

Requirements for a New Analytical Data Standard

Mark F Bean



CASS – Computational, Analytical and Structural Sciences

GlaxoSmithKline

Origins of Requirements

- March 2002 – two-day brainstorming meeting at Shimadzu in Columbia, MD
- Not codified until after AnIML was developed.
- Likely to be the first formal AnIML document available.
- Special thanks to Tony Davies for major contributions to the document.

Flexible, Strongly-Constrained Standard

- Flexible 
 - pH meter, alternating +/- ion switching LCMS, 2D NMR, derived data, etc..
- Constrained 
 - One and only one way of filling the file for a technique.
 - Technique definitions incomplete.

Hardware, Operating System, Vendor, and Software- Independence

- Inherent to XML 😊

Extensible 😊

- Standard, constrained manner of extending AnIML
- Must meet the changing needs of vendors, companies of users, and new technologies.
- Adequate to migrate data to AnIML from native formats without loss of information

Long-Lived, Human Readable

- AnIML file formats should still be readable in 60 years. 😊
 - XML is simple text (XML has wide support across computer industry)
 - Graphs are binary arrays encoded as Base64 (ASCII encoding format with wide support across computer industry)

Can Be Validated

- Syntax validation – via XML schema using standard tools from Microsoft, Sun, etc. 😊
- Semantic validation – based on technique definitions, but requires creation and support of technique validator. ⚠️

Supports Conversion from Prior Standards (ANDI and JCAMP)

- Technique definition dictionaries based on JCAMP and ANDI. 😊
- Not all JCAMP documents at the same level of polish or acceptance.
- AnIML committee itself will not provide converters. 😊

Can be Verified



- Supports electronic and digital signatures at any level of the document
- Verifiable audit trail
- Meets US-FDA 21CFRpart11 requirements

Supports Analysis Context (Metadata)

- Technique definition dictionary of terms. 😊
- Currently working on Sample and Location concepts. ⚠️

Technique-Constrained Software
Should be Able to Read Their
Technique Sections 😊

Supports Technique-Independent Viewers




- May add to the complexity of AnIML.
- How far to take this still under investigation.

Simple to Understand

- Should be easy for developers to implement.

Database Connectivity 😊

- Parse AnIML into database records and vice versa.

Raw and Processed Data
Must be Distinguishable 

Supports the Following Techniques

- **Phase 1:** IR, NMR, UV/vis, MS, chromatography, ion-mobility spectrometry, hyphenated techniques such as in multi-well plate high-throughput experiments
- **Phase 2:** electron magnetic resonance, near IR, crystallography
- **Phase 3:** other techniques