AnIMLs in the spectroscopic laboratory?

Tony Davies,[a] Peter Lampen,[b] Maren Fiege,[c] Thorsten Richter,[c] Thorsten Fröhlich[c]

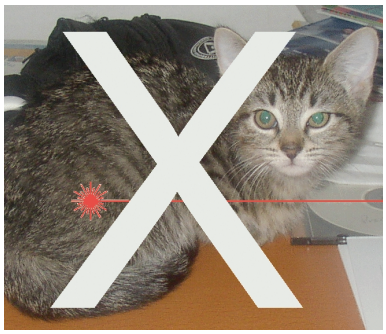
[a]External Professor, University of Glamorgan, UK. c/o Creon Lab Control AG, Europaallee 27-29 50226 Frechen, Germany

[b]ISAS, institute for Spectrochemistry and Applied Spectroscopy, Dortmund, Germany

[c]Creon Lab Control AG, Europaallee 27-29 50226 Frechen, Germany

No, we have not got a typing error in the title!

Nor are we reviewing lasers measuring the colour of kittens.



In this edition, we will look into efforts that have been gaining momentum this year to develop a new all-embracing data standard for analytical data based on XML. At the start of 2002, we reported on the decision of the IUPAC to play a leading role in the development of terminology for XML data standards.[1]

As we pointed out last year, it is not the use of XML itself that is interesting or novel, but the content stored within the XML files. In order to support standardisation in this field for the benefit of the community, the International Union of Pure and Applied Chemistry has decided to take an active role in helping to unify the various dictionaries and publicise their availability.

Since then, Dr Steve Stein at NIST has received funding to work on the ontologies already available within IUPAC (see project 2002-022-1-024 Standard XML data dictionaries for chemistry http://www.iupac.org/projects/2002/2002-022-1-024.html).[2]

In addition, the first of what is expected to be a range of new XML-based projects has just been given the go-ahead (2002-055-3-024 - XML-based IUPAC standard for experimental and critically evaluated thermodynamic property data storage and capture).[3]

Abbreviations

XML: eXtensible Markup Language

IUPAC: International Union of Pure and Applied Chemistry

NIST: National Institute of Standards and Technology, USA

LECIS: Laboratory Equipment Control Interface Specification

ANDI: Analytical Data Interchange protocols

IUCr: International Union of Crystallography

Collaboration with other organisations

It is vital that this IUPAC role extends to collaboration with other organisations active in the field of standards definitions and this is exemplified by the work being carried out on a radical new standard format for analytical data together with the ASTM in subcommittee E13.15.[4] Committee E13 is the ASTM committee responsible for Molecular Spectroscopy and Chromatography. E13.15 is the Analytical Data Management subcommittee.

If you follow Reference 4, you will see that this subcommittee has inherited responsibility for the old binary netCDF based ANDI standards from the Analytical Instrument Association in chromatography and mass spectrometry, as well as the LECIS standard for instrument connection. Unfortunately, there is no relevant information on the ASTM web site, but the work is being carried out (uniquely for ASTM) in the public domain through the use of the SourceForge.net organisation. SourceForge.net is "the world's largest Open Source software development website, with the largest repository of Open Source code and applications available on the Internet". SourceForge.net provides free services to Open Source developers and documents, drafts etc. for discussion are being placed in the public domain under the SourceForge AnIML project.[5] Do not be put off by the apparent lack of activity on the statistics page of the SourceForge project, as all information is being run from the homepage and not from the document upload areas. The project homepage can be found at http://animl.sourceforge.net/. Anyway, following this link, you will let the kitten out of the lab and see that AnIML stands for Analytical Information Mark-up Language. There is a mass of further information here as well.

AnIML file structure

In order to help push things along, the authors have been meeting in Europe to prepare proposals to the full ASTM E13.15 working meeting.

On an abstract level, the AnIML file has an onion-like shell structure (Figure 1). The more generic the content, the closer to the centre of the onion the information is to be found. The raw data is to be found in the Core, and we

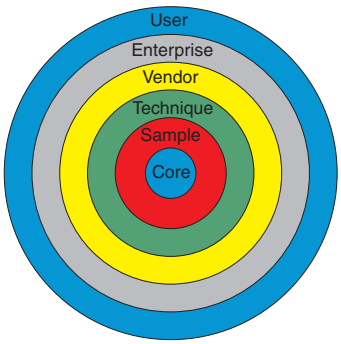


Figure 1. The Basic Shell Concept of the AnIML XML file.

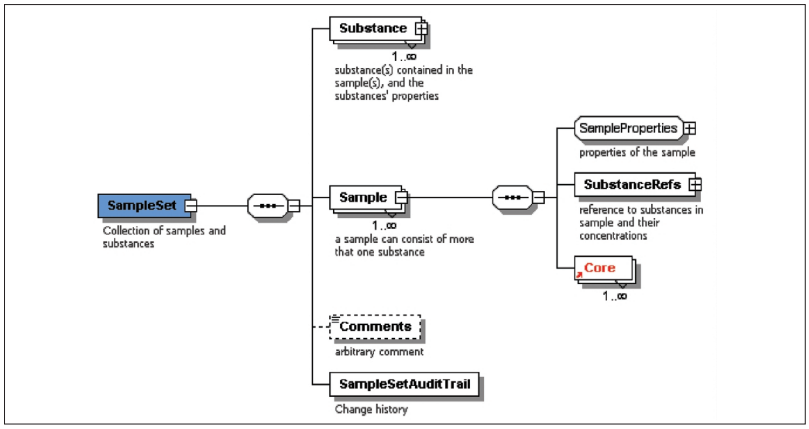


Figure 2. The AnIML Sample Layer proposal.

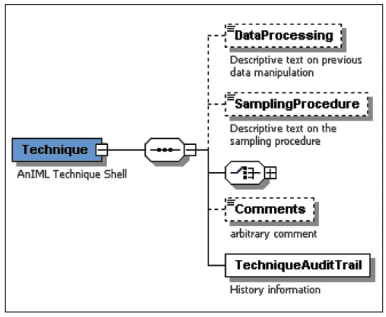


Figure 3. The Technique Layer basis.

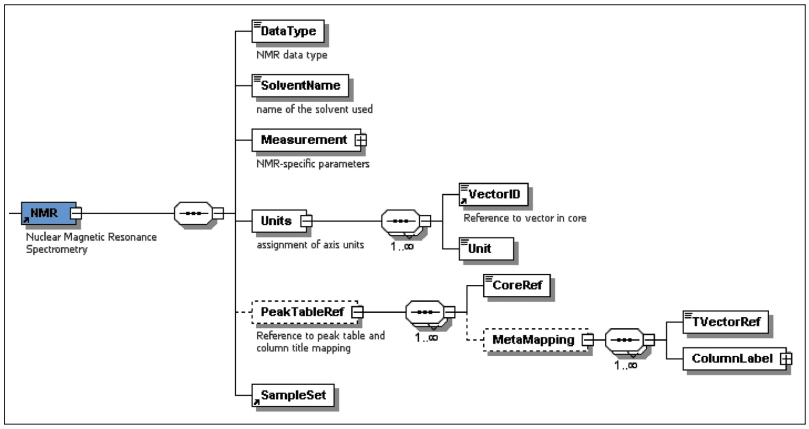
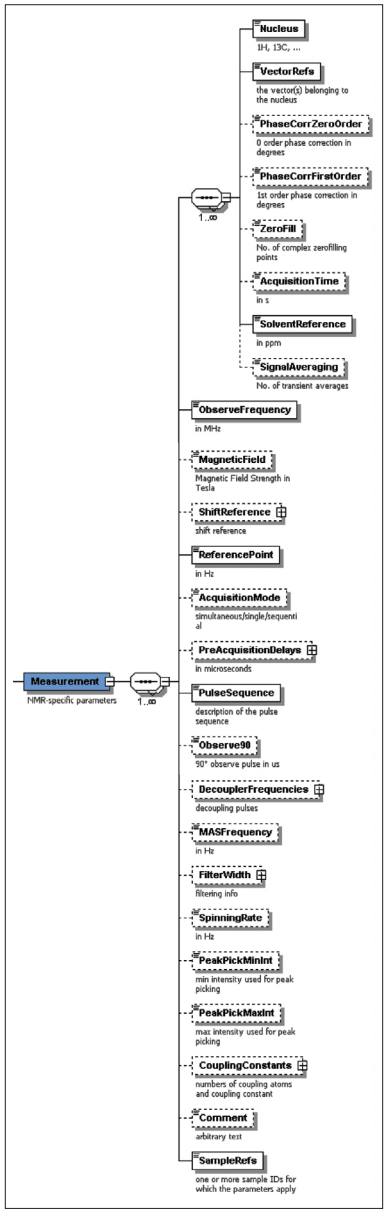


Figure 4. Some typical details of a Technique Layer, in this case NMR. Continued down right-hand column.



have proposed that an application should be capable of displaying all data found in the Core even if it has never heard of the analytical technique used for the measurement. The next layer, Sample, was added at the recent meeting in Philadelphia at the ASTM headquarters, as it is possible to have many data sets from a single sample (see Figure 2).

The Technique Layer contains information on the particular analytical technique used to provide the information, such as Infrared Spectroscopy, Mass Spectrometry etc. Here we have proposed the use of the ontologies already widely accepted in the JCAMP-DX range of standards and in the ANDI netCDF protocols (see Figures 3 and 4).

Beyond the Technique Layer, information will be stored which is only relevant to a particular vendor or instrument. Information that is specific to the measuring instrument is stored in the Vendor Layer; the Enterprise and User Layers contain information such as the company internal batch data, local working rules and any extra information that the operator wishes to add.

Don't re-invent the wheel!

This was the second point on the IUPAC "don'ts" list and is followed in this project by mapping the IUPAC JCAMP-DX and ANDI ontologies into the XML schema. Use of the JCAMP-DX data dictionaries as a starting point for the development of the technique-specific AnIML layers for IR, NMR, MS and IMS data will ensure conversion between JCAMP-DX and AnIML files in a well-defined way.

Table 1.

JCAMP-DX Label	AnIML Shell	AnIML Technique	AnIML Element	Ref.
##.IONIZATIONMODE	Technique	Ion Mobility Spectrometry	IMS.Measurement.IonizationMode	6
ANDI Label	**AnIML Shell**	**AnIML Technique**	**AnIML Element**	
injection-date-time-stamp	Technique	Chromatography	CHROM.Measurement.InjectionTime	7

```
<xs:element name="IMS">
  <xs:annotation>
    <xs:documentation>Ion Mobility Spectrometry</xs:documentation>
  </xs:annotation>
  <xs:complexType>
    <xs:sequence>
      <xs:element name="Measurement">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="IonizationMode">
              <xs:simpleType>
                <xs:restriction base="xs:token">
                  <xs:enumeration value="UV"/>
                  <xs:enumeration value="BR"/>
                  <xs:enumeration value="AL"/>
                  <xs:enumeration value="PD"/>
                  <xs:enumeration value="CD"/>
                  <xs:enumeration value="ESI"/>
                  <xs:enumeration value="LI"/>
                  <xs:enumeration value="LD"/>
                  <xs:enumeration value="SI"/>
                  <xs:enumeration value="SY"/>
                </xs:restriction>
              </xs:simpleType>
            </xs:element>
          </xs:sequence>
        </xs:complexType>
      <xs:element ref="SampleSet" id="Core:CoreID"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

Figure 5. An excerpt from the AnIML schema for IMS.

This looks something like Table 1 and appears in an extract from the draft AnIML XML schema proposed to the September meeting as in Figure 5.

Running before we can walk

All this is well and good, but we are still at the rough draft proposal stage and not anywhere near getting agreement across the industry on these proposals in the near future. Currently, the main effort is being made on getting agreement on the Core Layer, which is being worked on by NIST in Gaithersburg and should be available for further discussion by the next working meeting of E13.15 in December.

Within IUPAC, agreement to work on the technique layers was given during the recent General Assembly in Ottawa and further contacts will be made to groups such as the IUCr in Chester, UK, to discuss possible collaboration on the crystallographic data front (see Figure 6).

Conclusions

I am very pleased to see that, finally, most groups across the analytical sector have managed to come together in the

AnIML TECHNIQUE LIST - Proposal

Title	Abbreviation	Responsible Standards Organisation
VERSION 1.0		
INFRARED SPECTROSCOPY	IR	International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications (CPEP), Subcommittee on Electronic Data Standards
NUCLEAR MAGNETIC RESONANCE SPECTROMETRY	NMR	International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications (CPEP), Subcommittee on Electronic Data Standards
MASS SPECTROMETRY	MS	International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications (CPEP), Subcommittee on Electronic Data Standards (including labels from ASTM Standard E2077)
CHROMATOGRAPHY	CHROM	ASTM E13.15 with International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications (CPEP), Subcommittee on Electronic Data Standards (taken from ASTM Standard E1947)
ULTRA VIOLET & VISIBLE SPECTROSCOPY	UV/VIS	National Institute of Standards and Technology (NIST), Gaithersburg, USA (taken from SpectroML)
ION MOBILITY SPECTROMETRY	IMS	International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications, Subcommittee on Electronic Data Standards, with the International Society for Ion Mobility Spectrometry (ISIMS)
VERSION 2.0		
ELECTRON PARAMAGNETIC RESONANCE/ELECTRON SPIN RESONANCE SPECTROMETRY	EPR/ESR	International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications (CPEP), Subcommittee on Electronic Data Standards, EPR/ESR Task Group (Prof. Robert Lancashire)
NEAR INFRARED SPECTROMETRY	NIR	International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications (CPEP), Subcommittee on Electronic Data Standards, NIR & Chemometrics Task Group (Dr. Gerry Downey is Chairman of the International Council for Near Infrared Spectroscopy)
CRYSTALISATION	CRYST	International Union of Crystallography
VERSION 3.0		
CHEMOMETRICS DATA	CHEMOMETRICS	International Union of Pure and Applied Chemistry (IUPAC), Committee on Printed and Electronic Publications (CPEP), Subcommittee on Electronic Data Standards, NIR & Chemometrics Task Group (Dr. Gerry Downey is Chairman of the International Council for Near Infrared Spectroscopy)

Figure 6. Proposal for allocation responsibility of Technique Layer ontologies.

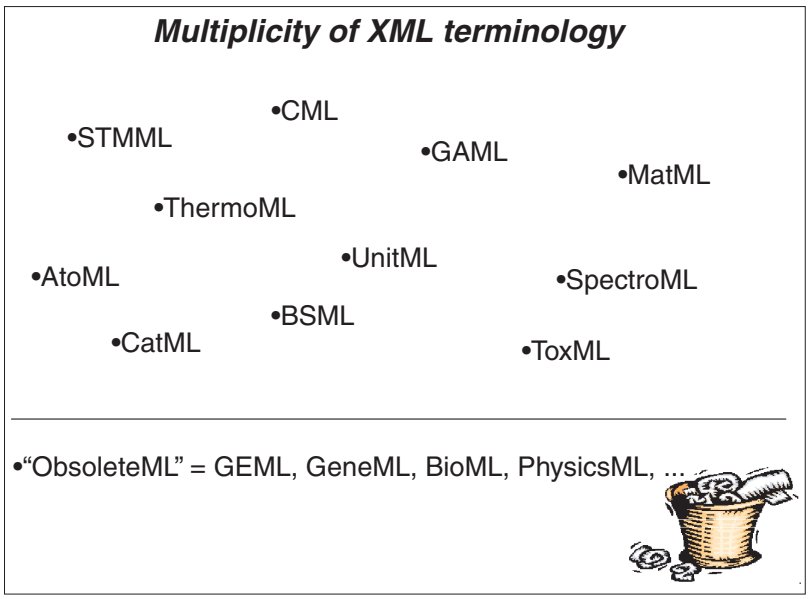


Figure 7. Multiplicity of different XML "standards" all with chemical content (after Steve Stein, NIST).

development of a single standard. As we have painstakingly reported over the years, these activities are always comparatively slow to come to fruition, not purely due to the complexity of the task but also due to the highly political nature of such cross-industry standardisation work. To finish, I would like to show a figure adapted from Steve Stein's presentation to IUPAC in Ottawa that is a good example of why this work is necessary. I think ObsoleteML speaks for itself (Figure 7)!

References

1. A.N. Davies, "XML in Chemistry", *Spectroscopy Europe* **14(1),** 22–24 (2002).
2. A combined meeting for two related IUPAC projects, the XML Data Dictionary Project and the IUPAC Chemical Identifier (IChI) Project (2000-025-1-800 http://www.iupac.org/projects/2000/2000-025-1-800.html), will be held at the National Institute of Standards and Technology (NIST, Gaithersburg, Maryland, USA) on 12–14 November 2003.
3. It is intended to create an XML-based dictionary for storage and exchange of thermophysical and thermochemical data, see http://www.iupac.org/projects/2002/2002-055-3-024.html for details.
4. http://www.astm.org/COMMIT/SUBCOMMIT/E1315.htm
5. http://sourceforge.net/projects/animl/
6. J.I. Baumbach, A.N. Davies, P. Lampen and H. Schmidt, "JCAMP-DX – A Standard Format for the Exchange of Ion Mobility Spectrometry Data (IUPAC Recommendations 2001)", *Pure Appl. Chem.* **73(11),** 1765–1782 (2001).
7. Standard Specification for Analytical Data Interchange Protocol for Mass Spectrometric Data, E 2077 – 00, ASTM, 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959, United States.