ACRL-STS Liaison Report from the 256th American Chemical Society (ACS) National Meeting and Exposition in Boston, Fall 2018

Josh Rose, ACRL-STS Liaison to ACS
September 2018

The 256th ACS National Meeting and Exposition in Boston took place from August 19th - August 23rd, 2018. The theme of the conference was *Nanoscience, Nanotechnology, and Beyond*. During the conference I attended technical division symposia put on by the Chemical Education (CHED), Chemical Information (CINF), and Computers in Chemistry (COMP) divisions. Below I have summarized a selection of talks from a CINF symposium and a CINF keynote talk, followed by two brief updates.

Technical Division Symposium

Chemical Information (CINF): "Chemistry Librarians of the Future"

Jeremy Garritano introduced the full-day, "Chemistry Librarians of the Future" symposium organized by Garritano, Dr. Vincent Scalfani, Dr. Leah McEwen - billed as an opportunity to talk about the future of chemistry librarians including what activities chemistry librarians will be doing, what skills will be needed, what partnerships might evolve, and more.

As part of the symposium, Dr. Nicholas Ruhs presented on "Reimagining chemistry librarianship: From the bench to the stacks", using experience at Florida State University Libraries to reimagine chemistry librarianship. Dr. Ruhs, who has a background in chemistry, described how everyone has some sort of domain-specific and non-domain specific knowledge that they can draw on to position the library as a research collaborator. Ruhs presented several examples of how this was being done at FSU including an ongoing partnership with FSU faculty to describe lab equipment that researchers use with unique universal identifiers for instruments and a citation format that could be included in publications and grant applications, with the goal of increasing research reproducibility of experimental results (the Research Data Alliance Persistent Identifier Interest Group was also mentioned during the question period). Other areas included research data services and assisting researchers to write data management plans, providing workshops on "Tools of the Trade" for researchers including training on ChemDraw, the Open Science Framework, SQL, and MatLab. Marketing efforts were highlighted as an important aspect of building partnerships. At FSU, these efforts include making postcards describing STEM library services, holding office hours near research centers on campus, attending departmental and faculty events (even allowing newly tenured faculty to select a book that was important to their career or experience for inclusion in the collection and displaying it at a newly tenured faculty celebration), and partnering with student groups and domain specific groups such as the ACS Chemistry chapter and the Graduate Women in Science organization.

Linda Humphries, Science Faculty Librarian at the University of Bath, provided an impromptu and informative overview of the state of science and chemistry librarianship and libraries in the UK based on experience and participation in Universities Science and Technology Librarians Group (USTLG). It is increasingly rare that institutions in the UK have a dedicated chemistry library, though a few still exist, Humphries said. Outside those, it's common for a chemistry librarian to be responsible for multiple subject areas - such as the whole of science or engineering - and there is a worrying trend of not having subject librarians at all but instead of teams organized around functional areas such as undergraduate teaching support or collections. Not very many science librarians have science degrees, which can also be problematic. Humphries described the experiences teaching chemical information skills to undergraduate and postgraduate (i.e. graduate) students at University of Bath as well as collections purchases, including research monograph e-book collections from Springer, Royal Society of Chemistry, and CRC Press. As well, online textbooks for undergraduates available on Bibliotech, a London-based startup focused on developing a modern etextbook platform, was described as a fantastic resource. In the future, developing a chemistry for non-chemists type course for new librarians without a chemistry background was seen as a worthy goal. Humphries was inspired by Judith Currano's "Chemistry for the Non-Chemist Librarian" Special Libraries Association (SLA) course, and ran a workshop in the UK - "Chemistry with Confidence" - suggesting that it would be great to have as a MOOC. Another goal described for the future is to encourage undergraduate chemistry students to consider careers in librarianship or publishing.

Judith Currano, Head of the Chemistry Library at the University of Pennsylvania (Penn), presented "Nobody else is doing it: Teaching opportunities for the chemistry librarian of the future". Currano discussed the teaching opportunities for the chemistry librarian of the future, operating on the assumption that nobody else is doing. How do you know what nobody else is doing? You observe what's happening, talk with students, talk with colleagues, note repeat questions that come up inside and outside the classroom. Currano presented four case studies of training designed and offered after determining that no one else was serving a need at the institution. These included training graduate students in research ethics at Penn, which over the last decade has developed into five hours of in-person training as well as a requirement for students to complete Citi Responsible Conduct of Research (RCR) modules. The training includes a talk by the graduate chair on ethics as a graduate student, case studies presented by Currano - with videos such as *The Lab* by the Office of Research Integrity, which prompt discussion by students and guidance from Currano. Other topics covered: plagiarism, data falsification, conflict of interest, corporate sponsorship, the sharing of research materials, bullying in the lab, and the ethical considerations of participants in the publication process. Another case study involved impact and impact metrics and having students be able to understand and define impact metrics like the CiteScore, Eigenfactor, SCImago Journal Rank

(SJR), Article Influence Score, and Scopus: Source Normalized Impact per Paper (SNIP), and be able to understand where the numbers come from, how they can be gamed, and how they can use these to determine where to publish. Groups of students complete and summarize readings on a discussion board and come to consensus about the advisability of using impact metrics to evaluate journals and authors. A third case study presented was in response to perceived need for undergraduate students to become better at information evaluation, building on previous work by Currano and Dr. Ken Foster, Biological Engineering Professor at Penn.

Dr. Ye Li, Scholarly Communications and Instruction Librarian at the Colorado School of Mines, presented a talk on "Workflows for scholarly communication and knowledge creation: Building partnership between researchers and librarians of the future". Dr. Li presented three premises:

- The future is really now. What are we getting ready for at this moment?
- Research and scholarship will become more collaborative, reproducible, impactful, diverse, interdisciplinary, and data intensive. Librarians can become partners in this information creation process - a difficult challenge but one worth pursuing.
- The liaison librarian "golden triangle" of roles - collections, reference, and instruction - can remain but the activities and content in these roles can change. For collections, this can mean instead of buying or licensing external content, librarians can help create "inside out collections" (i.e. creating unique collections from campus, hosting them in an institutional repository, and increasing access to them). Also, related to collections, librarians can get into the area of licensing software tools to support the research workflow.

Li then described five researcher learning outcomes that can guide librarians practices in the future. These included:

- Understand how data science applies to their research topics and can use proper tools, platforms, and good practices for research workflow to enable FAIR data sharing and reproducible research.
    - Readings related to this learning outcome were mentioned, including The National Academies of Sciences, Engineering, and Medicine consensus report titled *Open Science by Design: Realizing a Vision for 21st Century Research* (2018).
- Understand the diversity of research input and output and can share the variety of outputs at the point of action and comfortably interact with peers (and maybe the public), while protecting data privacy.
- Understand the global intellectual property landscape and can apply proper licenses and other legal tools to keep a balance between open science and IP protection.
- Understand how scholarly information flows across the academic boundaries and can employ their scholarly identity and other unique identifiers to enhance their research impact.

    - This can mean librarians promoting ORCID id's to researchers but it could also mean annotating other scholarly contributions that have a unique identifier, such as inChi strings for substances that a researcher synthesized, to connect and disambiguate a researcher's contributions in way that is machine-readable.
- Understand what machine learning can and cannot do with digesting information and can choose proper discovery tools based on information needs, critically evaluate the true source, and extract knowledge to move forward.
    - Databases are becoming black boxes and we no longer have the detailed documentation about their scope and search algorithms. Searches thus become very irreproducible. Librarians can either persuade discovery tools vendors to release their documentation or we can teach our researchers how to be able to search open data with script-based queries. Researchers do not need to become informaticians but they need to be able to have a conversation with emerging information systems in order to understand and modify their searches to suit their needs.

Dr. Ian Bruno, Director, Strategic Partnerships, at the Cambridge Crystallographic Data Centre (CCDC), presented on "Chemistry librarians and disciplinary data repositories working in partnership". CCDC compiles and distributes the Cambridge Structural Database (CSD), a repository of more than 950,000 experimentally determined small molecule - organic and metal-organic - crystal structures. About 80,000 datasets are deposited each year, many linked to journals but some directly deposited. Each data set is reviewed by a scientific editor to make sure it's in a state that is reusable for further research. CCDC also develops scientific software, undertakes research, and does educational outreach. The organization provides services to the community including deposition services and basic access to the structures so that one can download and reuse them, as well as free software tools and educational materials. A suite of software, called CSD enterprise, has applications for searching, analyzing, and reusing structures to apply them to scientific areas such as drug design and development. Those tools generate income to fund CCDC's core activities. CCDC tries to make these tools available to the academic community barrier free and as low cost as possible. Libraries, particularly in the U.S., have been instrumental in establishing campus-wide licenses that provide access to CSD. Dr. Bruno described examples of how CCDC has been able to reach out to researchers and make them aware of the CSD using the communication channels that libraries have. Bruno also described an initiative at the University of Cambridge, managed by the Office of Scholarly Communication, called Data Champions - volunteers from across the university's many departments who advise members of their groups on proper handling of research data. During the first wave of the program, three chemistry-aligned researchers took part. Outputs included an open data FAQs for chemists**.** Other opportunities and examples of CCDC partnering with libraries and librarians were shared. Resources mentioned during this presentation included the

Research Data Alliance - Libraries for Research Data Interest Group and r3data. Bruno was asked if CSD turns away researcher's deposits if they are not novel and a structure is already represented in CSD. The response was absolutely not because with databases like the CSD what is captured are the results of an experiment and every experiment is different and unique, even when looking at substances that have been investigated before. Polymorphisms and understanding the different forms of the same substance is very important, particularly for example, for drug manufacturing. As well, experiments may be done in different temperatures or pressures, which is all valuable information to have represented in the CSD.

Dr. Donna Wrublewski, Librarian for Chemistry & Chemical Engineering, Biology & Biological Engineering at California Institute of Technology (CalTech), presented on "Science librarians and the future of open science". Dr. Wrublewski, acknowledging co-authors Gail Clement, Head of Research Services and Librarian for Geology and Planetary Sciences, and Tom Morrell, Research Data Services Librarian, both at CalTech, asked how are open solutions working their way into the research lifecycle and how are librarians preparing to provide support using these solutions? Wrublewski described some ways that librarians at CalTech are doing this for their research community. Examples included helping researchers to maintain current awareness of their field(s) using open tools, becoming a publishing partner with a research group involved in the WormBase project, which annotates *C. elegans* genomic information, and was interested in publishing a data publication with the library as a partner, sponsoring two retreats where experts came to train library staff - one on evaluating data management plans, by Dr. Amanda Whitmire, and another by Dr. Sebastian Karcher, from Zotero, which allowed the library to discontinue its site-license to EndNote and support Zotero, and putting on Carpentry workshops for the CalTech community, and presentations and workshops for administrative staff who are supporting faculty and need to know about the NIHM deposit process in order to comply with funding requirements.

CINF Division Luncheon Keynote

Dr. Alex M. Clark, a scientist at Collaborative Drug Discovery and the founder of Molecular Materials Informatics, delivered the CINF division luncheon keynote "*Leveling Up Chemical Information for the Era of Big Data*." Dr. Clark discussed how, while cheminformatics does not deal with the scale of data that the term "big data" was coined to describe, chemical data is more inconvenient to work with and the vast majority of the chemical data that is generated is not machine readable. He argued that instead of chemists publishing their research in papers that only a tiny handful of people read and understand - citing his own work with aryl organometallic compounds as an example - in an ideal world, their research and the compounds that they generate would be published as machine readable data points that could be collected in a database that anybody could mine, extract from, and add to their models and visualizations.

Dr. Clark walked through a current, simplified, scientific process for a chemist publishing research that included performing the experiments (the hardest part), drawing the schematics, figures, and molecular structures for their results (a very time intensive process), using this information to create a human-readable, web-persistent document, which includes the introduction, methods, results, and conclusion, and then submitting this human-readable work for publication and distribution. While the publication process is entirely digital, it is still closely modelled on an early time when scientific results were printed and distributed on paper. Any machine readable data that is created is left out of the final publication. "That's a tragedy," he said, "because the amount of work that is actually put into it could so easily be repurposed to actually create something that is beautiful to humans and completely acceptable to machines."

ChEMBL, a manually curated chemical database of bioactive molecules, maintained by the European Bioinformatics Institute (EBI), and PubChem, particularly the data in PubChem submitted by the scientists themselves, were cited as examples of currently available, high-quality, machine readable chemical data.

Dr. Clark proposed that, in addition to the current efforts to make already published chemical information machine readable (i.e. "legacy data correction"), which he characterized as an important, if expensive, demanding, and monumental task, that there also be efforts to support or require publication of machine readable chemical data, thereby capping the legacy data correction problem to something finite.

Dr. Clark talked about projects related to the creation and publication of machine readable chemical data that he has worked on including Molecular Notebook, MolPress, an open source chemistry content plugin for WordPress, Assay Central, a project that attempts to amass high-quality drug target data to develop therapeutics for rare and neglected diseases that otherwise attract few resources for drug development, CDD Vault, a platform for drug discovery informatics, an IUPAC project, chaired by Dr. Leah McEwen, chemistry librarian at Cornell University Library, to create a data format for mixtures based on core InChi technology, and BioAssay Express, a web-based tool for annotating bioassay protocols using semantic web terms, to make assays machine readable.

Dr. Clark ended with a challenge for scientific publishers, who he described as have some leverage in changing the culture of science to promote machine readable data creation and publication, to include in the peer-review process a requirement for sharing machine readable data for substantive claims made in a paper.

Brief Updates

- Editors at the *Journal of the American Chemical Society (JACS)* announced that submissions of manuscripts that have been posted as preprints on ChemArxiv, arXiv, and bioRxiv will now be considered for publication.
- I met with Michael Qiu, Senior Global Library Relations Manager at ACS Publications. He is interested in and open to ideas for increasing science and technology librarians engagement with ACS and ways that ACS can encourage and support more chemistry librarians. Michael noted that ACS provides support to STS and CINF.

Contact your ACRL-STS Liaison to ACS

As the recently appointed ACRL-STS liaison to ACS, I am interested in learning how I can best serve ACRL-STS members in this role. Please feel free to contact me with questions, comments, or suggestions (jjr9@stmarys-ca.edu).