

A Graduate Program for Biological Information Specialists¹

Bryan Heidorn, Carole Palmer, and Dan Wright
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

UIUC GSLIS Technical Report UIUCLIS--2006/2+SCI

The nature of science communication is changing to include new forms of information collection, management, use, and dissemination, and biologists are generating new information at a staggering rate. Methods of information management, such as globally federated data sets in ecology and neuroscience not only alter the quantity of information that is available but also cause a qualitative shift in the nature of the scientific questions that can be asked and answered. For example, when information on the historic global distribution of species is combined with historic climatic data we can begin to answer questions about the future distribution of species under the pressure of climate change. This data transformation however is not free. It requires the collaborative effort of scientists with information specialists who both understand some aspects of the science and who are skilled in the associated computational and information management tasks. This new breed of information specialist must know how to collect, organize, access, use, and preserve information, freeing biological scientists to focus on biology.

Currently, information management tasks are performed either by biological scientists who are self taught in information management or by computer scientists with great computational training but limited understanding of the problem domain. Biological Information Specialists (BISs) will play an important role in improving information transfer and collaboration in science. They will allow biological scientists to concentrate on scientific problems and computer scientists to let go of projects when they move into implementation stages.

The project will build on ongoing local and national collaborations to support curriculum development and internships. Students, teachers and researchers working together will be able to develop new tools and methods for research and education, and these relationships will fuel new collaborative research opportunities. By working with preeminent partner institutions in biological informatics, we will be able to bring together and publish the best practices that have developed independently in the biology research projects around the country. The resulting curriculum will be published and made available on the web. Most importantly, the students will graduate and spread their training to research groups around the country.

Educational Aims

Why do we see a need for a masters degree program focused on scientific communication and rooted in Library and Information Science (LIS)? Simply stated, scientists should be able to spend time conducting science. Of course, doing science often involves information technology development, and, as Kling & McKim (2000) demonstrated, information systems should grow out of the needs and cultures of research communities. Information technology should not be based on vague notions that conflate distinct activities and interests of different research domains. However, there are many parts of information technology development and sustainability, as well as other information-based activities in the daily practice of science, that could benefit greatly if supported by specialists devoted primarily to the information, in service to the science.

In the future of informatics development, BISs will complement, not duplicate, the expertise of

¹ Text excerpted from the proposal for NSF award number 053456, A Graduate Program for Scientific Communication Specialists: Getting Past the Prototype in Biological Informatics

biological and computational scientists. They would serve as supportive team members to keep information technology development moving “past the prototype”, and they would continue to play their traditional research librarian roles in advancing research and scholarly communication. There is an extensive and ever growing universe of information resources, informatics tools, and scholarly communication options that need to be understood, assessed, and coordinated locally so that there can be more rational and equitable global development in terms of integration of data, literature, and information technologies. BISs will bridge these arenas of informatics development.

Bioinformatics programs are being developed across the country; however they do not cover in a comprehensive way the range of information issues that have emerged in the past decade. A few examples include data exchange standards, digital preservation, and electronic publishing. Moreover, they are not focused on the larger family of information problems that cut across all the sciences. While we understand the great need for discipline-specific specializations, it is important to recognize that the scientific enterprise will be more effective with well-trained professionals working to systematically advance information use at local and global levels.

Goals and objectives

Our primary goal is to design a program of graduate study that can serve as a model for training specialists in scientific communication for biological informatics. A secondary goal is to integrate this graduate training with ongoing bioinformatics research and practice to produce specialists that understand the research culture and can make substantive contributions to scientific discovery. The objectives are to:

- 1) develop curriculum for the specialization that builds logically on existing graduate programs at GSLIS and the core provided by the new campus-level bioinformatics degree;
- 2) establish internships at institutions and laboratories where students can develop and apply their growing expertise;
- 3) develop mechanisms for integrating course work into current informatics research at GSLIS, other UIUC departments, and partner institutions;
- 4) share the educational approach to biological informatics with other schools interested in developing similar specializations;
- 5) expand understanding of the role of informatics in scientific progress.

Broader impacts

We expect BISs to have important local impacts on the scientific enterprise while also contributing to global development and integration of information systems for scientific research and communication. Their levels of responsibility will vary depending on their placement, in a particular laboratory, research institute, academic department, or unit of a research library. However, their influences may be wide-reaching.

Local impacts will include: 1) Progress toward what we call the “getting past the prototype” problem. BISs will attend to implementation, evaluation, continual improvement, and sustainability of information and data systems. That is, they will make sure systems are responsive to the real needs of scientists and therefore assure greater adoption of technologies by target communities. 2) BISs will be prepared to better exploit existing data standards and work toward “long-lived data” and interoperability across the various scientific communities. They will be instrumental in building and integrating the increasing number of ontologies, taxonomies, digital libraries, indexing systems, and vocabularies associated with digital data and products.

Global impacts will include: More equitable access to publishing forums by scientists in general. Many of the larger, resource-rich research institutions are moving to an open access institutional repository model for disseminating research. Scientists at smaller institutions are likely to fall behind in this trend or be under-represented in such repositories. BISs can play an important role

on many campuses by developing and managing these new communication forums and assisting scientists in disseminating and preserving their intellectual property. They will stay up to date on and participate in data sharing and federation activities within and across scientific disciplines where preservation and archiving of data and results now requires ongoing communication and collaboration.

Background

Bioinformatics and biological informatics

What is bioinformatics? While there are many ways to characterize bioinformatics, we interpret the field broadly to include the range of biological sciences and a broad conceptualization of information. While bioinformatics is frequently associated with data mining and molecular modeling, all the biological sciences are moving forward with computational approaches, and they all are facing increasing problems related to finding, mobilizing, preserving, standardizing, sharing, and managing information and data. According to the NIH Biomedical Information Science and Technology Initiative (BISTI) documentation, bioinformatics is:

Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

This definition encompasses the range of biological sciences covered in our current science-oriented research projects in our college and the information problems that are at the core of our existing curriculum. It captures aspects of information production and use that are important in the entire cycle of scientific research and communication, from data collection to dissemination of research results.

The scope of this definition also includes biodiversity or biological diversity, a field often overlooked in discussions of bioinformatics. Biological diversity means “the variability among living organisms from all sources, including inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems” (Convention on Biological Diversity, Art. 2, para. 1). Informatics is as vital to biodiversity biologists as it is to molecular biologists.

Scope and limitations of current bioinformatics programs

In reviewing the current bioinformatics programs nationally, including the other programs being developed on our own campus, it is clear that the focus is not on service to science in the broad sense, or on the kinds of situations described above. But, these examples represent real concerns for the ongoing development and management of scientific research, since each year NSF and NIH invest heavily in these kinds of informatics projects with the aim of improving the practice of research for scientists at large. But, it appears that in many cases the scientists most involved in informatics development benefit while the typical scientist is often unaware of new developments, or at least unconvinced that there is anything for them in the informatics movement.

Based on the recent “Survey of Bioinformatics Programs in the United States,” by Hemminger, Losi and Bauers (2005) and a review of websites of existing programs, it is evident that almost all of the existing programs are heavily focused on computation and computer science. A few address information organization in terms of formal representation but do not consider problems such as metadata, thesauri construction, or indexing, for example. Several schools offer human-computer interaction types of courses, but only a couple offer courses that address user needs, scientific communication, or human information seeking behavior.

One program that has some parallels with our BIS model is the Certificate of Specialization in

Bioinformatics offered through SILS at UNC-Chapel Hill. They offer a slate of more traditional LIS courses, such as information retrieval, organization of information, and management of information agencies, to supplement biostatistics, biology requirements, and another group of LIS based electives. However, their program is an add-on to a masters in either library science or information science. Our program builds on the strengths of our existing masters option but is part of a separate campus-level bioinformatics masters degree program, requiring no additional certificate work.

The other program of note is the Master of Biomedical Informatics program at the Oregon Health Sciences University, an innovator in medical informatics education. While its focus is on healthcare and biomedical settings, and therefore the curriculum is more specialized than what we expect to design, the program is comparable in its main objectives: 1) To provide students with a theoretical and practical understanding of the role of information [in biomedical settings], 2) To provide students with a sound basis for implementing, developing, maintaining, and managing information resources and systems [in health care], 3) To provide students skills in the management of [biomedical] information, technology, and decision making.

Scientific communication focus

Our graduate program would train a different kind of professional, focusing on developing expertise to fill the roles we see lacking in bioinformatics research teams and in the coordination of information systems development in the biological sciences. The overarching focus of the program will be on scholarly, or more specifically, scientific, communication. In the discourse of higher education, discussions of scholarly communication often cover only the new ways in which research papers are being published and disseminated. And, there is no doubt that electronic publishing on the Internet is quickly changing the format and means of distribution of scholarly works. However, there are many other dimensions of the scholarly communication cycle that are also changing in important ways in terms of how information is identified and mobilized for research purposes. In this sense, we conceive of scholarly communication as the entire set of information activities involved in scholarly exchange, including how information is found, integrated, and disseminated to produce new research results and scholarly products (Palmer, in press). Therefore, under the main rubric of scientific scholarly communication, this program will concentrate on training professionals to support science by building expertise in three areas:

- 1) Evaluation and implementation of information systems
--user based assessment and continual quality improvement for the development of tools that work and are used.
- 2) Information acquisition, management, and dissemination.
--development of digital libraries, data archives, institutional repositories, and related tools.
- 3) Information organization and integration
--structuring information for optimal use and sharing, and standards development.

BISs will develop functional applications that are integrated with current science practice. They will also attend to the more global concerns such as standards development, data and literature federation, and equitable dissemination of research results.

References

Convention on Biological Diversity. (1992). *Convention on Biological Diversity*. Retrieved April 25, 2005, from <http://www.biodiv.org/doc/legal/cbd-en.pdf>

Hemminger, B.M., Losi, T., and Bauers., A. (2005). Survey of bioinformatics programs in the United States. *Journal of the American Society for Information Science and Technology*, 30(1), 12-13.

Kling, R., and McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science and Technology*, 51(14), 1306-1320.