STANDARD DATA FORMATS FOR ANALYTICAL SYSTEMS: STATUS AND CHALLENGES

Maren Fiege[1], Gary W. Kramer[2] and the ASTM E13.15 subcommittee
[1]Waters GmbH, Frechen, Germany, [2]NIST, Gaithersburg, MD, USA (ASTM E13.15 Subcommittee Chairman)

INTRODUCTION

Challenge: Create a standard format for all analytical data to overcome the shortcomings of existing standards

Requirements:

- Flexible enough to represent analytical chemistry data
  - UV/Vis, IR, Chromatography, NMR, MS, IMS...
  - Hyphenated techniques
  - Multi-sample techniques such as array-based assays (titer plates), kinetics experiments, analytical mapping
- Strongly constrained to ensure data interchange and interoperability and to enable creation of generic data viewers
- Simple to understand
- Extensible to satisfy current and future needs of vendors, corporate interests, users, and new technologies
- Sufficient metadata for result interpretation
- Sufficient metadata for reprocessing
- Conversion from prior standards (JCAMP-DX and ANDI)
- Platform independent
- Distinguish between raw, processed, re-processed, and simulated data
- Provide sufficient commonality so that technique-constrained software can read technique-specific sections of multi-technique files
- 21 CFR 11 compliance; electronic signatures
- Verifiable, validateable
- Long term stable, human readable

Design Goals:

- XML based format
- Network friendly; easy parsing and viewing
- Broad industry support through common data dictionaries
  - Proposed format uses data dictionaries from:
    - JCAMP-DX
    - ANDI (netCDF)
    - IUPAC Gold Book
    - ASTM Terminology
- Independent, separate techniques
- Sample and workflow tracking

HOW ANIML WORKS

What does "AnIML" mean?

"AnIML" is an acronym for "Analytical Information Markup Language".

Basics

AnIML is based on XML[1] (eXtensible Markup Language), a simple, very flexible text format. Each and every AnIML file adheres to the same "Core" XML Schema[2].

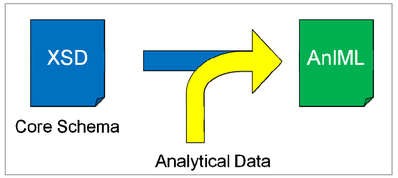

Figure 1. Core Schema.

The Core Schema provides a general structure for the data to be stored:

- Sample Information
- Measurement Data (curve and metadata)
  - Continuous and discrete data
  - Sparse and incomplete data
  - Non-plotted dimensions
  - Independent and dependent axes
- Audit Trail Information
- Digital Signatures

The "Technique Definitions", based on a separate "Technique Schema", provide technique-specific data dictionaries and a blueprint of how to arrange the data in the structure prescribed by the Core.

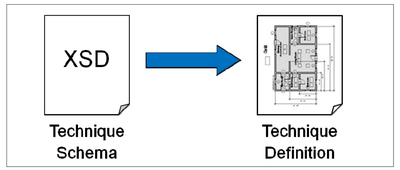

Figure 2. Technique Schema and Definition.

Challenge: Accommodate complex experiments

For complex experiments, multiple Technique Definitions can be used.

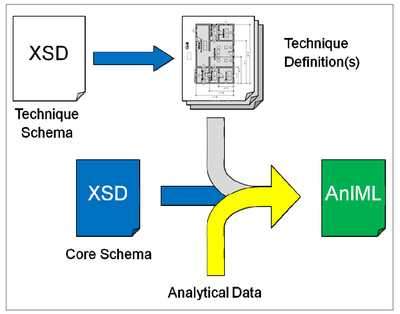

Figure 3. Bringing it all together.

Challenge: Accommodate vendor- or user-specific additions

The standard Technique Definitions only contain generally agreed-upon structures and metadata.

AnIML is flexible enough to accommodate additional data by creating Technique Extensions. Like Technique Definitions, they are based on the Technique Schema.

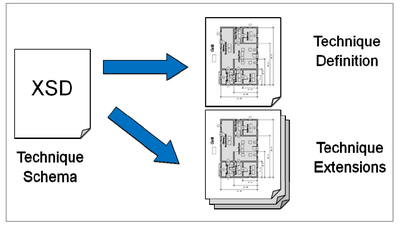

Figure 4. Technique Extensions.

COMPLIANCE ASPECTS

Validation of AnIML files

With the techniques described in the Technique Definitions and Extensions, AnIML files can be validated. A reference validation tool will be provided by ASTM and NIST.

- Syntactic Validation
  - Checks Document Against Schemas
    - Format
    - Element Completeness
    - Data Types
- Bounds/Limits Checking
  - Data ≤ or ≥ a Limiting Value
  - Data Between or Outside of Ranges of Values
- Semantic Validation
  - Correct Unit Types
  - Inclusion or Exclusion of Values in Sets
  - "Appropriateness"

Audit Trails

AnIML provides a full item-by-item audit trail including comments and a possibility to sign.

Digital Signatures

AnIML incorporates digital signatures according to the XML-DSig[3] standard by the W3C/IETF. Signatures can be applied to the entire file, or parts of it to protect it from undetected tampering.

- The XML document is "serialized"
- A cryptographic hash function is applied
- The result is encrypted by an asymmetric key algorithm
- A single-byte difference would invalidate the digital signature

EXAMPLE

The following is an example of an AnIML file containing data from a complex analytical technique:

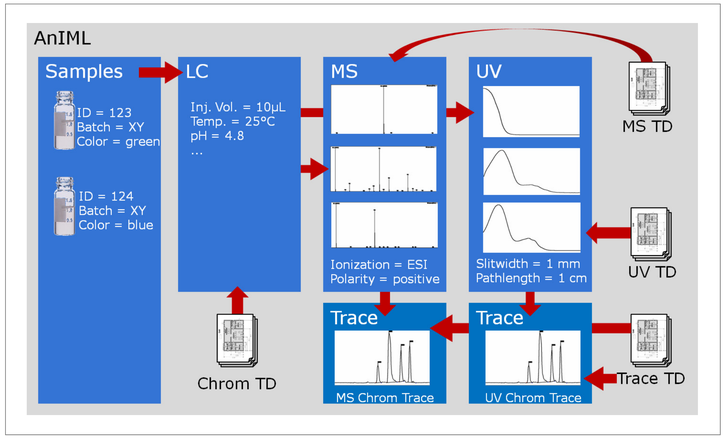

Figure 5. Complex LC-MS-UV experiment in AnIML (simplified).

STATUS

- Requirements document completed
- Core Schema completed
- Technique Schema completed
- Naming and Design Rules for Core and Technique Schemas completed
- Technique Definition for UV/Vis nearly completed
- Technique Definitions for Chromatography, IR, and MS in progress

Release of the first set of standard documents through ASTM is planned for 2009.

Initially supported techniques will include UV/Vis, infrared, NMR, mass spectrometry, and liquid and gas chromatography. More techniques will follow.

AnIML is a joint effort of the ASTM E13.15 subcommittee on Analytical Data and the IUPAC CPEP Subcommittee on Electronic Data Standards (SEDS)[4].

More Information:

Figure 6. Parties involved in the development of AnIML.

- http://animl.sourceforge.net
- http://www.animl.org

References

1. http://www.w3.org/XML/
2. http://www.w3.org/XML/Schema
3. http://www.w3.org/TR/xmldsig-core/
4. http://www.iupac.org/standing/cpep/wp_jcamp_dx.html

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS

720002760EN