

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

Comments by Dr. Robert S. Chen, Interim Director, CIESIN at Columbia University; Manager of the NASA Socioeconomic Data and Applications Center (SEDAC); and Secretary General, Committee on Data for Science and Technology (CODATA) of the International Council for Science (ICSU)

Top three issues in long-term preservation, management, and curation of digital scientific data and information:

1) *Digital rights management.* Clarification of rights to archive, use, and disseminate data and any applicable restrictions is essential to long-term data curation, and would be greatly facilitated by digital standards and technologies that permit data sources and users to quickly and easily specify and understand rights and restrictions in ways that meet their needs and concerns. This has been a key element in the success of the Creative Commons and needs active support in the realm of scientific data and information. Key issues that need to be addressed include protection of confidentiality, limitations on liability of data sources, use of data for humanitarian purposes, and definitions of appropriate uses (e.g., private sector vs. public sector research).

2) *New institutional partnerships.* A range of new partnerships is needed across disciplines, within and between universities, between sponsors and data managers, across the public and private sectors, and with the broader scientific community to establish appropriate and sustainable long-term data management structures covering all or most of science. Is a mix of disciplinary, cross-disciplinary, and institutional repositories going to evolve that can provide sustainable data curation and management in most fields? Can we identify gaps and find ways to fill them? Are there contingency plans when a particular institutional arrangement for a particular field encounters problems with sustainability? Are there ways to involve the private sector and/or the open source community in these arrangements to help infuse interoperable technologies and reduce costs without risking long-term sustainability or access? Should existing consortia, e.g., of universities, of libraries, or disciplinary data centers, be asked to take on long-term curation responsibilities or are new ones needed?

3) *Science education and community outreach.* Many scientists continue to use traditional approaches to data, i.e., developing custom datasets for their own use with little attention to long-term reuse, dissemination, and curation. Although there has been considerable progress in data stewardship for “big science” projects, even modest collaborative projects are inconsistent in their attention to data management and few individual scientists think beyond posting selected results and data on the Internet or submitting a final data product to a data archive if required to do so. Changing this sort of behavior will require a range of efforts, including investment in approaches to make data documentation, sharing, and preservation easier, establishment of an infrastructure to accept and assume responsibility for data (e.g., a local university depository or a disciplinary data center), and, perhaps most important of all, concerted efforts to educate current and future scientists to adopt better practices.