

## **Project Description: Academic Libraries in the Digital Data Universe**

The rapid and dramatic rise of information technology and ubiquitous networking has transformed the research enterprise. Central to this transformation are digital data collections. The life-cycle management challenges associated with these intellectual assets are substantial. Issues concerning infrastructure development, sustainable economic models, new organizational capacities, and managing fast-paced technological change require focus and new approaches that reflect the complexity of the evolving scientific, technical, institutional, and economic digital data framework. Increasingly, academic libraries are engaging in projects and initiatives to preserve these digital assets. Given the sheer scope of the expanding digital data universe, it will be important to explore new organizational and economic models to ensure the long-term preservation of and access to these digital assets.

## **Background: Statement of Need**

Recognizing the growing importance of digital data in research and education, NSF's Cyberinfrastructure Vision for 21<sup>st</sup> Century Discovery<sup>1</sup> describes a vision for the future "in which science and engineering digital data are routinely deposited in well-documented form, are regularly and easily consulted and analyzed by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved." A 5-year goal to achieve this vision is "to catalyze the development of a system of science and engineering data collections that is open, extensible, and evolvable... a national digital data framework consist[ing] of a range of data collections and managing organizations... simultaneously local, regional, national, and global in nature." A two-day workshop exploring these issues with members of the research, library, and commercial sectors represents one of the initial steps in achieving that goal.

A key factor in the resilience and sustainability of a national digital data framework is its diversity. By including in the framework organizations from the international, government, non-profit, commercial, and academic sectors, operating under a diversity of business models with a wide range of funding sources, no one sector bears the entire burden and the system is resilient to economic changes. To be successful, each of these sectors must be fully engaged. While this workshop focuses on the academic sector, other activities are planned to engage the other sectors and draw them together.

## **Rationale**

The NSF has a history and tradition of working effectively with universities and the academic community. Universities have a critical role in our knowledge society.

*Ever since their inception, universities have been occupied with the fundamental elements of what we now call "knowledge management," i.e., the creation, collection, preservation, and dissemination of knowledge.*

A. Oosterlinck, Knowledge Management in Post-Secondary Education: Universities  
<http://www.oecd.org/dataoecd/46/21/2074921.pdf>

University faculty are quickly adapting to the digital environment in pursuing their mission for the creation and dissemination of knowledge, producing vast amounts of digital data in the

---

<sup>1</sup> [http://www.nsf.gov/od/oci/ci\\_v5.pdf](http://www.nsf.gov/od/oci/ci_v5.pdf)

process. However, the changing digital environment presents a challenge to universities and the academic community in their traditional mission of preserving knowledge. Academic libraries are at the center of this shifting landscape.

*The Web has changed publishing forever and...a similar transformation is likely to affect university libraries. The logical role for a university library in ten years will surely be to...host and curate (digitally) all the research papers produced by the university. There is also the interesting issue of how much responsibility individual universities would undertake for hosting and curating the scientific data produced by their researchers.*

A. Hey and A. Trefethen, "The Data Deluge: An e-Science Perspective." in *Grid Computing—Making the Global Infrastructure a Reality*, F. Berman, G. Fox, A. Hey, eds., Wiley Press

While this quotation makes a commonly held distinction between research papers and scientific data, the future may see these become linked pieces in a comprehensive digital information environment. This is because the data underlying a publication can be as important as a starting point for work by others as is the publication itself. In an important sense, a publication is an element of metadata, providing a means for finding and understanding the underlying data—so long as publication and data can be linked. Thus, solutions for preserving digital publications must not be undertaken in isolation from those for preserving digital scientific data rather than the array of assets produced in the research process that need to be addressed with attention to the relationship between components.

### **Workshop Objective**

Academic and research libraries operate in close proximity to researchers and educators who use and produce digital data. While they have had their traditional role in the realm of scientific publications, they are equally important stakeholders in planning for a national digital data framework.

*It is to the research library community that others will look for the preservation of...digital assets, as they have looked to us in the past for reliable, long-term access to the "traditional" resources and products of research and scholarship.*

Association of Research Libraries (ARL) Strategic Plan 2005–2009

The objective of this workshop is to engage this critical stakeholder community in the planning process.

Questions to be addressed at the workshop include the following:

- What role do the academic libraries envision for themselves in a digital data framework that provides for preservation of digital publications, digital data, and the links between them?
- What partnerships/coalitions among academic libraries and with other sectors (government, international, non-profit, and for-profit) could facilitate the creation of such a framework?
- Are there opportunities to test sustainable models for digital data preservation organizations, including consortia and partnerships?
- What resources would be required to enable such tests?

One outcome of the workshop will be a published report addressing these questions.

**Workshop Implementation**

The workshop, planned for July 27–28, 2006, will be held at NSF to allow for the participation of NSF staff and representatives from other federal agencies with interests in digital scientific data. A total of approximately 30 to 35 external participants is anticipated for the workshop. Participants will be diverse and represent a broad spectrum of academic institutions. Representatives from the international, commercial, and non-profit sectors will also be included.