TONY DAVIES COLUMN

How standard are your standards?

Antony N. Davies[a,b] and Robert Lancashire[c]
[a]Strategic Research Group – Measurement and Analytical Science, Akzo Nobel Chemicals b.V., Deventer, the Netherlands
[b]SERC, Sustainable Environment Research Centre, Faculty of Computing, Engineering and Science, University of South Wales, UK
[c]The Department of Chemistry, The University of the West Indies, Mona, Kgn 7, Jamaica

Berlin, February 2017… and high level representatives of a number of consortia are together on the podium during the SmartLab Exchange conference discussing the latest developments in standards, which include initiatives in some spectroscopic fields. In the last ten years or so, several well-funded initiatives, strongly supported by the pharma companies, have formed consortia to deliver to their members utilities and "standards". Of course, this column greatly welcome initiatives towards more standardisation,. However, as again shown in Berlin, the use of the term "standard" itself is being widely used as a marketing tool and risks delivering the wrong message to those hearing presentations around these initiatives for the first-time.

International standards bodies

One of the very first bodies to recognise and adopt standardisation in the support of international trade was the International Electrotechnical Commission (IEC). I like the definitions they use:

"A standard is a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context."

"International Standard"

"An International Standard is a standard adopted by an international standards organization and made available to the public." The definition given in all IEC standards reads: "A normative document, developed according to consensus procedures, which has been approved by the IEC National Committee members of the responsible committee in accordance with Part 1 of the ISO/IEC Directives."

The IEC is a not-for-profit, quasi-governmental organisation, founded in 1906, whose members are National Committees, and they appoint experts and delegates from industry, government bodies, associations and academia to participate in the technical and conformity assessment work of the IEC.

ISO, the International Organization for Standardization, was the result of an international meeting in 1946 when delegates from 25 countries met at the Institute of Civil Engineers in London. They decided to create an international organisation "to facilitate the international coordination and unification of industrial standards". With remarkable speed, by 23 February 1947, the new organisation was ready and ISO began operations. Today, the ISO Central Secretariat is based in Geneva, Switzerland.

They cover many of the areas outside of the remit of the IEC and nicely improve on the IEC's rather oblique *"…optimum degree of order in a given context"* to something everyone can identify and understand "fit for purpose".

"A standard is a document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose."

Now these organisations were established because of a similar perceived need and threat to international trade and development.

The International Union of Pure an Applied Chemistry goes back even further. IUPAC was formed in 1919 by chemists from industry and academia, who recognised the need for international standardisation in chemistry. The Union was formed to handle standardisation of weights, measures, names and symbols and is essential to the "well-being and continued success of the scientific enterprise and to the smooth development and growth of international trade and commerce".

The International Association of Chemical Societies (IACS) had met in Paris in 1911 and produced a set of proposals for the work that the new Association should address, including:[3]

- Nomenclature of inorganic and organic chemistry;
- Standardisation of atomic weights;
- Standardisation of physical constants;
- Editing tables of properties of matter;
- Establishing a commission for the review of work;
- Standardisation of the formats of publications;
- Measures required to prevent repetition of the same papers.

In modern times this work of these international standards bodies has necessarily evolved and extended into the digital domain as the delivery of the scientific content has moved into this environment.

TONY DAVIES COLUMN

New industrial standardisation efforts

The following bodies presented initiatives at SmartLab. Their initiatives will impact spectroscopic data handling and the presenters were brave enough to sit for a panel discussion and subsequent Question and Answer session.

SiLA: "Standardization in Lab Automation"

Founded in 2008, driven by Roche, Novartis and Actelion, SiLA is a non-profit organisation whose documentation is only available to members. Downloads of their documents require registration as a "Personal Member" or an upgrade to "Corporate Member". On the content front, the ontologies and taxonomies are agreed amongst members.[4]

SiLA is based on HTTP/2, the successor of the Internet standard HTTP, which is likely to exist for decades as well (HTTP is from 1999). HTTP/2 is an Internet Engineering Taskforce (IETF) standard.

- 2009: Project-based internal implementations of the standard. Release of SiLA Specifications V1.0
- 2010: 1st public implementations of device interface standard. Release of SiLA Specifications V1.1
- 2011: Started to evaluate existing data standards; shared feedback with AnIML
- 2012: SiLA began working on data standard. Release of SiLA Specifications V1.2
- 2013: PoC of updated AnIML standard. Release of SiLA Specifications V1.3
- Oct 2016: SiLA 2 roadmap officially announced (go-live planned for mid-2017)

Proteomics Standards Initiative

HUPO, the Human Proteomics Organisation, develops data format standards for proteomics, looking at both data representation and annotation standards. They aim to involve data producers, database providers, software producers and publishers etc. Very much aimed at driving the public deposition of analytical data in this field. Main activities are:

- Formats: usually an XML schema (but also tab-delimited files)
- Controlled vocabularies—currently around 2600 terms which are usually an Open Biomedical Ontologies OBO-style, hierarchical controlled vocabulary precisely defining the metadata that are encoded in the formats.
  - Minimum information (MIAPE) specifications: format-independent specification of minimum information guidelines.
  - Databases and tools: software implementations to make the standards truly useful.
  - Community interaction to ensure deposition of data in public repositories.

As they are focussed on mass spectrometry-based proteomics, they have developed the following PSI Standard File Formats for MS:[5]

- mzML (MS data)
- mzIdentML (Identification)
- mzQuantML (Quantitation)
- mzTab (Final Results)
- TraML(SRM, Selected Reaction Monitoring)

This organisation intends in the next five years to focus on improving adoption/support, especially in vendors' software, to move on to handle mass spectrometry metabolomics data. They wish to finalise compatible formats with genomics data, e.g. proBed and proBAM (applicable for proteogenomics studies) and start working with the structural biology community (since MS proteomics is being increasingly used in that context).

ASTM Subcommittee E13.15 on Analytical Data

XML standardisation effort started in this form in 2003.[6] In recent years, it has been driven hard and championed by the very patient Burkhard Schaefer. Recent publications are available in chromatography and mass spectrometry.[7]

- E1947-98(2014) Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data
- E1948-98(2014) Standard Guide for Analytical Data Interchange Protocol for Chromatographic Data
- E2077-00(2016) Standard Specification for Analytical Data Interchange Protocol for Mass Spectrometric Data
- E2078-00(2016) Standard Guide for Analytical Data Interchange Protocol for Mass Spectrometric Data

Allotrope example: semantics provides common meaning

In 2012, the Allotrope Foundation was launched funded by the member companies, mainly from the pharma industries, through annual subscription. The governance of Allotrope Foundation is achieved via consensus within the Foundation, and is administered by a Board of Directors comprised of two individuals from each member company.[8]

- 2012: Allotrope launched, scope and strategy defined
- 2013: Initiate software development, evaluation of existing standards
- 2014: Feasibility studies and POCs, ADF design, testing and due diligence
- 2015: API and taxonomy development, V1.0 released internally, first deployments inside member companies
- 2016: ADF/API updates, V1.1 released internally (March 2016), API testing, Minimum Viable Product (MVP) defined, V1.2 release for internal beta testing (Nov 2016)

In the field of standards development (document standards, metadata and test data), they evaluate existing data standards and define appropriate controlled vocabularies and ontologies to use with the Allotrope Framework. They intend to provide within their membership group test datasets for use in development and map the metadata associated with the test data to existing standard definitions.

The Allotrope Foundation has a number of V1 taxonomies under development. The member companies take on the initial deployments of individual technical areas. These include gas chromatography, Karl Fischer, liquid chromatography, mass spectrometry, nuclear magnetic resonance spectroscopy, thermogravimetric analysis, ultraviolet spectroscopy, capillary electrophoresis, cell counter, cell culture analyser, blood gas analysis, balance and pH.

continued on page 25

continued from page 17

The first release of a technique will be an ADF format file for HPLC-UV detection planned as below:

- Allotrope Data Format (ADF) Instance Data (end Q2 2017)
- Allotrope Foundation Ontologies (AFO) Classes and Properties (end Q3 2017)
- Allotrope Data Models (ADM) Constraints (end Q4 2017)

Conclusions

As it is clear to see, there is currently a number of overlapping well-funded alliances driving forward "standardisation" in their own interest areas. Unfortunately, much of the development paperwork is only visible to paying members or by paying for the documentation.

This issue was highlighted during a very good question and answer session at Smartlab[9] which included some pretty strong statements around the lessons learnt from the failure of previous lab connectivity initiatives. This was highlighted as being due to the closed nature of those developments. The general conclusion was that in order for any standards initiative to succeed they should be "…extremely open!" It is well worth repeating that we welcome the strength of these initiatives but it is clear that if these are not to end up in the graveyard of failed "standards" more care needs to be taken that the recognised standardisation bodies are involved to ensure that multiple different uses of the same controlled terminology take place and that movement between these different formats is ensured.

This is not new—the longevity of the Adobe PDF standard—initially defended and a company confidential format has been ensured by its documentation and adoption as an ISO standard—with Adobe relinquishing control and rights to the standard.[10]

References

1. International Electrotechnical Commission (IEC), www.iec.ch/index.htm
2. International Organization for Standardization (ISO), www.iso.org
3. International Union of Pure and Applied Chemistry (IUPAC), iupac.org/who-we-are/our-history/
4. SiLA—Standardization in Lab Automation, www.sila-standard.org
5. G. Mayer, L. Montecchi-Palazzi, D. Ovelleiro, A.R. Jones, P.A. Binz, E.W. Deutsch, M. Chambers, M. Kallhardt, F. Levander, J. Shofstahl, S. Orchard, J.A. Vizcaíno, H. Hermjakob, C. Stephan, H.E. Meyer and M. Eisenacher, "The HUPO proteomics standards initiative—mass spectrometry controlled vocabulary", *Database (Oxford)* 2013 bat009 (2013). doi: dx.doi.org/10.1093/database/bat009
6. T. Davies, P. Lampen, M. Fiege, T. Richter and T. Frohlich, "AnIMLs in the spectroscopic laboratory?", *Spectrosc. Europe* **15(5),** 25–28 (2003). bit.ly/2lbWpRz
7. ASTM Subcommittee E13.15 on Analytical Data, www.astm.org/COMMIT/SUBCOMMIT/E1315.htm
8. Allotrope Foundation, www.allotrope.org
9. Smartlab Exchange Europe 2017, smartlabexchangeeurope.iqpc.co.uk/
10. ISO 32000-1, Document Management—Portable Document Format – Part 1: PDF 1.7. www.iso.org/iso/news.htm?refid=Ref1141