

# Requirements Specification

for

## AnIML - Version 1.0

Prepared by Mark F. Bean and Antony N. Davies

ASTM/IUPAC

Revision Date: October 13, 2005



### Revision History

Name	Date	Reason For Changes	Version
Mark Bean	22. Mar 2005	Initial Draft	1
Tony Davies	26-30. Sep. 2005	Updated to include results of discussions July/August 2005	2.00
Tony Davies	13 October 2005	Better linking of technical to business requirement – e-Signatures requirements clarified. Text edits	2.01

# Table of Contents

<b>1. Introduction</b> .....	<b>1</b>
1.1 Purpose.....	1
1.2 Intended Audience and Reading Suggestions.....	1
1.3 Scope.....	1
1.4 References.....	2
1.4.1 E1947–98 for Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data.....	2
1.4.2 E 1948–98 Standard Guide for Analytical Data Interchange Protocol for Chromatographic Data.....	2
1.4.3 E 2077–00 Standard Specification for Analytical Data Interchange Protocol for Mass Spectrometric Data.....	2
1.4.4 E 2078–00 Standard Guide for Analytical Data Interchange Protocol for Mass Spectrometric Data.....	2
1.4.5 JCAMP-DX for IR, Applied Spectroscopy 42(1), 1988, 151-162.....	2
1.4.6 JCAMP-DX for Chemical Structures, Applied Spectroscopy 45(1), 1991, 4-11.....	2
1.4.7 JCAMP-DX for NMR, Applied Spectroscopy 47(8), 1993, 1093-1099.....	2
1.4.8 JCAMP-DX for Mass Spectrometry, Applied Spectroscopy 48(12), 1994, 1545-1552.....	2
1.4.9 JCAMP-DX v. 5.01 (IUPAC Recommendations 1999), Pure Appl. Chem. 71(8), 1999, 1549-1556.....	2
1.4.10 JCAMP-DX for IMS (IUPAC Recommendations 2001), Pure Appl. Chem. Vol. 73(11), 1765-1782, 2001.....	2
1.4.11 JCAMP-DX NMR Pulse Sequences (IUPAC Recommendations 2001), Pure Appl. Chem., Vol. 73(11), 1749–1764, 2001.....	2
1.4.12 JCAMP-DX for Electronic Magnetic Resonance Spectrometry EMR, EPR, ESR, Ratified August 2005, Beijing IUPAC GA, Submitted for Publication.....	2
1.4.13 JCAMP-DX V.6.00 for LC/MS – Liquid Chromatography / Mass Spectrometry Hyphenated Methods, Ratified August 2005, Beijing IUPAC GA, Submitted for Publication.....	2
1.4.14 <a href="http://www.jcamp.org">http://www.jcamp.org</a> .....	2
1.4.15 <a href="http://animl.sf.org">http://animl.sf.org</a> .....	2
1.4.16 SpectroML <a href="http://www.mel.nist.gov/div826/msid/sima/03_spectro.html">http://www.mel.nist.gov/div826/msid/sima/03_spectro.html</a> .....	2
<b>2. Overall Description</b> .....	<b>3</b>
2.1 Product Perspective.....	3
2.2 Product Features.....	3
2.3 User Classes and Characteristics.....	3
2.4 Operating Environment.....	4
2.5 Design and Implementation Constraints.....	4

2.6	User Documentation .....	4
2.6.1	AnIML Requirements Document .....	4
2.6.2	AnIML core schema standard .....	4
2.6.3	AnIML technique schema standards for:.....	4
2.6.4	AnIML core guide.....	4
2.6.5	AnIML technique instance documents (for each technique named above), .....	4
2.6.6	AnIML technique guide (for each technique named above). .....	4
2.7	Assumptions and Dependencies .....	4
<b>3.</b>	<b>System Features .....</b>	<b>5</b>
3.1	Flexible Strongly-Constrained Standard .....	5
3.2	Simple to Understand .....	5
3.3	Extensible .....	5
3.4	Long-Lived.....	5
3.5	Human Readable.....	5
3.6	Can be Verified.....	6
3.7	Can be Validated .....	6
3.8	Adequate Read Speed for Viewers.....	6
3.9	Database Connectivity .....	6
3.10	Handle Analysis Context (Metadata).....	6
3.11	Supports Conversion from Prior Standards (ANDI and JCAMP).....	6
3.12	Supports the Following Techniques .....	7
3.13	Hardware, Operating System, Vendor, and Software-Independence.....	7
3.14	Mechanism for Distinguishing Raw and Processed Data.....	7
3.15	Technique Constrained Software should be able to read their Technique Sections.....	7
<b>4.</b>	<b>External Interface Requirements.....</b>	<b>8</b>
4.1	User Interfaces .....	8
4.2	Hardware Interfaces .....	8
4.3	Software Interfaces .....	8
4.4	Communications Interfaces .....	8

# 1. Introduction

## 1.1 Purpose

This document covers the requirements for version 1.0 of an XML (eXtensible Markup Language) based standard for interchange, storage, and viewing of analytical chemistry data.

The standard will be called AnIML (Analytical Information Markup Language).

The document will also define what computing technologies/components are required to work with AnIML so as to be able to flag early an impending need to migrate the format in the future.

## 1.2 Intended Audience and Reading Suggestions

The document was prepared for the ASTM E-13.15 Committee (Analytical Data Management) formed on October 15, 2002 (Providence, RI, USA) and the IUPAC Subcommittee on Electronic Data Standards. It is intended for AnIML developers and can be used as the basis for public presentations of AnIML and to enable implementers to check their solutions against the business/user requirements.

## 1.3 Scope

**The AnIML project will develop a definition for a standard analytical data format using the eXtensible Markup Language.**

When ratified by the ASTM, this standard will supersede the current ASTM standards and guidelines. See 1.5.

The Data Dictionaries required will draw wherever possible on the existing ASTM and IUPAC/JCAMP-DX standards.

## 1.4 References

- 1.4.1 E1947–98 for Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data
- 1.4.2 E 1948–98 Standard Guide for Analytical Data Interchange Protocol for Chromatographic Data
- 1.4.3 E 2077–00 Standard Specification for Analytical Data Interchange Protocol for Mass Spectrometric Data
- 1.4.4 E 2078–00 Standard Guide for Analytical Data Interchange Protocol for Mass Spectrometric Data
- 1.4.5 JCAMP-DX for IR, Applied Spectroscopy 42(1), 1988, 151-162
- 1.4.6 JCAMP-DX for Chemical Structures, Applied Spectroscopy 45(1), 1991, 4-11
- 1.4.7 JCAMP-DX for NMR, Applied Spectroscopy 47(8), 1993, 1093-1099
- 1.4.8 JCAMP-DX for Mass Spectrometry, Applied Spectroscopy 48(12), 1994, 1545-1552
- 1.4.9 JCAMP-DX v. 5.01 (IUPAC Recommendations 1999), Pure Appl. Chem. 71(8), 1999, 1549-1556
- 1.4.10 JCAMP-DX for IMS (IUPAC Recommendations 2001), Pure Appl. Chem. Vol. 73(11), 1765-1782, 2001
- 1.4.11 JCAMP-DX NMR Pulse Sequences (IUPAC Recommendations 2001), Pure Appl. Chem., Vol. 73(11), 1749–1764, 2001
- 1.4.12 JCAMP-DX for Electronic Magnetic Resonance Spectrometry EMR, EPR, ESR, Ratified August 2005, Beijing IUPAC GA, Submitted for Publication
- 1.4.13 JCAMP-DX V.6.00 for LC/MS – Liquid Chromatography / Mass Spectrometry Hyphenated Methods, Ratified August 2005, Beijing IUPAC GA, Submitted for Publication
- 1.4.14 <http://www.jcamp.org>
- 1.4.15 <http://animl.sf.org>
- 1.4.16 [SpectroML http://www.mei.nist.gov/div826/msid/sima/03\\_spectro.html](http://www.mei.nist.gov/div826/msid/sima/03_spectro.html)

## 2. Overall Description

This section will provide an overall description of AnIML; detailed requirements are to be found in section 3.

### 2.1 Product Perspective

This is the first version of a series of standards and when ratified by the ASTM, this standard will supersede the current ASTM standards and guidelines E1947–98, E 1948–98, E 2077–00, and E 2078–00. These initial standards will be developed and published through ASTM E13.15. In the long-term it is hoped that the specific technique committees will be able to host these standards where the necessary combination of IT skills with technique competence is present.

### 2.2 Product Features

The series of standards under the general name of AnIML (Analytical Information Markup Language) will define a standard for the interchange and archiving of analytical data.

The standard and systems implementing it must be capable of meeting all the legal requirements placed on users commonly known as compliance (GLP, cGMP, FDA 21 CFR part 11) including electronic signatures. (REQ 5, 7-10, 13-15, 19)

AnIML files will be written in XML. (REQ 6, 18)

Data stored in the current IUPAC/JCAMP-DX formats and the ASTM netCDF based formats must be capable of migration to an AnIML file without loss of information. (REQ 4, 16). All of the necessary files required to parse and validate an AnIML file must be freely available over the Internet.

The core of the AnIML standards must consist of the irreducible minimum content common across a particular technique and industry. All vendor-specific/instrumentation-specific content is the responsibility of the individual vendors to define and maintain and not subject of these documents. The mechanism by which such content is to be included within a valid AnIML file is the subject of these standards. No content which must appear in a vendor-neutral form in the core can be superseded by vendor specific modification. (REQ 1, 2, 4, 17, 20)

### 2.3 User Classes and Characteristics

The standards defined under the heading AnIML should be equally applicable to users in any group from academia to industry, from research to quality control, from clinical to heavy chemistry and environmental. Qualitative and quantitative analyses as well as chemical structure studies must be supported (REQ 1, 2, 11).

The standards must be simple enough that non-computer specialists can understand, deploy and implement them in their own environments (REQ 3).

## 2.4 Operating Environment

AnIML is designed to be platform neutral and be capable of deployment on any computer system supporting XML and XML Schema.

## 2.5 Design and Implementation Constraints

The AnIML standards must be capable of being deployed fully compliant with current regulatory requirements for electronic records such as FDA 21 CFR part 11.

The standards must allow adequate read speed for complex full-file hyphenated technique files such as LC/MS (as opposed to reports). (REQ 11, 12)

## 2.6 User Documentation

Over the complete period of the project, ASTM E13.15 will deliver the following documentation as part of the AnIML development project. (REQ 17)

### 2.6.1 AnIML Requirements Document

### 2.6.2 AnIML core schema standard

### 2.6.3 AnIML technique schema standards for:

*2.6.3.1 Infrared Spectroscopy*

*2.6.3.2 Nuclear Magnetic Resonance Spectrometry*

*2.6.3.3 Mass Spectrometry*

*2.6.3.4 Chromatography*

*2.6.3.5 Ultra Violet & Visible Spectroscopy*

*2.6.3.6 Ion Mobility Spectrometry*

*2.6.3.7 Electron Paramagnetic Resonance/Electron Spin Resonance Spectrometry*

*2.6.3.8 Near Infrared Spectrometry*

*2.6.3.9 Crystallization*

*2.6.3.10 Chemometrics Data*

### 2.6.4 AnIML core guide

### 2.6.5 AnIML technique instance documents (for each technique named above),

### 2.6.6 AnIML technique guide (for each technique named above).

In addition, example files for each technique defined must be provided.

## 2.7 Assumptions and Dependencies

The AnIML Schema must validate in common software packages handling XML, such as XMLSpy.

## 3. System Features

### 3.1 Flexible Strongly-Constrained Standard

AnIML must be flexible enough to represent a wide range of analytical chemistry data such as pH meter measurements, alternating positive-negative ion switching LC-MS with simultaneous PDA and ELS detection, 2D NMR, well micro-titer plate measurements, multi-dimensional data sets, etc.

The manner of representing such data needs to be strongly constrained to permit data interchange and creation of generic data viewers.

**REQ-1:** Flexible

**REQ-2:** Strongly constrained

### 3.2 Simple to Understand

For AnIML to be a successful standard, it must be relatively easy to understand and implement.

**REQ-3:** Simple to Understand

### 3.3 Extensible

AnIML must be extensible to cover changing needs of vendors, companies of users, and new technologies. Such extensions must be adequate to migrate data to AnIML from native formats without loss of information, although other properties (e.g. speed) may need to be sacrificed.

**REQ-4:** Standard, Constrained Manner of Extending AnIML

### 3.4 Long-Lived

AnIML file formats should still be readable in 60 years.

**REQ-5:** Longevity

### 3.5 Human Readable

AnIML must be human readable; it is not a binary file format and should not need special software or instructions to understand its content. However, it is reasonable to adopt ASCII encoded binary formats as containers for measurement data if needed.

**REQ-6:** Human Readable

### 3.6 Can be Verified

AnIML, to be human readable, must be text, and text can be altered. To avoid falsification of results, it must be possible to verify that the contents of an AnIML file have not been altered, or, if altered, there must be an audit trail of what changes were made, when, by whom, and why. AnIML must be capable of meeting current regulatory requirements for electronic records such as US-FDA 21CFRpart11.

**REQ-7:** Support Electronic and Digital Signatures

**REQ-8:** Verifiable

**REQ-9:** Audit Trail

### 3.7 Can be Validated

AnIML documents must conform to a standard, so it must also be straightforward to demonstrate that an AnIML document is conformant.

**REQ-10:** Conformance to standard can be validated.

### 3.8 Adequate Read Speed for Viewers

AnIML will often be deployed as an archive format, but AnIML files should also be able to be viewed in acceptable read-times by simple viewing software even for complex hyphenated technique data sets such as in LC/MS experiments.

**REQ-11:** Acceptable viewing of spectra even from LC/MS data.

### 3.9 Database Connectivity

It should be straightforward to parse AnIML into database records and vice versa.

**REQ-12:** Database-AnIML interchange

### 3.10 Handle Analysis Context (Metadata)

The increasing importance of well plates and automated systems in analytical chemistry demands that AnIML files need to track some plate or vial information. The analysis context should be sufficient for interpretation and reprocessing of data.

**REQ-13:** Track analysis context (metadata)

**REQ-14:** Sufficient metadata for interpretation of results

**REQ-15:** Sufficient metadata to permit reprocessing

### 3.11 Supports Conversion from Prior Standards (ANDI and JCAMP)

Although AnIML itself cannot supply converters, the structure of AnIML must be able to hold any data held in prior formats without loss.

**REQ-16:** Structure adequate to hold ANDI netCDF, and IUPAC/JCAMP Data .

### 3.12 Supports the Following Techniques

Data dictionaries are needed for the following prioritized techniques. These must include information required for complete interpretation of the data sets. The development of the standards will be split into three phases.

Phase 1: IR, NMR, UV/Vis, MS, Chromatography, Ion-Mobility Spectrometry, Hyphenated Versions of These, Multi-Well Plate High-Throughput Experiments

Phase 2: Electron Magnetic Resonance, Near IR, Crystallography

Phase 3: Chemometrics

**REQ-17:** Supports common analytical techniques in extensible manner

### 3.13 Hardware, Operating System, Vendor, and Software-Independence

AnIML should work on any system that can work with XML. High Priority

**REQ-18:** Software Independence

### 3.14 Mechanism for Distinguishing Raw and Processed Data

The AnIML standards must include metadata, allowing viewers to clearly indicate whether the data is as-measured or subsequently processed (audit trails).

**REQ-19:** Data type records

### 3.15 Technique Constrained Software should be able to read their Technique Sections

Software designed to read AnIML files may only recognize certain techniques to read and process those technique sections. However, when data exists within the AnIML file which a particular software package is not designed to read, the presence of this content should be clearly indicated at the time a file is read.

**REQ-20:** Technique Constrained Software Should be Able to Read Their Technique Sections

## **4. External Interface Requirements**

### **4.1 User Interfaces**

The AnIML specifications do not attempt to set rules for application software with the exception of paragraphs 3.14 and 3.15 above.

### **4.2 Hardware Interfaces**

There are no hardware interface requirements in AnIML.

### **4.3 Software Interfaces**

AnIML does not specify software interfaces.

### **4.4 Communications Interfaces**

Communications Interfaces are not part of this project.

## Appendix A: Glossary

ASTM	ASTM International (formerly American Society for Testing and Materials)
IUPAC	International Union of Pure and Applied Chemistry
JCAMP-DX	Joint Committee on Atomic and Molecular Physical Data – Data Exchange
JCAMP	The Joint Committee was sponsored and staffed by representatives of the following organizations: <ul style="list-style-type: none"><li>• American Chemical Society</li><li>• American Physical Society</li><li>• American Society of Mass Spectrometry</li><li>• American Society for Testing and Materials</li><li>• Coblenz Society</li><li>• Optical Society of America</li><li>• Society for Applied Spectroscopy</li><li>• Spectroscopy Society of Canada</li></ul>
LC/MS	Liquid Chromatography / Mass Spectrometry
GLP	Good Laboratory Practice
cGMP	Current Good Manufacturing Practice
FDA 21 CFR part 11	Rules from the USA Food and Drug Administration governing the equivalence of electronic and paper records.
PDA	Photodiode array detector
ELS	Evaporative Light Scattering detector
2D NMR	Two Dimensional Nuclear Magnetic Resonance Spectrometry experiment
IR	Infrared
NMR	Nuclear Magnetic Resonance
UV/Vis	Ultra-violet and Visible
MS	Mass Spectrometry
netCDF	Network Common Data Format ( as opposed to CDF – Common Data Format)
XML	Extensible Mark-up Language