

Academic programming for chemical informatics: models and preliminary plans for the Chemistry Library

Jeffery Loo and Mary Ann Mahoney
Chemistry and Chemical Engineering Library
January 26, 2010

Contents

Instructional programs.....	1
Chemical informatics research programs.....	2
Chemical informatics scientific services	3
Preliminary Chemistry Library plans for chemical informatics programming.....	4

Instructional programs

Here are recognized instructional programs and expertise for chemical informatics:

Jean-Claude Bradley at Drexel University

- He teaches a semester-long class on Chemical Information Retrieval (CHEM367 and CHEM767).
 - Primary topics = Searching chemical literature databases (SciFinder, Bielstein, etc.), searching for chemical properties, structure searching, the organization of chemical information, open access and chemistry, searching for links to biological and interdisciplinary information, Web 2.0
 - Syllabus = <http://getcheminfo.wikispaces.com/syllabus>
 - Lectures are screencasted (slides and audio)
= <http://getcheminfo.wikispaces.com/>
- He also gave a lecture on open notebook science that outlines web 2.0 and science, free publicly available chemistry databases, strategies for tagging molecules and reactions with various identifiers, open access resources, and structural representation.
 - Screencast available here: <http://www.scivee.tv/node/5553>

There is an extensive teaching and research program in chemical informatics at the School of Informatics and Computing, Indiana University.

- <http://cheminfo.informatics.indiana.edu/>
- The focus is on Integrative Drug Discovery Informatics:
 - Offers a Master's degree and PhD in Cheminformatics
 - Details about their programs are here: <http://cheminfo.wikispaces.com/>
- The I571 Chemical Information Technology class has a wiki with the syllabus, lecture notes, assignments:
 - <http://i571.wikispaces.com/>

- This course gives a “wide overview of cheminformatics, including basic algorithmic and database techniques, and includes perspectives from industry and academic practitioners on topics such as virtual screening, computational toxicology and data mining.”

The Indiana Cheminformatics Education Portal (<http://icep.wikispaces.com/>) provides resources for developing an educational program. It includes:

- cheminformatics learning materials on 2D and 3D structure searching and representation, reaction searching and representation, data mining, visualization, QSAR, and similarity searching
- tips for developing curricula and programs
- details on a courseshare program for chemical informatics among Big 10 Universities

The Chemical Information Sources Wiki is a comprehensive source on understanding and retrieving chemical information

- http://cheminfo.informatics.indiana.edu/cicc/cis/index.php/Main_Page
- For example, there is a listing of chemical information instructional materials: http://cheminfo.informatics.indiana.edu/cicc/cis/index.php/Clearinghouse_for_Chemical_Information_Instructional_Materials

HS Rzepa taught Chemical Information Technology at Imperial College

- Class resources here: <http://www.ch.ic.ac.uk/local/it/lectures.html>
- Topics covered include computer and networking basics, chemical structure and chemical property searching, wikis, bibliographic software and searching, integrated information and value-added environments.

Here is a presentation of international programs in chemical informatics:

- http://infochim.u-strasbg.fr/chemoinformatics/PDF_presentation/Willett_OrbernaiTalk.pdf

Chemical informatics research programs

Research programs tend to focus on drug design and chemical biology:

- Here is a Google search for chemical informatics research programs: [http://www.google.com/search?hl=en&as_q=research&as_epq=&as_oq="chemical+informatics"+cheminformatics+chemoinformatics](http://www.google.com/search?hl=en&as_q=research&as_epq=&as_oq=)

Here is a sample of research projects:

- Karypis Lab, Department of Computer Science & Engineering at the University of Minnesota
 - <http://glaros.dtc.umn.edu/gkhome/project/cheminfo/overview>

- BCIRC, a bioinformatics and chemical informatics research center, residing in the Shoichet Laboratory of the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF)
 - <http://bcirc.docking.org/>
- The Chemical Informatics and Cyberinfrastructure Collaboratory (CICC) at Indiana University
 - This NIH-funded project combines grid computing with chemical informatics. Details are here: <http://www.chembiogrid.org/index.html>
- Chemical Biology Platform, Broad Institute, MIT and Harvard
 - The informatics team manages data collection and integration from chemical synthesis through screening and data analysis. Other research areas include electronic lab notebooks, data warehouses, automated lab management system, and developing the public database ChemBank for small molecules and small molecule screens.
 - <http://www.broadinstitute.org/science/platforms/chemical-biology-platform/chemical-biology-platform>
- Chemical Informatics Program, Scripps Research Institute
 - Services for work flow, procedures, and business rules and in integrating discovery informatics with the operational informatics infrastructure to facilitate discovery processes and developing integrative applied cheminformatics methods.
 - <http://www.scripps.edu/news/sr/sr2008/tri08schurer.html>
- The Carolina Exploratory Center for Cheminformatics
 - This group promotes multidisciplinary, multi-institutional collaboration among researchers in computational chemistry, chemical biology, datamining, computer science, and statistics to address critical issues in Cheminformatics in the context of Molecular Libraries Initiative (MLI) at NIH
 - <http://ceccr.unc.edu/>
- The Rensselaer Exploratory Center for Cheminformatics Research
 - <http://reccr.chem.rpi.edu/>
 - They conduct projects in data generation, machine learning and data mining, property-prediction, and applications.

Chemical informatics scientific services

Here are some groups that provide scientific services to academic departments:

- Core C: Library Design, Analysis & Purification. A scientific services group at the University of Kansas Center for Chemical Methodologies and Library Development
 - http://www.cml.d.ku.edu/core_c.shtml
 - A chemical informatics data specialist conducts data management activities like creating databases, managing tools for project management

and knowledge transfer, *in silico* pre-screening and evaluation of chemical libraries, and facilitating data sharing initiatives.

- Here was a project at Indiana University developing a chemical informatics and cyberinfrastructure collaboratory funded by the NIH:
 - <http://www.chembiogrid.org/documents/internal/CICCresearchplan.pdf>
- The CADD Group Chemoinformatics Tools and User Services
 - A group at the National Cancer Institute: <http://cactus.nci.nih.gov/>
 - The purpose of this service is to provide structures, data, tools, programs and other useful information to the public.

Preliminary Chemistry Library plans for chemical informatics programming

What are the Chemistry's Library's plans?

The Chemistry Library's chemical informatics contributions will focus on libraries' traditional strengths in:

1. resource instruction (i.e., resource guides and information literacy instruction), and
2. organizing seminars and events.

What would library resource instruction cover?

It would focus on:

- chemical information sources - particularly, PubChem, SciFinder, Gmeilin, Bielstein, Science of Synthesis
- structure and reaction searching
- data resources and retrieval
- scholarly communications
- bibliographic management (such as EndNote and Refworks)

What would the seminars and events be like?

The seminars would cover pressing issues in library and information science as it applies to chemistry. This includes:

- scholarly communication and digital publishing
- open access resources for chemical data and information
- interdisciplinary links between chemical information tools and other fields

These seminars would hopefully address many of the issues surrounding new scientific communication models for chemistry that are detailed in this white paper prepared by an international workshop of chemists, societies, publishers and information science

groups: http://ecommons.library.cornell.edu/bitstream/1813/14150/4/whitepaper_fina_l.pdf

This white paper is summarized in this Nature Chemistry article:
<http://www.nature.com/nchem/journal/v1/n9/full/nchem.448.html>

Sample topics would include:

- PubChem, <http://pubchem.ncbi.nlm.nih.gov>
- ChemSpider, <http://www.chemspider.com/>, the publicly available database of chemical structures and properties. Here are some presentations on the resource: <http://www.slideshare.net/AntonyWilliams>
- Web 2.0 and eScience. This includes issues like structured data, data manipulation, and structure file formats.
- Zinc, <http://zinc.docking.org/>, a free database of commercially-available compounds for virtual screening
- Other open access databases for chemistry. See this article by Antony Williams: <http://dx.doi.org/10.1016/j.drudis.2008.03.017>

How would the Chemistry Library's role evolve?

These notes are preliminary ideas for services. We are open to different approaches after these preliminary plans are established and refined and a foundation is built. Possibilities include curriculum-integrated instruction and librarian team teaching with faculty, and more.

In general, science libraries are moving towards issues on the "reuse of data, data management skills, cyberinfrastructure planning, interinstitutional collaboration, incorporation of smaller-scale e-science activities, and discussions of values and policies" (Clifford Lynch, 2008, <http://www.arl.org/bm~doc/rli-262-science.pdf>). An important point is that: "Rather than imagine that science librarians will have to become experts in each of these areas ... many individuals may become proficient at one or two of these newly valuable skills."