Investing and trading in strategic resources for academic data services: A case study

Angela Hamilton
Acting Deputy Chief Librarian and Liaison Librarian for Biological Sciences
University of Toronto Scarborough Campus Library

Deena Yanofsky
Data Librarian, University of Toronto Scarborough Campus Library

Abstract:

While the meaning of "data" and its role in the research process has shifted numerous times over the past few decades, the role of the academic data librarian has remained focused on helping faculty and students access, use, and preserve all aspects of this distinctive class of information in support of the creation of new knowledge. At the same time, rapid advances in information technology, the explosive growth of digital data, and new methods of scholarly communication have collectively transformed data services in libraries, by impacting their operations, functions, and even their mission. Today, data librarianship is a field of knowledge that can require specific expertise in network technologies, information architecture, metadata, data structuring, text and data mining, data cleaning, manipulation, analysis and visualization, coding, data metrics, data management and data curation, data literacy and data ethics. As a consequence, library services related to research data continue to be a developing area where the responsibilities and practices of data librarianship within academic libraries are generally not yet firmly established.

This paper offers an overview of the strategic planning and establishment of research data services within a Canadian university campus library. The University of Toronto Scarborough (UTSC) Library is located on the eastern edge of the City of Toronto, one of three campuses that make up the larger University of Toronto system. Primarily an undergraduate community, students, faculty, and staff have access to library resources and services across all three

campuses, including a centrally located Map and Data Library. The main challenges facing the UTSC Library are: 1. to develop high-quality research data services that would meet the current needs of faculty and students in alignment with the pre-existing services of the main campus Library; 2. to build awareness and use of these services; 3. to generate support, excitement, and engagement of librarians and staff within the UTSC Library; 4. to work with librarians and staff to build the knowledge and skills needed to ensure the scalability and sustainability of UTSC's data services; 5. to be flexible enough to respond to future development opportunities in an area that is continuously shifting and expanding.

The authors of this paper will describe the process of developing research data services at the UTSC Library, beginning from the creation (proposal stage) of a data librarian position, hiring and staffing, development of a mandate in alignment with the Library's mission, to the first stages of implementation and assessment. We will discuss how these activities integrate into the wider context of the UTSC Library's user services and functional departments, specifically the Digital Scholarship Unit (DSU) and the Liaison Librarian Program. Additionally, as one of three campuses which span the Greater Toronto Area, we will discuss the role of data services at a smaller, undergraduate campus in relation to University of Toronto Libraries main campus data services and, in particular, what distinct value and services a smaller-sized library data service can contribute to the research community, as well as future developments in the field of data librarianship.

Introduction

In today's research environment, the advance of information technologies and new methods of scholarly communication have both increased the demand and reach of data and raised new challenges for the library community. Data services in libraries have long been associated and valued in hard sciences, business, and social science research because replication and validation are basic principles of the scientific method. With the growing trend towards computational-based research and the multidisciplinary nature of many contemporary disciplines, such as the digital humanities, academic and research libraries now face a new data-intensive environment of study that encompasses all fields, not just sciences. As the types of data vary from discipline to discipline, researcher to researcher, the definition of data expands, taking the form of text, audio, images, video, spatial data, computational code, and more. Moreover, recent mandates from government and private funding agencies for data management, open access publication, and data sharing are compelling researchers to reexamine their research and publishing practices. In this increasing data-driven environment, academic libraries face a number of serious challenges to meet the growing demands of students, faculty, and researchers, including finding ways to promote the development of globally competitive, critically thinking, data literate citizens.

Many academic libraries are already engaged in data-related activities, while others are examining a range of roles and services to further support data-intensive research across their campuses. This case study reports on early results of the strategic planning and establishment of data services for the University of Toronto Scarborough Library (Canada). In this paper, the authors will introduce a range of options for developing a data services model that fits the context of its campus, and the associated challenges of working within that context. Beginning from the creation (proposal stage) of a data librarian position, hiring and staffing, development of a mandate in alignment with the Library's mission, to the first stages of implementation and assessment, the authors will discuss how these activities integrate into the wider context of the Library's user services and functional departments, specifically the Digital Scholarship Unit and the Liaison Librarian Program. As one of three University of Toronto campuses which span the Greater Toronto Area, we will also examine the role of data services at a small, undergraduate campus in relation to a larger, central campus library – the third largest academic library system in North America (behind Harvard and Yale). In addition, we will highlight the distinct value a small-sized library data service, like that at the University of Toronto Scarborough Library, can provide across multiple disciplines, departments, service units, and institutions in a dynamic and rapidly evolving data-rich environment.

Background

The U of T Scarborough (UTSC) is one of two University of Toronto campuses established in the 1960s in anticipation of the huge cohort of post-World War II baby boomers graduating from high school and looking toward higher education (University of Toronto Scarborough, n.d.). Today, the UTSC campus serves a diverse student body of over 12,000 undergraduate students, and approximately 700 supervised graduate students. At the centre of campus is the Academic Resource Centre, which brings together a number of academic support services, including the UTSC Library. The Library offers an extensive reference and research service, where students and faculty can receive research assistance in person, by telephone, through email, and through the online Ask a Librarian chat service. In addition to the local services and a physical collection of over 400,000 volumes, Scarborough campus students and faculty also have full access to the entire U of T Library system, which includes 44 libraries, over 12 million print volumes in 341 languages, 1,500,000 electronic resources in various formats, 28,000 linear metres of archival material, and 500 terabytes of data.

While many universities have the benefit of being served by a system of libraries, such as that at U of T, the challenge facing the UTSC library was to find a balance of services that would serve Scarborough campus users *on site* – physical or virtual – at the same time careful not to duplicate, overstep, or overpromise services to our users when they could be (possibly better) served by other special libraries and services across the system. This was of particular concern for us with the development of a UTSC Library data service, given the existence of a larger, well-established Map and Data Library located on the downtown campus (45 minutes away). Staffed by qualified and professional librarians, technicians, and students, the Map and Data Library (MDL) at the University of Toronto provides a range of services, supporting students, faculty, and researchers through reference and consultation, curricular and co-curricular instruction,

3

training, collection development, research data management (RDM), online tools, websites, and research guides. Although these specialized services are available to the entire U of T community, and MDL staff regularly collaborate with colleagues across the library system, including the two suburban campuses, Scarborough (UTSC) and Mississauga (UTM), there are limits to the kinds of support the MDL can provide to the remote campuses. It is a real challenge for UTSC students and faculty, for example, to access or request curricular and co-curricular workshops, to visit the MDL reference desk for assistance, or to book an appointment with a data librarian. Of equal concern for UTSC Library staff, was the fact that services available through the Map and Data Library could not be tailored to meet the growing needs of students and researchers at the Scarborough campus. Librarians at UTSC had previously built a successful Information Literacy Program and Digital Scholarship Unit by focusing on the needs and requirements specific to the UTSC community. When it came to building a data service, it was extremely important to develop services that could build on and help advance these highly valued Library services and resources.

The Challenge: Narrowing the Knowledge Gap

Until recently, the challenges and concerns listed above had little impact on users as most faculty and researchers tended to only come to the UTSC campus to teach classes; the majority of researchers and faculty would conduct their research programs from offices at the central St. George campus (downtown Toronto) and could therefore rely on the services and resources offered at the Map and Data Library. And as a primarily an undergraduate campus, data-related questions from students were few and far between. Over the past few years, however, there has been a growing shift in the research culture at UTSC. In an effort to encourage research on campus, the Office of the Vice-Principal Research developed specialized services and supports for faculty researchers; laboratories were physically located on campus; researchers began enlisting a significant number of UTSC students as research assistants. At the same time, the UTSC Library established the Liaison Librarian Program which provided faculty with quick, tailored responses to problems and needs relating to a range of issues, including research support and in-class instruction. In a very short amount of time, Liaison Librarians built a reputation on campus as trusted partners and advisors to researchers and faculty. And as researchers became more ably supported by the Library and services at UTSC, they began to shift their entire research portfolios to the Scarborough campus, including their expectations for the same level of data-related support and resources they received from the downtown campus.

At this time, changes in the broader national and international research culture were also taking place. Funding agencies in the UK, US, and EU began taking steps towards the idea, and in some cases expectation, that data gathered and assembled for research purposes should be readily accessible to the public, especially when such research activities receive funding from the public sector. In 2016, Canada's three federal research granting agencies — the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC) — adopted the *Tri-Agency Statement of Principles on Digital Data Management*. The Statement

outlines the agencies' overarching expectations for researchers, research communities, and research institutions, stating that "research data management is necessary at all stages of the research project lifecycle, from design and inception to completion and beyond" (Government of Canada, 2016). The UTSC Library, building on the efforts of Liaison Librarians to meet the needs and expectations of researchers on campus, now began deliberating over how to address funder mandates for data sharing plans. This new role, which includes providing researchers with "guidance to properly manage their data" created both an opportunity for librarians to develop new skills and presented a challenge for the Library to build complementary services to address the needs of our users while being sustainable and scalable for a library and campus of our size (Government of Canada, 2016).

In an early effort to address some of the new data-intensive user needs, many Liaison Librarians participated in workshops and training opportunities on data-related topics, but the need and opportunity to apply these skills was infrequent or still beyond the basic abilities of librarians at the UTSC Library.[1] Faculty and students continued to be referred to less convenient services at the main campus, and reference support for data and statistics remained separate from the rest of the embedded Liaison Librarian Program. Other researcher needs related to research data management were addressed by functional specialists at the UTSC library, including staff

working in the Digital Scholarship Unit as well as the Scholarly Communications Librarian; however, given their limited knowledge, combined with heavy competing workloads, research data management planning and support services remained a secondary priority for these units.

The Databrarian Solution: Part I

By 2017, the UTSC Library had developed a complement of librarians with a smattering of basic data skills and knowledge of data-related issues. It was also clear by this date that we were still missing a cohesive vision or leadership for library data services on campus. In an effort to bring together this poorly served area and provide more specialized support and coherency to data services, Library management decided that a Data Librarian was needed. Reflecting back on this decision, however, just what role this librarian would fill in our relatively new research data environment was still hazy, at best.

Although data librarianship as a field has been around since the 1970s, it is not an area of specialization that is easily defined as it encompasses a varied collection of overlapping but

---

[1] This is not to say that the UTSC Librarians were in any way unwilling or unable to develop these skills. It was (and continues to be) an issue of resources and planning. Liaisons are considered subject experts; their commitment to professional development tends to focus on areas that support research, learning and teaching in a specific academic subject or related field. Data librarianship, the authors will argue below, is a set of specialized skills and responsibilities that cannot be easily (or ably) assumed by librarians with numerous other roles and responsibilities.

distinct roles.[2] Most data librarians share a common set of base skills, including knowledge of file formats, documentation and metadata standards, and disciplinary research practices, but beyond that, as Thompson and Kellam point out in the introduction to their influential text, *Databrarianship: The academic data librarian in theory and practice*, the data librarian's job "can take a myriad of forms" (2016). Job descriptions for data-related LIS roles can require specific expertise in data discovery, data instruction techniques, data structuring, text and data mining, data cleaning, manipulation, analysis and visualization, coding, data metrics, data curation, and data literacy, among others. Between different disciplines, the term *data* may be used in a variety of ways depending on field and context. A computer scientist may use the term to refer to a sequence of bits; a statistician can consider data as a set of natural numbers; a humanities researcher might use the term to refer to the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen. The skill sets required of data librarians are as diverse as the needs of our faculty, researchers, and students. As a consequence, library services related to research data continue to be a developing area where the responsibilities and practices of data librarianship within libraries are generally not yet firmly established.

The Unicorn Problem

Although the UTSC librarians felt strongly that we needed a data librarian, in many ways we needed that data librarian to be able to tell us what we needed (and perhaps just as importantly, what we did not). Librarians held discussions with internal stakeholders, including UTSC library staff, faculty, and librarians from other units including the Map and Data Library at the central Library. Each stakeholder represented a diversity of interests, making the process of writing a clear and cohesive job posting neither smooth nor simple. Some stakeholders put priority on the technical skills; the ability to do a lot of the data manipulation and statistical analysis work. In contrast, other stakeholders put priority on teaching and reference ability; they wanted someone that could do what liaisons do but for data and statistics as well as someone that had the interpersonal and communication skills to serve as a leader for data and statistics related issues in the library. The list of required skills and experience was unwieldly; we wanted someone who could do data, GIS, data mining, teaching, outreach, understood the data infrastructure, metadata schema, coding, and that same person needed to be able to show academic strengths and fit within the current library culture. The first job ad was an amalgamation of all these wants and unfortunately resulted in a failed search. A second search was run without making adjustments to the job ad and also resulted in a failed search. There were a number of good candidates identified in each process but none met the long list of requirements we had set out. The challenge for us was pushing against the idea that we would be able to find the perfect data librarian who could step in and serve as a program planner, as a liaison librarian for data and statistics for all disciplines, and as a technical librarian who also

---

[2] Dates may vary, depending on what definition is used to describe a field that is characterized by a diversity of interests. The authors use the Conference on Data and Program Library Services held in Toronto in 1974 to mark the emergence of data librarianship as a professional field. It was this meeting that led to the founding of the International Association for Social Science Information Services and Technology (IASSIST) – largely recognized as the most influential organization in the development of the profession and its identity.

conducted data manipulation and analysis. Without a vision for data services at UTSC it was hard to know what to prioritize. Ultimately, we chose to remove the more technical aspects of the original postings, and focus on finding someone who had solid foundational skills to work with the liaison librarian program and who was eager (or at least willing) to learn more. We hoped we could work together to identify the direction of data at UTSC and they could engage in professional development to fill in the gaps in their knowledge and skills as we determined what was needed here at UTSC.

Education options for data librarians remain limited even as demand for advanced data reference, management, and curation competencies is increasing in the job market. In reviewing available information of course descriptions from LIS-oriented data programs, Harris-Pierce and Liu found that significant skill sets in areas such as metadata and digital preservation are being covered, but in varying degrees (2012). More recently, a study from the National Research Council (U.S.) found that while many of the principles and skills covered in conventional degree programs are integral to data management and curation education, courses continue to be too general in nature, with inadequate attention given to the specific knowledge and skills needed for management of digital data (Committee on Future Career Opportunities, 2015). Consequently, the current pool of data librarians looking for work is small and the skills and experiences are varied. And while we had numerous internal discussions about the necessary hard skills and soft skills that we thought would make a successful Data Librarian for UTSC, it was challenging lining up our expectations with the reality of available candidates looking for academic positions. This situation forced us to reconsider our main objectives for hiring a Data Librarian: instead of looking for the (near) perfect candidate with a set of skills and experiences that made a one-to-one relationship with our job posting, we learned that we needed to be flexible and willing to consider different possible future scenarios for the Library based on the candidates that we attracted. We also had to be willing to consider this individual as a long term investment; we accepted that the successful candidate would be someone that would need support (time and money) for them to develop into our ideal Data Librarian. Instead of searching for our unicorn, we became prepared to help support the growth and development of a Data Librarian capable of meeting the specific needs and challenges at UTSC, over time.

The Databrarian Solution: Part II

The successful candidate for the position was ultimately a mid-career librarian with a robust set of professional librarian skills, strong foundational data skills, and a desire to develop and evolve those skills as needed. But hiring our Data Librarian was only the first challenge we faced. Our next concern was the proposition of essential data services for the UTSC campus.

In order to develop a model of service, we began by considering three central questions:

i. What is this thing *we* call data services?

In the 2018 IFLA Global Vision Report Summary on top ten highlights and opportunities for libraries, the number two challenge identified for the library field of the future is to "update our traditional roles in the digital age" (IFLA, 2018). The enduring value and role of libraries is to support learning, literacy and reading meaningfully. In a data-driven age, this means that libraries must adapt continuously. Services, collections and practices must develop to meet changing user expectations.

While the complexity of the data can serve as the basis for identifying the technical skills required by data librarians, it is not at all clear what services and practices to prioritize for a group of users as diverse and changing as the UTSC campus. Consultations with faculty, researchers, and students have resulted in a host of desirable services and collections, from teaching basic Excel skills to students to conducting big data analysis for faculty. In order to move forward, we had to find a way to balance the needs of our users with the priorities and campus role of the UTSC Library. Taking into account present needs, but also looking to the future and the growing needs of a data-intensive UTSC campus, we decided to focus on what many data librarians consider the essential core of databrarianship, a commitment to ensuring access to data, be they primary research data created by our institution's researchers or secondary data used for analysis (Thompson & Kellam, 2016; Semeler, Pinto, & Rozados, 2017;).

For the moment, this approach allows us to focus on data literacy skills, rather than individual tools or products. The data librarian supports reference, teaching, and collection activities in the context of helping users (at all levels) become data-literate citizens who can read, understand, create and communicate data as information.

ii. Do we train or teach?

Across academic disciplines and throughout the private sector, we are recognizing a growing need for data-literate graduates from all backgrounds. As digital data are regarded as the raw materials of the knowledge economy, the issue of training, in particular, is becoming increasingly important for all areas of society (Ridsdale et al., 2015). How can academic libraries best equip graduates with the knowledge, understanding, and skills required for the data-rich knowledge economy?

Initially, we thought that the data librarian position would support capacity building among the UTSC liaison librarians. By training the liaisons who directly assist researchers, instructors, and students with their research and data needs, we hoped to create a model of service that was both scalable and sustainable. In the short-term, however, this approach has not been well-received or successful. With only a small unit of 8 liaison librarians (with some liaison work also conducted by functional librarians and administrators) for a multi-disciplinary campus of over 12,500 students, Library staff carry an already heavy workload of diverse subjects and functional areas. And not all data-related skills and knowledge are required or used on a regular basis by librarians in the course of their day-to-day responsibilities. The desire to train and build

new capacity within the Library was quickly tempered by a range of interest-levels, time commitments, and aptitudes.

An alternative solution is to have the data librarian take responsibility for all reference, teaching, and consultations related to data. We are conscious, however, that this approach will quickly become unscalable and unsustainable. Our goal is, therefore, to explore a hybrid model that will focus on training liaison librarians gradually, using opportunities for collaboration as a way to progressively build confidence and skills. For example, when the librarian for Sociology is asked to give a workshop on Canadian census data, the data librarian offers to teach the workshop with the liaison librarian in attendance. The next time the workshop is requested, the data librarian may offer to co-teach with the liaison librarian. And the next time, the expectation is that the liaison librarian will have the confidence and willingness to teach the workshop on his own.

iii. If we build it, do we have the staff and resources to support them when they come?

The growth of data intensive research, coupled with funding mandates for data management plans and government open data has led many academic libraries to develop services in support of their faculty, staff, and students. Practically speaking, however, only a small percentage of libraries have dedicated librarians to provide data services. According to a survey conducted in 2014, for example, individual discipline librarians or staff members are the largest group providing informational/consultative research data services in libraries (Tenopir, et al., 2014). This situation raises significant concerns with respect to the Library's current capacity to support a highly specialized area, which continues to increase in demand.

Working from lessons learned with other services that the UTSC Library has built, only to be unable to keep up with demand, we are proceeding with caution. As we develop our data services model, we want to ensure it will be flexible, sustainable, and scalable for the current research data landscape, and the challenges of the future. In order to achieve this balance, we have decided to build services and support for data-related research and teaching at UTSC in partnership with academic departments, programs, offices, and other experts on campus. In addition to supporting liaison librarians with reference and teaching activities, the data librarian actively works with faculty and staff to design, build, and in some cases implement, data projects and initiatives. In the Faculty of Management, for example, the data librarian has created a series of mandatory data workshops to teach co-op students to access, manage, evaluate, and apply data in support of evidence-based analysis. In the Faculty of Arts, the data librarian is working with staff to create a similar program tailored to the needs of students in the humanities and social sciences. Working in collaboration with researchers and faculty in the Department of Physical & Environmental Sciences, the data librarian has developed a project to support RDM knowledge sharing and best practices for the campus. Other partners and collaborators include, the Office the Vice-Principal Research, Registrar's Office, Digital Scholarship Unit, and the Map and Data Library (central library). This is a partnership model that relies on a common set of goals and understandings for the entire campus, tailored to the mandate of the university. It is worth noting that while this approach provides a more robust

services model than one solo data librarian can provide on her own, our current partnerships are driving increased demand across UTSC campus. And as the demand continues to grow, we anticipate the need to build out our service model, dedicating additional staff and resources to data services in the near future.

Conclusion

In the data-rich environment of today's academic ecosystem, data librarianship is a field of knowledge that can require specific expertise in network technologies, information architecture, metadata, data structuring, text and data mining, data cleaning, manipulation, analysis and visualization, coding, data metrics, data management and data curation, data literacy and data ethics. As a consequence, library services related to research data continue to be a developing area where the responsibilities and practices of data librarianship within academic libraries are generally not yet firmly established.

Data services are complex and will naturally grow more and more complex as we move into this data-rich environment. At the University of Toronto Scarborough Library, while we have taken steps to develop data services and support for faculty, researchers, students, and staff, uncertainty and insufficient resources remain. What we have learned, however, is that there is no one single, clear answer or path forward. There are many possible futures for data services at UTSC, and academic libraries in general.

References

Government of Canada. (2016). Tri-agency statement of principles on digital data management. Retrieved from http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html

Committee on Future Career Opportunities and Educational Requirements for Digital Curation. National Research Council. (2015). Preparing the workforce for digital curation. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK293663/doi:10.17226/18590

Harris-Pierce, R. L., & Quan Liu, Y. (2012). Is data curation education at library and information science schools in North America adequate?. *New Library World*, *113*(11/12). https://www.emeraldinsight.com/doi/full/10.1108/03074801211282957

International Federation of Libraries and Associations. (2018). Global Vision Report Summary. Retrieved from https://www.ifla.org/ES/node/11905

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H. Bliemel, M., Irvine, D. … Wuetherick, B. (2015). Strategies and best practices for data literacy education: knowledge synthesis report. Retrieved from http://hdl.handle.net/10222/64578

Semeler, A., Pinto, A., Rozados, H. (2017). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science, 00*(0) https://doi.org/10.1177/0961000617742465

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., … Lundeen, A. (2015). Research data services in academic libraries: Data intensive roles for the future? *Journal of eScience Librarianship, 4*(2). http://dx.doi.org/10.7191/jeslib.2015.1085

Thompson, Kristi Anne and Kellam, Lynda. (2016). Introduction to databrarianship: The academic data librarian in theory and practice. *Databrarianship: The academic data librarian in theory and practice*. https://scholar.uwindsor.ca/leddylibrarypub/47

University of Toronto Scarborough. (n.d.). About us. Retrieved from https://www.utsc.utoronto.ca/aboutus/historical-timeline