New Collaborative Relationships:

The Role of Academic Libraries in the Digital Data Universe

Position Paper from Johns Hopkins University

(Sayeed Choudhury, Robert Hanisch, Alex Szalay)

The National Science Foundation's Cyberinfrastructure Vision for 21st Century Discovery describes a broad-based effort that is transforming the manner in which scientists, social scientists and engineers (and perhaps even the humanists) conduct research, teach and learn, and disseminate their research findings and publications. Projects such as the Virtual Observatory (VO) provide ample evidence of data-driven scholarship, which offers both challenges and opportunities for academic research libraries, especially in the realm of data curation. Even with ambitious new efforts to create large corpora of digitized text such as Google Book Search or the Open Content Alliance, libraries still represent a core element of the preservation picture.

Given the scale and complexity of even a single cyberinfrastructure-based project such as the VO, it is not reasonable to assume that a single library or organization can manage the entire range of data curation needs. Rather, libraries must find ways to work together with an array of organizations such as other libraries, supercomputing centers, museums, archives, publishers, and corporations. For different stages and applications of data, various organizations will need to identify appropriate roles and develop systems that interface—technically and organizationally—with a range of partner institutions. Such a complex array of relationships and technological infrastructure may benefit from examination and leadership from the highest levels of the academic and corporate community. However, at this stage, there are major research and development questions that remain unaddressed. In this current environment, it's essential to develop prototype systems that demonstrate both technical and organizational infrastructure to support data curation. These prototype development efforts will help us better understand appropriate technologies, potential costs, and organizational relationships that will be necessary to support cyberinfrastructure-based projects and programs.

At Johns Hopkins University, we are working with a network of libraries, publishers, scholarly societies, and corporate partners to develop a repository-based system that will support an end-to-end process for capturing, curating, preserving, and providing access for the long term to derived data that is cited in electronic publications. We are prototyping such a process and system. Our goals include assessing the scientific impact of this new approach to astronomical data, as well as working out sustainable business models for increasing the value of data in this way. Our prototype phase focuses on astronomy because of the technological maturity of electronic publications and data management in this discipline, and because of the wide access to digital data archives, and the unique, established relationship between the astronomers and libraries at Johns Hopkins University.