

# The Project Place

Information about new, current, and complete IUPAC projects and related initiatives. See also [www.iupac.org/projects](http://www.iupac.org/projects)

## Herding AnIMLs (no, it's not a spelling mistake): Update on the IUPAC and ASTM Collaboration on Analytical Data Standards

by Tony Davies

Working on international standardization projects is difficult. Working on international standardization projects where two standardization bodies with their own rules, guidelines, and working practices are collaborating together is doubly difficult. The recent 2007 IUPAC General Assembly (GA) in Torino, Italy, gave the IUPAC CPEP Subcommittee on Electronic Data Standards (SEDS) the occasion to invite their ASTM partners in the Analytical Information Markup Language (AnIML) project to participate in a joint meeting. With both groups well represented by their respective officers and interested parties, some very important decisions were made.

During the 2001 GA in Brisbane, the clear need for IUPAC to establish itself as the international standardization body for chemistry in the digital age was documented and addressed with the initiation of the XML in Chemistry initiative. Several projects have arisen out of this decision, including the newly available XML versions of the Gold Book online and the successful completion of the XML Data Standard for Thermochemical Information ThermoML. Those involved with the third major project—the joint development of the Analytical Information Markup Language (AnIML) with the ASTM International Subcommittee E13.15—have recently made strong progress following a period in which a lot of background work was necessary.<sup>1</sup>

### Why Create an Analytical Information Markup Language?

Historically, the availability of “rival” standards in the field of analytical instrumentation has only served to confuse users and vendors alike. The partial completion of the old Analytical Instrument Association netCDF-based binary standards for mass spectrometry and chromatography, and their abandonment of work on an infrared spectroscopy standard ended a period of conflicting and misleading presentations at international meetings around the world.<sup>2</sup> This had hampered work on the IUPAC JCAMP-DX ASCII standards and led to a dispersion of scarce talent between the two standardization activities.<sup>3</sup> When the SEDS subcommittee realized that both their experts and those of the ASTM were working simultaneously on a new XML-based data exchange standard for analytical



data, it was clear that the previous situation had to be avoided at all costs.<sup>1</sup>

The successful completion of such an exercise, with the widespread deployment of well-documented internationally recognized standard data formats, will constitute clear advantages. Users will benefit from the greatly simplified long-term storage and retrieval of analytical data, and the more advanced techniques of data mining and knowledge generation. Proving conformity to regulatory compliance demands record retention, which is far simpler if data are in a vendor-neutral standard format. Practically, exporting data from a computer associated with a specific instrument in the laboratory to your office desktop, where often further data analysis and reporting is expected to take place, will be made much simpler. With the increasing use of process analytical technologies and design for manufacturing strategies in the pharmaceutical sector, often data from many diverse instrument types and vendors have to be brought together in a single analysis package where the actual “results” will be computed.<sup>4</sup>

Instrument vendors and third-party software houses have also realized major cost savings by adopting the standard formats, not only for their own internal use between systems, which often originate from different software development groups on different continents, but also so they can claim their products are “compliant ready.” Nowadays, a vendor’s products will not sell if it cannot clearly demonstrate how they will smoothly integrate into the existing laboratory IT environment and workflows of their customers. Stand-alone island solutions are a thing of the past. Often major contracts are awarded for software solutions that better integrate with existing IT infrastructures as the return on investment is easier to prove to senior management. The deployment of standard formats is a major contributor to such advanced integration.

## The Project Place

### Recent Progress

Much work has already been achieved by bringing IT specialists up to speed on the complexity of analytical chemistry data types and formats. This process sometimes yields interesting results in which multidimensional data types, such as liquid chromatography UV/mass spectrometry hyphenated experiments, need to be stored. Conversely, the chemistry experts have had a lot to learn about the intricacies and capabilities

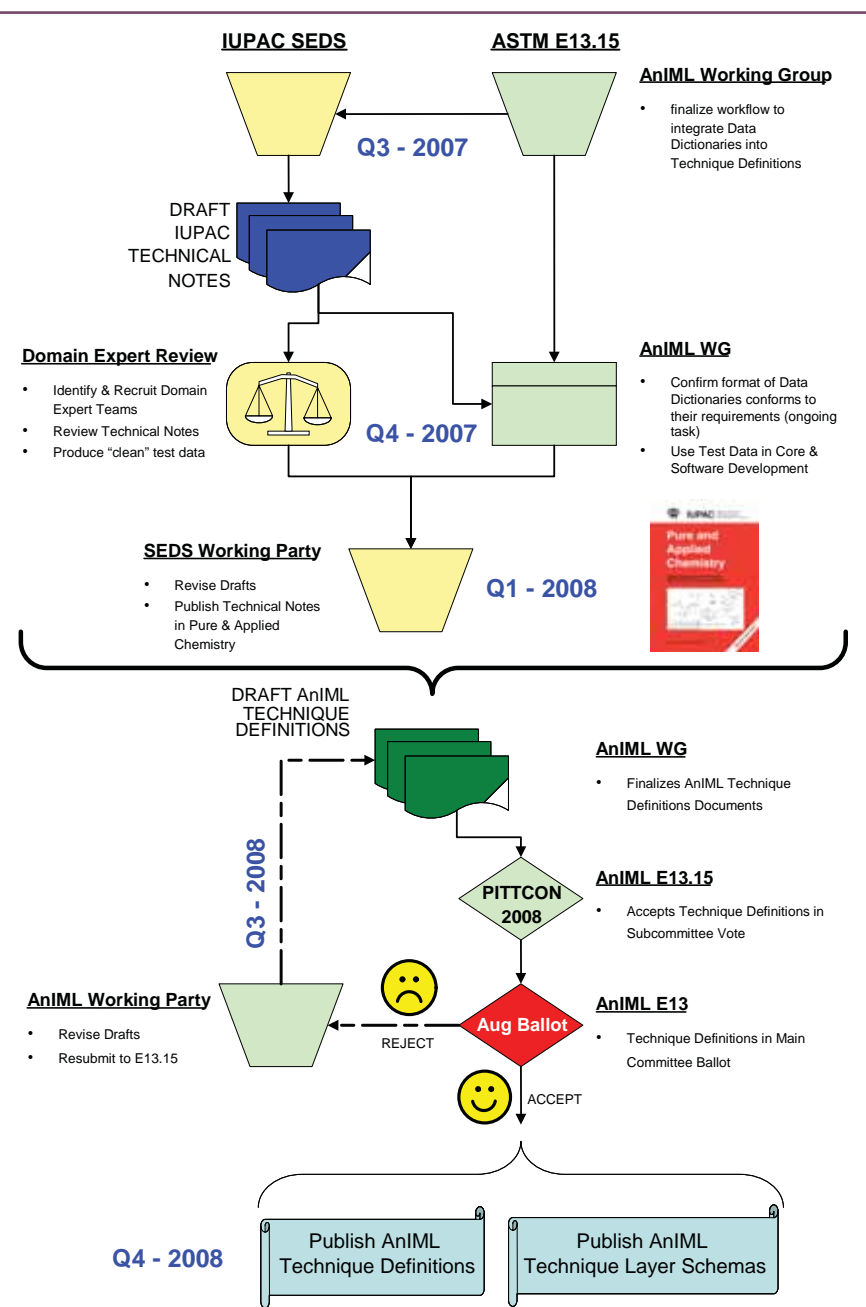
of the XML language, including the recent introduction of naming standards. Fortunately, some vendors and users have been gaining experience in converting large volumes of data from very different historical legacy analytical systems into vendor-neutral XML files.

In order to accommodate the different demands of a very diverse user group, difficult decisions have been made. For example, the expectations involving

audit trails and electronic signature capture that a user working in a fully regulated pharmaceutical industry company might have can be met, but significantly increase the complexity of the structure of the AnIML file.

Even though prototype systems have been available since 2004 that support early alpha versions of the AnIML standard and despite numerous lectures and seminars held around the world, not much has actually been published or finalized. This is one of two major issues that we successfully addressed in Torino. The second issue involves completing the streamlining and clarification of the competencies of the various members of this project. The plan is for ASTM to concentrate on the technical aspects of the new standard and for IUPAC to standardize the terms and data dictionaries.

### Overview of workflow to finalize AnIML standards.



### Next Steps for Completing the Standards

There has been some confusion and criticism from some task group members over the fact that the requirements document, generated relatively early in the project, has remained an internal document. It was agreed at this meeting that the document should be published as a IUPAC technical report since it essentially sets the goals and boundaries for the AnIML format and is a key document for any new vendor or user coming across AnIML for the first time. The document had been reviewed by the authors, both prior to and during the meeting. Its adoption will appear on the agenda of

## The Project Place

the next full task group meeting after which it will be submitted for publication.

As agreement on the data dictionaries is an essential precursor to the finalization of any of the standards, the initial versions will draw extensively, and almost exclusively, on the IUPAC/JCAMP-DX and ASTM ANDI standards that have already been published. Bearing this in mind, it was agreed to use the domain or technique-specific knowledge available within IUPAC to draft technical notes for the data dictionaries in chromatography, mass spectrometry, infrared spectrometry, nuclear magnetic resonance spectrometry, and the other so-called phase-one techniques (NIST will document the UV-Vis Data Dictionary). These drafts will be completed and made available during the rest of 2007. The drafts will be reviewed by technical experts and provided to the development teams who will integrate them into the IT technical documentation of the standard to identify any issues that need to be resolved before publication.

Formal adoption of the AnIML standards themselves will follow the ASTM process during 2008, and with tentatively a formal adoption planned to start at the ASTM E13.15 Business Meeting during the Pittcon Conference in spring 2008, with formal ASTM adoption at the end of 2008 provided no members veto the adoption.

### Appeal

As you can see, as with all such standards development processes, we are extremely reliant on volunteers from the scientific community. If you feel like you can contribute, we would very much encourage you to come forward, even if it is only as a reviewer of the Technical Notes as they are made available.

Participants at the Torino meeting were Mohan Cashyap (GlaxoSmithKline, Ware, UK), Tony Davies (ALIS, Analytical Laboratory Informatics Solutions Ltd, Dortmund, Germany, and SEDS chair; <antony.n.davies@gmail.com>), Maren Fiege (Waters Informatics, Frechen, Germany, and ASTM E13.15 member), Gary Kramer (NIST, Gaithersburg, USA, and ASTM E13.15 subcommittee chair), Peter Lampen (c/o ISAS, Institute for Analytical Science, Dortmund, Germany, and SEDS secretary), Robert Lancashire (University of the West Indies, Kingston, Jamaica, and CPEP and SEDS member), and Dave Martinsen (ACS, and ASTM E13.15 secretary). We would like to thank Ben Mezoudj of Adobe in Germany for ensuring that the Adobe Connect eConferencing facility was working and available for the meeting.

### References

1. R.J. Lancashire and A.N. Davies, "The Quest for A Universal Spectroscopic Data Format," *Chem. Int.*, 28(1) 10-12, 2006 <[www.iupac.org/publications/ci/2006/2801/3\\_lancashire.html](http://www.iupac.org/publications/ci/2006/2801/3_lancashire.html)>.
2. A.N. Davies, "Data Transfer Standards—The Unidata netCDF Standard," *Spectroscopy Europe*, 4(5), 36-39, 1992. (The ASTM standards and associated guides "Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data" and E2077-00 "Standard Specification for Analytical Data Interchange Protocol for Mass Spectrometric Data" are available from the ASTM website for a fee. To locate them, search the Standards page with the keyword netCDF <[www.astm.org](http://www.astm.org)>.)
3. R.S. McDonald and P.A. Wilks Jr., "JCAMP-DX: A Standard Form for the Exchange of Infrared Spectra in Computer Readable Form," *Applied Spectroscopy*, 42(1), 151-162, 1988. (This was the first JCAMP-DX standard to be published. All of the JCAMP-DX standards are available from <[www.iupac.org/jcamp](http://www.iupac.org/jcamp)>.)
4. P. Lampen and A.N. Davies, "JCAMP-DX to ORIGIN Utility Tools for Making Spectra Available to Chemometricians," *Spectroscopy Europe*, 16(5), 28-30, 2004 <[www.spectroscopyeurope.com/td\\_col.html](http://www.spectroscopyeurope.com/td_col.html)>.

 [www.iupac.org/standing/cpep/wp\\_jcamp\\_dx.html](http://www.iupac.org/standing/cpep/wp_jcamp_dx.html)

## Multiple Uses of Chemicals: Clear Choices or Dodgy Deals?

Chemical processes determine much of who we are. Chemistry in the brain provides our sense of being and helps determine how we act. The oxygen we breathe is created by the chemical processes in plants, and the consumer products we buy are the result of chemical processes. Chemistry is central to life on earth, and many major global issues today require chemical solutions, including reducing pollution of the earth's atmosphere, providing clean water, improving food supplies, and developing new medicines. Chemistry thus has enormous potential to contribute positively to global well being.

But chemistry has not always been used for the common good. Ninety years ago, during World War I, chemists and chemical engineers used their knowledge to perfect weapons that were based on the physical and toxic properties of chemicals. The use of chlorine, phosgene, and mustard gas in WWI resulted in some 1.3 million casualties, and it is now estimated that 90 000 died soon after exposure. Many others died years later from lung injuries sustained on the battlefield after inhaling the various chemical warfare gases and aerosols.