Scholarly Publishing across the Disciplines: Storing and Sharing Research Data

By Bonnie J. M. Swoger

Introduction
In the past few years, articles about data management, preservation, and sharing have become increasingly prominent in the academic press (e.g. Gibney & Van Noorden, 2013, Poline and Poldrack, 2013, Tenopir et al., 2011). Researchers are making new discoveries due to changes in computing power that enable the analysis of larger and larger datasets (Kambatla, 2014). Publishers explore how to cite data and make data available to researchers (e.g. Silva, 2014). Funding agencies are increasingly requiring grant applicants to develop data management plans and to make data from publicly funded research available (e.g. Holdren, 2013).

In order to gather some basic information about how Geneseo faculty fit into these larger trends, we asked faculty some optional questions about how they manage and share their research data.

Of the 87 faculty members we spoke with, 59 answered our questions about data. Faculty who didn’t respond either felt the topic didn’t pertain to their research and scholarship (most commonly in the humanities) or simply didn’t have enough time to answer these optional questions.

Our questions were not intended to provide a comprehensive summary of data related activities on campus. Instead, we asked focused questions intended to provide some general information about sharing and storing research data that could help library staff understand the practices and potential needs of Geneseo faculty in these areas.
What is data?
We didn’t ask faculty to define “data.” However, when we asked faculty about the kinds of data they are creating, we learned about how faculty in the disciplines perceive “data.” We also learned about the wide variety of data resources used and created by Geneseo scholars.

As expected, the type of data used and created by faculty varied by discipline. Faculty in the sciences were more likely to have computer code or data related to specific analytical equipment, whereas social sciences and humanities faculty were more likely to use interview transcripts or surveys. Science faculty are creating genetic sequence data, chemical structure data, and geospatial data for use in geographic information systems (GIS).

Social science faculty were more likely to qualify a description of their data with “quantitative” or “qualitative,” suggesting that they used a wider variety of data types than faculty in the sciences. Science faculty used very little qualitative data and were therefore less likely to make the same distinction.

Across all disciplines, faculty were using images, photos and videos in their work. Faculty in the sciences used microscope images and field photos. Social science and humanities faculty used historical images or videos of interviews. Some faculty used specialized software to take measurements from their images while others used them largely for illustration or teaching purposes.
In a digital environment, work products of all kinds become data. Our interviews suggest that faculty perceptions of data are growing beyond the traditional image of spreadsheets and numbers.

**Storage, backup and preservation**

Geneseo faculty are using a variety of tools and strategies to store and backup their data. The most common techniques were local backups to desktop hard drives, flash drives or secondary computers. One faculty member specifically mentioned Time Machine, although the program is likely being used by others that mentioned desktop hard drive backups. Others had data saved to CDs or DVDs. Collaborators or student researchers often had secondary copies of data, and a few relied on disciplinary archives.

Faculty across the disciplines kept paper copies of research data as archival backups. Original field notes, lab notebooks, behavioral observations or interview notes were kept in addition to printed-out copies of transcripts, chemical analyses or data tables. Faculty used three ring binders, carefully organized bookshelves, piles of paper on work surfaces or storage containers to store their paper copies.

Ten faculty members used data storage services located outside of their offices or labs. Geneseo Files spaces were used most often, one faculty member indicated they backed up data remotely to storage space at another institution and one faculty member mentioned cloud services (Dropbox, Drive, etc.). Because we did not ask for details about faculty computer backup practices, we feel that these reported numbers are an underestimation of the faculty using these services.

Several faculty admitted to having no backups for digital data, and some admitted to being disorganized or unsure about where to start.

A couple of faculty mentioned catastrophic computer failures that resulted in data loss as a result of the lack of data backup plans or practices.

After hearing about faculty’s mixed experiences and practices with general data backup and storage, we asked if they had concerns about the long-term availability of their research data. Of the 59 faculty who answered our data related questions, 22 (37%) expressed concerns about the long-term availability of their data.

The most common concern was changing file formats or physical storage media. Faculty are aware that some file formats are more suitable for long-term storage than others, but they are not always sure about the most stable formats. Several faculty raised concerns about using specialty software or analytical equipment that create unusual file formats. Some faculty are preparing for potential data loss by converting to more stable formats (when possible) or printing hard copies of data.
We spoke with several faculty who had - somewhere in their office - data stored on obsolete (or nearly-obsolete) storage media such as floppy disks or even punch cards. Faculty were divided about the need to constantly update these formats. In some fields, the research moves so quickly that old data simply becomes out-of-date. In others, older data maintains its importance to the field and needs to be accessible.

As scholarly publication moves away from print media and toward online distribution, preservation of research data is also dependant on the digital preservation of the formal publications in which the data appears. Several faculty noted this and expressed concerns about the preservation of scholarly journals in digital formats.

Although we didn’t ask if they would like assistance with storage and preservation issues, several faculty members expressed a need for assistance in understanding data preservation concepts or managing their data.

Sharing data and data security
Funding agencies (Holdren, 2013) and journals are calling for increased access to researcher data, beyond the graphs and charts included in published results. Data sharing norms vary widely among disciplines, and researchers are just beginning to navigate the fallout from required data management plans. Although evidence suggests that articles with freely available data are cited more often, many researchers don’t see much benefit in sharing data more widely, especially in consideration of the cost and time associated with managing and sharing data (Piwowar et al., 2007, Piwowar & Vision, 2013).

Geneseo faculty are generally protective of their data and rarely share it publicly outside of the formal publication process (publications, presentations, workshops, etc.). In our interviews with science faculty, past experiences by some led to concerns about data theft, and several faculty members expressed concerns about the potential of being scooped. Faculty expressed concerns about the possibility that a researcher at a large research institution would be able to complete work on a project in just a few months that might take a researcher at Geneseo several years to complete due to vast differences in resources (especially personnel).

Thirteen of the 59 faculty members who answered questions about their research data had security concerns about their data (see figure 2).
Social science faculty did not mention the idea of being scooped, but expressed concerns about protecting the sensitive personal information that they collect as a part of their research. Several social science faculty members mentioned laws or ethical restrictions that require them to secure the personal data they collect. For a few faculty members, this means destroying data when they are done with a project.

When Geneseo faculty share their data, they often reported sharing it with collaborators or students (as part of class or lab assignments). A few faculty in the sciences are contributing their data to well-established disciplinary repositories for scientific data.

Although faculty are not publicly sharing their data, a quarter of the respondents indicated that they would share data directly with another researcher if asked. Some faculty qualified that statement suggesting that they would only share if they have completed their own research.

Other data issues: The Data Issues Exploratory Group
In addition to the faculty interview questions related to data, Milne Library staff have been exploring data related issues through informal conversations with Geneseo staff and internal professional development. In the summer of 2013, Milne Library gathered a group of interested faculty and staff to discuss their data related interests and concerns. Informally called the Data Issues Exploratory Group, this group of faculty and staff discussed a wide range of data-related issues.

Many of the concerns discovered in our faculty interviews were also discussed with the Data Issues Exploratory Group. DIEG participants expressed concerns about data storage best practices, had questions about the distinctions between local, network and cloud based data storage solutions, and wondered about the role of SUNY Geneseo in the preservation of faculty
research data. The DIEG also had questions about sharing and publishing data.

The informal discussion in the DIEG meetings also revealed other important data related considerations. Faculty had many questions (and a few answers) about the use of data and data analysis tools with students, including best pedagogical practices. There was a lot of interest in sharing information across campus related to common tools and procedures for data analysis. Faculty wanted to know what resources were available to them.

Institutional policies and procedures were of interest to some. Participants wondered about policies and procedures regarding intellectual property and the role of institution in preserving data once the person responsible for collecting the data is no longer at Geneseo.

The DIEG discussions were informative and interesting and helped provide a wider view of faculty concerns than our interview questions.

**Conclusions and Recommendations**

In our efforts to learn more about the data storage and sharing practices of Geneseo researchers, we learned that practices and knowledge vary widely between disciplines and among individual researchers.

Existing services could assist many faculty with their data storage and backup concerns. Faculty could benefit from the knowledge and expertise of Geneseo staff in CIT and Milne Library regarding data storage and preservation. Plans are underway to develop workshops about these issues as a part of the Geneseo Professional Development Series.

Armed with better knowledge of file formats and media types, faculty that need to convert data from one format to another still need to find the time to do the conversion. It may be possible to train departmental work-study students or other student employees to assist with some of this work.

As we seek to provide professional development for faculty and staff at Geneseo, library staff recognize several areas in which we need greater knowledge. A small group of librarians is currently exploring best practices in data archiving and preservation in an effort to preserve Milne Library information and provide assistance to the campus as necessary. Milne Library is also exploring how the library can assist campus faculty, staff and students acquire datasets. Some data is freely available online, but other datasets can be quite expensive. The library is working with campus departments to explore how the campus can make the best use of limited resources.

While the ad hoc Data Issues Exploratory Group was able to shine a light on the data-related concerns and issues of Geneseo faculty and staff, the loosely-organized group is not well situated to take action on many of the concerns that require interdepartmental collaboration. It
may benefit the campus to consider forming official advisory groups to assist with some of the data related issues and concerns.
Works Cited


Download a .PDF at http://libguides.geneseo.edu/AcrossTheDisciplines