

Supporting Biological Information Work: Research and Education for Digital Resources and Long-lived Data

Carole L. Palmer, Melissa H. Cragin & P. Bryan Heidorn
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
{clpalmer, cragin, pheidorn @uiuc.edu}

ABSTRACT

New practices are emerging in all stages of biological research, from data collection through dissemination of results. Through a series of cooperative projects with biologists working in data-intensive and informatics-based domains, we have documented requirements for digital libraries, tool development, and data management techniques to support contemporary scientific practice. This research is now serving as the foundation for a new biological informatics master's program to train scientific information specialists to manage and integrate scientific information and tools to support scientific problem solving and communication.

Categories and Subject Descriptors

K.3.2 [Computers and Education]: Computer and Information Science Education – *Information Systems Education*.

General Terms

Management, Design, Standardization

Keywords

Digital libraries, cyberinfrastructure, long-lived data, standards

1. BIOLOGICAL INFORMATION WORK

A number of recent reports on cyberinfrastructure and eScience initiatives indicate a shortage in qualified professionals to manage the increasing stores of scientific data [1]. Information and communication technologies are impacting all modes of scientific inquiry, from the management of big science to the conduct of daily bench work. Shared digital resources, including digital libraries, data mining and visualization tools, and data repositories, are being developed across the biological domains and there is a growing awareness of the need to work toward the integration of data across biological scales.

For several years our research teams have been building a concentration in information technology and digital library development for domains in the biological sciences. Through a series of joint projects with biologists on collaborative technology development and information problems in scientific discovery, we have studied the particulars of the trends in data and information use in biology [2]. At most research sites, information tasks and systems development are either performed by biological scientists

who are often self-taught or by computer scientists who have limited training in the biological domain. Across projects, we have continually observed and documented cases where the expertise of scientific information professionals could have supported and advanced the conduct of research.

Additionally, libraries and museums are increasingly being called upon to become the long-term curators and stewards of the range of intellectual products and research data. The contributions made by specialists in scientific information and communication will be most effective if they are realized across the laboratories, academic departments and research centers, museums, and libraries that together provide the infrastructure for science.

2. SCIENCE INFORMATION EXPERTISE

Areas of information support that have emerged as priorities to complement the expertise of biological and computer scientists include digital library and repository development, data curation and preservation, ontology and standards development for interoperable systems, and literature-based discovery. To respond to the qualitative changes in biological research and the specific workforce gaps identified in our research, we are developing a comprehensive master's level training program in scientific communication, as part of a campus-wide bioinformatics initiative at the University of Illinois at Urbana-Champaign. The program will train a new generation of Library and Information Science professionals to serve in scientific research environments.

3. ACKNOWLEDGEMENTS

This project has been supported by Information and Discovery in Neuroscience, NSF-0222848 (CL Palmer, PI), and A Graduate Program for Scientific Communication Specialists: Getting Past the Prototype in Biological Informatics, NSF-IIS-0534567 (C.L. Palmer and P.B. Heidorn co-PIs).

4. REFERENCES

- [1] See for example, National Science Board (September 2005): NSB-05-40, Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century. <http://www.nsf.gov/pubs/2005/nsb0540/>
- [2] Projects include Carole Palmer's Information and Discovery in Neuroscience (NSF IIS-0222848), Bryan Heidorn's Georeferencing Museum Specimen Sources (Moore 2005-2929-00), and Heidorn and Palmer's, Internet Environment for BioDiversity Survey Collaboration and Verification (NSF BDI-011391).