1 Introduction

Promotion of data management has become a goal for many academic libraries that plan for the future and take part in strategic planning activities (1). Health sciences academic libraries should be aware of this initiative and should try to take part to be involved in data curation and archiving at the institutional level (2). The University of Toronto Libraries (UTL) strategic plan lists digital scholarship as one of its Innovative Inquiry goals, which includes helping researchers and scholars navigate through an increasing amount of digital information and data (3). Little was known at the Dentistry Library, University of Toronto, about the research data needs and practices of the Faculty’s laboratories. We assumed that data management occurs but knew very little about the specifics. In order to fulfill UTL’s strategic goal to promote digital scholarship, the Dentistry Library decided to learn more about the data management practices of local laboratories in order to understand what data services can be provided, if any.

In February 2014 there was a call for proposals for Toronto Academic Libraries Internship (TALint) positions. These internship opportunities were to be offered to Faculty of Information students by U of T Libraries. The Dentistry Library, under the encouragement of the former Director of Science Libraries, U of T, submitted a proposal for an internship project. The project would entail doing background research and creating a questionnaire for Faculty of Dentistry laboratory directors that would help us gather information on local data management practices. The results would help us support data management with existing services or improve or add to our services to meet data management needs. An intern was hired for a 2-year term, from September 2014 to July 2016.

This project is a program evaluation study. Therefore, following consultation with the University of Toronto’s Office of Research Ethics, it was determined that this study does not require a complete ethics review. The sharing of this data in the form of a poster or publication will be done within the framework of quality improvement. The results will benefit the local context.

In this study, we are seeking to answer the following questions:
- What are the current data management practices in the Faculty of Dentistry laboratories?
- How can the library support these practices?

2 Literature Review

The following databases where searched for literature: MEDLINE; EMBASE, CINAHL, ERIC, LISTA, LISA, and Cochrane. Search strategies were adapted for each database, but the MEDLINE search strategy was used as the starting point of the review (See Appendix A). The following literature review will describe data management in laboratories and data management services offered by academic health libraries.

2.1 Data Management in Laboratories

Automated data management in laboratories dates back to the beginning of computing. The benefit of automated data sharing was evident to those involved in health care. In 1969 Nordschow and Shipton (4) discussed the potential of a data management system improving patient care due to increased availability and sharing of the patient records in a diagnostic laboratory to patient care areas. Others went further to seeing the benefit of sharing between organizations that serviced the same patients, such as Dyer et al (5) who developed a data management system catering specifically to virology laboratory data within 17 hospitals. Technicians would log results of tissue tests and connect these to
specific patients. All hospitals in their system would have access to the results (5). The data was password protected and the program had the ability to create tables summarizing the patient test results. Litzkow, Ingram, and Lezotte talk about the large amounts of paper being used in hospitals and how a laboratory records retrieval and archive system could change that (6). They talk about the challenges of sharing and misplacing paper records and how a microform archival system and magnetic tapes accessed by an online retrieval system would improve these issues (6). In the 1980s Stillwell talked about the need of laboratory database management systems, which were different than the standard Database Management System (DBMS) available at the time, due to the need for statistical analysis or mathematical modeling of scientific data (7). He describes a program that he custom built, called TAB, that included the ability to create tables in addition to storing data. The ability to communicate scientific data in a way that could be easily understood was evident.

Fink talks about the necessity of computerized data management in biochemical laboratories at the Clinical Biochemistry Department of West Middlesex Hospital using a system developed in house (8). His main concerns however, were the lack of guidelines for such a system and the wider collaboration that is necessary within different hospitals (9). Hammond talks about the inception of a laboratory information system (LIS) in 1979 at the Clinical Chemistry Laboratories at University of North Carolina Hospital, and its development into an online system by the mid-1990s (10). All the examples thus far have been local hospital laboratory solutions. Moving through the 1990s, many laboratories viewed the increased usage of the Internet as an opportunity to share and store laboratory data beyond their institutions. Storage was becoming a major issue, and Kasten talks about an early solution of optical disk archiving in the early 1990s, which allowed him to store up to 6 years of laboratory data, in particular write once, read many (WORM) (11), the predecessor to CD-ROMs. The International Federation of Clinical Chemistry (IFCC) and the International Union of Pure and Applied Chemistry (IUPAC) created guidelines for training members in clinical laboratory management (12). Within these guidelines, laboratory results as a means for diagnosis rather than data management are discussed (12). Pascal talks about the ethics of research data management and the importance of data policies to ensure quality and accuracy of data, thus maintaining scientific integrity (13) He also discussed that poor data management in the laboratories can mislead young scientists, and therefore good data practices are critical. The institutions in which the lab is situated should be responsible for training and ensuring that the lab is following good data practices (13). Although much has been said in the literature about standardizing laboratory data management, it has yet to materialize. Many laboratories continue to operate in silos and with their own local guidelines for data management. Wiley ran a survey in 2014 that asked scientists why they are reluctant to share data, and at 42% of respondents the most popular answer was intellectual property concerns (14).

2.2 Data Management by Academic Institutions

Data management is not just a challenge in hospital laboratories, but academic laboratories as well. Academic institutions are trying to solve their laboratory data management challenges by offering data repositories where scientists can deposit their data in a secure and reliable place. One of the challenges of repositories is meeting the variety of data needs found within a single academic institution.

Popular tools to build data repositories are DSpace and Dublin Core (DC) (15). Kurtz did an analysis of three universities using both DSpace and DC for their repositories and although there are some challenges such as inconsistency within the DC subject element, personal names as subject headings and capitalization, for the most part the records were consistent with DSpace created elements in all three universities (15). Consistency makes the repositories easier to use within institutions (15). Although
there are many examples of repositories in academic institutions, there still remain a reluctance for
archiving laboratory data (14, 16).

2.3.1 Changing roles/future of academic libraries

Libraries have a natural disposition to provide information services, and data management is included in
these services. The literature on data management services offered by academic libraries is vast. This
vastness is also true of writings on services at science libraries in particular. In 2011 the National Science
Foundation (NSF) begun requiring researchers to submit a two-page data management plan as part of
each funding proposal (17). Although librarians in certain subject specialties, such as in data and map
libraries, have been involved in data management for a very long time, the NDF mandate has increased
the need for science librarians to try and provide services in this area. Purdue University created the
Distributed Data Curation Center (D2C2) in 2006 to bring researchers together to solve data sharing and
management options (17). Yet, science librarians still feel unsure about their role in data management.
Antell et al. created a 16-question survey for science librarians to answer questions about their
involvement in data management. 60.1% (n=165) of respondents indicated that they provide data
management support, while 17.8 % said that it is being planned (17). The same survey asked librarians
an open ended question about what skills they need to work on to help scientists with data
management, of which the most common response, 17.1% was “the data management cycle” (n=333). It
is evident that librarians think that an understanding of the data management cycle is necessary in order
to provide support.

The challenge is that within the sciences, each discipline has its own data culture. Scientists in some of
the smaller disciplines or the “small data” projects, don’t have incentives to share data (18). Another
barrier is notions of security and control: once data leaves their labs or computers, researchers are
concerned about loss of control that could lead to misinterpretation (18). Technological barriers of
storing and sharing data are also evident, as is format and data types within disciplines (18). MacMillan
goes on to say that librarians have a role to play in solving these issues, but as part of a team of other
stakeholders that include researchers, journal publishers, repositories, and institutions (18). Acquiring
this depth of knowledge needs to be made a pre-requisite before new infrastructures or services for
research data are developed (19). Conducting data interviews with researchers is one approach towards
achieving this foundational understanding (19-21).

2.3.2 Future liaison roles between libraries and Faculty

The consensus among the literature is that data management is a great opportunity for academic
librarians to get involved. Charbonneau (22) provides a few examples of areas where health science
librarians may wish to improve their skills, such as: data interviews, data management planning, funder
compliance; data literacy instruction; data ethics; data citation and attribution; data sharing; data
publishing; emerging data metrics; and data curation and preservation (22). Some of these skills are
traditional, such as storage and retrieval of data; others are new, such as negotiating storage solutions
and consultation services for research data curation (22). It is evident that some librarians are already
equipped with these skills, but most have to learn new skills to accomplish these tasks. Gold summarizes
the situation well in saying that social sciences and GIS data services are well developed in the academic
libraries, but services for bioinformatics data are not (23). Purchasing of scientific data sets is another
task that may be asked of science librarians, but that is not a traditional responsibility (23). Again, this
task is well established in the social sciences and GIS, but new to sciences.
McGowan surveyed library directors to determine their plans to adjust their libraries to beyond traditional library mandates (24). Eighty library directors answered the survey and 54.9% responded that a formal liaison role is taking part in systematic reviews along with researchers (24). This involvement in research teams will provide librarians with a direct opportunity to offer data management plan services, when available.

2.3.3 Repository projects led by libraries

There are examples where libraries have led data management initiatives in universities. The Scott Memorial Library at Thomas Jefferson University launched an institutional repository project in 2005 (25) aimed to house faculty publications, but which then included other institutional documents. Other libraries have led institutional repositories, but encountered roadblocks such as staffing and funding (26). Salo describes how the open-access movement promised libraries that faculty would populate the repositories, but this is not always the case (26).

2.4 Gaps in the literature

Very little is published about libraries conducting data interviews with laboratory managers to ask about data needs (19, 21), although it is mentioned as a method to get involved in the data management discussion with faculty (20). The data interviews by Lage, Losoff and Manes are similar to this project in terms of methods (21). They scheduled interviews with 26 researchers (74% response rate) for 15-30 minutes, but which lasted a bit longer than that, which suggests interest in the topic (21). They interviewed faculty from varied sciences background, but none from health sciences.

Literature exists on electronic data notebooks, there is little to outline what the needs are for small data laboratories in dentistry. This study aims to fill the gap in the literature by conducting data interviews with laboratory directors at the Faculty of Dentistry and asking them what their research data needs are.

Once the interviews are conducted and the data analyzed, it will be evident whether institutional solutions are already available to the laboratories or not. Also, it will be evident whether the library can help in meeting the needs of small data laboratories in Dentistry or not.

3 Methods

Interview questions were used to gather qualitative information on data management practices at the Faculty of Dentistry laboratories. In-person interviews were conducted, of approximately one hour in length. The set of interview questions are an adaption of the Data Curation Profiles Toolkit, which includes a series of questionnaires to be asked in interviews.a The Toolkit is the result of a collaborative research project undertaken by the Purdue University Libraries and the Graduate School of Library and Information Science at the University of Illinois Urbana-Champaign.b After checking in with one of the creators of the Toolkit, he informed us that it is meant for all to use and can be modified as needed. The Toolkit is just one of many data management surveys and interviews that have been created, due to the growing interest in gaining this kind of information about research data practices. A number of similar packages were considered before we decided to use the Toolkit, because the Toolkit questions seemed most relevant and useful for our needs.

---

a [http://datacurationprofiles.org/](http://datacurationprofiles.org/)
b For more information, see [http://datacurationprofiles.org/overview](http://datacurationprofiles.org/overview)
The Dentistry Librarian and the TALint Intern conducted in-person interviews with Faculty of Dentistry lab directors and staff. Questions pertaining to data management activities were asked. The interviews were transcribed and coded for themes. The answers, once coded and analyzed, were used to assess whether the library can help with data management processes, such as suggesting data management software or providing assistance in creating data management plans systematically.

4 Findings/Results

In total six interviews were conducted. Two of the responses are administrative managers: one of the clinics and one of the IT department. The other four are managers of laboratories and they are appointed as faculty. The subject areas were oral microbiology, cell biology, orthodontics, and restorative dentistry.

The first theme that emerged is that lab managers are open to the idea of receiving research data management support from the library. In particular, they are interested in receiving assistance for archiving and preserving their data. They are interested in having data preserved anywhere from two to five years, to indefinitely.

The second theme is that individual labs said they felt isolated from other labs and departments in the faculty. That is, they are unaware of the software, hardware, systems and methods that other labs use to manage their data. They feel they are left on their own to develop systems and methods, and a few even rely on help from family members to set up backups and to research and troubleshoot new software.

Lab managers all said that if data were to be stored in a repository or managed on their behalf, they would need to maintain some control over the data, or else be guaranteed of the safety of their data. A few of our respondents mentioned that there can be a lot of competition within their field of study and that secrecy is essential prior to publication. All respondents were interested in at least having the option of password protecting or otherwise limiting access to their data as needed.

Related to the theme of isolation, lab managers and staff are using their own (oftentimes unique or custom built) hardware and software to store their data. For example, one respondent’s use of a personal iMac computer means that the lab data is not stored on the faculty server. One lab is responsible for budgeting for all of their own hardware and software, so they are relying on mostly open source software and are independent troubleshooting and setting it up.

Finally, it was suggested by many respondents that it would be helpful to establish standardized workflows and processes to guide research data management in order to encourage dental faculty to get involved. Yet data management practices and standards vary a lot across the dentistry research disciplines, or have not yet been established.

5 Discussion

From the different themes that emerged from the interviews, we can say that if the library plans to support research data management, the following have to be considered:

- Data security is critical, including having the option of different access settings for different users
• Allowing lab managers to maintain control over their own data is recommended
• Metadata and data management planning expertise is lacking
• Preservation/archiving support is wanted
• And any data management platform we consider must be able to handle multiple data types and formats, because Dentistry faculty work with a variety of file types, including videos, images, Word, and Excel files, etc.

6 Limitations

We were limited by our low response rate, which resulted in a sample size of just 6 individuals. Because of the low response rate, we opted not to tally the quantitative part of the survey, and instead we focused on qualitative data found in answers to the open ended survey questions. These answers were coded and resulted in the five themes listed above.

7 Recommendations/Next Steps

Based on the implications of our findings, and given the tools we currently have available at U of T, at this time, if we were to offer RDM support, we would utilize Dataverse for data storage/archiving, and the Portage Network for data management planning, training and policy development. Data management plans (DMPs) are not currently required by our respondents’ funding bodies, but based on national and international trends in research funding, especially the recent Tri-Agency Statement of Principles on Digital Data Management, we anticipate that this is something they will be required to do in the near future.

The good news is that we do not have to do this alone. We have colleagues at U of T that we can call upon for help, especially the Research Data Management Librarian. We will stay involved, especially now that we’ve be the ones to start this conversation about research data management with our faculty, so we understand the kinds of data they have and their needs. It is strongly recommended that the Dentistry Library reach out to the Research Data Management Librarian or any other group at U of T as soon as possible, to share the findings of our project and to be more aware of RDM initiatives at U of T. It will also enable Dentistry Librarians to avoid duplication of effort, by pointing Dentistry researchers to UTL- or U of T-central initiatives and resources when appropriate, without having to produce our own such initiatives or resources on-site.

Beginning immediately or in the very near future, the Dentistry Library can support digital scholarship with our colleagues by providing:

• DMP training (if faculty can be convinced of the benefits of DMPs)
• Dataverse training
• Metadata support/training

These services could be offered by appointment, at the library service desk (requiring all library staff to have some level of training in these tools), or at lunch and learn or other such workshops. Dentistry Librarians and library staff may wish to seek additional training in order to be prepared to offer these

---

^https://onesearch.library.utoronto.ca/researchdata
services. The Portage Network, Scholars Portal, RDM Librarian at UofT, and the RDM Working Group at U of T may all be good resources for training and keeping on top of trends in this rapidly changing area.

Within the Faculty of Dentistry, it is recommended that, when possible, the library conduct outreach at events such as the annual Research Day. In the past the library has been involved in a research impact projects for the Dental Research Institute, which has potential to engage other members of the faculty in further dialogue about RDM. It may be helpful to look to other libraries, such as University of Alberta Libraries, for ideas for outreach events or activities. Finally, the Dentistry IT department is looking into a secure cloud based repository and hope to have something set up soon. This would be a great option for storing research data, because security is a priority in the faculty, and it will help to have a system that is maintained by the local IT team.

If time and resources allow, the Library might consider these initiatives:

- Creating an information packet and/or orientation sessions (in-person and/or virtual, i.e. videos, online tutorials, LibGuides) for new researchers joining the faculty and for new library staff
- Further research into individual dentistry journals our faculty are publishing in, and their RDM requirements (or lack thereof); as well as the funding requirements for grants awarded to our faculty; and keeping on top of changes in both of these areas
- Training or consulting with researchers interested in promoting their research via social media or a personal website (when appropriate)
- Independently, or in partnership with Dentistry’s IITS/Scholars Portal/UTL, offering extensive assistance or even taking on duties relating to the appraisal, ingest and preservation of datasets on behalf of researchers – which will require more extensive training in metadata and data preservation principles and best practices

8 Conclusion

At the very least, the Dentistry Library must be prepared to maintain current awareness of the rapidly changing world of research data management, at the Faculty-wide, university-wide, national and international levels. If resources are scarce, the Library can serve as a referral point to RDM services and tools available through Scholars Portal, Portage, UTL and the Dentistry Faculty’s IITS. However, if the Library wishes to and is able to take on a more active role in RDM, the survey responses suggest that there is some interest from faculty. This interest is expected to only grow with time, as more funding agencies and journal publishers come to require more evidence of data management planning and the publication of datasets.

9 Acknowledgements

Thank you to the Data Curation Profiles Toolkit team for allowing us to adapt their toolkit questions. Thank you also to the University of Toronto Libraries and the iSchool at the University of Toronto for creating the TALint program and making this project possible.

---

*e* [https://portagenetwork.ca/](https://portagenetwork.ca/)

*f* Scholar’s Portal Day 2015 – see webinars and white papers

*g* [https://www.dentistry.utoronto.ca/research/contact](https://www.dentistry.utoronto.ca/research/contact)

*h* [https://www.library.ualberta.ca/research-support/data-management/rdm-events](https://www.library.ualberta.ca/research-support/data-management/rdm-events)
10 References

### 11 Appendix A

<table>
<thead>
<tr>
<th>#</th>
<th>Searches</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>institutional repositor?.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>digital repositor?.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>Laboratories/</td>
<td>17746</td>
</tr>
<tr>
<td>4</td>
<td>Laboratories, Dental/</td>
<td>2171</td>
</tr>
<tr>
<td>5</td>
<td>laboratory?.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]</td>
<td>412392</td>
</tr>
<tr>
<td>6</td>
<td>(&quot;good automated laboratory practices&quot; or &quot;GALP&quot;).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]</td>
<td>984</td>
</tr>
<tr>
<td>7</td>
<td>&quot;electronic laboratory notebook&quot;.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]</td>
<td>27</td>
</tr>
<tr>
<td>8</td>
<td>ELNs.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>3 or 4 or 5</td>
<td>420955</td>
</tr>
<tr>
<td>10</td>
<td>1 or 2 or 6 or 7 or 8</td>
<td>1048</td>
</tr>
<tr>
<td>11</td>
<td>9 and 10</td>
<td>45</td>
</tr>
</tbody>
</table>