



Developing Data Literacy Competencies to Enhance Faculty Collaborations

Don MacMillan

Librarian for Biological Sciences, Physics, Astronomy & Mathematics
University of Calgary, Taylor Family Digital Library, Calgary, Alberta,
Canada
macmilld@ucalgary.ca

Abstract

In order to align information literacy (IL) instruction with changing faculty and student needs, librarians must expand their skills and competencies beyond traditional information sources. In the sciences, this increasingly means integrating the data resources used by researchers into instruction for undergraduate students. Open access data repositories allow students to work with more primary data than ever before, but only if they know how and where to look. This paper will describe the development of two IL workshops designed to scaffold student learning in the biological sciences across two second-year courses, detailing the long-term collaboration between a librarian and an instructor that now serves over 500 students per semester. In each workshop, students are guided through the discovery and analysis of life sciences data from multiple sites, encouraged to integrate text and data sources, and supported in completing research assignments.

Key Words: Data literacy; bioinformatics; collaboration; science librarians

1. Introduction

This case study describes how life science data was successfully incorporated into an information literacy (IL) program in order to better align instruction with student and faculty needs. The librarian and instructor had developed a successful workshop in Biology 311-Principles of Genetics, a second-year course in 2006 (MacMillan, 2010). The instructor noted that while most students in the genetics course took Biochemistry 393, Introduction to Biochemistry, in the following semester, they were not linking knowledge gained in one course to the other as well as she expected. Many undergraduate biology programs continue to teach genetics and biochemistry from different philosophical or pedagogical perspectives which inhibits students from fully understanding or appreciating the content that is common to both subjects (Barrette-Ng & MacMillan, 2014). These two courses, however, were taught by the same instructor, so there were opportunities to encourage students to develop a more integrated understanding of both disciplines. In 2010 the librarian and instructor collaborated on a second IL workshop for the biochemistry course that would foster transfer of knowledge between the two subjects, and introduce students to other domain-specific data resources. Students learned how similar data informed research in genetics and biochemistry, and were able to practice skills gained in navigating complex gene databases as they explored resources on proteins. The primary goal of lab was to remove some of the artificial barriers that exist between these two disciplines by having students examine the molecular basis of genetically inherited diseases. Specifically, students in Biology 311- Principles of Genetics, used relevant data sources to investigate the molecular basis of their genetically inherited disease topic, and in Biochemistry 393, during the winter semester, they examined the structural origins of the molecular basis of their disease topic with related data. Both the genetics and biochemistry information literacy sessions were developed with the same guiding principle: to introduce students to the benefits of using 'real-world' tools early in their academic careers. They were both structured to provide scaffolded practice exploring selected bioinformatics resources, followed by assignments that required integrating information from these resources with other course content to answer questions about the role of genes and proteins in genetically-inherited diseases. As with the first genetics session, working with the biochemistry class enabled the librarian to develop new data competencies and gain a deeper understanding of the two disciplines. This has contributed to a more robust and sustainable

IL program that aligns more closely with faculty and student needs. The case study will provide background information on the resources included in the sessions, activities used to guide students in their exploration of these resources and suggestions for integrating bioinformatics data into information or data literacy instruction.

2. Literature Review

Researchers work with scholarly publications, data, and discovery tools in an integrated workflow. This suggests that academic librarians need to develop some familiarity and competencies with the scholarly communication patterns of the discipline in order to better integrate data and information literacy to educate the next generation of scientists. Providing an early introduction to bioinformatics data resources enables students to see science in action in fields where the published article may be almost secondary to the supporting data. Research data in these disciplines are in many cases more important than the articles they are linked or associated with (Akers & Doty, 2013; Castelli, Manghi, & Thanos, 2013). In a report by the National Academy of Sciences entitled *BIO 2010: Transforming Undergraduate Education for Future Research Biologists*, a key recommendation was the “need to involve students in working with real data and tools that reflect the nature of life sciences research in the 21st century” (as cited in Ditty et al., 2010, p. 1). Similarly, the *American Society for Biochemistry and Molecular Biology (ASBMB)* endorsed the use of bioinformatics and related data resources as a core competency for students in molecular biology courses (Boyle, 2004; Voet et al., 2003).

Bednarski, Elgin and Pakrasi (2005) were among the earliest to incorporate bioinformatics data into undergraduate courses. Bednarski developed an inquiry-based laboratory project which introduced undergraduate students to several key bioinformatics databases including the *Online Mendelian Inheritance in Man (OMIM)*, *GenBank*, and *Protein Data Bank (PDB)*, in order to study the effects of mutations on protein structure, function and human physiology. Students indicated they appreciated the investigative nature of these exercises and had a better understanding of how researchers develop hypotheses about genetic disease topics and share their scientific results. The authors also concluded that “the oral and written reports gave students a chance to practice their skills in communicating scientific information, and

the collaborative nature and group work in the lab gave students opportunities to both defend their ideas and learn from their peers” (Bednarski et al., 2005, p. 216).

Further evidence of the benefits of incorporating bioinformatics resources within undergraduate information or data literacy instruction to enhance student understanding of molecular processes has been reported by a number of recent studies (Badotti et al., 2013; Dymond et al., 2009; Weisman, 2010; Wightman & Hark, 2012). Best practices for implementing similar programs have been delineated (Badotti et al., 2013; Bednarski et al., 2005; Miskowski, Howard, Abler, & Grunwald, 2007; Via et al., 2011; Weisman, 2010), while others have discussed the programmatic role librarians can provide in supporting molecular biologists as they navigate the myriad data resources (MacMullen & Denn, 2005). Effective collaborations between librarians and the life sciences faculty have also been described in the literature (Bowden & DiBenedetto, 2001; Brown & Krumholz, 2002; Callinan, 2005; Dinkelman, 2010; Ferrer-Vinent & Carello, 2008; MacMillan, 2010; Winterman, 2009).

A common theme among many of these studies are the benefits associated with collaborating with the teaching faculty, including the potential to enhance librarians’ domain knowledge and expertise in the subject area. Many of the authors also provide valuable guidance on how to use bioinformatics tools more effectively. Familiarity, if not expertise in using bioinformatics databases is now considered a critical competency in life sciences librarians, as articulated in two recent reports. One prepared for the *Association for Research Libraries* (ARL) examined the role of liaison librarians in research libraries and recommended more sustainable engagement strategies that shift the focus to what users do (e.g., research, teaching, and learning) rather than on what librarians do (e.g., collections and reference) (Jaguszewski & Williams, 2014). In the second study published by *Ithaka S+R*, the author posits that “an emerging theme in the development of the liaison model is to shift the focus away from the work of librarians to that of scholars and to develop engagement strategies based on their needs and success indicators” (Kenney, 2014, p. 4).

In a study that examined the data information literacy needs of students and the research faculty, Carlson, Johnston, Westra and Nichols (2013) concluded that librarians can leverage teaching opportunities by integrating data into training and that a major “objective of this collaboration was to increase

the capacity of librarians to engage with students and faculty” (p. 215). The authors of this study also determined that most researchers in this study thought that offering some type of data literacy was required for their students (Carlson et al., 2013). Other studies have also concluded that academic librarians are well positioned to support data literacy and data competencies (Calzada Prado & Marzal, 2013; Gray, 2004; Schield, 2004; Stephenson & Schifter Caravello, 2007).

3. Context

The development of the two IL sessions described in this paper rests on two contexts. The first is the rapidly changing information environment in the life sciences and the second is the institutional setting where the classes take place.

The completion of the Human Genome Project in 2001 resulted in a deluge of life science data, particularly from large-scale sequencing projects that generate data pertaining to the structural and functional aspects of genes and proteins (Hey, Tansley, & Tolle, 2009). The discovery and analysis of this type of data has been facilitated by the emergence of bioinformatics databases and the development of a robust data infrastructure including disciplinary repositories that provide quality control and sustainable preservation.

Furthermore, researchers in the life sciences are increasingly being mandated to adhere to funder and/or publisher mandates to disclose the results of their research, including data, in an appropriate disciplinary data repository so that findings from any publically funded research is accessible and reproducible to others (Smith et al., 2011; Vickers, 2011).

In addition to providing sustainable data preservation and discoverability, bioinformatics databases are a unique form of scholarly content, integral to the scholarly communication paradigm in the life sciences. They are now built into the work of researchers, journal publishers, libraries and funding agencies intent on making research results more widely available (Bourne, 2005).

The accessibility of life science data is further enhanced by the interoperability between key publishers and selected data repositories. For example,

scientific publishers including Elsevier, Wiley, Springer and Nature are increasingly mandating that authors submit any data associated with a manuscript at an appropriate data repository such as the National Center for Biotechnology's (NCBI) *GenBank* and *Protein Data Bank* as a condition of publication. Increasingly, the article and associated data are linked bi-directionally which enabled students in this case study to begin their research on their disease topic from either the article or appropriate data repository.

While sophisticated, bioinformatics resources are relatively user-friendly, enabling inexperienced users to navigate and comprehend their potential if the right conditions are met (MacMillan, 2010). Fortunately, at the University of Calgary, the collaboration between the instructor and the librarian has met those conditions in developing workshops that not only introduce students to content-rich resources, but foster the integration of that content across disciplines.

The University of Calgary has a student population of approximately 31,000 students and offers over 200 undergraduate, graduate and professional degree programs. Information literacy support is provided for a variety of undergraduate programs including the biological sciences. When the author first worked with the biological sciences department, the main contact point with students was a first-year course with an assignment that introduced students to library resources in order to locate a peer-reviewed article on a research topic. However as students in first year did not need to do much research outside of the library assignment, this may not have been the most effective place for IL instruction, and the sessions were eventually dropped from the course in 2010. An instructor in the second year genetics course, however, did require her students to integrate information resources in their work, and so began the collaboration that is detailed in the case study. Information literacy instruction in these second-year courses is extended in third and fourth year in biochemistry and cellular biology courses where instruction includes reminders about bioinformatics content but focuses on bibliographic resources and information management.

The initial collaboration between the instructor and the librarian on an information literacy session for Biology 311-Principles of Genetics began in 2006, with the goal of providing a more authentic inquiry-based learning environment, and has been refined annually ever since to incorporate new resources,

changes to interfaces and greater understanding of where students have the greatest difficulty. In 2010 this collaboration was extended to Biochemistry 393-Introduction to Biochemistry, so that students would better understand the content that is common to both disciplines by investigating the effects of mutations on different levels of protein structure, function, and genetically inherited diseases (Barrette-Ng & MacMillan, 2014). Lessons learned in developing Biology 311 were applied to the new class and both model authentic pathways that replicate researcher workflow, incorporate scaffolded steps and hands-on practice that leverage the interoperability of different bioinformatics resources. Both are assessed through short term assignments and larger projects that require further independent exploration of the resources.

4. Case Study

This case study will outline each of the classes, and describe the information resources and activities. Details are provided so that others may adapt the work to their own contexts.

4.1. Biology 311-Principles of Genetics

This is a required course for biological sciences majors and covers topics such as Mendelian inheritance, sex determination, molecular genetics, and the structure and function of genetic material. The course is offered during the fall semester with enrollment in the fall ranging between 500–550 students. In mid-October students attend an information literacy workshop during their scheduled lab time. Typically three lab sections are combined for a total of 70–75 students who are accompanied by the Graduate Teaching Assistants for each lab. The session takes place in the library's computer classrooms which have 48 workstations and students generally work with their lab partners. The goal of this lab is the investigation of the molecular basis of genetically inherited diseases.

The teaching objectives for this lab include:

- Use PubMed to find a peer-reviewed article on a specific genetically-inherited disease topic.
- Locate a patent that pertains to a gene or mutation believed to cause a genetically-inherited disease.

- Use the *OMIM* database to determine the mode of inheritance of a genetically-inherited disease as well as the causative gene(s) and mutation(s).
- Use *NCBI's Gene* database to obtain data on gene architecture including mRNA and protein length.

Students work through a 20 question laboratory workbook during the session and then must revisit the resources to prepare and present a scientific poster using life science data that explains why some individuals inherit genetic diseases. During each IL session students are introduced to a series of exercises that focus on a specific resource used to answer a specific question and/or teaching objective listed above, related to the disease they have chosen to research. Students select their own disease topic which may enhance their motivation; many choose to research a disease that affects people they know which is likely to increase their engagement with the assignment.

The librarian provides a live demonstration of each resource using the *CFTR* gene that is implicated in Cystic Fibrosis. As each resource or bioinformatics database is introduced and demonstrated students are given the opportunity to practice with their own examples and additional time is available at the end of the session. The lab workbook provides a structured outline of the various bioinformatics tools including screen shots and explanatory notes that the students can use during and after the scheduled session. The questions in the workbook are progressively more complex and demanding, requiring students to knit together their understanding of the different resources as they complete the exercises. Following the session, each student is required to answer a series of questions each week about their chosen disease topic, which are integrated into a poster and oral presentation during the final week of term. The entire project is worth 6% of the overall course grade. A course related *LibGuide* (<http://libguides.ucalgary.ca/content.php?pid=55723&sid=603862>) with links and brief explanations of all of the resources used for this assignment provides additional structure for each session and a convenient entry point to each resource as students work on subsequent assignments and their posters. The librarian provides support while students work on the in-class assignment and is available for further consultation as needed before the assignments are due.

The resources in the BIOL 311 library session are described below in the order they are presented in the sessions.

Genetics Home Reference

(<http://ghr.nlm.nih.gov/about>)

This is the National Library of Medicine's web site for information about genetic conditions including the genes and chromosomes implicated with specific disease. It contains information on over 1000 genetically-inherited diseases or syndromes and over 1200 genes. Most students are able to retrieve data and background information on their disease topic such as the inheritance pattern of the disease, the name of the gene(s) implicated with their disease and the cytogenetic location or specific chromosome where a gene is located. As many diseases have more than one causative gene(s) or mutation, students were asked to investigate only one gene and mutation for their assignment and presentation.

Online Mendelian Inheritance in Man (OMIM)

(<http://www.omim.org/>)

OMIM is a comprehensive and authoritative database of human genes and genetic disorders hosted by the NCBI, and maintained by McKusick-Nathans Institute of Genetic Medicine at Johns Hopkins University School of Medicine and updated daily. *OMIM* utilizes NCBI's *Entrez* search and retrieval system, familiar to *PubMed* users, which provides seamless access to over 40 molecular and literature databases including *PubMed*.

OMIM provides background information on diseases with hyperlinks to related journal articles in *PubMed*. Students need to select the correct entry or record from the *OMIM* results. All entries in the *OMIM* database are preceded by symbols designating the molecular status of each entry. An asterisk (*) denotes data on a gene of a known genetic sequence, a pound sign (#) or plus sign (+) indicate a descriptive entry with no known genetic locus or known sequence. Other symbols usually indicate an ambiguous genetic origin with less relevant or minimal data, so students are advised to select an *OMIM* entry or record preceded by an asterisk or pound symbol if necessary as they would provide more information. Students who select various cancers or Alzheimer's disease which are characterized by a complex etiology, numerous causative genes or mutations and possible environmental factors, often have to check or corroborate their findings with related *PubMed* articles or other data sources used in this exercise to ensure that they have identified the most relevant *OMIM* entry. Each *OMIM* record is well organized and includes a *Table of*

Contents comprised of approximately twenty sections such as *Description*, *Gene Structure*, *Mapping* and *Biochemical features* as well as an *External Links* section with links to other data sources covering *DNA*, *Genome*, *Gene Info*, *Protein* and *Clinical Resources*. Students then use the *Gene Info* link on the *OMIM* record as a gateway to *NCBI's Gene* database to locate data pertaining to the gene's architecture and function for the next part of their assignment.

NCBI Gene

(<http://www.ncbi.nlm.nih.gov/gene/>)

Gene provides extensive gene-specific data including gene structure or architecture such as mRNA and protein length. Students use the *Gene* database to better understand gene structure and the effect mutations have on related functions. Like *OMIM*, *Gene* uses *NCBI's Entrez* discovery tool and includes a *Table of Contents* and *Related Information* for convenient access to the data students needed to answer questions. Questions included the length of the wild-type protein which could be found using the "*RefSeqProteins*" link, the length or size of the mature mRNA (e.g., *RefSeqRNAs*) for their gene, and the number of exons and introns (e.g., *RefSeqGene*). Exons are nucleotide sequences that code for proteins while introns do not. Students also identify the type of chromosome where the gene is located (e.g., metacentric, telocentric, or acrocentric), and locate an ideogram or schematic diagram (e.g., *Map Viewer*) displaying this information. All of this data is available from the *Related Information* section on the *Gene* page.

PubMed

(<http://www.ncbi.nlm.nih.gov/pubmed>)

Each student has to obtain one review article on their genetic disease topic to provide background information and supplement and contextualize the data in their poster presentations.

Google Patents

(https://www.google.com/?tbs=pts&gws_rd=ssl)

Patents provide a unique source of scientific information particularly in the life sciences where many researchers often seek patent protection to recover research and development costs associated with research conducted on genes, proteins, laboratory techniques and related diagnostic tools. Patents in these fields frequently contain pertinent molecular information including

images that help students with their research. Students locate one patent that relates to the molecular origins of their disease topic such as the causative gene or protein, or a diagnostic method used to detect the mutant gene or protein. Patents that relate to therapies or treatments for their disease topic are not used for this assignment.

As this is likely the students' first introduction to patents and they only need to locate one, *Google Patents* is used to access patents from the *United States Patent and Trademark Office*. *Google Patents* provides a user-friendly interface for searching patents from the major patent issuing agencies. Students answer questions regarding the patent number, filing date, and details about the patent's "Claims" as well as a discussion question regarding the ethics of patenting life forms such as genes and patents and whether they agree that their chosen patent should have been granted.

By completing the lab workbook, students gather much of the information they need for their poster presentation. Students had several weeks to develop the poster, during which time they can refer to the lab workbook and LibGuide, or consult with the librarian.

4.2. BCEM 393-Introduction to Biochemistry

This course, like Biology 311 which is a pre-requisite, is required for all biology majors, and is offered in the winter semester. Whereas the Biology 311 information literacy workshop focuses on the molecular basis of genetically-inherited diseases, the library lab in Biochemistry 393 concentrates on developing a structural understanding of the molecular basis of these diseases, particularly the effects of mutations on protein structure and function. This course covers the structure and function of carbohydrates, amino acids, proteins, lipids, coenzymes and enzymes including an introduction to metabolism and energy transduction. The information literacy workshop is integrated into Lab 2, "Gaining a structural perspective on the molecular basis of inherited diseases", and is usually held during the first week of February.

The teaching objectives for this lab include:

- Use the *Protein Data Bank* to locate the protein's three-dimensional coordinates.

- Use *ProtParam* to calculate the isoelectric point of a protein.
- Run a Blast search to identify homologous or similar proteins and use *ClustalW* to align them.
- Use *PyMOL* to investigate the effect of a mutation on protein structure and function.

Students select a human genetically-inherited disease topic from a list of eleven approved disease topics chosen because they have associated files and designated mutations that would work with the appropriate bioinformatics resources in this assignment. In order to focus students' attention and effort on using the data, the instructor does more advance work, including identifying appropriate topics and articles, and pre-loading corresponding *PyMOL* files on the course website. Like the work students complete for their Biology 311 project, students in this class are introduced to several bioinformatics databases that serve as data repositories for protein structural information where they examine the impact of mutations on protein structure, function and specific diseases.

Students begin their research with a 'seed' or primary article pre-selected by the instructor to provide background and structural information on their chosen protein without discussing the mutation they will be studying. Students need to develop and test a hypothesis about the mutation based on the data they collect from the various bioinformatics sources in order to complete their assignment and presentation.

The librarian leads the students through the exercises using the "K-Ras" protein as an example, and screen shots are included with accompanying text in the lab workbook. Students start their research with familiar resources introduced in Biology 311, such as *Genetics Home Reference*, *PubMed* and *OMIM* to locate background information about their disease including etiology and enzyme(s). They then explore specific protein data repositories to retrieve information about their chosen protein.

These data resources are highly integrated with each other, enabling students to link seamlessly between each source to gather required information. For example, students using the *Protein Data Bank (PDB)* to obtain structural information about their chosen protein can also retrieve the 'Primary Citation' article associated with that protein by using the *PubMed ID* or *Digital Object Identifier (DOI)* which is hyperlinked to *PubMed* and/or

the publisher's site. Students obtain information about the size of the protein (e.g., number of amino acids) associated with their chosen topic. The size or length of the protein can vary slightly among different bioinformatics databases but students are shown how to check the actual protein size by selecting the *UniProtKB* number hyperlink which automatically directs them to *ProtParam*, another protein structural repository, for the most current information. The *ProtParam* protein repository is subsequently used to locate protein sequence data.

Protein Data Bank (PDB)

(<http://www.rcsb.org/pdb/home/home.do>)

The *Protein Data Bank (PDB)* is a portal for three-dimensional structural information on over 100,000 proteins and nucleic acids. A major component of the project for this lab was having students locate the 3D coordinates for the protein associated with their genetically-inherited disease topic. Because a protein's function is intricately linked to its structure, any change or degradation in a protein's three-dimensional structure will result in a loss of protein function. Consequently, protein structural information is critical to understanding a protein's function. For this portion of the lab students are guided through a series of questions on their selected protein's structure, including protein size, molecular weight, and number of amino acids. Each *PDB* record also has a unique *UniProtKB* reference link that redirects the user from the *PDB* record to the corresponding *UniProtKB* entry, where students are able to locate more structural information about their protein and perform a *BLAST* search to locate similar proteins to their selected protein.

UniProtKB

(<http://www.uniprot.org/>)

UniProtKB is a clearinghouse for sequence and functional information about proteins. Students use *UniProtKB* to find data including the molecular weight, isoelectric point and number of amino acids for their protein.

BLAST

(<http://www.uniprot.org/blast/>)

Students performed a *BLAST* search to locate any similar or homologous proteins and measure the degree of similarity between different protein sequences.

This enables students to determine what function a certain protein fulfills among various organisms. Students use *BLAST* to determine if there are any protein sequences from *Gallus gallus* (Chicken), *Mus musculus* (Mouse), or *Danio rerio* (Zebrafish) similar to their chosen protein. By examining their results students deduce what function their chosen protein might perform. Students also conduct a sequence alignment between these four organisms using the *ClustalW* program which is embedded within the *BLAST* search results. They can then determine the best match between the selected sequences.

PyMOL

(<http://www.pymol.org/>)

PyMOL is an open source molecular visualization program that enables users to manipulate and view protein structures in three-dimensions. Students used *PyMOL* to study the effects of a single amino acid substitution on protein structure and function. *PyMOL* is downloaded in advance on the classroom workstations and each student has access to a *PyMOL* file consisting of a pre-selected view of each protein structure through their course management software pack. After uploading their specific *PyMOL* file students can manipulate and rotate their file using different function keys in order to view their protein structure from different perspectives. Students then insert their assigned disease-causing mutation in the protein by substituting an amino acid. For demonstration purposes the librarian uses the "K-RAS" protein to demonstrate mutation *G60R*, where *Gly-60* (Glycine) was substituted by *Arg-60* (Arginine). Using their own protein examples, students complete a series of exercises that investigate the impact of mutations on the three-dimensional structure of their protein. Their observations and collected data are incorporated in the poster presentation.

Cologne University Protein Stability Analysis Tool (CUPSAT)

(<http://cupsat.tu-bs.de/>)

The final data source presented in this session is *CUPSAT*, which enables students to predict changes in protein stability when point mutations are introduced. This type of mutation changes a single nucleotide base pair. Students carry out a number of tasks and include the results in their research.

Students work in pairs during this workshop and while most students complete the assigned questions during the session, they have approximately

5 weeks to finish it and prepare their poster for the oral presentation at the end of term. Students also complete a series of independent exercises at regular intervals throughout the term which integrate course content into the work they do on the poster presentation.

5. Results

Incorporating disciplinary data from gene and protein data sources within undergraduate IL has benefits for librarians and students. For the librarian, developing familiarity with life science data provided an opportunity to engage with the discipline more deeply. Developing competencies through the discovery and analysis of molecular biology data has informed the librarian's domain expertise in genetics and biochemistry, in turn providing valuable contextual knowledge when demonstrating and answering questions about the various data resources. Gaining knowledge of bioinformatics resources has also translated into an improved alignment of library services with faculty requirements, and has resulted in many productive conversations with faculty members about their data use and needs. This in turn has led to a more collaborative, sustainable and rigorous information literacy program. The value of this collaboration with faculty members extends beyond the classroom, and has resulted in numerous presentations to both library and biology audiences.

As mentioned earlier in this paper, librarians, particularly science librarians, need to maintain currency in patterns of scholarly communication, to better align their professional activities more closely with the research and learning needs of students and faculty members. Jaguszewski and Williams posit that "an engaged librarian seeks to enhance scholar productivity, to empower learners, and to participate in the entire lifecycle of the research, teaching, and learning process" (Jaguszewski & Williams, 2013, p. 4). This need to develop new skills in order to take on new roles is a dominant theme in current literature (Auckland, 2012; Kenney, 2014).

Students make significant learning gains through the experiential learning provided in the library sessions. Using bioinformatics databases to answer specific molecular biology questions ensured that students were able to extend and integrate learning throughout the academic year, progressing from easy

to more complex questions, while at the same time understanding concepts that underline both disciplines. In both courses, bioinformatics tools facilitated the discovery and analysis of life sciences data, illustrating the role of gene and protein structure in determining related functions. Students can learn about these topics in a lecture or from a textbook but using *PyMol* to substitute an amino acid and visualize the impact on protein structure and function in three-dimensions provides a more engaged, interactive and authentic learning experience. Using these databases in conjunction with other sources such as *PDB* and *PubMed* allowed for a more integrated understanding of how various sources of information serve researchers. Using a *GenBank* record, a patent, and a scholarly article that refer to the same data demonstrates how these resources complement each other, and how students can correlate information found across the resources to better understand it. Starting from genetics resources to solve problems in biochemistry underscores the deep connections between the two disciplines and encourages deeper transfer of knowledge.

In considering the gains students make through this integration beyond the simple exposure to new resources, it is possible to relate them to the threshold concepts in information literacy currently under discussion by the ACRL. Many of the Knowledge Practices described in the Framework document are supported by the design of the library session and accompanying assignments (Gibson & Jacobson, 2014). Students are encouraged to develop a deeper understanding of how knowledge in the discipline is created and deploy that understanding to search and research more effectively, in order to synthesize information from a variety of sources into a scholarly artefact. Students develop an understanding of bioinformatics data in BIOL 311 through the library session and subsequent practice. They then deepen that understanding in BCEM 393, through exercises deliberately designed to foster knowledge transfer, and again have multiple opportunities to practice extracting information from data.

Students who participated in these sessions gained hands-on experience with critical tools in their discipline, developed an appreciation for how knowledge in genetics and biochemistry is created and integrated based on incremental discoveries, and how new bioinformatics tools facilitate the synthesis of information into knowledge. Their instructor has noted that students have developed a more integrated understanding of biochemistry and genetics, in part through seeing how researchers use similar data to understand questions in both domains. The librarian frequently attends the poster sessions

and has observed students demonstrate an ability to use the bioinformatics data in conjunction with more conventional text resources to explore genetic factors in disease. Some of the Graduate Teaching Assistants, now working with the classes, were early participants in the library sessions and frequently comment on the usefulness of early exposure to the databases.

Throughout this collaboration, assignments and activities have been continuously revised in response to new resources, new interfaces and observations of where students have been challenged to understand the material or the processes to access and link material. From experience and the literature, the following suggestions for incorporating bioinformatics data emerge. Many of these are common to IL as a whole, but worth reconsidering and relating to new content:

- Be prepared to put in substantial time developing familiarity with the bioinformatics databases and subject content in order to develop exercises and demonstrations and to anticipate problems and student questions.
- Be prepared to revise constantly as the tools change frequently.
- Instruction in bioinformatics data should be course integrated, with a graded assignment of sufficient weight so that students take it seriously.
- Develop sequential steps from simple databases and operations to more complex ones.
- Develop specific questions for each resource that highlight the relevant data in each.
- Scaffold exercises to integrate just-learned material with new steps.
- Explicitly link the content to the course, and the content of different resources to each other to help students develop a structure for the new knowledge.
- Students learn by doing, allow hands-on interactivity & flexibility, and multiple opportunities to practice.
- Consider integrating text and data resources so students see how the same information is used in different ways.

6. Conclusion

This case study presents the development of two linked IL sessions that introduce undergraduate students to bioinformatics resources. While each of the

classes provides benefits to students in terms of enhanced content knowledge, linking the courses through their foundational information resources helps students understand genetics and biochemistry as integrated fields of inquiry. This has clear benefits for the students, and indeed for the librarian. Moving beyond the consideration of IL as isolated experiences, to a more integrated curriculum allows for a deeper engagement with student learning and a closer alignment with the needs of current and future researchers. This closer alignment has led to more effective collaborations, not only with the instructor in these courses, but also with other members of the department. Developing an understanding of the role of bioinformatics databases has allowed the librarian to work with current researchers at a different level, as well as providing future researchers with necessary skills.

References

- Akers, K.G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5–26. doi: 10.2218/ijdc.v8i2.263.
- Auckland, M. (2012). *Re-skilling for research: An investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers* (p. 115). London: Research Libraries UK. Retrieved January 7, 2015, from <http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf>.
- Badotti, F., Barbosa, A.S., Reis, A.L.M., do Valle, I.F., Ambrósio, L., & Bitar, M. (2013). Comparative modeling of proteins: A method for engaging students interest in bioinformatics tools. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 42(1), 68–78. doi: 10.1002/bmb.20721.
- Barrette-Ng, I., & MacMillan, D. (2014). Gaining a structural perspective on the molecular basis of inherited diseases. In *Tested Studies for Laboratory Teaching: Proceedings of the 35th Workshop/Conference of the Association for Biology Laboratory Education (ABLE)* (pp. 225–244). Calgary, Alberta: Association for Biology Laboratory Education.
- Bednarski, A.E., Elgin, S.C.R., & Pakrasi, H.B. (2005). An inquiry into protein structure and genetic disease: Introducing undergraduates to bioinformatics in a large introductory course. *Cell Biology Education*, 4(3), 207–220. doi: 10.1187/cbe.04-07-0044.
- Bourne, P. (2005). Will a biological database be different from a biological journal? *PLoS Computational Biology*, 1(3), 179–181. doi: 10.1371/journal.pcbi.0010034.

- Bowden, T.S., & DiBenedetto, A. (2001). Information literacy in a biology laboratory session. *Research Strategies*, 18(2), 143–149. doi: 10.1016/S0734-3310(02)00071-X.
- Boyle, J.A. (2004). Bioinformatics in undergraduate education. *Biochemistry and Molecular Biology Education*, 32(4), 236–238. Retrieved January 7, 2015, from <http://onlinelibrary.wiley.com/doi/10.1002/bmb.2004.494032040376/pdf>.
- Brown, C., & Krumholz, L.R. (2002). Integrating information literacy into the science curriculum. *College & Research Libraries*, 63(2), 111–123. doi: 10.5860/crl.63.2.111.
- Callinan, J.E. (2005). Information-seeking behaviour of undergraduate biology students. *Library Review*, 54(2), 86–99. doi: 10.1108/00242530510583039.
- Calzada Prado, J., & Marzal, M.Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and concepts. *Libri: International Journal of Libraries & Information Services*, 63(2), 123–134. doi: 10.1515/libri-2013-0010.
- Carlson, J., Johnston, L., Westra, B., & Nichols, M. (2013). Developing an approach for data management education: A report from the data information literacy project. *International Journal of Digital Curation*, 8(1), 204–217. doi: 10.2218/ijdc.v8i1.254.
- Castelli, D., Manghi, P., & Thanos, C. (2013). A vision towards scientific communication infrastructures. *International Journal on Digital Libraries*, 13(3–4), 155–169. doi: 10.1007/s00799-013-0106-7.
- Dinkelman, A.L. (2010). Using course syllabi to access research expectations of biology majors. *Issues in Science & Technology Libraries*, Winter. doi: 10.5062/F49Z92TJ. Retrieved January 7, 2015, from <http://www.istl.org/10-winter/refereed3.html>.
- Ditty, J.L., Kvaal, C.A., Goodner, B., Freyermuth, S.K., Bailey, C., Britton, R.A., ... Kerfeld, C.A. (2010). Incorporating genomics and bioinformatics across the life sciences curriculum. *PLoS Biology*, 8(8), e1000448. doi: 10.1371/journal.pbio.1000448.
- Dymond, J.S., Scheifele, L.Z., Richardson, S., Lee, P., Chandrasegaran, S., Bader, J.S., & Boeke, J.D. (2009). Teaching synthetic biology, bioinformatics and engineering to undergraduates: The interdisciplinary Build-a-Genome course. *Genetics*, 181(1), 13–21. doi: 10.1534/genetics.108.096784.
- Ferrer-Vinent, I.J., & Carello, C.A. (2008). Embedded library instruction in a first-year biology laboratory course. *Science & Technology Libraries*, 28(4), 325–351. doi: 10.1080/01942620802202352.
- Gibson, C., & Jacobson, T.E. (2014). *ACRL Framework for information literacy for higher education* (p. 26). Chicago, IL: Association of College & Research Libraries. Retrieved January 7, 2015, from <http://acrl.ala.org/ilstandards/wp-content/uploads/2014/11/Framework-for-IL-for-HE-draft-3.pdf>.
- Gray, A.S. (2004). Data and statistical literacy for librarians. *IASSIST Quarterly*, 28(2/3), 24–29. Retrieved from http://www.iassistdata.org/sites/default/files/iq/iqvol282_3gray.pdf.

Hey, T., Tansley, S., & Tolle, K. (2009). *Jim Gray's fourth paradigm and the construction of the scientific record* (pp. 177–183). Redmond, WA: Microsoft Corporation. Retrieved January 7, 2015, from http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf.

Jaguszewski, J.M., & Williams, K. (2013). *New roles for new times: Transforming liaison roles in research libraries*. Washington DC: Association of Research Libraries (p. 17). Retrieved January 7, 2015, from <http://www.arl.org/storage/documents/publications/NRNT-Liaison-Roles-final.pdf>.

Kenney, A.R. (2014). *Leveraging the liaison model: From defining 21st century research libraries to implementing 21st century research universities*. Ithaca (p. 11). Retrieved January 7, 2015, from <http://www.sr.ithaka.org/blog-individual/leveraging-liaison-model-defining-21st-century-research-libraries-implementing-21st>.

MacMillan, D. (2010). Sequencing genetics information: Integrating data into information literacy for undergraduate biology students. *Issues in Science & Technology Libraries, Spring*. doi: 10.5062/F44F1NNK. Retrieved January 7, 2015, from <http://www.istl.org/10-spring/refereed3.html>.

MacMullen, W.J., & Denn, S.O. (2005). Information problems in molecular biology and bioinformatics. *Journal of the American Society for Information Science and Technology, 56*(5), 447–456. doi: 10.1002/asi.20134.

Miskowski, J.A., Howard, D.R., Abler, M.L., & Grunwald, S.K. (2007). Design and implementation of an interdepartmental bioinformatics program across life science curricula. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology, 35*(1), 9–15. doi: 10.1002/bmb.12.

Schild, M. (2004). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly, 28*(2/3), 6–11. Retrieved January 7, 2015, from <http://www.statlit.org/pdf/2005SchildIASSIST.pdf>.

Smith, A., Balazinska, M., Baru, C., Gomelsky, M., McLennan, M., Rose, L., ... Kolker, E. (2011). Biology and data-intensive scientific discovery in the beginning of the 21st century. *Omics: A Journal of Integrative Biology, 15*(4), 209–212. doi: 10.1089/omi.2011.0006.

Stephenson, E., & Schifter Caravello, P. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences. *Reference Services Review, 35*(4), 525–540. doi: 10.1108/00907320710838354.

Via, A., De Las Rivas, J., Attwood, T.K., Landsman, D., Brazas, M.D., Leunissen, J.A.M., ... Schneider, M.V. (2011). Ten simple rules for developing a short bioinformatics training course. *PLoS Computational Biology, 7*(10), e1002245. doi: 10.1371/journal.pcbi.1002245.

Vickers, A.J. (2011). Making raw data more widely available. *BMJ (Clinical Research Ed.)*, 342(May04_2), d2323. doi: 10.1136/bmj.d2323.

Voet, J.G., Bell, E., Boyer, R., Boyle, J., O'Leary, M., & Zimmerman, J.K. (2003). Recommended curriculum for a program in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education*, 31(3), 161–162. doi: 10.1002/bmb.2003.494031030223.

Weisman, D. (2010). Incorporating a collaborative web-based virtual laboratory in an undergraduate bioinformatics course. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 38(1), 4–9. doi: 10.1002/bmb.20368. Retrieved January 7, 2015, from <http://www.readcube.com/articles/10.1002/bmb.20368>.

Wightman, B., & Hark, A.T. (2012). Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 40(5), 310–319. doi: 10.1002/bmb.20637. Retrieved January 7, 2015, from <http://onlinelibrary.wiley.com/doi/10.1002/bmb.20637/pdf>.

Winterman, B. (2009). Building better biology undergraduates through information literacy integration. *Issues in Science & Technology Libraries, Summer*. doi: 10.5062/F4736NT6. Retrieved January 7, 2015, from <http://www.istl.org/09-summer/refereed1.html>.