



August 2011

# Planning for Data Curation in the Small Liberal Arts College Environment

Sarah Goldstein

Mount Holyoke College, [sgoldste@mtholyoke.edu](mailto:sgoldste@mtholyoke.edu)

Sarah K. Oelker

Mount Holyoke College, [soelker@mtholyoke.edu](mailto:soelker@mtholyoke.edu)

Follow this and additional works at: <http://jdc.jefferson.edu/scitechnews>

 Part of the [Physical Sciences and Mathematics Commons](#)

[Let us know how access to this document benefits you](#)

## Recommended Citation

Goldstein, Sarah and Oelker, Sarah K. (2011) "Planning for Data Curation in the Small Liberal Arts College Environment," *Sci-Tech News*: Vol. 65: Iss. 3, Article 4.

Available at: <http://jdc.jefferson.edu/scitechnews/vol65/iss3/4>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in *Sci-Tech News* by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

## SciTech Division Contributed Paper

*The following paper was presented at the 2011 Special Libraries Association Annual Conference in Philadelphia, PA. Other contributed papers will appear in a later issue of SciTech News.*

### Planning for Data Curation

*Presented by Sarah Oelker  
Mount Holyoke*

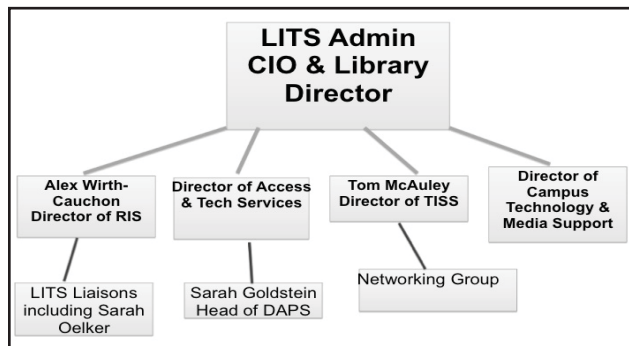
#### Introduction

Like libraries at many smaller schools, Mount Holyoke College's (MHC) Library, Information, & Technology Services (LITS) has been supporting their faculty in all manner of pedagogical and research endeavours. Questions about the data management plan (DMP) requirements for NSF grant applications generated a spate of internal meetings within LITS as well as a process of reaching out to the MHC Sponsored Programs Office. Our meetings with faculty on this topic tended to have two parallel and related tracks: one was "can you help us write a DMP" and the other was "I have all this stuff, how can I store it/share it/access it/develop it?"

#### I. Who We Are

We are a small, all-female liberal arts college in western Massachusetts with approximately 2,500 students and about 250 faculty. Library, Information, & Technology Services (LITS) is a blended library/IT group that merged in 1996, and which currently has a staff of sixty-five people. We do not have one office or department specifically for data/scholarly publishing, nor do we have a data services librarian position. Our institutional repository is a DSpace installation called IDA (Mount Holyoke College, 2009); it currently contains one collection of research data from a faculty member's NSF grant that expired in 2009.

MHC is part of the Five Colleges Inc., a consortium that includes Smith College, Hampshire College, Amherst College and the University of Massachusetts at Amherst. Thus, we have a built-in set of networks, committees, colleagues, and inter-institutional contact. Five Colleges, Inc. helps coordinate shared resources for teaching throughout the consortium; any student enrolled at any of the schools may take



*Figure 1: Library, Information, & Technology Services organizational chart*

classes at any other school. We also maintain a shared library catalog and circulation policy, as well as a shared library depository. Initiatives are underway on many levels to expand our shared digital resources.

Mount Holyoke usually defines itself as a small liberal-arts women's college, but on our campus we sometimes refer to ourselves as a "mini-Research I." We share our focus on research and a high level of federal support with these larger institutions, and we maintain very well-equipped science facilities for an institution of our size. With no graduate students in the sciences and very few postdoctoral fellows, we nevertheless maintain very competitive research groups. Over the past ten years, more than fifty National Science Foundation (NSF) grants have been awarded to MHC faculty in various disciplines (M. Caris, personal communication, June 3, 2011). Mount Holyoke provides undergraduate research experience to a large number of undergrads each year, including intensive hands-on lab work in the curriculum, and it sends many of its students on to graduate work in the sciences. Unsurprisingly, MHC was founded by a chemist: Mary Lyon pioneered learning via lab experiments, instead of solely by lecture and rote learning. Between

# WHERE YOU GO WHEN YOU'RE SERIOUS ABOUT U.S. GOVERNMENT SCIENTIFIC AND TECHNICAL RESEARCH - [www.ntis.gov](http://www.ntis.gov)



U.S. Department of Commerce

**National Technical Information Service**

NTIS offers a selection of digital products, with 30,000+ new titles added annually, to meet the needs of your organization:



U.S. Department of Commerce

**National Technical Reports Library**

- Online subscription access to 2+ million metadata records
- Links to more than 600,000 digitized full-text reports
- Easy to use interface
- Unlimited access to recent and legacy reports
- Go to [www.ntis.gov/products/ntrl.aspx](http://www.ntis.gov/products/ntrl.aspx)

## **Bibliographic Database**

- Comprehensive database resources with 2+ million metadata records (most include abstracts)
- Database available through a variety of well-known commercial services or lease directly from NTIS
- Value-added products from commercial services support powerful search and retrieval
- Go to [www.ntis.gov/products/ntisdb.aspx](http://www.ntis.gov/products/ntisdb.aspx)

## **NEW:** **Selected Research Service**

- Full text reports in your choice of subject category
- Includes the full bibliographic content in both xml and html formats
- Updates available via secure FTP site
- Subscribers download content and can provide access on their Intranet
- Go to [www.ntis.gov/products/srs.aspx](http://www.ntis.gov/products/srs.aspx)

National Technical Information Service • U.S. Department of Commerce • 5301 Shawnee Rd • Alexandria, VA 22312 • 703-605-6000

1910 and 1969 Mount Holyoke college graduated more women who went on to obtain doctorates in the physical sciences and engineering than any other American institution (Tidball & Kistiakowsky, 1976). In chemistry, particularly, Mount Holyoke graduated 93 women between 1920 and 1980 who went on to obtain doctorates in chemistry, the most women to do so from any institution, single-sex or coeducational, in the United States for that period (Hall, 1985).

## II. How We Started, Or, "The New Machine"

In March of 2010, a new LITS department was created, called Digital Assets & Preservation Services (DAPS), headed by Sarah Goldstein. This department was charged with centralizing the production of digital content, managing the campus repositories, and developing long-term digital preservation strategies for unique, archival, and scholarly digital materials. DAPS has four staff members, including Sarah Goldstein, a metadata and systems librarian, a digitization specialist, and a visual resources curator. Goldstein was new to the world of scientific data management, but identified it as an important part of her department's fledgling mandate.

During the summer of 2010, LITS staff had internal meetings to discuss NSF's data management plan requirements that were about to take effect. Sarah Goldstein and the LITS Director of Research and Instructional Services (RIS), Alex Wirth-Cauchon, met to discuss ways of supporting the requirement. Wirth-Cauchon's RIS team included the subject liaisons, who were already familiar with many of the science faculty and their projects and could help provide Goldstein with the background and language necessary to develop data management plans. Wirth-Cauchon and Goldstein subsequently met with then-Associate Dean of Faculty for Sciences Craig Woodard, who works closely with the MHC Office of Sponsored Research. Their conversation covered data management plans, but also data storage. Some of science faculty were outgrowing their departmental storage solutions and other departmental resources and wanted LITS assistance with better solutions.

Around this time, Alex Wirth-Cauchon coined the phrase "building a new machine" to refer to the new DAPS department and the new col

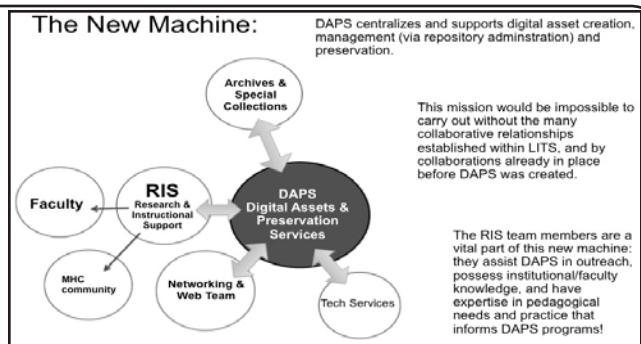


Figure 2: The New Machine

laborations within LITS necessary to manage digital assets. In the case of research data management and storage, this phrase seemed particularly apt. In our small college environment, there was simply no way that one person or even one department could hope to build the infrastructure to support meaningful research data management.

To understand the growing storage needs, we undertook an informal survey of digital research data within our institution. The LITS Liaisons reached out to faculty in the hard sciences departments<sup>1</sup>, asking the following questions:

1. Approximately how much digital research data (in any format, whether images, video, or quantitative data, in megabytes) do you currently have? (If you need to use larger units, please indicate the units you use)
2. In the course of your work, including any anticipated grant-funded projects, approximately how much data do you expect to generate between now and June of 2012?
3. Have you lost any information recently that wasn't backed up?

We invited them to email or call with their responses and asked the Associate Dean of Faculty for Sciences to encourage participation. We compiled the answers in a spreadsheet and shared the results with Tom McAuley, the Director of Technology Infrastructure and Systems Support (TISS), the LITS group responsible for the campus network architecture.

<sup>1</sup> It was agreed that the first round of surveying would be done with the natural and physical sciences, because these departments were expected to have the highest and most urgent needs, and that social sciences and humanities data would be collected later.

The results were enlightening and offered an interesting view into how our faculty were managing their data. Current data management practices at MHC were a hodge-podge of technology, people, and needs. Some faculty were making use of available network storage, some were purchasing extra hard drives, some were doing both. Many had done well on their own, but some wanted more assistance. While the survey did not unearth any emergencies, and our data management requirements are small compared to larger institutions, we are still required to provide long-term care for the data being produced. If we were in fact building a “new machine” of overall digital asset management, this was an excellent area for internal collaboration. Our survey gave us a sense of the direction we wanted to move in, and reinforced the work DAPS and the Networking group were already doing on developing the functionality of our repository tools (CONTENTdm and DSpace) as well as laying down practices and workflows for later digital preservation.

Around the same time, a meeting was held that included the Associate Dean of Faculty for Sciences and the sciences department chairs, along with Alex, Sarah, and Tom from the TISS group. The group from LITS wanted to get some feedback on how best to deploy resources and offer assistance. Help for faculty who were writing DMP seemed to be the most pressing topic, and the LITS group offered to work closely with any faculty PI (Principal Investigator) who had questions or concerns. Subsequently, we decided to approach the storage issues separately while our TISS group made improvements to the network architecture. We followed up our science chairs meeting with a LITS blog post (Goldstein, 2010) on data management plans and the NSF, including outside resources and places to find examples and templates. Sarah Oelker, a RIS team member and science liaison, was tapped to work with Sarah Goldstein to form our fledgling data management “response team.” Our internal collaborations were now in place and communication on longer term issues, such as improved network architecture, were ongoing and relatable to our overall digital curation goals.

### III. Learning On The Fly

The rest of the fall and winter passed quietly into Christmas, but one day in early January came requests for help from two MHC grant

teams, one of Physicists, and another of Political Ecologists. Both needed a DMP written by the end of the day. Could the folks in LITS help write the plans? Did we have a template that could just be filled in? Unfortunately, we did not. Sarah Goldstein and Sarah Oelker decided the best thing to do was come up with a boilerplate set of text on how DAPS could support data management. This was given to the PIs to incorporate into their DMP. We also provided the PIs with examples of other DMPs available online, including templates posted by research institutions. Within a few hours, the PIs had crafted the plans, we all helped review and refine them, and they were added to the applications.

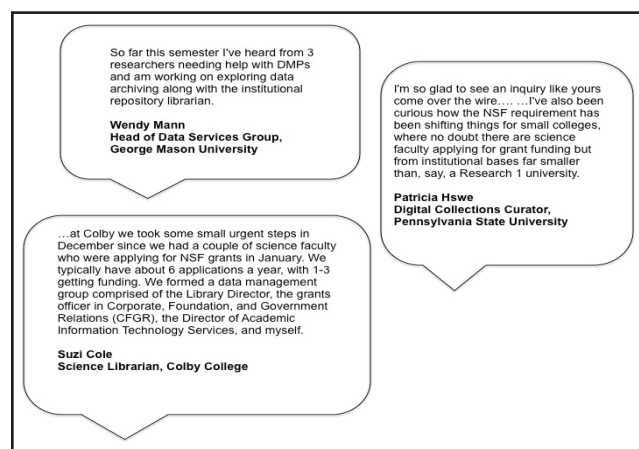
This exciting 24 hours spurred the development of some expanded LITS-based resources including a LITS web page (Mount Holyoke College, 2011) with links to DMP examples and templates. Following up on a request for a custom-made DMP template, we created one based on excellent templates from University of Chicago (University of Chicago, 2010) and the University of Virginia (University of Virginia Library). Still in the works are plans to archive and make available the DMPs of MHC faculty who agree to share them with colleagues, thus creating a “library” of browsable DMPs for other PIs to use.

This experience encouraged us to explore further efforts at external collaboration. We turned to colleagues at our Five Colleges institutions, in particular, our friends at UMass Amherst. As the local research university, there was already some specific groundwork in place for data management along with their successful IR, ScholarWorks. UMass had formed a Data Working Group in 2010 (University of Massachusetts Amherst Libraries, 2010), and they were more than happy to meet with us in March of 2011 and share what they had learned so far. It was heartening to find out that, save for being a few steps ahead and with more staff in place to respond, UMass was not yet too far ahead of MHC, especially in terms of information gathering, outreach to faculty, and resource development. We had each spent some time bringing together a cross-departmental team or task force, conducting informal surveys and interviews, analyzing the results, disseminating what we hoped was helpful information to target constituents, developing online resources, and assisting in the writing



of DMP. All of this was very heartening for a small liberal arts college to know: not only were we not hopelessly behind compared to a larger neighbor, we were not playing catch up in our overall strategy or awareness of good tools to share with faculty. It appeared that our smaller size and our merged Library-IT organization also provided us advantages, making it easier to get the right people talking to each other. The groups decided to meet again informally and to engage colleagues from the other three institutions to join us the next time. We left the meeting feeling a little more confident in our efforts thus far and reassured that we now had a larger and more local group to call upon with questions and initiatives.

During March 2011, we also surveyed some of our peer groups via their listservs, focusing on the Association of Research Libraries data sharing support group, the Oberlin Group Science Librarians list, and the Oberlin Group Directors list. We briefly described how our organization worked, the activity we had seen and the templates we had created, and asked them to share their own experiences with data curation and faculty outreach.



*Figure 3: Responses from our colleagues in the Oberlin Group and the Association of Research Libraries*

On each list, the responders were happy to hear from another small school on this topic. Many were at similar points in the process: they were looking to each other for templates, and many had suggestions about web resources and templates they had looked to for advice, most commonly the University of Virginia and University of Chicago resources we had used. Some of them had encountered their first re-

quests for DMP help during the fall 2010 round of grant deadlines, but many had, like us, not heard from any faculty PIs until January 2011, when the winter deadlines were approaching. Multiple libraries reported emerging opportunities for informal collaboration with other units in their institutions, whether in IT or in a Grants or Sponsored Research office, to compare notes and provide assistance. Each campus had their own unique needs: different configurations of offices, whether Library and IT were merged or not, and the flavor of the relationships between librarians and faculty, between faculty and IT, between Library and IT, or between the grants or sponsored research office determined the best path. In almost all cases, respondents said that they were themselves mostly at the beginning of the same process, essentially feeling their way forward as we were.

There were other similarities at various institutions which echoed what we heard from our Five College peers: faculty are just-in-time people, juggling teaching, research, and grant writing, along with the rest of their lives. They need just-in-time assistance on DMP that is easy to follow and straightforward, and they expect that those assisting them will have a good grasp of the data management concepts with which they are unfamiliar. We anticipate that as faculty get accustomed to writing DMPs, they will get better at them, and will begin to give us feedback about what makes a better plan, and how their data needs are changing. We know we will need to be prepared by sharing information with campus partners and talking with our colleagues at other institutions about what to expect.

#### IV. Our Next Steps

We continue to work with our Office of Sponsored Research and the Associate Dean of Faculty for Sciences, as well as Tom McAuley in TISS, in approaching a strategic long-term approach to data storage and backup infrastructure. DAPS sends out a quarterly email reminder to the sciences chairs that help is available for NSF or other data management plan requirements. If the Office of Sponsored Research hears from a faculty member about getting a grant ready, they now make sure to cc DAPS and the appropriate RIS liaison and urge them to contact us and use our resources (we are currently assisting a third faculty PI on her NSF application). DAPS is also working with

the staff in the TISS Networking group on improving our DSpace repository to ready it for any research data collections that come out of grant requirements. These are small steps to be sure, but they represent cohesion within the LITS departments on how to provide services.

One very important thing we learned was that requests for help almost always come last minute and must be accommodated. We are working to make sure that the DMP templates we are working on today are better than the DMP templates of three months ago or six months ago. Our initial failure to have a DMP template or boilerplate text ready when it was most needed was humbling and an excellent motivator to dive in, create those resources, and then work to make them better. DAPS is also working out a test model for a digital preservation workflow which will help inform our practices in archiving research data.

We intend to reconvene with our UMass Amherst colleagues late this summer or early fall and include a larger group from the other Five Colleges. We want to discuss options for bringing in outside experts to hold a future Five Colleges session on data management for faculty as well as staff. We hope that this will lead to even more shared efforts at helping faculty PIs better understand the role of the DMP and the importance of data curation. While an option for a shared repository is not yet viable, the sharing and preservation of digital collections is a current set of consortial initiatives being seriously discussed among our institutions.

Most importantly, we feel that we're better prepared to respond to needs as they arise than we were at this point even just last year: additionally, we know we have the ability to reach out within our own school, our own consortium, and beyond with confidence. We also know that in many cases, there's no need to recreate a small universe of resources and tools related to data management. Many things your faculty may find helpful are already online, easily accessible, and ready to be shared.

We know from our own experience that the disadvantages for a small college in dealing with data curation are still real: lack of experience among staff, lack of infrastructure or tools, "silos" of responsibility, budget shortfalls, etc. We imagined we might be hopelessly unprepared for providing assistance, and we weren't even

sure what kinds of assistance we could or should be offering. But by continually reaching out we discovered we were certainly not an outlier, and realized we either had expertise close by via a peer or colleague, or genuinely good prospects for longer-term planning and projects if we made the effort to collaborate. We also learned that no effort or attempt at outreach or communication to either a peer, a colleague, or a faculty PI was too small and was likely to be very appreciated.

## V. Your Next Steps

We urge our colleagues at smaller schools to adopt a policy of cooperation and collaboration. Who would you talk to in your Office of Sponsored Research or similar office? How well connected are you with the people at your institution who manage networked storage solutions? This varies by institution, and we know that at MHC we are lucky to be part of the same division as our networking colleagues. Our survey of digital research data was a tool that helped continue a dialogue with our networking group, and it gave us a much better sense of how to plan for future allocations of storage, at least in the short-term. It may be helpful to create a list of the people who, in your community of professionals at peer schools, would be likely to have similar data curation concerns. It may be that each institution has complimentary staff skills; for instance, if School A has a metadata and systems specialist and School B has a research librarian with a needed subject expertise, perhaps there's an opportunity to connect and help each other with DMP requests or how to deal with a specific dataset file type. Even if you can't build a perfect storage solution or repository tool right away, opening up communication to faculty, administrators, and your peers may help you provide assistance to the most urgent needs first while gathering momentum for longer-term planning down the road.

At MHC the immediate proactive work will be in continually building a network of resources and people who can help in the just-in-time phase, while we expand and improve our repository tools and network capability at a more modest pace. This is not happening by accident: the National Science Foundation expects institutions to improve incrementally at managing our researchers' data, and while they may not have directly intended libraries to strengthen their ties within their institutions, we can do that in

the process. In the longer term, we know that the "new machine" at MHC needs to grow into

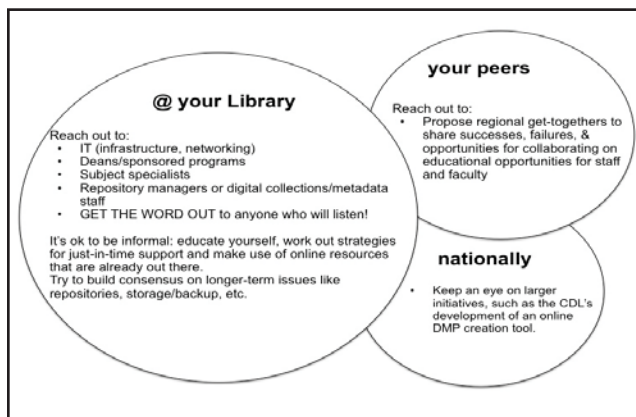


Figure 4: How to build your own New Machine

something that addresses the ever-growing need for reliable storage and backup, for scalable systems to manage access to data, and shared strategies for preservation. More importantly, when we are visible, collaborative, and communicative in this process we can, hopefully, reinforce the value of libraries, repositories, and data curation at our institutions.

## References

Goldstein, S. (2010). LITS :: Data management and the national science foundation. Retrieved 7/18/2011, 2011, from [https://pub.mtholyoke.edu/journal/LITS/entry/data\\_management\\_and\\_the\\_national](https://pub.mtholyoke.edu/journal/LITS/entry/data_management_and_the_national)

[edu/journal/LITS/entry/data\\_management\\_and\\_the\\_national](https://pub.mtholyoke.edu/journal/LITS/entry/data_management_and_the_national)

Hall, A. E. (1985). Baccalaureate origins of doctorate recipients in chemistry: 1920-80. *Journal of Chemical Education*, 62(5), 406.

Mount Holyoke College. (2011). (a). Data management :: LITS :: Mount Holyoke College. Retrieved 7/18/2011, 2011, from [http://www.mtholyoke.edu/lits/learn/data\\_mngmt.html](http://www.mtholyoke.edu/lits/learn/data_mngmt.html)

Mount Holyoke College. (2009). (b). MHC institutional digital archives: Home. Retrieved 7/18/2011, 2011, from <https://ida.mtholyoke.edu/jspui/>

Tidball, M. E., & Kistiakowsky, V. (1976). Baccalaureate origins of american scientists and scholars. *Science*, 193(4254), pp. 646-652.

University of Chicago. (2010). Template for NSF data management plan.

University of Massachusetts Amherst Libraries. (2010). Digital initiatives - UMass amherst library. Retrieved 7/18/2011, 2011, from <http://www.library.umass.edu/about-the-libraries/digital-initiatives/>

University of Virginia Library. Scientific data consulting group: NSF data management plan templates. Retrieved 7/18/2011, 2011, from <http://www2.lib.virginia.edu/brown/data/NSFDMP.html> ❖