clinical Experience (FCE) options available to medical students. Since then the course has run 6 times with a total of 9 students. There have been generally only 1 or 2 students enrolled in the course during each session, resulting in a low return on investment for the two instructors who spend a combined 14 hours lecturing that week plus time preparing. Medical student education is changing as more students skip in-person lectures to watch the lecture recordings online. After completing a course in online instruction, the FCE instructors transformed the in-person class content into online modules. The lightning talk presents how best practices in online instruction were applied to the first iteration of the online course that ran November 5-9th, 2018. Subsequent student feedback and future improvements that will be made to the course include offering the class to our Graduate School of Biomedical Sciences and Graduate School of Nursing, adding further functionality features, and updating assessments, which will be discussed further during the presentation.

### Computational Reproducibility Platform and Executable Repository
*Travis Hewgley (Code Ocean)*

5 min

Code has become a crucial component of the scientific life cycle. Streamlining the research workflow to allow researchers to develop, access, and preserve reproducible code and data is critical to solving the reproducibility crisis and speeding up the pace of science. Code Ocean, winner of ALPSP's Innovation in Publishing Award in 2018, is a cloud-based open access computational research and reproducibility platform. We allow researchers to create, collaborate on, and execute code. Universities and labs also benefit from the tools we have to empower researchers.

Questions and discussion

20 min

| | |
|---|---|
| 3:25 p.m. – 3:40 p.m. | Break |
| 3:40 p.m. – 4:40 p.m. | **Institutional and Domain Repositories: Competitors, Complements, Collaborators? Discussion panel** |

*Jake Carlson, University of Michigan*
*Thu-Mai Christian, Odum Institute for Research in Social Science*
*Wendy Kozlowski, Cornell University*
*Sebastian Karcher, Qualitative Data Repository*

45 min

With the increasing recognition of data as a first-class resource object, data repositories have proliferated. Re3data (https://www.re3data.org/) currently lists 2,226 data repositories. The two dominant types of repositories are domain, or disciplinary repositories (1,850 of those listed on re3data) and institutional repositories (559 in re3data). This panel seeks to explore and debate the relationship between these types of repositories. Our goal is to have this discussion in a spirit of shared goals around public access to data and the appropriate curation of data.

- With different repositories vying for data and funding, largely from the same sources, to what degree are domain and institutional repositories in *competition*? For example, should dissertation data for a researcher at a given institution be more appropriately deposited into an institutional repository or a domain repository?
- Given their different strengths, can they *complement* each other's activities and if so, how? Is the oft-cited division of labor where institutional repositories accept data where no suitable domain repository exists a viable model or does it raise as many questions as it solves?
- And given shared goals of promoting data sharing and preserving and disseminating research data, can institutional and domain repositories *collaborate* effectively. What are some models of such collaboration and what potential challenges?

Questions and discussion

15 min

---

| | |
|---|---|
| 5:00 p.m. – 6:30 p.m. | **RDAP Reception** |

Hosted and sponsored by University of Miami

Welcome from Charles Eckman, Dean of Libraries, University of Miami

Light appetizers will be served

---

# Day 2 | May 16 2019

## Shalala Student Center
## 1330 Miller Drive
## Coral Gables, FL.

---

9:00 a.m. – 9:30 a.m.      **Coffee and light morning food**

---

9:30 a.m. – 9:35 a.m.      **Opening announcements**

*Tisha Mentnech, RDAP Summit co-chair*

---

9:40 a.m. – 10:40 a.m.      **Community and data in the world panel**

Data Organization, Sharing Infrastructure, and the
Development of Resident Participation in Biological
Conservation
*Ali Krzton, Auburn University*

15 min

The world is currently experiencing a biodiversity crisis severe enough to
have been compared to a mass extinction event.  In many of the most
critically important regions for conservation, research is conducted by
people who are not from the area, especially scientists from Europe and
North America.  These visiting scientists rely on the assistance of local
residents to do their work, often training them to be competent technicians
who can monitor environmental conditions and collect data on their own.
However, residents rarely have the opportunity to participate in the full
scientific process, especially with regards to analyzing and interpreting
data.  The purpose of this presentation is to examine aspects of data
management, such as metadata standards and data sharing practices that
could be changed with an eye to improving residents' ability to interact
meaningfully with conservation data pertaining to their local landscapes.
Robust data sharing infrastructure that makes data available online in a
timely fashion would also facilitate the training of students at regional
universities, allowing them to practice analyzing data and participating in
scientific communication on local hardware. Ideally, residents of the
landscapes we aim to protect will have meaningful opportunities to
become scientists in their own right and take ownership of local

**Research Data Access and
Preservation Summit 2019**

conservation initiatives, enhancing our ability to protect biodiversity worldwide.

### Distributed, Dynamic, Accessible: Cooperative Planning for a Community-Created Data Rescue Toolkit
*Reid Boehm, Johns Hopkins University (with Mara Blake, JHU, Fernando Rios, University of Arizona)*

15 min

This presentation will describe an IMLS Planning grant focusing on community driven work. Securing sustainable access to government earth science data is integral to the interests of citizens. As the need for data rescue came into the spotlight in the past two years, it also illuminated diverse aspects of the research data management community and the stakeholders we serve. The numerous efforts presented us with new opportunities to learn about how we as information professionals can effectively work together to co-create methods that will allow for the most effective participation. We are learning more about how to plan for flexibility to meet the dynamic needs of a distributed society. Here we will highlight the Data Rescue Toolkit planning project and the design of our Fall 2018 community meeting with its theoretical basis in Participatory Action Research (PAR). We will share meeting outcomes, lessons learned, and present overarching questions as examples of those which are inherent in distributed collaborative work with shared public data.

### Responding to Data Sharing Dynamics in Community-Based Research
*Nora (Eleanor) Mattern, University of Chicago (with Aaron Brenner, University of Pittsburgh)*

15 min

This presentation focuses on research data access and sharing in the specific context of community-based research. Community-based research refers to partnerships between academic investigators and community-focused organizations that aim to understand and address community needs and challenges. In contrast to data sharing within a solely academic context, the collaborative nature of community-based research introduces unique dynamics around study roles and credit, ownership of research outputs, and access to data during a project and in the long-term. We argue that the development of a plan between

university and community-based researchers at the initiation of a research partnership can foster meaningful and equitable arrangements around data collection, management, and preservation.We will present an initial effort at the University of Pittsburgh Library, in collaboration with the University's Community Engagement Centers, to develop a resource to guide the development of data agreements between partners in community-based research.

### Data as a tool for community engagement in south Florida
*Jill Krefft, Florida International University (with Jennifer Fu, Jamie Rogers, and Sheyla Aquilar De Santana, Florida International University)*

15 min

Florida International University (FIU) works closely with the local community to develop tools and resources that improve access to data that support learning and research-related issues in south Florida. This presentation will focus on collaborative projects at FIU that have brought wider access to data through curation, visualization, and engagement with the community.

The Sea Level Rise App, developed by FIU Library GIS Center, supports the Eyes on the Rise community initiative (http://www.eyesontherise.org/). Another example is the data management and dissemination support for the Regional Environmental Monitoring and Assessment Program (R-EMAP). Both examples serve the scientific and layman community simultaneously by developing a data management workflow on the backend and web geo-visualization that can be digested easily by the public.

The panel will also discuss the development of an open access institutional research data repository that collects and curates research data created by the FIU community. Through development of a local Dataverse repository, we seek to fill a need for findable, usable and openly accessible datasets that support our local and global communities.

---

10:45 a.m. – 11:45 a.m.    **Researcher perspectives panel**

Taking a Hard Look at Our Research Data Support Through A Critical Lens

*Hilary Davis, North Carolina State University Libraries (with Susan Ivey, North Carolina State University Libraries)*

15 min

In support of the growing research data management needs for our researchers and students, the North Carolina State University Libraries has utilized a cross-departmental team approach to offer a variety of services and resources targeted to specific research data needs of our community. This team has a revolving membership of individuals from across the library, which helps to train a large number of librarians on data management best practices, while also allowing the Libraries' research data support to more easily scale to meet these ever growing needs. Some of the primary resources the team created are checklists that can be used at the pre-award, post-award, and project close-out stages or when a member of a research team departs.  In this presentation, we will describe our efforts to audit our pre-award, post-award, and data exit checklists to see what kinds of expectations we are placing on researchers in terms of data literacy, technical literacy, and policy literacy. We will introduce the ways in which we intend to interrogate our checklists, and by extension, our services. Our goal is to dig deeper to understand how and when we are introducing bias and exclusivity in our services and actively make changes so that we can be better stewards and better advocates for our diversity of researchers.

Growing Skills in Data Management with researchERS: Emerging Research Scholars, a Program for Undergraduate Researchers
*Trisha Adamus, University of Wisconsin-Madison (with Cameron Cook, Tobin Magle, Casey Schacher, Heather Shimon, University of Wisconsin-Madison)*

15 min

Many universities are encouraging an undergraduate research experience where students engage in research projects in the classroom and in research labs. Yet research data management is often not included as part of these opportunities. This creates a perfect time to introduce students to threshold concepts involving good practices across the data life cycle and to plant the seed that data is something to cultivate and steward.

At UW-Madison, we piloted researchERS: Emerging Research Scholars, a program for undergraduates from all disciplines to learn data management skills. Focusing on core concepts as well as data ethics, reproducibility, and research workflows, the format of the program included four workshops and one field trip to a local business per semester. This presentation will discuss how we assessed the program, key lessons learned, and our recommendations for continuing the program.

Peer Review of Research Data Submissions Study. How can we improve the curation of research datasets to enhance reusability?
*Clara Llebot Lorente, Oregon State University (with Steve Van Tuyl, Oregon State University)*

15 min

Librarians working in research data management have been developing best practices to ensure that published datasets are reusable. At Oregon State University we apply these concepts to curation of datasets submitted in our institutional repository, ScholarsArchive@OSU. However, we don't really know if the datasets in our repository are really reusable by domain experts. Curation workflows are designed by librarians based on librarian's recommendations, but research data is extremely specialized, and such workflows are rarely evaluated by researchers. In this project we used domain expert level peer-review to evaluate the reusability of the datasets in our institutional repository, with the goal of informing our curation methods and ensure that the limited resources of our library are being used effectively to maximize reusability of research data. In this presentation we will share how the domain expert evaluation of the reusability of a dataset compares with the librarian's evaluation (ours). We will also share what are the aspects of the dataset submission that, in the opinion of the reviewers, influenced more strongly the reusability the dataset, and ideas about how librarians can help improve these.

Questions and discussion

15 min

11:45 a.m. – 12:15 p.m.     **Poster session**

Teaching Research Data: ACRL Framework Approach towards Carpentries Instruction
*Ari Gofman, Tufts University*

This presentation will use the ACRL Framework as a lens for understanding the strengths of teaching (Software/Data/Library) Carpentries workshops for dealing with research data by reflecting on the experiences of NESCLIC and identifying replicable best practices.

Community-developed infrastructure for curating research data
*Elizabeth Coburn, University of Minnesota (with Lisa Johnston, University of Minnesota, Cynthia Hudson Vitale, Penn State, Joel Herndon, Duke University, Wendy Kozlowski, Cornell University)*

Building and sustaining institutional capacity for research data curation services and support is a challenge many Libraries face. Given limited staffing individual institutions may struggle with obtaining and sustaining necessary subject or format curation expertise. To address this need, the Data Curation Network (DCN) has developed a collaborative two-pronged approach to engage the library community and extend capacity for curating research data. This approach consists of the development of a shared data curation service as well as a data curation training program to increase the curation capacity of the broader community. Bridging the two is the CURATE(D) model, a theoretical framework of curation steps developed by the DCN and aimed at helping curators apply consistent levels of curation, regardless of a dataset's particular characteristics.

Shared data curation service
Funded for a three-year implementation phase commencing in June 2018, the DCN has dedicated its first year to developing and testing workflows and critical infrastructure. We use a role-based workflow that incorporates differing levels of responsibility between the "DCN Coordinator," the "DCN Curator" and the "Local Curator" as well as Jira (an issue-tracking and agile project management system) to track and document the work performed by each role in the curation of every dataset. We will discuss our results of implementation phase year one.

Data curation training program
Our hands-on workshop program, in which attendees share their subject and functional expertise in a peer-to-peer training model, will be discussed, including the publication of the first round of data curation primers. Data curation primers are community created documents meant

to jump-start the curation process by detailing specific curation considerations required to appropriately curate different subject areas, file formats, or functional data types. Copies of the first published primers will be available for handout.

### Building a community of practice: Situating data management in the transition to electronic lab notebooks
*Heather L. Coates, IUPUI University (with Erin D. Foster, Elizabeth C. Whipple, Indiana University)*

This poster will describe an information management initiative sponsored by the School of Medicine Research Administration. The 4-month pilot involved the rollout of an electronic lab notebook (ELN) system alongside tailored support for developing operational data management plans. Administrative drivers included improving lab efficiency and reducing institutional liability. The pilot primarily focused on the benefits of ELN use and formalizing data management practices. A key goal was to improve the collection of and access to experimental information and improve information management workflows. Integrating data management plan support as part of the ELN pilot contextualizes how the ELN fits into, and can potentially improve, existing lab workflows and highlights the unique information management needs of each research lab.

### Research Data Management Librarian Academy
*Jean P. Shipman, Elsevier (with Elaine Martin, Harvard Medical School)*

Several librarians recently partnered with Elsevier to study the need for a Research Data Management (RDM) Librarian Academy to offer online training. The team includes Harvard Medical School, Tufts Health Sciences, MCPHS University, Boston University School of Medicine and Simmons College. This poster will report on the needs assessment and inventory. It also will outline the online modules being developed that target needs identified through the assessment.

### Beyond the DMP: Research data management and other library-based support and services are referenced throughout grant proposals
*Jen Ferguson, Northeastern University*

Librarian assistance with data management plans (DMPs) for grant proposals has become a fairly common activity. While this service is

offered to support researchers, it also has the potential to 'give back' to libraries and librarians. DMPs and grant proposals can serve as rich sources of potentially actionable information both for frontline staff supporting researchers, and for administrators who are increasingly tasked with assessing and asserting the value of libraries.

This poster reports on an analysis of the content of 150+ full grant proposals. Results indicate that when librarians supporting research data management lack access to grant proposals, or have access only to DMPs, they - and their libraries - miss out on information and opportunities to support their institution's researchers. This knowledge could help librarians in RDM-supporting roles, as well as administrators, advocate for increased access to DMPs and full grant proposals.

## Developing Metadata Analytics as a Data Service for Humanities Scholars
*Jian Qin, Syracuse University*

The processing, digitization, and organization of special collections in academic libraries have produced very large amounts of digital images, text, and audio/visual recordings as well as metadata describing the physical collections and their digital surrogates. Even though rich metadata has been created for special collections at item, object, and collection levels, access to these special collections is still largely dependent on traditional search and browse functions. With today's data science and linked data technologies, it is possible for data service librarians to use metadata analytics as a way to develop novel services for humanities scholars. The term "metadata analytics" refers to the practice that applies data science and other computational methods to analyze metadata descriptions generated from information organization and digitization processes for the purposes of developing new library services and supporting research and learning. This presentation will first define metadata analytics and then use the special collections' metadata in Syracuse University Library as an example to illustrate the methods and steps in developing metadata analytics. Metadata analytics is different from the general conception about research data access and preservation in that librarians are the producers and researchers are the consumers of data products generated from metadata analytics, as to what is generally recognized: researchers produce data and librarians curate them. As such, metadata analytics is a proactive research practice that produces data products for researchers and students to more effectively discover and identify phenomena, patterns, relationships, and trends that would have been covert to scholars otherwise. Such data services would need to be built on careful planning and sustainable operationally and financially.

This presentation proposes the new concept of metadata-analytics-based data services and hopes to stimulate discussion in the RDAP community, which would offer valuable input to MLIS education of future data librarian workforce.

## Hidden in the Archives: Finding and exposing reusable data
*Julia Kelly, University of Minnesota (with Shannon Farrell and Kristen Mastel, University of Minnesota)*

Before researchers collected the majority of their research data in digital formats, they were collecting data in analog formats, such as lab notebooks, field books, and data sheets. These data, although they may not be digitized or machine-readable, may have utility for future research, particularly longitudinal studies that track environmental conditions. Although it may not be part of their official collection policies, we were curious about whether our University Archives contained raw data from individual researchers, departments, or centers and if so, if it was potentially reusable.

We will report on our findings on a sample of the collections, including how much data we found, what condition it was in, and how well it could be interpreted. We hope to raise awareness about the existence of raw data in archives and encourage its reuse by current researchers.

## Modeling the FAIR Metrics Landscape
*Marijane White, Oregon Health & Science University (with Robin Champieux, Oregon Health & Science, Kimberly Robasky, Payal Mehndiratta, RENCI, Lily Winfree, Open Knowledge International)*

Researchers and organizations that support science increasingly consider the FAIR Data Principles to be the gold standard for the management and sharing of data and research resources. There are many parallel efforts to identify recommended practices and metrics to improve and measure the FAIRness of data. We sought to delineate the rubrics and frameworks that describe practices and criteria for interpreting and measuring FAIR data as well as the relationships between these rubrics, their criteria, and the original FAIR principles.

This poster describes our work to create a semantic linked data model based on the schema.org vocabulary that describes the characteristics of the FAIR metrics landscape. We integrated the data we collected about the various rubrics with this model to create a small knowledge base that

we interrogated with SPARQL queries in order to identify which components of the FAIR principles have been sufficiently addressed with metrics and the nature of those metrics, which components have yet to be sufficiently addressed, where the community agrees and disagrees in their interpretations of FAIRness, and where the community has extended and reinterpreted the FAIR principles.

Our goals are to provide an evidence-based identification of existing metrics that reflect community agreement on the measurability of FAIR components, the components that have been neglected by these metrics, and the characteristics of data and research resources that the community agree are important to reuse but have not been adequately addressed in the original FAIR principles.

## How Open? Reflections on Building an Open Data Toolkit for Diversity Scholarship
*Rachel Woodbrook, University of Michigan (with Laura Sanchez-Parkinson, National Center for Institutional Diversity, Karen Downing and Jake Carlson, University of Michigan)*

The Library at the University of Michigan has established relationships with several research centers on campus. One of these is the National Center for Institutional Diversity (NCID), which supports research and interdisciplinary community building around diversity scholarship (e.g., issues related to identity, culture, representation, power, oppression, and inequality). Recently, NCID approached the library as a potential partner in modeling and promoting best practices for open data, such as planning for appropriate openness of valuable diversity data with nuanced ethical considerations. The result of these conversations was a new project, an open data toolkit created specifically for diversity scholarship. This poster will reflect on the process of collaborating across campus institutions to build a resource in support of an interdisciplinary network of diversity scholars, and the potential for extending this as a model for researchers beyond the network.

Open data now has an established place in our discourse about information accessibility, and discussion of data ethics beyond IRB and deidentification has a history as well. More recently, emerging fields such as data justice and critical data studies have begun to coalesce. These are helping to bring explicit conversations on power dynamics, autonomy, equity, and inclusion/exclusion regarding data practices into more broadly visible arenas. Nevertheless, many scholars have not had easy access to guidance on adopting a publicly accessible data agenda that takes a holistic view of these considerations.

NCID works with publicly engaged scholars who have led the way in best practices such as collaborating with subject communities in research design and data gathering. While open data resources do already exist, this project is an attempt to combine general data management best practice and open data considerations with the applied expertise and narratives of active diversity scholars to build a resource that addresses considerations for responsible open data practice across the research lifecycle.

## UC San Diego: Keeping up with The Carpentries
*Reid Otsuji, UC San Diego (with Stephanie Labou, Ryan Johnson, Claire Mizumoto, and Cyd Burrows-Schilling, UC San Diego)*

The demand for training in research automation and computational tools for reproducibility is only going to to increase, as interest spreads from STEM fields to social science domains. By leveraging lessons learned in the past few years we have progressively moved forward to expand our Carpentry community and data training efforts. Since the beginning of our data training development efforts built on The Carpentries workshops in 2016, the library recently developed a joint partnership with campus Research IT Services to develop a plan and implement resources to support our growing Carpentries community and workshop activities. Our community now has a large qualified instructor pool, with diverse expertise in terms of lesson content and disciplines. We have worked to establish a growing sense of community among instructors and a support system for the learners we teach. The poster we are presenting will outline and trace the significant Carpentry community building developments of UC San Diego Library's data services research training support ecosystem in the form of Software, Data, and Library Carpentry workshops for our academic research community.

## Rebooting the Research Data Management Program: The First Two Years
*Renee Walsh, University of Connecticut (with Jennifer Chaput, University of Connecticut)*

An overview of how a RDM program was relaunched at an academic library.

## Developing, Delivering and Redesigning Metadata and Data Documentation Workshop for Graduate Students

*Sai Deng, University of Central Florida Libraries*

This poster will present the development, delivery and redesign of a metadata and data documentation workshop for university graduate students and researchers. The original workshop was in lecture style, and it covered large amount of information such as the status of data documentation and management as revealed by a campus survey, research data documentation basics, general and domain metadata standards, and data documentation practices in different disciplines. The redesign looks at students' experiences in previous workshops and also what the Metadata Librarian has gained from national endeavors such as the Inter-university Consortium for Political and Social Research (ICPSR)'s Data Fair and the Data Curation Network's workshop. The redesign has adjusted the overall learning methods to include more interactions, such as short questions, discussions on published datasets, students' own needs and cases in data documentation. It is expected that the redesign will be more in line with the graduate students' research and be more beneficial to their academic pursuits.

## From the Ground Up: Building a roadmap of research priorities for a national research data management community
*Shahira Khair, University of Victoria (with Melissa Cheung, University of Ottawa, Dylanne Dearborn, University of Toronto, Lyne Da Sylva, Université de Montréal, Kathy Szigeti, University of Waterloo)*

Launched in 2015, the Canadian Association of Research Libraries (CARL) Portage Network is dedicated to the shared stewardship of research data in Canada through fostering national communities of practice for research data, and building national research data services and infrastructure. Portage has developed a large network of expertise comprised of Expert Groups, each with mandates related to specific areas of research data management (RDM). The objective of the Research Intelligence Expert Group (RIEG) is to provide the Portage Network with research intelligence on the state of RDM in Canada for a variety of related topics through the development and oversight of targeted studies designed to gather supporting evidence.

In order to set research priorities for RIEG, a high-level roadmap was developed to bridge gaps in our knowledge about RDM practices, developments, communities, and policies in Canada. To understand the existing landscape, an environmental scan was conducted, guided by a custom taxonomy developed by RIEG. Topics including data curation,

data sharing and reuse, services, policy and law, and information technology were examined from an international and Canadian context to determine gaps and priority areas. This poster will offer an overview of the development of this roadmap, detailing how it helped to formalize recommendations for action, prioritize research tasks, and set objectives for RIEG and the Portage Network moving forward.

### Changing the One for All Model of an RDM Fundamentals Workshop
*Sophia Lafferty-Hess, Duke University (with Jennifer Darragh, Duke University)*

Academic research data management (RDM) education initiatives equip various campus audiences with concepts and practices that help make their research more efficient, reproducible, and shareable. At Duke University, our education program is one pillar of our core suite of services and we have been lucky to have broad reach with one particular group - graduate students. In partnership with the Graduate School we offer various workshops that can fulfill student's Responsible Conduct of Research (RCR) credit requirements. Our workshops are often attended by students from across disciplines and therefore we have tended towards broadly relevant high-level topics. However, our pedagogical practice of trying to teach to all disciplinary communities in the room, resulted in a missed opportunity to create more meaningful content for disciplinary communities. We also recognize that in our effort to speak to all, we may have potentially shown bias towards certain disciplines in our language and examples.

In response to this need, we implemented a revised "Research Data Management 101" series designed for three specific disciplinary groups - Humanists, Scientists, and Social Scientists. In collaboration with subject specialists, we developed content that contained examples and discussion opportunities that (we hoped) resonated more directly with each disciplinary group. This presentation will provide an overview of the curriculum, lessons learned, and an assessment of how we might continue to make our curriculum more accessible not only to various disciplinary communities but to other communities we serve.

### The Pulley Ridge Data Curation Experience
*Timothy B Norris, University of Miami (with Chris Mader, University of Miami)*

The curation and preservation of scientific data has long been recognized as an essential activity for the reproducibility of science and the advancement of knowledge production. While investment into data curation for specific disciplines and individual research institutions has advanced the ability to preserve research data products, data curation for "big" interdisciplinary science remains relatively unexplored terrain. This presentation reports on the (geographic) data curation process for such a scientific endeavor. In 2011 the University of Miami Center for Computational Science (CCS) was invited to collaborate as data curators on a multi-year trans-disciplinary NSF funded research project located in the Gulf of Mexico. The CCS was specifically tasked to build an online decision support resource that includes a data repository, a map-based data exploration tool, and a map- and data-based story telling tool. The presentation will report on the entire curation process with a focus on data sharing, metadata creation, repository development, and the development of tools to synthesize and share the data across disciplines. The presentation concludes with a reflection on the successes and failures of engaging with communities that span data producers (researchers), data curators, and data users (resource managers and decision makers).

## Designing tools with researchers in mind to overcome barriers to reproducibility
*Travis Hewgley, Code Ocean*

Accessing and preserving code and algorithms associated with research is critical in solving the reproducibility crisis and advancing research discoveries. We present an open access, interoperable, executable repository that supports computational reproducibility of code for researchers and preservation of code for libraries. The platform does two key things: (1) it integrates the metadata, code, data, and dependencies into a 'compute capsule', ensuring that the code is easily rerun; and (2) it provides access to and exporting of all inputs. This computational research platform employs container technology to execute code in the cloud, making it possible to develop, edit, or download the code and collaborate to visualize, save or download output. Users or reviewers can upload their own data and test the effects of changing parameters or modification of the code. When published, the platform provides a DOI for all capsules to facilitate attribution and a permanent connection to any published work.

| 12:15 p.m. – 1:15 p.m. | **Lunch, poster session continued** |

| | |
|---|---|
| 1:20 p.m. – 2:20 p.m. | **RDAP Association Business Meeting** |
| | *Erica Mehan Johns, President, RDAP Association* |

| | |
|---|---|
| 2:20 p.m. – 2:35 p.m. | **Break** |

| | |
|---|---|
| 2:35 p.m. – 3:35 p.m. | **Closing keynote** |
| | *Valencia Gunder, Founder and President of Make The Homeless Smile Miami/Atlanta* |

| | |
|---|---|
| 3:40 p.m. – 3:50 p.m. | **Closing announcements** |
| | *Tisha Mentnech, RDAP Summit co-chair* |

# Day 3 | May 15 2019

# Workshops

# Newman Alumni Center
# 6200 San Amaro Drive
# Coral Gables, FL

| | |
|---|---|
| 8:30 a.m. – 9:00 a.m. | **Coffee break** |

| | |
|---|---|
| 9:00 a.m. – 12:00 p.m. | **Morning Workshops** |
| | Diversifying, Expanding, and Enhancing a Research Data Management Training Clearinghouse through Community Engagement and Action<br>*Karl Benedict, University of New Mexico* |
| | Research data management (RDM) training has been identified as a significant need by researchers asked to identify challenges to effective RDM, and the development and delivery of RDM training has been a growing area of focus for the data management, access, and preservation |

communities. The recognition of this need and rapidly growing availability of RDM training materials led to the creation of the Data Management Training Clearinghouse (DMT Clearinghouse - http://dmtclearinghouse.esipfed.org) by the Earth Science Information Partners in collaboration with the USGS, DataONE and others over the past two years. This year, a three-year Institute of Museum and Library Services (IMLS) sponsored project was initiated to enhance the Clearinghouse. The goals of the project are to 1) diversify the collection of materials to include additional discipline-oriented materials (with a particular focus on the social sciences); 2) examine and improve the underlying metadata model and search tools to ensure effective discovery and understanding of DMT content within the Clearinghouse; 3) develop an assessment capability within the Clearinghouse to provide feedback to training material creators, users, and instructors; and 4) expand the diversity of Clearinghouse users (both in terms of instructors and learners) through targeted outreach to communities for whom familiarity with the Clearinghouse would be useful. This workshop will provide hands-on experience with the clearinghouse with the goal of both increasing familiarity with the clearinghouse and its content, and set the stage for workshop participants to join a growing community of contributors and reviewers for content in the clearinghouse.

Objectives:
1. Increase the visibility of the DMT resources that are discoverable through the Clearinghouse to maximize the use and impact of existing training materials
2. Enable the workshop participants to work with the clearinghouse as contributors to and reviewers of the growing collection of training materials registered in the clearinghouse
3. Engage with the data management, access, and preservation communities attending RDAP to identify gaps in existing and registered training materials within the clearinghouse
4. Recruit community members for participation in the project as members of the editorial team, metadata and assessment working groups, and participants in the planned usability testing of planned clearinghouse enhancements.


Turning your Poster or Presentation into a Paper – JeSLIB
*Thea Atwood, University of Massachusetts Amherst*
*Kristin Lee, Tufts University*

For individuals considering submitting to the JeSLIB RDAP 2019 Special Issue.

Objectives:
1. Prepare your presentation for submission to the special RDAP issue of the Journal of e-Science Librarianship.
2. Receive guidance and working time to develop an outline for submission to JeSLIB.

---

12:00 p.m. – 1:00 p.m.       Lunch break (on your own)

---

1:00 p.m. – 4:00 p.m.        **Afternoon workshops**

Includes coffee break

Train-the-Trainer: Developing a Research Data Management Workshop to Support Graduate Student NSF Doctoral Dissertation Research Improvement Grant Proposals
*Andrew Creamer, Brown University*
*Hope Lappen, New York University*
*Sam Simas, Bryant University*

Each year thousands of graduate students in the Social Sciences and Humanistic Social Sciences write their first National Science Foundation (NSF) proposal for a Doctoral Dissertation Improvement Grants (DDRIG). These student researchers are in fields such as Anthropology, Political Science, History of Science, and Science and Technology Studies, and are interested in conducting research with an emphasis on archival, ethnographic, and international fieldwork. This workshop is a "Train-the-Trainer" and participants will develop the basis for a workshop for your campus to serve the needs of these students to prepare them for 1) writing a Data Management Plan that captures their activities in the contexts of archival and ethnographic fieldwork 2) writing a Broader Impacts section of a proposal that connects with their research objectives 3) make students aware of the bottlenecks encountered by researchers when preparing their first NSF DDRI proposals 4) selecting and appraising data for publication, sharing, and preservation within the contexts of data derived from archival and ethnographic fieldwork.

Objectives:
1. Teach graduate student researchers to navigate Research.gov and FastLane and provide overview of solicitation, supplementary document requirements, and public access compliance requirements, including depositing in NSF-PAR

**Research Data Access and
Preservation Summit 2019**

2. Point out common pitfalls for graduate students navigating and complying with solicitation and PAPPG
3. Conduct an evaluation of students previously funded DDRIs regarding their projects' data collection and documentation needs and lessons learned to incorporate into the workshops
4. Pull together a resource list of experts and offices to support data storage, data security, intellectual property, and ethics to support their research goals
5. Consult on directorate-specific Data Management Plans for their proposed DDRI projects that integrate library research data management resources among others, and help students locate and integrate data documentation standards utilized by the repositories and journals in their field
6. Consult on the development of students' Broader Impacts plan for their DDRI projects

## Preparing data and code for reproducible publication using container technology
*April Clyburne-Sherin, Code Ocean*

This is a step-by-step, practical workshop on how to prepare research code and data for computationally reproducible publication. The workshop starts with some brief introductory information about computational reproducibility, but the bulk of the workshop is guided work with code and data. Participants move through best practices for organization, documentation, automation and dissemination of research data and code using Binder and JupyterLab - free, open source technologies.

Objectives:
1. Understand the barriers to reuse of published research code and data.
2. Practice organizing, documenting, automating, and disseminating reproducible code and data.
3. Explore and assess tools and resources to aid reproducible publication.