

“Issues in Digital Data Curation Leading Us to the Need for New Partnerships”
Tyler O. Walters, Georgia Institute of Technology Library and Information Center

Among the top three issues effecting the growth and future of digital data curation involves: 1) Culture and policy frameworks; 2) Technology integration and data curation tools; and 3) Economic sustainability of our partnership and consortial models.

The expense and enormity of the challenge to ensure the preservation, curation, and overall management of scientific digital data lends itself to a multi-institutional solution. However, in order to build effective partnerships, we must first begin with the single organization. Raising awareness of the need and benefits to managing digital data will be paramount. Universities will need to create programs and incentives that **embed digital data curation in university culture**, becoming an integral part of research projects. The initial **change driver can be policy**. Granting agencies such as NSF and NIH (DOE could be another source for a data curation policy) are considering data access and management policies that may drive universities with agency-funded research projects in this direction. In response, universities will need to make a commitment to the use, re-use, and maintenance of digital data. However, pragmatics also will dictate that not all data can be sustained. This situation requires designing criteria for selecting which sets of digital data will and will not become a long-term university responsibility. Libraries can contribute to the selection process by adapting archival appraisal theory as well as other parameters to judge which research materials are worthy of long-term accessibility. This combination of policy setting, awareness raising, cultural engineering, and selection criteria building, will become essential components in the rise of digital data curation programs. Libraries can assist with developing and implementing this agenda and be equal partners with scientists, advanced technologists, and policy makers. New partnerships and services akin to Genbank, operated by the NIH / NLM's National Center for Biotechnology Information, may need to be established. They may take on a discipline-based alignment, such as several universities with deep interests in astrophysics forming a consortium to provide data curation services for their home institutions' related data sets.

Much work remains in areas such as **improving technology integration and building reliable data curation tools**. Data does not reside in just one information system; therefore, integration between systems is critical. Universities have research data residing in many applications such as databases (commercial and open), digital asset management systems, content management systems, and repositories. Curating this data consequently may include moving it from one system to another, linking it between systems, and migrating it to a central system. The exact future architecture is undetermined and many universities and consortia may take divergent paths. There will be common format and metadata portability issues. Ongoing development projects such as the Global Digital Format Registry (GDFR) can play an essential role in matching formats (including database programs and protocols) and providing information on software that can read certain databases and data formats, and recommend migration paths. Data curators will need tools for data and metadata extraction, database emulation, data provenance tracking, and to document the origin, use, and re-use of data. Partnerships and consortia can play a key role. They can promote the further development of new data curation technologies as well as standards and technical protocols to ensure interoperability and data migration. They can be vital to maximizing present resources and strategies in generating these new technologies through their synergistic activities.

If the work described above progresses, then the **economic sustainability of these types of partnerships and consortia** will become mandatory. The bottom line – funding and revenue streams need to be established. A mixture of funding sources will best guarantee the success of these new entities and help them to not become too reliant on any one source of funds. Partner dues, seed and project monies from grant agencies and private foundations, revenues from a variety of service and consulting fees, and several other creatively produced sources of funding are examples of cooperative ways to sustain the new partnerships. Those interested and vested in digital data curation should explore deeply new and dynamic models of organizational and economic sustainability. This is an opportunity to reinvent ourselves for the better as we face inherently new challenges in managing complex research objects such as data sets.