A Vision and a Necessity for the 21st Century

by Jeremy G. Frey

The growing importance of chemical information challenges IUPAC to fulfill its mission to address global issues involving the chemical sciences, given that in the modern digital world all manufacture, research, teaching and learning is now assisted by computer systems. In this article I wish to argue the case for "Digital IUPAC."

I strongly believe that IUPAC must, as a matter of urgency, take a lead in providing machine-readable (i.e. computable and understandable) representations of chemical information as well as structure, using standards that IUPAC defines and standards that other international authorities agree to use.

Chemical information is a uniquely important area with considerable commercial and financial significance. Traditional providers of chemical information include the major chemistry societies, e.g. Chemical Abstracts (ACS), major information companies such as Accelrys, ACD/Labs, PerkinElmer, Thomson Reuters, and others, as well as the information provided by the chemical suppliers. New digital providers such as ChemSpider (RSC), with an agenda that is at least partially open now, complement these providers.

The chemical information business is large and significant for both industry and academia. It is also international. The public resources devoted to information provision in the bioscience areas, and the resulting degree of international cooperation and agreement, makes chemical information provision look much less significant than it really is.

Traditionally, major parts of chemical information services have revolved around chemical structure representation, with a great deal of the research in this area being undertaken by the commercial players. In the IUPAC arena the digital future for structure representation is well provided for by Chemical Nomenclature and Structure Representation Division (Division 8) forefront research, and by the development of Chemical Nomenclature and Structure Representation based on the IUPAC International Chemical Identifier (InChI), funded to a considerable degree by the National Institute of Standards and Technology (NIST).

Looking at the wider data and information agenda, the Royal Society report "Science as an open enterprise" argues the absolute necessity for intelligent access to the data on which scientific conclusions are based. Intelligent openness is fundamental to the whole progress of science. In the modern digital world intelligent access really requires that this access can be mediated by computers. For computers to "understand" the information and to be able to manipulate it to deliver rational results from complex searches, the data, the quantities, the concepts, and the relationships between these items all have to be encoded in a machine-readable manner.

With the rapid increase in the use of the web to disseminate and manipulate chemical knowledge, there is a major requirement for clarity in both the computer-interpretable and the human-interpretable meaning of the concepts of chemical terminology. The web interface to IUPAC must not and cannot be regarded as solely for human consumption. The IUPAC identifier www.iupac.org must lead to computer-readable resources that provide computer systems with the context that the trained human takes for granted but which the computer systems can only deploy if it is provided specifically.

I believe there has been a disturbing lack of recognition of the immense value and importance of www.iupac.org as the necessary and authoritative root domain name for all computationally accessible representations of chemical information.

The comprehensive conversion of IUPAC's knowledge base of standards and definitions from human-readable to computer-readable form is essential. It is vital that this conversion be done now as a matter of extreme urgency, if IUPAC is to maintain its role as the international authority for the chemical sciences. If computers cannot find and use the information provided by IUPAC, that information will effectively cease to exist for the "Wikipedia generation." If we do not do undertake this digital conversion now, others will do it, potentially poorly and without international agreement: IUPAC will be sidelined.

It is essential that IUPAC becomes fully engaged with developments in the representation of chemical concepts and relationships in a computer-readable and actionable form, namely the ontologies and the semantics for chemistry on the Web. IUPAC runs a

Goodbye CPEP, Hello CPCDS

In August 2013, the IUPAC Council was presented with a proposal for modifying the Committee on Printed and Electronic Publications (CPEP). The proposal was put forward by CPEP chair, David Martinsen. The two motions proposed were passed and starting in January 2014, IUPAC includes a Standing Committee on Publications and Cheminformatics Data Standards (CPCDS), whose chair is a member of the Bureau. On 1 January 2014, Bonnie Lawlor, Executive Director of the US National Federation of Abstracting and Information Services (NFAIS), succeeded David Martinsen as Chair, becoming the 1st CPCDS chair.

major risk that the standards for the representation of chemical information (other than structures) will be set by outside groups from other disciplines, without adequate consultation across the whole of the diverse areas of the chemical sciences.

The hugely successful IUPAC work on InChI, ThermoML, and the work with Wikipedia on validated chemical information (driven by the Polymer Division and the Committee on Chemistry and Industry (COCI) and others) and the XML driven version of the Gold Book, show what IUPAC can achieve; this is only the start of what is needed. The technology is understood; we now need the will to move IUPAC forward to "Digital IUPAC." Common standards are essential: they lower the overall cost, increase access, lower barriers to entry, and increase exploitation. IUPAC must be at the forefront of the cause of common agreed standards for machine-readable information.

IUPAC has an opportunity to play a major role, working with both the academic, industrial, and governmental communities, in setting the standards that will enable chemical information provision to move fully into the digital present, let alone the digital future. If IUPAC does not undertake this transformation now, others will. IUPAC will become a minor player, and wither. If it grasps this future, IUPAC's impact will grow rapidly as its standards become embedded in the core digital infrastructure used by industry, commerce, and research in all of the chemical sciences.

References & bibliography

Semantic Web

Feigenbaum L, Herman I, Hongsermeier T, Neumann E, Stephens S. The Semantic Web in Action. *Scientific American*, 2007, 297:90-97

Frey JG. The value of the Semantic Web in the laboratory, *Drug Discovery Today*, 2009, 14:552-61, http://dx.doi.org/10.1016/j.drudis.2009.03.007

Chemical Information

Bird CL and Frey JG, Chemical information matters: an e-Research perspective on information and data sharing in the chemical sciences *Chem. Soc. Rev.*, 2013, 42:6754-6776, http://dx.doi.org/10.1039/C3CS60050E

Bird CL, Willoughby C, and Frey JG, Laboratory notebooks in the digital era: The role of ELNs in record keeping for chemistry and other sciences, *Chem. Soc. Rev.*, 2013, 42:8157-8175, http://dx.doi.org/10.1039/c3cs60122f

IUPAC

InChI: http://www.iupac.org/inchi

IUPAC Gold Book: http://goldbook.iupac.org and http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Chemistry/Gold_Book_workgroup

Extension of ThermoML: The IUPAC standard for thermodynamic data communications (IUPAC Recommendations 2011) Michael Frenkel, et al, *Pure Appl. Chem.*, Vol. 83, No. 10, pp. 1937–1969, 2011. http://dx.doi.org/10.1351/PAC-REC-11-05-01, and http://www.iupac.org/namespaces/ThermoML/

Researchers and Data

The Fourth Paradigm: Data-Intensive Scientific Discovery, http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Royal Society report on "Science as an Open Enterprise", http://royalsociety.org/policy/projects/science-public-enterprise/report/

Collaborative yet independent: Information practices in the physical sciences, http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/physical-sciences-case-studies-use-and-discovery-

Researchers of Tomorrow: the research behaviour of Generation Y doctoral students, JISC, 2012, http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow,

Clive Cookson, Dealing with the data deluge, Financial Times, 1 Feb 2013, http://www.ft.com/cms/s/2/ff439394-6a73-11e2-a7d2-00144feab49a.html#axzz2g5EAeQFG

Matthew P. Long and Roger C. Schonfeld, Supporting the Changing Research Practices of Chemists -- Research Support Services: Chemistry Project (2013), http://www.sr.ithaka.org/research-publications/supporting-changing-research-practices-chemists

Acknowledgements

I would like to thank Colin Bird and Fabienne Meyers for their help in improving this article. My views have been formed both as a practicing chemist and information researcher and have been brought into focus over the last decade with work funded by the RCUK e-Science programme (EPSRC grant GR/R67729, EP/C008863, EP/E502997, EP/G026238, BBSRC BB/D00652X), the EPSRC National Crystallography Service (Tender RCUK /D/EPSRC/Facilities/XRC/10), the HEFCE and JISC Data Management Programme and the University Modernisation (UMF), and most recently the RCUK Digital Economy Theme as part of the IT as a Utility Network+ funding (EPSRC EP/K003569). These views could not have been honed without con

(continued on page 18)

PhosAgro/UNESCO/IUPAC Research Grants in Green Chemistry – Call for Applications

Research in green chemistry and allied areas in biochemistry, geochemistry, biotechnology, ecology and healthcare give young scientists ample opportunity to demonstrate their inventiveness and provide important input to sustainable development. With this in mind, UNESCO's International Basic Sciences Programme (IBSP) and PhosAgro, the largest producer of phosphate-based fertilizer in Europe, in close cooperation with IUPAC, launched the Green Chemistry for Life Project in 2013.

Over the course of five years, the project will offer research grants of up to USD 30000 to scientists aged 35 years or less with an innovative research project that respects the 12 principles of green chemistry to help them implement their project. In addition to seeking to harness talents of young scientists for the advancement of green chemistry and the use of its fruits, the Project sets out to raise awareness—among decision- and policy-makers, industrialists, and the public at large—of the vast opportunities green chemistry offers to meet pressing societal needs.

1st call for applications

The deadline for applications for the first round of research grants for young scientists is **28 February 2014**.

The PhosAgro/UNESCO/IUPAC Partnership in Green Chemistry for Life aims to generate and apply new scientific knowledge in green chemistry through the promotion of the activity in this area of young scientists. It is also to reinforce the research capacities of the participating institutes and cooperation between them.

For more information see:
www.unesco.org/new/en/natural-sciences/science-technology/basic-sciences/chemistry/green-chemistry-for-life/