Data Management: Graduate Student’s Awareness of Practices and Policies

Sarika Sharma
School of Information Studies
Syracuse University
skhar01@syr.edu

Jian Qin
School of Information Studies
Syracuse University
jqin@syr.edu

ABSTRACT
This poster presents the preliminary results from a survey on the data management practices of graduate students. We investigate what graduate students know about data management from the technical aspects such as the familiarity and use of tools and technologies along with their participation and familiarity with data standards and policies. Preliminary data shows that graduate students lack awareness of data management policies, technologies, and practices when it comes to storing, accessing, and sharing data. The aim of this work is to provide insight on the data literacy needs of graduate students.

Keywords
Data management skills, graduate students, e-Science

INTRODUCTION
Graduate students play an important role in research settings. They are the workforce-in-training that perform many research tasks from data gathering to data analysis. As future researchers, graduate students will be working along digitally supported scientific research such as e-science and cyberinfrastructure endeavors. These digital environments entail collaborative work, data sharing, and the use of computational tools. The fourth paradigm in scientific research emphasizes sound data management that covers basic skills in storing, managing, accessing, and preserving data. Graduate students will have to cope the greater amounts of data along with the changing culture of data sharing. In order to support these activities, libraries and information centers are trying to create educative programs to properly teach data literacy in the context of scientific research.

Current research on data management looks at both faculty and graduate student’s basic skills around data such as data collection, analysis, storage, and preservation (D’ignazio and Qin, 2008; Creamer, Morales, Crespo, Kafel, and Martin, 2012). Recent studies show that graduate students are being delegated greater responsibility with data yet they are underprepared to take on such responsibility (Carlson Fosmire, and Nelson, 2011). They also have greater autonomy to curate datasets not only for themselves but also for research teams (Carlson et al. 2011). Graduate students are becoming data stewards. As this trend rises, graduate students will need to understand how to properly manage scientific data in order to avoid negative affects to their scientific outputs.

Libraries and information centers though are challenged by the precise needs and services they could provide to support graduate students. The challenge stems from varying levels of graduate student’s skill and the diverse cultural norms of data based on discipline. Many students exhibit a wide range of practices (Carlson et al., 2013). Some student’s display self-taught skills while others mimic skills they learned elsewhere when assigning metadata, standardizing documenting practices, and/or maintaining version control (Carlson et al., 2013).

Some studies have proposed the link between data literacy initiatives as way to prepare students for data management activities (Qin and D’ignazio, 2010 and Carlson et al., 2011). Proposed data literacy initiatives discuss data literacy to incorporate both the social and technical aspects of data. Qin and D’Ignazio, 2010 provide a framework that extends data literacy to include data awareness. They describe metadata as an important teaching concept for data literacy. Similarly, Carlson et al., 2011 compared the ACRL (Association of College and Research Libraries) guidelines for information literacy to that of data literacy and found places for integration. The data literacy guideline focuses on data need, access, evaluation, use of data, and ethics of data. Both studies recognize the role of being aware of cultural climate and norms in one’s discipline as a facet of data literacy.

Data literacy serves as a good platform to understand data management but little is known about what constitutes a strong data literacy program that translates into strong data management skills. In order to explore this, we use the framework proposed by Qin and D’Ignazio, 2010 and
Carlson et al., 2011, in order to investigate graduate student’s data literacy awareness as a way to understand data management preparedness. The data literacy framework provides the assessment for the awareness of metadata, technologies, and policies around data. Our study was interested in learning about whether graduate students are aware of the types of data they use, kinds of data they work with, and the tools and technologies that support the management of their data. The study was driven by the following research question: Are graduate student’s familiar with data management policies, trends, and technologies? Baseline understanding of graduate student’s awareness of data literacy may provide insight on their preparedness for data management.

METHODS
To broadly understand graduate students awareness of data literacy a survey was deployed to graduate students at the master’s and doctoral level at a mid-size research university. The survey consisted of questions that spanned from data mediums, data types, data formats, data management technologies, and data management policies.

The survey was created with the help of a data librarian on campus. After several iterations, including a pilot that was presented to colleagues, the questionnaire concluded with fifteen questions. Respondents answered the survey using a Likert scale, multiple choices with fill in options, yes/no questions, and unsure selections. The survey was deployed through SurveyMonkey. Approximately one thousand and two hundred students were sent the survey of which one hundred and seventy-three students responded. Low response rate could not be controlled. Emails for participation were sent out twice, yet even with a broad dissemination, we received a low response rate.

FINDINGS
The one hundred and seventy-three students we surveyed come from a range of disciplines ranging from the social sciences, natural science, and health sciences, and engineering. We report on their familiarity of data types, data management practices, management standards, and data policy.

Type of data
All of our respondents were involved in some sort of data collection whether it was physical data or digital data. Data formats ranged from analog to digital, audio, video, mixed and artwork but the majority of respondents worked with digital data. Accordingly, all of our respondents were involved in some tasks related to data management. The most prevalent task graduate students were involved in included data analysis, followed by data visualization, data conversion, and lastly, data cleaning. The most frequent size of data used by graduate students was between 1MB to 99MB, yet approximately twenty percent of the respondents indicated not knowing the size they use.

Data Management: storing, accessing, and preserving

Most of the graduate students from our sample either heard about data management from their peers or through their research groups.

The use of tools and technologies were based on personal preferences. In order to preserve and maintain data for future access, our respondents primarily used personal computer filing systems. The second most popular software system was cloud storage systems, followed by removable storage. Content management systems were the least used technology for storing data.

Management Standards
Only a small percentage of our respondents were familiar with data management standards in their discipline. When asked about metadata standards most graduate students were not aware of the standards in their discipline. The majority maintained the context of their data based on their own practices and planning followed closely by discussions in the research lab, and examples from peer researchers. Most students were neutral or disagreed when asked about whether they have participated in submitting datasets to community/national/international repositories.

Policy
Of the one hundred and seventy-three students only forty were aware of policies that relate to data management. Only forty respondents were aware of the funding research policies and data sharing.

DISCUSSION
Using the framework on data literacy provided an approach to study data management preparedness. From our data, a large number of our participants displayed a lack of awareness for some of the basic skills required to be data literate. Many respondents were able to describe of their data type but some respondents were unable to indicate their file size. A large percentage of our respondents were also unaware of the term metadata. Many students were unaware of data sharing in their discipline signifying a lack of awareness of the cultural norms in their discipline. It was also evident that students were unaware of policies related to data management.

The level of awareness did not differ between disciplines, rather we did not notice any difference between graduate students from the natural sciences or the social science, students from all disciplines displayed a lack of understanding for data management policies and data sharing trends. This signifies that data literacy awareness may be conducive to all graduate students regardless of discipline.

Because our sample displayed a lack of awareness for some of the basics of data literacy, including technical and cultural aspects of data, it may signify an under appreciation for the value of data. In order to foster better data literacy, graduate students may initially need to learn about the important role that data plays in eScience
scientific work. Without the basic understanding of why data literacy matters, students may not take notice of policies, standards, and trends. Awareness of both social and technical factors of data may play a role in the overall ethos of graduate students towards data literacy. The next step of this study will to understand how the results on data literacy inform data management preparedness.

CONCLUSION
Our research study focused on graduate students’ awareness of data management. Preliminary results display a lack of awareness around current data management practices such as the data policies and standards. Our next goal will be to conduct a full analysis on the data and provide further insight on graduate students.

REFERENCES


