

## Planning Cooperative Data Curation Services

Susan Wells Parham, Research Data Librarian

Sara Fuchs, Digital Initiatives Librarian

Georgia Institute of Technology Library and Information Center

The National Science Foundation implemented its requirement that all grant proposals include a two page data management plan in January, 2011. At research institutions across the U.S., individuals from multiple, diverse departments realized that they would need to collaborate with other institutional groups to support this mandate. At Georgia Tech, the library has emerged as a focal point for these various campus groups. We were well positioned for this role due to a strong track record of developing effective, collegial relationships with other GT units. Two programs in particular supported us in this capacity: 1) prior outreach regarding data management (including an institute-wide research data assessment); and 2) a well-developed suite of scholarly communication services built upon SMARTech, our institutional repository. In this presentation, we will discuss institute-wide organizational issues surrounding the development of coordinated research data curation services.

In late 2010, the library's Research Data Project Team conducted a wide assessment of campus research data outputs based upon the *Data Asset Framework (DAF)*, an assessment tool developed by HATII at the University of Glasgow in conjunction with the Digital Curation Centre.[\[1\]](#) Our goals were to discover the types of data assets created and held by researchers; how the data were managed, stored, shared, and reused; and what researchers' attitudes were toward data creation, sharing, and preservation. As we tested and implemented the assessment survey, we discovered another, unintended outcome: promotion and outreach regarding research data curation issues.

As we gathered information from faculty about such matters as data format standards, data storage and back-up, and data management plans, we simultaneously drew their attention to the importance of these issues. In fact, during our survey pilot study, more than one of our faculty testers informed us that they were unaware of data management plans and many data management issues before reviewing the survey. Through the assessment pilot study and survey implementation, we started a conversation with researchers on campus about data curation. We additionally published articles about research data curation in a number of campus publications, attended campus workshops about grant funding, and taught classes on research data management.

The library has a history of building strong campus relationships around scholarly communication services. We have been running the university's DSpace-based institutional repository, SMARTech (Scholarly Materials and Research at Tech), since 2004. Currently, the repository has over 30,000 items. In the course of recruiting this material, we have formed productive relationships with a variety of campus constituencies. The department in charge of the repository, Scholarly Communication & Digital Services, now offers a suite of services for

faculty members that both generate content and facilitate the ingestion of content into the repository. These include:

- Lecture recording service. Lectures, symposia, and speaker series are recorded and the video files are uploaded to the repository.
- Journal support. We provide journal hosting software to encourage the creation of open access journals. The journal articles are uploaded to the repository.
- Conference support. In conjunction with Distance Learning and Professional Education, we provide conference hosting software that supports the peer review process of abstracts and papers. Conference proceedings are uploaded to the repository.
- Copyright research. We research copyright permissions for traditionally published articles and advise authors on how to retain copyright so that their material can be deposited into SMARTech.
- NIH Policy Access Policy support. In conjunction with the Office of Sponsored Programs, we answer questions about the NIH Public Access Policy, and deposit manuscripts to PubMed Central and SMARTech on behalf of authors.

SMARTech contains a variety of material from across campus. All theses and dissertations are archived in SMARTech, and we recently completed a retrospective scanning project for our bound theses and dissertations. SMARTech preserves the final project reports of grant-funded research which we receive from the Office of Sponsored Programs. Conference proceedings from the School of Aerospace Engineering, the School of Textile and Fiber Engineering, the College of Architecture, the Library & Information Center, the Space Systems Design Lab, and the Ivan Allen School of Public Policy are archived in SMARTech. As illustrated, we work closely with multiple schools, departments, research centers and interdisciplinary programs to provide access to and preserve their scholarly content; these services are used as a means of educating faculty members about author rights and open access issues.

Initial results from our research data assessment revealed that many faculty members are creating data sets that exist in accessible formats, are relatively small in size, and can be made publicly available for an indefinite amount of time. These criteria fit our collecting policies for SMARTech, and we are in the process of developing recommendations and policies for the data sets we are able to accept at this time.

The library's combination of research data outreach and existing repository services provided us with a foundation for working with diverse groups on campus. Despite prior promotional efforts, the NSF data management plan requirement confounded many researchers and professional support staff. Among other elements, the plan asks that researchers detail metadata standards and determine methods for data sharing and preservation. It became obvious from both the data assessment results and responses from faculty and research staff to the NSF requirement that coordinated data curation services were needed on campus.

We soon began working with colleagues from the Office of Sponsored Programs, Office of Information Technology, and IT directors from the individual colleges and research centers. Campus advocates for the library's assuming a central role in the development of research data

services quickly emerged, including an Associate Dean from the College of Engineering and the IT directors from a number of colleges. Faculty members with previous experience working with SMARTech and the Library's Scholarly Communication and Digital Services Department were also essential in sharing information about current repository services for published reports of research data.

We currently wrestle with the complication of working with a diverse group of individuals on a campus known for its disconnected nature, particularly in the areas of information technology and research support. Georgia Tech only recently developed the Office of the Executive Vice President for Research, and there is long history of the library's omission from larger campus conversations. For example, no information regarding the library was included in a campus-wide IT study conducted in 2010, and a recent survey implemented by the Executive Vice President for Research regarding research support on campus includes no questions regarding library services.

As we examine requirements for coordinated research data curation services, several questions have arisen across campus:

- Do all NSF-funded research data need to be shared with other researchers? With the world?
- Which raw data should be retained, and for what length of time?
- What is the difference between data retention and long-term access to data?
- What cost models can be developed for data curation services?
- For budgeting purposes, can data management costs be viewed as F&A costs?
- Are there existing services that can be leveraged on campus?
- What network and server infrastructures must be developed?
- What repository architectures must be modified or developed?
- What additional services and positions will need to be developed?

When we first began meeting with other groups on campus, our assumption was that the library would become the information hub for research data curation; in other words, we would collate and disseminate information about data and metadata standards, disciplinary repositories, and related topics. We also hoped to bring stakeholders together, and to provide information about various campus services, including those offered via SMARTech. What has emerged instead is a plan to recommend coordinated campus-wide data curation services under the umbrella of the library, and it is our campus partners who are proposing that the library manage these services. While still in the early stages of this partnership, we can discuss the organizational setting which brought us to this point, and our initial recommendations.

---

[1] Digital Curation Centre. "Data Asset Framework." Accessed February 28, 2011.  
<http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework>