

New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe

Comments by Frank Rack – formerly Director, Ocean Drilling Programs, Joint Oceanographic Institutions; currently, Executive Director, ANDRILL (Antarctic geological Drilling) Program

My views on the topic of the role of academic libraries in the preservation of digital data collections are primarily derived from a geosciences perspective, based on experiences from participating in and directing scientific drilling programs. These programs, which have enjoyed a long history of sustained support from NSF and international partners, have produced a substantial volume of physical samples (e.g., over 330 kilometers of sediment and rock cores) and data (both analog and digital) that are organized into structured collections using relational data base management systems as well as archives of unstructured data. These data holdings are expanding rapidly with the increasing acquisition of visual imagery and specialty data sets (e.g., volumetric imaging using X-ray CT and NMR/MRI) that are common to biomedical domains. Scientific drilling projects and programs cross NSF organizational boundaries spanning Earth, Ocean, and Polar Programs. These activities are now expanding to include the collection of observational time series across time and space with requirements for data streaming and remote user control of sensors and other resources. New models of sharing, storing, analyzing, and archiving data are required. Academic libraries have an important role to play in this new data universe.

Academic libraries as collaboration centers for research, education and public outreach: Academic libraries are naturally the center of the campus knowledge management and information exchange process and can be enhanced with physical resource investments to become “amplified collaboration environments” serving as a focal point for cyberinfrastructure access on academic campuses. The past NSF investment in connecting academic campuses to the high-end research networks, such as Internet-2 and National Lambda Rail can be enhanced with campus-wide dark fiber networks that support broadband connectivity to campus buildings and facilities, like libraries, that may be remote from the IT-2 or NLR node, but are centers of learning. These dark fiber networks are owned and operated by the academic IT infrastructure and provide capabilities to link researchers, educators, and students in a distributed collaborative environment dedicated to knowledge management and information sharing supported by data discovery, analysis, and visualization tools that can be accessed through the campus library system. The University of California at San Diego has already taken this step and is currently operating a dark fiber network on their campus.

These nodes could provide connectivity to the ACCESS GRID for collaborations with virtual research groups across campus or at other institutions and provide links with the DATA GRID for accessing computing resources and visualizations that are either pushed to the site or pulled from the site for research, educational or outreach activities. Investing in library infrastructure (hardware and software plus skilled staff) creates synergies with both local and remote research groups and encourages partnerships among researchers,

academic staff and students through opportunities for training and participation in focused demonstrations of the research outcomes using visualizations that can also be shared with the broader academic community and the general public. Academic libraries should be encouraged to establish collaborative relationships with local research groups and existing centers of excellence who provide digital content in exchange for data aggregation, public access, archiving and preservation services (either locally or remotely through networked data centers that provide tools and web services that can be accessed easily by trained users), including publication of data.

Academic libraries as partners in preservation of analog reference collections and legacy data: Academic libraries are uniquely positioned to play an important role in the preservation of analog reference collections and legacy data for projects that are aligned with the research and education mission of a particular academic institution or the mission of local centers of excellence on each campus. The network of academic libraries should coordinate with each other to collaborate on content preservation efforts that build on their strengths while minimizing duplication of effort on a national (or better yet, and international) scale. Domain specific groups of researchers, working as community agents, could work with designated academic libraries to digitize analog collections for long-term preservation with appropriate metadata. This type of partnership would combine the traditional strengths of library professionals with the opportunities provided by collocated teams of researchers in a coordinated way to support a large-scale, interoperable, networked architecture for data discovery, information sharing and knowledge creation. The academic libraries would play a fundamental role in supporting research and educational goals within the context of a pervasive cyberinfrastructure that would require partnerships between federal, state and local (academic institution and local community) for investments that leverage technology to provide access to knowledge resources and training and outreach to stakeholders at all levels.

An example from scientific ocean drilling is the need to transform 40 years of analog “Proceedings” of the Deep Sea Drilling Project (DSDP) and Ocean Drilling Program (ODP) into electronic format through document scanning and OCR. These volumes weigh approximately 2000 pounds as a set and contain both data and metadata that could serve a wide community of users if they were readily available across the network. Plans to undertake this scanning/digitization project has been made and pilot studies have begun to transform this pile of paper into digital content through a partnership between the Texas A&M University Digital Library and the ODP Science Services group located at Texas A&M. Similar collections of key reference materials should be identified by specific domain science communities and prioritized for digital access in partnership with academic libraries.

The evolution of a distributed, networked, partnership among academic libraries, technology centers and research groups would require strategic planning and phased investments that leverage existing programs and initiatives to create new opportunities and enhance synergies among all parties. The prior NSF investments in establishing point-source academic infrastructures (e.g., Internet-2 and NLR connections/nodes) should be leveraged by establishing a process to encourage the construction of dark fiber

campus-wide networks connecting libraries to research centers to support data sharing, open access, and preservation/archiving of data that can be provided to researchers, educators and communities of learners. The infrastructure investments should be combined with opportunities for the development of web service architectures to support data discovery, analysis and visualization to create a transformational environment of innovation that would enhance knowledge creation and dissemination and stimulate learning.

Observational data in the future will be streaming from thousands to millions of field sensors that will require scalable visualization resources to allow humans to readily understand and comprehend the significance of these data. Academic libraries have a unique opportunity to establish a strategic role for themselves as centers for data integration, analysis and visualization, combined with a strong education and outreach mission. In order to realize this dream, academic libraries will have to form innovative partnership with a variety of research groups and broad-based communities of educators and technologists who can translate these challenges into coordinated action plans that capitalize on the opportunities promised by this transformation. Academic libraries will become next generation centers for learning, education, and public outreach, and will need to provide training to a broad range of users to articulate the significance of the new world view that accompanies this change in the data universe.