Trinity University
Digital Commons @ Trinity

Library Faculty Research

Coates Library

2013

When Data Curation Isn't: A Redefinition for Liberal Arts Universities

Megan Toups
Trinity University, mtoups@trinity.edu

Michael Hughes
Trinity University, mhughes@trinity.edu

Follow this and additional works at: https://digitalcommons.trinity.edu/lib_faculty

🟡 Part of the Library and Information Science Commons

This Post-Print is brought to you for free and open access by the Coates Library at Digital Commons @ Trinity. It has been accepted for inclusion in Library Faculty Research by an authorized administrator of Digital Commons @ Trinity. For more information, please contact jcostanz@trinity.edu.

When Data Curation Isn't: A Redefinition for Liberal Arts Universities
MEGAN TOUPS and MICHAEL HUGHES
Coates Library, Trinity University, San Antonio, TX, USA

Address correspondence to Megan Toups, Coates Library, Trinity University, One
Trinity Place, San Antonio, TX 78212, USA. E-mail: mtoups@trinity.edu

ABSTRACT. Data curation is one way that libraries are extending traditional
services to meet the changing needs of patrons. Requirements from research
funders have placed increased pressure on grant recipients to create Data
Management Plans and to securely store raw data. Research universities have
stepped up to provide comprehensive data support services. Despite discrepancies
in funding and staff, smaller institutions can similarly provide robust services by
focusing on their strengths, such as interdepartmental collaboration, flexibility, and
rapid turnaround time. This article details how librarians at Trinity University
adapted the larger practice of curation to meet local data management needs.

In the summer of 2012, when Trinity University's librarians began planning

data curation services, Big Data had already advanced from marketing jargon to

emerging discipline (Lohr, 2012), and a concomitant effort to manage and preserve

raw data was in full swing. This suite of practices, variously termed digital or data

curation, entails "the active and ongoing management of data through its lifecycle of

interest and usefulness to scholarship, science, and education" (University of Illinois

at Urbana-Champaign, 2012, para. 1). Interest in data collected during research,

traditionally overshadowed by published results, grew after the National Science

Foundation (NSF) began requiring applicants to include with their proposals a data

management plan (DMP) detailing how research data will be collected, managed,

stored, and disseminated. Most NSF grants go to research universities (Drutman,

2012), and the most data-intensive projects occur on the national and international

scale (e.g., 1000 Genomes Project, CERN's large hadron collider). So why should the

library of a small liberal arts university support data management and preservation

for faculty and student researchers? This article details the effort undertaken to

answer this question: the inquiries made, research conducted, and trial services offered to test the viability and establish the need for adapting the data curation paradigm to the particulars of our residential campus.

LITERATURE REVIEW

The practice of data curation in higher education took off in 2004 when the United Kingdom's Digital Curation Centre was established to prevent "duplication of effort in research data creation" and to "[enhance] the long-term value of existing data by making it available for further high quality research" (Digital Curation Centre, 2013, para. 3). Academic libraries took up the challenge shortly thereafter (Boock & Chadwell, 2010; Schmidt et al., 2011; Peters & Dryden, 2011), seeking to capitalize on the potential captured in Sayeed Choudhury's claim that "data sets are the new special collections" (Palmer et al., 2010, slide 3). Purdue University provides an early example. Under the direction of Dean James Mullins, the libraries began extending to research data "the knowledge that librarians have: the ability to collect, organize, describe, curate, archive, and disseminate data and information," resulting in library participation on nine "multidisciplinary proposals" (Brandt, 2007, p. 365), some of which were funded by NSF and National Institutes of Health (NIH) grants. Early collaborations also took place at Johns Hopkins University, where the Sheridan Libraries and the National Virtual Observatory developed "a data curation prototype system that connects digital archiving [using an institutional repository] and electronic publishing systems" in order to securely store portions of the "compound objects"—research papers, theses, gray literature, derivative datasets—produced by astronomy researchers (Choudhury, 2008, p. 215).

But positioning the library to curate research data is not as simple as depositing datasets into institutional repositories (IRs), as if datasets were equivalent to traditional research outputs, such as journal articles. The would-be curator must understand the nature of the data and how future researchers are likely to use it in order to ingest, store, and provide access accordingly (Salo, 2010). The problems posed by data, in all their variety and specificity, are compounded by

the methods from which they are generated. Salo highlights the problem of "small science," the research of small teams or individual scientists engaged in "hyperlocal" projects. In the aggregate, these projects are in all probability bigger than Big Data, but they lack standardization, and are subject to different—or indifferent—modes of collection, storage, and preservation (Salo, 2010, para. 5). Whatever the challenges of refitting IRs in order to properly receive, describe, and present curated data, the local nature of much research, particularly undergraduate research, suggests that "institutions without local Big Data projects are by no means exempt from large-scale storage considerations" (Salo, 2010, para. 5). Research at Trinity University is emblematic of these considerations. It may occur on a smaller scale, but the findings of Trinity scholars still influence the wider conversation of scholarship. In point of fact, Trinity faculty have received grants from many prestigious foundations, including the NSF, the NIH, the Mellon Foundation, and others. This "small science" generates research data that are every bit as fragile as data produced anywhere, at any scale.

Armed with these considerations and reinforced by commentaries encouraging academic libraries to lead by example (Ogburn, 2010; Walton, 2010; Heidorn, 2011), a working group of Trinity librarians searched the literature for examples of non-research libraries taking up the data challenge. Many liberal arts universities are at the so-called early majority stage in the diffusion of data curation, their position captured by a folksy adage invoked by Everett Rogers: "Be not the first by which the new is tried, nor the last to lay the old aside" (Rogers, 1995, p. 284). Consequently, there are few studies or practice papers concerning the adaptation of data curation services to the unique needs of researchers working in such environments. The Department of Library, Information, & Technology Services (LITS) at Mount Holyoke College provides one example of such an effort. LITS had already formed a sub-department to handle digital assets and preservation services (DAPS) when faculty began to voice concerns over NSF data management requirements, specifically the writing of a data management plan (DMP) and the resulting need for a secure repository to store research data. Adopting a "learning on the fly" approach, DAPS staff surveyed faculty about their digital research data,

began offering writing assistance on DMPs, assembled a web page with additional resources on DMP creation, and leveraged their small size and lack of bureaucratic red tape to begin collaborative conversations with the Data Working Group at the University of Massachusetts at Amherst (Goldstein & Oelker, 2011).

This leading-edge approach was reinforced by evidence from another study, which not only argues for library involvement on data curation issues but demonstrates the success that comes from interpersonal engagement rather than impersonal marketing. In "The Problem of Data," a detailed study of curation practices by university researchers, a respondent captured the essence of this approach: "I also need someone to tell me that it's in my interest to do it and kind of prod me and help me do it. Both urge me and help me to do it at the same time" (Jahnke, Asher, Keralis, & Henry, 2012, p. 15). At Trinity, the liaison librarian model is a core strength of the library. Capitalizing on the relationships liaisons have already established, and on the opportunity to provide "badly needed, real-time professional support" (Jahnke et al., 2012, p. 16), proved essential to slotting data curation into the suite of services already provided by the library. Below, we detail how Trinity's librarians investigated and initiated data curation services for the campus.

INVESTIGATING DATA CURATION AT COATES LIBRARY

In August 2012, we formed a team of librarians to investigate what data curation services, if any, Coates Library could offer. This team consisted of two science liaison librarians, the university archivist, and the head of discovery services. Members were chosen for their interest in the topic and for their broad skills, knowledge, and expertise, including experience with open access, institutional repositories, archives, and scientific laboratory practices.

Once formed, we began our investigation by conducting an environmental scan of the library literature in order to understand the services other libraries were offering. In addition, we contacted Trinity's Coordinator of Research Programs for insight into how best to approach research faculty regarding these issues. The coordinator worked directly with research faculty on grants and is closely involved

in undergraduate research programs, making her a natural collaborator. On her advice we designed focus groups targeted at faculty from different disciplines on campus, including researchers who had already submitted DMPs for NSF grants.

Based on the coordinator's recommendation, the team set up two focus groups with faculty and one-on-one interviews for those unable to attend a focus group. We chose ten faculty members from the following departments: Biology, Religion, Classical Studies, Art & Art History, Business Administration, and Chemistry and ensured that each participant was heavily involved in research. In focus groups and one-on-one meetings, the team asked questions regarding:

- the data created, including types, formats, and sizes
- where data is stored
- how data is managed, or not managed
- priorities of data curation, i.e., which issues are more urgent than others
- other data concerns not identified by the interviewer

Many of the questions used in our focus groups were modified from those developed by Jeanine Scaramozzino and her colleagues during their investigation of faculty data curation behaviors and attitudes at California Polytechnic State University, San Luis Obispo (Scaramozzino et al., 2012).

FACULTY INPUT

Focus group participants gave us valuable information on which to base data curation services and confirmed our suspicion that such services were needed. During our conversations, faculty told us they had never thought of the library as a partner in addressing their data concerns. One participant explained that the library's culture of service makes it a natural fit for increasing access to and preserving data. Naturally we agree with this assessment, which is evocative of the traditional work we do in service of the library's mission.

Such feedback was instrumental in marketing our services once we arrived at the implementation stage. Below, we highlight some of the information gleaned from these conversations.

Data—Types, Formats, and Sizes
Faculty members interviewed had anywhere from a couple of gigabytes to 14 terabytes of data, with most having less than one terabyte. A number of participants indicated that their need for storage was likely to grow in the future. Data created by faculty covered the gamut of formats—images, video, audio, paper-based notes, databases (e.g., Filemaker Pro, GIS), statistical data, and Word and Excel documents.

Storage Location of Data
Storage locations varied greatly, even for a single faculty member. Some faculty kept everything on one computer. Some stored data in multiple places: home computer, work computer, USB thumb drives, cloud-based storage, external hard drives, Trinity's servers, and on physical media such as paper and whiteboards. Not surprisingly, many participants voiced a desire for seamless syncing of their data.

Current Management of Data
Overall, faculty believe they are managing data well, but their research is not without management issues. Half of our participants noted that they had lost data because of hard drive failure, file corruption, or by misplacing a USB drive. Irregular file naming and backup also threatened data integrity, especially where student researchers were concerned. For example, it was noted that multiple students might work on a project at any given time, and those students might rotate off or graduate, leaving a chaos of non-standard file names and types. One participant, a biology professor, noted that such practices open a pedagogical opportunity, that is, a chance to intervene at the early-career stage in order to help students develop good data management practices, particularly when sharing lab and storage space with multiple researchers.

DATA SUPPORT SERVICES

After consulting with faculty and the Coordinator of Research Programs, we decided on a small-scale approach to data curation, one that fit the size and mission of Trinity University while allowing us to grow over time. We named our new program "Data Support Services," a label that eschews jargon that might confuse patrons and

one that more accurately describes the kinds of services we are currently prepared to offer. We arrived at this name after participating in a discussion on naming conventions that took place on ACRL's Digital Curation Interest Group (DCIG) Listserv. Many DCIG members felt that "data support" was a more illustrative and less confusing term than data or digital curation, which sometimes carries an ambiguous meaning, particularly for nonspecialists. The core of Data Support Services is a three-pronged approach, including:

- Data management plan writing and consultation
- Data management education
- Opening our institutional repository to "homeless" data, i.e., data without an obvious, disciplinary-based repository such as the National Center for Biotechnology Information (NCBI) or the Inter-University Consortium for Political and Social Research (ICPSR)

In addition to these three services, we consult with faculty on a case-by-case basis for other data-related projects, and determine, depending on staff expertise and time, whether we can assist them. For example, the library is now collaborating with biology faculty on a vertebrate hormone database, which is explained in greater detail below.

The approach to data support services that we designed is very similar to the approach advocated by the University of Massachusetts-Amherst Libraries Data Working Group (Reznik-Zellen et al, 2012). UMass's libraries advocate an approach they call "Tiers of Research Data Support Services," which focuses on three tiers of service—education, consultation, and infrastructure. We were pleased to learn that our homegrown approach mirrored the structure at UMass, which is home to larger and more visible research libraries.

IMPLEMENTATION OF DATA SUPPORT SERVICES

By Fall 2012, we had completed our initial work on data curation and were ready to introduce data support services (DSS). Our next goals included advertising the service to the wider campus; training liaisons about data curation, our services, and how to engage their faculty; and gathering information about the data

management needs and practices of additional faculty beyond the ten involved in our focus groups.

To advertise DSS, we created a LibGuide (http://libguides.trinity.edu/dss) containing a summary of our services, links to additional resources, and contact information for each librarian involved in the program. In addition, we briefed the Coordinator of Research Programs and the Associate Vice President for Academic Affairs about our services so that they could direct faculty to us. The next round of advertisements for the program involved training liaison librarians to make their faculty aware of DSS by engaging them on the phone, sending an email, or dropping in for a quick chat. These one-to-one activities were consolidated in the spring semester of 2013 when our library director sent an email to the entire faculty containing information on the new service and a link to the LibGuide for more information.

To train the library liaisons, we held an hour-long workshop in January 2013 that covered:

- *A definition of data curation*. We defined data curation and explained the logic behind our redefinition.
- *The services offered by Coates Library*. We outlined our initial plans for DSS.
- *Brief overviews of DMPs, best practices for managing data, and using the institutional repository for data*. We described each topic and detailed exactly what the library is committed to offering in terms of assistance.
- *Ways to engage faculty*. One way to approach faculty is to talk with them about their research projects and student mentees. We also pointed liaisons to our LibGuide for further resources.
- *Role-play on engaging faculty about these topics*. We designed a roleplaying game to help liaisons get comfortable talking to faculty about data issues. Each member of the data team played a faculty member while liaisons played themselves. Liaisons were given notecards with a question on one side related to data management, such as "Where do you store your data?" or "Do you have student researchers? What kind of projects do they work on?" On the back of the card, we suggested follow-up questions or offered supplemental tips and information to pass on to the faculty. The point of the game was to help liaisons internalize questions in order to more effectively

engage their faculty, and to gain facility in continuing the conversation beyond an initial question.

At the end of the workshop, liaisons were tasked with talking to faculty in their areas, excluding those we had already spoken to, and reporting their findings to us at the end of the spring semester. As of July 2013, we have heard from 12 additional research faculty. Their current data management practices, or lack thereof, and their questions about the library's data services will help us to further target our services to the evolving needs of faculty. One geosciences professor mentioned that "developing a training session for student researchers could be really useful. Students beginning research generally have no idea of the different types of information they will be gathering through the course of their research, much less how to organize that information. I rarely think of it until it is too late, when I am trying to navigate through my students' disorganized digital folders to find something" (K. Surpless, personal communication, April 28, 2013). This same professor partnered with the library on a data management workshop for undergraduate research students held over the summer of 2013. Another professor, in modern languages and literatures, mentioned that he "would like to get some help to develop a systematic data management plan" (C. Ardavín, personal communication, March 23, 2013). These comments, among others, indicate the need to continue expanding data support services on campus.

UPTAKE OF DATA SUPPORT SERVICES

Our services are still in their infancy, but already we have seen signs of uptake from faculty. For example, the Associate Vice President for Academic Affairs directed a chemistry professor to us for help with a DMP, a requirement of her NSF grant application. We identified areas of her DMP that needed clarification and made suggestions on how to improve it. She incorporated our revisions, submitted her NSF grant, and was successfully funded.

A biology professor who had attended one of our focus groups approached the data team for help in creating a collaborative vertebrate hormone database. She

and her colleagues at Cornell University and the University of St. Thomas were applying for a grant to fund the development of the database, but the researchers needed technical expertise: a partner who could build the database itself. One of our librarians, the Technology Interface & Assessment Coordinator, has coding expertise and background in database creation. Accordingly, we enlisted his help and committed to developing the hormone database. Although the grant was not funded, we are pressing ahead with the collaboration and will build the database in-house using library resources.

We continue to consult with faculty members about data management, opening up hidden areas for collaboration. In July, one of the science liaisons again teamed with the Coordinator of Research Programs to conduct a data management workshop for undergraduate summer research students. At Trinity, many students work closely with faculty on research projects during a ten-week summer program that provides a stipend, free housing, and tuition credit. These students often produce publishable research, and many go on to graduate school, making them prime candidates for instruction on the importance of backing up data, creating good metadata, and exercising best practices for data organization.

We plan to continue talking with faculty about their data management needs in order to gain a more nuanced appreciation of the ways our program can help them. Our next steps include working with faculty interested in depositing their data in Trinity's institutional repository (bepress' Digital Commons); introducing new faculty to data support services at our library's new faculty orientation (Fall 2013); and developing a lunchtime seminar series where faculty can discuss their research tools and strategies and share with each other their useful data management practices.

CONCLUSION

Trinity's data support services is still in its early stages, but already we have learned many valuable lessons that will help small liberal arts institutions interested in pursuing similar programs. For example:

- Collaboration is essential. Our program's success rests on the knowledge contributions from other campus officials, such as the Coordinator of Research Programs, the Chief Information Technology Officer, and others involved in the intersection of learning and technology. Of course we also relied on the first-hand experiences of the faculty to whom these services will be directed. Their needs and interests made us aware of service gaps, some of which we could address and others that were better suited to existing relationships with Information Technology Services.
- Build on existing relationships with faculty who are open to working with the library, and engage your liaison librarians in these conversations.
- Leverage the small size of your institution for quick development and tailoring of services.
- Commit to small services that have room to grow and tier those services to constituents' needs.
- Remember to include humanities and social sciences faculty in the development of services.
- Train liaison librarians to discuss data management with their faculty and to recognize data support needs.
- Use third-party services to fill in resource gaps, such as bepress' institutional repository software (i.e., Digital Commons).
- Take note of students enrolled in summer research projects and extend data education and services to them. Early-career interventions are crucial to developing good research habits in graduate school and beyond.

The most important thing, however, is to engage faculty in conversations about their research, their students' research, and their data management needs. As Yasmeen Shorish reminds us, "by proactively engaging with faculty, libraries of all sizes can build closer relationships and help educate faculty on data documentation and organizational best practices" (Shorish, 2012, p. 263). Our experience bears out the truth of her recommendation: one-to-one conversations with faculty drove most of

the uptake of our services and informed the types of services we offer. Moreover, faculty have mentioned on multiple occasions their surprise that the library offers help on such issues. These conversations were absolutely essential in creating awareness among faculty of the library's capability and desire to help them with data management. This is an area of service that meets a need of both faculty and students. It also resonates deeply with the mission of academic libraries to support the research needs of their campuses. Large research institutions may have more resources and staff, and their need for data curation may be greater. But we at smaller institutions are poised to learn from their pioneering work, borrow accordingly, and tailor data support services to the local needs of our patrons.

REFERENCES

Boock, M., & Chadwell, F. (2010). Steps toward implementation of data curation services at Oregon State University Libraries. Retrieved from http://hdl.handle.net/1957/20593

Brandt, D. S. (2007). Librarians as partners in e-research: Purdue University Libraries promote collaboration. *College & Research Libraries News*, 68(6), 365– 396.

Choudhury, G. S. (2008). Case study in data curation at Johns Hopkins University. *Library Trends*, 57(2), 211–220. The Johns Hopkins University Press. Retrieved from Project MUSE database.

Digital Curation Centre. (2013). What is digital curation? Retrieved from http://www.dcc.ac.uk/digital-curation/what-digital-curation

Drutman, L. (2012). How the NSF allocates billions of federal dollars to top universities. *Sunlight Foundation Blog*. Sunlight Foundation. Retrieved from http://sunlightfoundation.com/blog/2012/09/13/nsf-funding/

Goldstein, S., & Oelker, S. K. (2011). Planning for data curation in the small liberal arts college environment. *Sci-Tech News*, 65(3), 4. Retrieved from http://jdc.jefferson.edu/scitechnews/vol65/iss3/4

Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 51(7/8), 662–672. doi:10.1080/01930826.2011.601269

Jahnke, L., Asher, A. D., Keralis, S., & Henry, C. (2012). *The problem of data*.

Washington, DC: Council on Library and Information Resources. Retrieved from http://www.clir.org/pubs/reports/pub154/pub154.pdf

Lohr, S. (2012, August 11). How big data became so big. *New York Times*. Retrieved from http://www.nytimes.com/2012/08/12/business/how-big-data-became-sobig- unboxed.html

Ogburn, J. L. (2010). The imperative for data curation. *portal: Libraries and the Academy*, 10(2), 241–246. doi:10.1353/pla.0.0100

Palmer, C., Cragin, M., MacMullen, J., Chao, T., Renear, A., Dubin, D., ... Auvil, L. (2010). The data conservancy: Research on data curation and repositories. Center for Informatics Research in Science and Scholarship. Retrieved from http://groups.lis.illinois.edu/guest_lectures/showcase10/palmer.ppt

Peters, C., & Dryden, A. R. (2011). Assessing the academic library's role in campus-wide research data management: A first step at the University of Houston. *Science & Technology Libraries*, 30(4), 387–403. doi:10.1080/0194262X.2011.626340

Reznik-Zellen, R., Adamick, J., & McGinty, S. (2012). Tiers of research data support services. *Journal of eScience Librarianship*, 27–35. doi:10.7191/jeslib.2012.1002

Rogers, E. M. (1995). *Diffusion of innovations*. New York: Free Press.

Salo, D. (2010). Retooling libraries for the data challenge. *Ariadne*, 64. Retrieved from http://www.ariadne.ac.uk/issue64/salo

Scaramozzino, J. M., Ramirez, M. L., & McGaughey, K. J. (2012). A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries*, 73(4), 349–365. Retrieved from http://crl.acrl.org/content/73/4/349.abstract

Schmidt, L., Ghering, C., & Nicholson, S. (2011). Digital curation planning at Michigan State University. *Library Resources & Technical Services*, 55(2), 104–118.

Shorish, Y. (2012). Data curation is for everyone! The case for Master's and Baccalaureate institutional engagement with data curation. *Journal of Web Librarianship*, 6(4), 263–273.

University of Illinois at Urbana-Champaign. (2012). Specialization in data curation. Retrieved from http://www.lis.illinois.edu/academics/programs/ Specializations/data_curation

Walton, G. (2010). Data curation and the academic library. *New Review of Academic*

Librarianship, 16(1), 1–3. doi:10.1080/13614531003640427