

Bioinformatics education in an MLIS program: the McGill experience

Joan C. Bartlett

Abstract: Program objective – The objective of this course (GLIS691 – Bioinformatics) was to provide formal bioinformatics education within a master of library and information studies (MLIS) program. As bioinformatics becomes increasingly integral to biomedical research, there is a need for librarians to expand their practice into the domain of bioinformatics, supporting the efficient and accurate use of these complex resources. We developed this course, the first such course offered in a Canadian library school, in response to the demand for librarians to be able to support bioinformatics information needs. Setting – The course was offered in the winter term of 2005 in the Graduate School of Library and Information Studies, McGill University. Participants – Course participants were MLIS students. Program – The course took a library and information science perspective to bioinformatics. The goal was to provide students with the skills and knowledge to provide information services in the domain of bioinformatics and to collaborate in the design and development of bioinformatics resources. This included understanding the field of bioinformatics and the range of resources, the needs and requirements of user groups, practical searching skills, the creation of resources, and the role of the librarian. Conclusions – This course represents one approach to providing formal bioinformatics education for librarians. Librarians who are knowledgeable and proficient in bioinformatics will be able to expand the role of the library into this domain; apply their knowledge, skills, and expertise in a complex, chaotic information environment; and develop the essential role of the librarian in the domain of bioinformatics.

Introduction

Bioinformatics is a complex, dynamic, and emerging discipline that has a significant impact on biomedical research. It has been defined as “the computer-assisted data management discipline that helps us gather, analyze, and represent [biological] information” [1], and can be seen to have three main objectives: (1) to organize data in such a way that researchers can access information, (2) to develop tools and resources to aid and support data analysis, and (3) to apply the tools to analyzing and interpreting the data in a biologically meaningful way [2].

Bioinformatics revolves around primary, biological data, such as genetic or protein sequence information. The resources fall into two broad categories: (i) databases of primary biological data (e.g., GenBank) and (ii) software tools that manipulate and analyze the data (e.g., BLAST). Frequently, the two functions are integrated in a single resource. Bioinformatics resources generally don't take the form of text-based, bibliographic information and, as such, represent a nontraditional type of information for library and information studies (LIS).

According to the 2005 edition of the *Nucleic Acids Research* annual database issue, there were over 700 individual

bioinformatics resources [3]. This only includes those that are publicly accessible; therefore, the actual number is higher. These resources tend to be complex, dynamic, and non-standardized. For a scientist, navigating this range of resources is a challenge, particularly for laboratory scientists, for whom bioinformatics analysis is a valuable tool but who don't use bioinformatics resources on a regular basis. The challenge is not only to know how to use a resource (in an environment in which multiple resources may be needed to solve a single problem), but also to be aware of the types of resources available and the types of questions and problems they can address.

While strongly linked to biology and computer science, bioinformatics is, at its heart, an information-based discipline, involving many of the same aspects of the information life cycle that have long been the domain of library and information science. As such, librarians and information professionals have a clear role to play in the domain of bioinformatics. Areas of contribution include providing reference and information services in the same way that users of other information resources are supported, supporting and providing bioinformatics education for scientists, and bringing library and information science principles (such as the use of controlled vocabularies) to the design and development of bioinformatics resources. There is currently a demand for librarians to be able to support bioinformatics information needs. However, in Canada, the opportunities for bioinformatics education within library and information science are rather limited. Currently,

J.C. Bartlett. Graduate School of Library and Information Studies, McGill University, 3459 McTavish Street, Montréal, QC H3A 1Y1, Canada (e-mail: joan.bartlett@mcgill.ca).

none are offered at the master of library and information studies (MLIS) level. It is in this environment that McGill University's Graduate School of Library and Information Studies (GSLIS) launched a course in bioinformatics for MLIS students.

Goals and objectives

The GSLIS course in bioinformatics was first offered in the winter term of 2005 as a Special Topics course. The course was envisioned to parallel existing, advanced reference courses in other disciplines and was developed with the understanding that students would apply the skills learned in the course to work in a health sciences or scientific library environment. As such, there were two main motivations. The first was to provide future librarians with the skills and knowledge to provide reference and information services in the domain of bioinformatics. The second was to complement existing advanced courses in Health Sciences Information and Science and Technology Information so that students would have an additional set of skills and knowledge at their disposal when working in either a health sciences or science and technology environment.

The course objectives were the following:

- (1) To understand the basic biology and genetics that underpin bioinformatics
- (2) To understand the content and use of a range of bioinformatics resources
- (3) To understand the role of library and information science, and other disciplines (e.g., computer science, biology) in bioinformatics
- (4) To be prepared to provide reference services and support to bioinformatics information users.

Course content

Within the framework of these objectives, the course content covered a variety of topics. The first was Genetics 101. Given that students were neither required nor expected to have a biology or genetics background, it was essential to cover enough of the fundamental biology that underpins bioinformatics so that the remainder of the material covered in the course would make sense and fit within the appropriate scientific context. The material included was that which would be covered at either the senior secondary school or early (first or second year) undergraduate level. The material was presented from the perspective of the storage, flow, and manipulation of genetic information as it moves through the biological system.

The lecture started at the level of the cell and then worked through the genetic system discussing chromosomes, genes, DNA, RNA, and protein. It also covered the processes of transcription and translation. During an in-class exercise, the students worked with a DNA sequence and followed the genetic code to translate it to the corresponding amino acid sequence and to identify the open-reading frames.

A class on the information behaviour (including information needs, information seeking, and information use) of scientists, particularly with respect to bioinformatics resources, also provided a foundation for the course. An understanding

of the user group is as important as an understanding of the resources.

Discussion of a range of bioinformatics resources began with an examination of the content and structure of the data, specifically a DNA (gene) sequence record from the GenBank database and an amino acid (protein) sequence record from Swiss-Prot. This allowed the students to become familiar with the nontextual data contained in the resources and the particular issues and challenges (e.g., keeping track of different versions of gene sequence) of managing such data.

Within the confines of a 12-week course, it was clearly not possible to discuss the hundreds of bioinformatics resources currently available or to even discuss a representative sample of each type of resource. The range and diversity is too broad. Instead, the focus was on two of the major sources for bioinformatics resources that are among the most widely used by scientists. One was the National Centre for Biotechnology Information (NCBI), which is part of the US National Library of Medicine (NLM) and is often the first choice for bioinformatics information among North American scientists. The other key site discussed was the European Bioinformatics Institute (EBI), which has a similarly prominent place for European researchers. All three classes on resources took place in a computer lab, with a mixture of lecture, demonstration, and hands-on practice.

NCBI hosts a vast range of resources, including the GenBank database, one of the three major DNA sequence repositories, and BLAST, search software that finds sequences based on similarity. Through the Entrez interface, many of the different resources are linked, making it simple to navigate among related records. For instance, from a gene sequence record, one can link to the corresponding protein sequence record, the Online Mendelian Inheritance in Man (OMIM) record discussing the clinical implications of the gene, and to the PubMed records of articles discussing the gene. One lecture was spent exploring the range of resources available through NCBI and using the Entrez system to link and navigate amongst these resources.

A second class was devoted to BLAST searching. If a scientist only knows about one bioinformatics resource, it is usually BLAST. The lecture covered the various types of BLAST, how to run a BLAST search and interpret the results, and some of the caveats to consider with respect to BLAST.

A similar approach was taken to the range of resources from EBI and the related Swiss-Prot resources. The Swiss-Prot database and related protein analysis tools such as ExPASy and InterPro are considered superior to the protein resources of NCBI. Therefore, the focus was on the protein resources from EBI.

There are many ethical issues in relation to bioinformatics. Awareness and consideration of these ethical issues is important, even though they may fall outside the domain of library and information science. Therefore, one seminar-style class was devoted to a discussion of some of these issues. The implications of the information obtained through bioinformatics analysis (e.g., prenatal screening, genetic profiling) are one such issue. Another was the debate over freely shared versus private-for-profit access to genetic information.

Finally, the course considered the various disciplinary perspectives to bioinformatics. Two faculty members from the McGill Centre for Bioinformatics (one from biology, the other from computer science) discussed the field of bioinformatics from the perspective of their particular disciplines. Complementing these two guest lectures were discussions of the role of library and information science with respect to bioinformatics, examining the various roles that librarians and information professionals can and should play in this field. In one class discussion, students identified and articulated some of the unique skills and expertise that librarians and information science professionals possess and how these skills complement those of other professionals.

Evaluation

Evaluation was based on three assignments. The first was to prepare an annotation of two bioinformatics resources, one database (e.g., GenBank) and one analytical software tool (e.g., BLAST). This provided each student with the opportunity to explore in detail resources that would not necessarily be covered in the course. The annotations not only included the scope of the resources (e.g., information provided, database content, input/output format, search features), but also addressed issues of usability and utility of each resource.

The second assignment required students to complete search tasks in three of the resources discussed in class: OMIM, various elements of NCBI Entrez, and BLAST. Grading considered not only the search outcomes, but also the students' annotations and descriptions of the search process. They needed to demonstrate that they understood their actions and the rationale behind their results.

The final assignment, a term paper and class presentation, provided students with the opportunity for an in-depth exploration of a particular aspect of bioinformatics. Through the presentations, the class could also learn about each topic. The range of topics was very diverse, including ontologies and knowledge representation, information retrieval, data visualization, clinical implications of bioinformatics, and bioinformatics education and training for both librarians and scientists.

Students

Six students enrolled in the class. This small size was not unexpected for the first offering of a new, very specialized course. All were MLIS II (second year) students. Three of the students had a librarianship focus to their program; the other three had an information science perspective. Two of the students had previously taken the Health Sciences Information course, and no one had a biology background. The different perspectives of the students provided an interesting balance between those interested in resource creation and development and those interested in providing library and information services. Overall, the performance of the students was very good. The search assignments, in particular, were extremely well done.

Future directions

In the spring of 2005, the GSLIS Bioinformatics course was formally approved by the McGill Senate and re-designated as GLIS673 – Bioinformatics in LIS (the course title was modified to distinguish it from an existing course in biotechnology also entitled Bioinformatics). The course is now included in the GSLIS calendar as a regular course offering and is currently scheduled for winter 2006. Future offerings of the course will be expanded in scope to include more of the information science elements (e.g., information retrieval, data visualization) identified as areas of interest by the students.

The course is also listed on the site of the McGill Centre for Bioinformatics (MCB). In the fall term of 2005, the MCB will launch its graduate option. This will provide graduate students in participating departments the opportunity to include a formal bioinformatics specialization to their program, through a series of interdisciplinary courses offered by the Centre. We are currently exploring the possibility of including GSLIS among the participating departments. This would provide masters and doctoral students with an opportunity to study both library and information science and bioinformatics. While this opportunity is likely to appeal to a small minority of LIS students, it does provide an opportunity for a multidisciplinary education that is currently not available elsewhere in Canada. Graduates would be uniquely qualified to bring LIS expertise to the field of bioinformatics.

Conclusions

We have found that there is interest in a bioinformatics course within an MLIS program and that the absence of a strong biology background was not a barrier to student success. I believe that the presence of this specialized course has the potential to complement the program of study of MLIS students with diverse interests. In particular, given the range and diversity of issues and topics now faced by health sciences librarians, I believe that courses such as this (and other specialized courses, such as the course in Consumer Health Information offered by the University of Western Ontario) can only enhance the traditional, single course in health sciences librarianship. Having a selection of specialized, complementary courses will provide a stronger foundation in health sciences librarianship and provide new graduates with a stronger, broader, and more comprehensive set of knowledge and skills.

References

1. Persidis A. Bioinformatics. *Nat Biotechnol.* 1999 Aug;17:828–30.
2. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Method Inform Med.* 2001;40:346–8.
3. Galperin MY. The molecular biology database collection: 2005 update. *Nucleic Acids Res.* 2005 Database Issue;33:D5–D24.