

# Integrating Research and Teaching for Data Curation in iSchools

**Amelia Acker**

*University of Texas at Austin, School of Information, USA.  
aacker@ischool.utexas.edu*

**Adam Kriesberg**

*Simmons University, School of Library and Information Science, USA  
adam.kriesberg@simmons.edu*

**Devan Ray Donaldson**

*Indiana University Bloomington, Luddy School of Informatics, Computing and Engineering, USA.  
drdonald@indiana.edu*

**Andrea Thomer**

*University of Michigan School of Information, USA.  
athomer@umich.edu*

**Nicholas Weber**

*University of Washington, Information School, USA.  
nmweber@uw.edu*

## ABSTRACT

The quickly changing nature of information science and technology creates unique and remarkable challenges in terms of developing curriculum focused on building data competencies. Faculties responsible for teaching current developments in information studies have the unique burden of needing to continuously update our curricula without sacrificing our broader teaching goals. This panel features diverse perspectives on teaching data curation skills in five US-based schools of information at the undergraduate and graduate levels. Panelists will present their unique perspectives on pedagogical approaches in courses dedicated to data curation, digital preservation, description and access standards, as well as data access and interchange. Topics introduced will range from flipped classroom techniques, finding messy datasets, common pitfalls, hands-on labs, cloud based tools, data carpentry labs, and sequencing learning objectives to match stages of the data life cycle. This panel will give ASIST conference participants an opportunity to see a range of junior faculty, each with IMLS funded research projects related to data curation, share their experiences of teaching data competencies in the classroom.

## KEYWORDS

Data education, data curation, data science, curriculum development, pedagogy.

## INTRODUCTION

If our society is to address the grand challenges facing us in the 21st century, we will need the work of information professionals to accomplish our goals. Truthful, authentic, well-described, and accessible information is vital to building a more sustainable world where humanity can live in a closer equilibrium with our environment and develop the planet more equitably. Many of these future leaders in the information field are currently sitting in our classrooms, both physical and virtual. In courses addressing data curation, digital preservation, information access, and metadata, instructors are building on the foundation provided by the cultural heritage sector to confront the challenges of the day while addressing broad questions of sustainability. Any field which considers the life of information beyond the lives of those who created it, which is concerned about uses not yet imagined for data collected today, and whose most important users may not have been born yet is fundamentally concerned with sustainability. As instructors teaching data curation, digital preservation, archives, and related topics, we must equip our students to build sustainable organizations when they enter the workforce and to act as stewards for information under their care.

## BACKGROUND AND MOTIVATIONS

The educational mission of Information Science and Technology has never been more relevant. And, the economic, political, and societal implications of an informed public is a challenge iSchools are particularly well equipped to

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/pr2.285](https://doi.org/10.1002/pr2.285)

meet. But at the same time, the digital ecosystems in which we interact, work, and spend a majority of our waking hours are rapidly evolving. These changes have, in turn, significantly reshaped expectations about how to educate information professionals for meaningful careers. Students must be versant in a number of digital information management strategies across a variety of institutional settings, and simultaneously competent in working with diverse and complex digital objects. Continuing to meet these challenges requires significant investment of time, energy, and innovation from iSchool faculty.

As five Assistant Professors designing and delivering courses related to data we believe it is important to begin sharing and exchanging ideas about what is effective in combining our own research and teaching experiences. We take seriously the integration of our research in the classroom and seek to not only deliver high quality classes, but do so informed by the empirical research agendas in which we are each engaged. We also recognize that sharing approaches to curriculum development is particularly important in our current climate where shifts in course delivery necessitate quick revision, and adaptation to both online and in-person modalities.

## **PANEL FORMAT AND PRESENTATION ABSTRACTS**

We have assembled a panel of early-career faculty who both teach and conduct research in the areas of data technologies in information institutions. Amelia Acker will begin the session with a brief introduction to the current landscape of data curation pedagogy and then introduce the speakers. Then, all five panelists will introduce the courses they teach and principles they incorporate in their teaching at different Schools of Information throughout the United States. Devan Ray Donaldson will discuss training new students who are unfamiliar with digital curation in libraries; Adam Kriesberg will talk about injecting digital curation perspectives into archives and cultural heritage courses. Andrea Thomer will talk about developing scalable project-based digital curation curriculum; Nic Weber will describe approaches to integrating infrastructural thinking into data curation, and Amelia Acker will discuss teaching metadata standards and data structures with hands-on data practicums. Nic and Andrea will close the panel by synthesizing common trends across each of the five presentations. We will then open the floor to audience questions and comments, which Adam and Devan will moderate. In total the session will run 90 minutes.

### ***Developing Core Competencies for Data Curation Work***

How do you prepare future librarians for digital/data curation work when, before entering your classroom, the majority of your students have never even heard of the term? In this presentation, Devan Ray Donaldson will discuss answers to this question in the context of his graduate-level digital curation course (Z586), a course he has taught annually at Indiana University since 2016. He will discuss the evolution of the class. Specifically, from 2015-2017, he structured the course around the Digital Curation Centre (DCC) Curation Lifecycle Model, where each week of the course corresponded to a different component of the model (e.g., conceptualize; create or receive; appraise and select; ingest; preservation action; store; access, use, and reuse; transform, etc.) (Higgins, 2008). Additionally, the course centered on presentations from an international cadre of guest speakers who are leaders in their respective libraries, archives, and museums to enable students to learn about the real-world challenges that practitioners face at each stage of the curation lifecycle. More recently, from 2019-Present, Dr. Donaldson has structured the course around *Digital Curation for Libraries and Archives*, a digital curation textbook written by Stacy T. Kowalczyk that covers topics including: curation strategies and models; formats; metadata for curation; data assurance (i.e., data quality), repositories, data management; disaster planning; digital curation assessment and planning; considerations in curating research data, etc. (Kowalczyk, 2018). Additionally, the course now focuses on preparing students for data librarianship by assigning them to scientists from across a range of disciplines to learn about their data practices and support them in the creation and execution of data management plans, including the use of institutional data repositories to support long term data preservation and access (Donaldson, 2019).

### ***Connecting the Dots and Expanding Horizons in the MLIS Curriculum***

Historically, iSchool courses on archives, records, and digital preservation have not confronted the challenges presented by digital data head-on. An expanded focus on digital records, digitization standards, and technical skills

necessary to succeed as a digital preservationist has been a boon to MLIS curricula, but has often overlooked how cultural heritage and other information organizations curate and manage data. Drawing on his research experience working with social scientists, zoologists, archaeologists, agricultural scientists, and humanists as well as an archival perspective on data curation, Dr. Kriesberg's presentation will discuss how he has worked to inject digital curation thinking into archives courses, as well as how he has worked to broaden the scope of digital curation courses to include more readings and assignments exposing students to the multi-dimensional problems of working with data. The goals of this endeavor are to connect disparate sections of the MLIS curriculum and demonstrate to students that archivists, preservationists, and digital curators are often asking similar questions of different types of material. Each of these traditions have much to offer the other, and, through critical analysis and the ability to ask important questions, Dr. Kriesberg works to expand his students' professional horizons and show them how their skills can be useful in a range of professional roles beyond those traditionally pursued by MLIS students.

### ***Building scalable, sustainable digital curation curriculum***

In all iSchool classes, instructors must strike a balance between teaching technical skills, and teaching the theoretical foundations of our field – all in a way that is sustainable from semester to semester, and scalable in growing class sizes. In digital curation classes specifically, this often entails a balance between teaching specific tools and best practices alongside broad strategies for setting curation policy and rapidly prototyping new workflows. Andrea Thomer will discuss her work developing sustainable and balanced digital curation curriculum at the University of Michigan School of Information, specifically focusing on an assignment series in her introductory Digital Curation class. Students complete weekly hands-on “labs” that iterate between technical and critical thinking exercises. These skills are put into practice first through a critical data paper assignment, in which students must evaluate the risks to, and potential reuses of, a publicly available dataset. Students then build on this evaluation to develop a mock grant proposal, in which they plan for further curation or project development. This series of assignments gives students an opportunity to integrate technical, theoretical, project management, and prototyping skills – and has been highly reusable from semester to semester.

### ***Open Infrastructure Imaginaries***

Instruction in an LIS setting is often rooted in the possible and potential rather than the practical. In teaching students enrolled in time-bounded professional degree programs, instructors necessarily attempt to expose students to best case scenarios and to be competent or versant in idealized best practices. This is, just as often, to their detriment - students learn to imagine how things should be, rather than how they are likely to be upon entering their first professional position. At the University of Washington's Information School we view this imaginary as a feature rather than a bug. We have designed a multi-course data curation curriculum that gives students practical skills useful to their first professional position as curators while simultaneously helping them to become reflexive about necessary long-term tradeoffs they will face over their careers: Data Curation 1 focuses on identifying and learning about community standards for managing, preserving, and preparing data for meaningful reuse. Based on a student's background or interest, Data Curation 1 allows for multiple pathways through the curriculum - including the curation of public and private sector, humanities, and scientific data. We follow up this initial course with Data Curation 2 where students begin to apply their emerging skills to practical problems that are faced in transforming, integrating, and preparing various types of data for publication on the web. Both courses ask students to create a curation protocol that explicitly documents data transformations and then reflect on the implications these curatorial interventions have on long-term storage and accessibility. In both courses we attempt to shift a student's perspective from the material realities of curating a specific set of digital objects to imagining the infrastructural realities that will both constrain and afford various forms of end-user interaction. We believe that this imaginary - between what is possible and what is practical - is key to realizing the potential of transparent and accountable institutions that can benefit from the publication and reuse of openly accessible data.

### ***Building Up A Spice Rack: Teaching Data Skills with Hands-On Labs***

Whether our students intend to work in a library, high tech internet company, a user research firm, a scientific lab or

a digital publishing platform, the application of metadata technologies are threaded throughout all contemporary information work. While representing data for access is an essential component of training information professionals, teaching data skills for a range of information professionals remains a challenge for LIS educators. Motivated by the need to teach conceptually-grounded data skills, Dr. Amelia Acker has designed a signature course *Theories and Applications of Metadata* for the University of Texas School of Information. This course builds on her research agenda concerning the emergence, standardization, and transmission of data created and collected with platforms. Students are introduced to a variety of web-based, open source tools as well public datasets to understand the value of semantically enhanced data for reuse and sharing. By the end of the course, participants are familiar with contemporary descriptive standards, format registries, data interchange, and the implementation of access tools through a series of hands-on labs. These cumulative hands-on labs and skills-based practicums allow students to build up a “spice rack” of tools, metadata concepts, data resources, and semantic models that prepares them to recommend and implement technical solutions to solve real-world problems in information organizations. In this panel contribution, Acker will discuss the successes and challenges of teaching data skills in a lab style format to students who plan to work in a wide range of data intensive realms.

## PANELISTS' BIOGRAPHIES

Dr. Amelia Acker is an Assistant Professor of Information in the School of Information at the University of Texas at Austin. She holds a PhD and MLIS from the University of California, Los Angeles. Dr. Acker studies the emergence and standardization of new information objects and data traces from mobile ICTs. Currently, she is researching data cultures, information infrastructures, and digital preservation contexts that support cultural memory. She teaches MSIS courses that cover topics in information science, including archives, science & technology studies, mobile ICTs, standards and infrastructure.

Dr. Devan Ray Donaldson is an Assistant Professor of Information Science in the Department of Information and Library Science in the Luddy School of Informatics, Computing, and Engineering at Indiana University Bloomington. Dr. Donaldson's research interests include: digital repositories, data sharing practices, data librarianship, mass digitization, research data management, trust, security, users' perceptions of archives and archival content, and perceptions of electronic health record data. His research has been funded by the University of Michigan, Indiana University, the Alfred P. Sloan Foundation, the Institute of Museum and Library Services, the Regenstrief Institute, and the United States Department of Energy. He holds a PhD in Information from the University of Michigan, a MSLS from the University of North Carolina at Chapel Hill, and a BA in History from the College of William and Mary in Virginia.

Dr. Adam Kriesberg is an assistant professor in the Simmons University School of Library and Information Science. His research focuses on digital preservation, digital curation, data management, and public sector information, and he has experiences teaching a range of courses in the areas of archives, records management, information ethics, and digital curation. Dr. Kriesberg holds a PhD from the University of Michigan School of Information.

Dr. Andrea Thomer is an assistant professor of information at the University of Michigan School of Information. She conducts research in the areas of scientific data practices, data curation, and computer-supported cooperative work. She is especially interested in long-term data curation and infrastructure sustainability; the creation, enactment, interpretation, and transformation of (meta)data standards; and issues of data provenance, reproducibility, and integration. Her teaching focuses on digital curation and knowledge organization.

Dr. Nic Weber is an assistant professor at the University of Washington. He holds a PhD and MLIS in Information Science from the University of Illinois. His research addresses the development and long-term maintenance of data infrastructures that facilitate transparent and accountable governance of public institutions. His current work

focuses in particular on the development of public interest technologies for local governments, and data privacy in the context of open science.

## REFERENCES

Donaldson, D. R. (2019, September). Scientists, Institutional Repositories and Data Management Practices. Poster presented at Association for Library and Information Science Education Conference (ALISE'19), Knoxville, TN.

Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 1(3), 134-140.

Bauin, S., & Rothman, H. (1992). "Impact" of journals as proxies for citation counts. In P. Weingart, R. Sehringer, & M. Winterhager (Eds.), *Representations of science and technology* (pp. 225-239). Leiden: DSWO Press.

Borgman, C.L. (Ed.). (1990). *Scholarly Communication and Bibliometrics*. London: Sage.

Buckland, M., & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science* 45, 12-19.

Hoppe, K., Ammersbach, K., Lutes-Schaab, B., & Zinssmeister, G. (1990). EXPRESS: An experimental interface for factual information retrieval. In J.-L. Vidick (Ed.), *Proceedings of the 13th International Conference on Research and Development in Information Retrieval (ACM SIGIR '91)* (pp. 63-81). Brussels: ACM.

Kling, R. & Elliott, M. (1994). Digital library design for usability. Retrieved December 7, 2001 from <http://www.csdl.tamu.edu/DL94/paper/kling.html>. List references in APA style.

## COPYRIGHT

The standard copyright permission is included. This may be to be modified should you wish copyright to be retained by someone other than the authors.