

Todd Vision, The National Evolutionary Synthesis Center

Position Paper, *The Role of Academic Libraries in the Digital Data Universe*, Sept 26-27 2006, Arlington VA

Three major issues in the preservation, management and curation of digital scientific data:

- 1. The respective roles of scientific publishers and institutional libraries.** Currently, the responsibility for management of data accompanying scientific publications falls to publishers in the form of supplemental data collections. Academic journals have little incentive to invest in the establishment and maintenance of digital data repositories that can be used for anything beyond minimal documentation of published reports. If there were a uniform and reliable system for digital data management that was hosted by researcher's home institution, this could potentially supplant the current system whereby unstructured supplemental data is deposited at publication. However, such a system would need to be available widely, and not just at elite institutions, in order to be a viable alternative for publishers.
- 2. The untapped value of raw data to the researcher.** In many scientific fields, difficult-to-obtain and essentially irreproducible datasets (such as long-term field observations in ecology) are analyzed and reanalyzed by the same research group over a period of many years, resulting in multiple publications, each containing only such statistical summaries of the data as are necessary to support the claims in the publication. Such datasets may continue to grow and accrete value over time. Understandably, the researcher feels entitled to an indefinite term of exclusive use, and there may be no single moment at which he or she would be comfortable providing even limited access to the full dataset. How common is this situation, and how worthwhile is it to invest in a system of scientific data preservation that would exclude such unique and valuable data collections? Or is there a way to manage such data while protecting, or alleviating, the researcher's concerns of exclusivity.
- 3. The burden of metadata curation.** Digital data is nearly useless without extensive and high-quality metadata, both for resource discovery and for interpretability of the data itself. The researcher knows the data, but doesn't necessarily have sufficient expertise in information science to provide quality metadata curation, while librarians have the opposite problem. What incentives can be provided to researchers to undertake the burden of careful metadata curation, how can this task be made more manageable to nonexperts, and what incentive can be provided to institutions for QC of metadata produced by researchers?