

DRAFT

Requirements Specification

for

AnIML - Version 1.0

Prepared by Mark F. Bean and Antony N. Davies

ASTM/IUPAC

Revision Date: November 28, 2005

Revision History

Name	Date	Reason For Changes	Revision
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
Tony Davies	17 October 2005	Changes requested during E13.15 telephone conference 14 Oct. added	2.02
Tony Davies	11 November 2005	Changes made from E13.15 TC 28 Oct 2005	2.03
Tony Davies	28 November 2005	Changes made following E13.15 EAS meeting Nov. 13 th 2005 Reformatted to add ASTM-conform paragraph numbering.	2.04

Table of Contents

1. Introduction	1
1.1 Purpose	1
1.2 Intended Audience and Reading Suggestions	1
1.3 Scope	1
1.4 References	2
2. Overall Description	3
2.1 Overview	3
2.2 Product Perspective	3
2.3 Product Features	3
2.4 User Classes and Characteristics	4
2.5 Operating Environment	4
2.6 Design and Implementation Constraints	4
2.7 User Documentation	4
2.8 Assumptions and Dependencies	5
3. System Features	6
3.1 Flexible Strongly-Constrained Standard	6
3.2 Simple to Understand	6
3.3 Extensible	6
3.4 Long-Lived	6
3.5 Human Readable	6
3.6 Can be Verified	7
3.7 Can be Validated	7
3.8 Can be Viewed	7
3.9 Database Connectivity	7
3.10 Handle Analysis Context (Metadata)	7
3.11 Supports Conversion from Prior Standards (ANDI and JCAMP)	8
3.12 Supports the Following Techniques	8
3.13 Hardware, Operating System, Vendor, and Software-Independence	8
3.14 Mechanism for Distinguishing Raw and Processed Data	8
3.15 Technique Constrained Software should be able to read their Technique Sections	9
4. External Interface Requirements	9
4.1 User Interfaces	9
4.2 Hardware Interfaces	9
4.3 Software Interfaces	9
4.4 Communications Interfaces	9

1. Introduction

1.1 Purpose

1.1.1 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.

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

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

1.1.4 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.

1.2 Intended Audience and Reading Suggestions

1.2.1 The document was prepared for the ASTM E-13.15 Committee (Analytical Data) formed on October 15, 2002 (Providence, RI, USA) and the IUPAC Subcommittee on Electronic Data Standards.

1.3 Scope

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

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

1.3.3 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.mel.nist.gov/div826/msid/sima/03_spectro.html
- 1.4.17 Generalized Analytical Markup Language (GAML) no longer on Thermo Web Site but still available under XML.ORG at http://www.xml.org/xml/registry_searchresults.jsp?industry=12&keyword=&update_date=7200&schema_type=0
- 1.4.18 ISO 11179 RFC 2119: Key words for use in RFCs to Indicate Requirement Levels. Scott Bradner, 1997. (See <http://www.ietf.org/rfc/rfc2119.txt>)
- 1.4.19 XML Namespaces: <http://www.w3.org/TR/REC-xml-names>
ISO 11179: <https://fed-xml-ndr.core.gov/servlets/ProjectDocumentList?folderID=614&expandFolder=614&folderID=613>
Other naming recommendations from existing NDR projects of OASIS UBL, UN/CEFACT and the Environmental Protection Agency.
 - UBL: Universal Business Language NDR
 - EPA: Environmental Protection Agency NDR
 - UNC: UN/CEFACT NDR

2. Overall Description

2.1 Overview

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

2.2 Product Perspective

2.2.1 This is the first version of a series of standards. 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 ASTM 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 domain expertise is present.

2.3 Product Features

2.3.1 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.

2.3.2 The standard and systems implementing it must be capable of meeting 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)

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

2.3.4 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).

2.3.5 An AnIML file consists of a CORE which is the same regardless of the particular analytical technique originally used to create the data (the same for all analytical data sources). This CORE is then extended to cover terms and conventions which are Technique Specific such as Infrared Spectroscopy or Chromatography. An AnIML file can also store content which is user or vendor specific.

2.3.6 The CORE of the AnIML standard file with the relevant Technique Extension will be the basis for the exchange of data and must at least consist of the agreed minimum common content to allow for parsing and processing of the data. All additional vendor-specific/instrumentation-specific content is the responsibility of the individual vendors to define and maintain and not the 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. AnIML file content defined in the CORE and its standardized technique extension cannot be superseded by user or vendor specific modifications to ensure compatibility between installations. (REQ 1, 2, 4, 17, 20)

2.4 User Classes and Characteristics

2.4.1 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).

2.4.2 The documentation for the standards shall be simple enough for a developer to understand, deploy and implement AnIML and for users to extend in their own environments (REQ 3).

2.5 Operating Environment

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

2.6 Design and Implementation Constraints

2.6.1 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.

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

2.7 User Documentation

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

2.7.2 AnIML Requirements Document

- 2.7.3 AnIML core schema standard
- 2.7.4 AnIML technique schema standards
- 2.7.5 AnIML core guide
- 2.7.6 AnIML technique instance documents
- 2.7.7 AnIML technique guide (for each technique defined).

2.7.8 In addition, example files for each technique defined must be provided. All of the necessary files required to parse and validate an AnIML file must be freely available over the Internet.

2.8 Assumptions and Dependencies

2.8.1 The AnIML Schema must validate in common software packages handling XML, such as XMLSpy (2005 and later), Visual Studio and the W3C Schema Validator.

3. System Features

3.1 Flexible Strongly-Constrained Standard

3.1.1 **REQ-1:** Flexible enough to represent analytical chemistry data. AnIML must be flexible 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, multi-well micro-titer plate measurements, multi-dimensional data sets, derived data, simulations etc.

3.1.2 **REQ-2:** Strongly constrained. The manner of representing such data needs to be strongly constrained to permit data interchange and creation of generic data viewers.

3.2 Simple to Understand

3.2.1 **REQ-3:** Simple to Understand. For AnIML to be a successful standard, it must be relatively easy for developers to understand and implement.

3.3 Extensible

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

3.4 Long-Lived

3.4.1 **REQ-5:** Longevity. AnIML file formats should still be readable in 60 years.

3.5 Human Readable

3.5.1 **REQ-6:** 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.

3.6 Can be Verified

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

3.6.2 **REQ-8:** Verifiable.

3.6.3 **REQ-9:** Audit Trail.

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.

3.7 Can be Validated

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

AnIML documents must conform to a standard, so it must also be straightforward to demonstrate that an AnIML document conforms to a standard. A published list of validated files will be made available.

3.8 Can be Viewed

3.8.1 **REQ-11:** Viewing of spectra even from LC/MS data.

AnIML will often be deployed as an archive format, but AnIML files should also be able to be viewed even for complex hyphenated technique data sets such as in LC/MS experiments.

3.9 Database Connectivity

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

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

3.10 Handle Analysis Context (Metadata)

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

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

3.10.3 **REQ-15:** Sufficient metadata to permit reprocessing
The analysis context should be sufficient for interpretation and reprocessing of data.

3.11 Supports Conversion from Prior Standards (ANDI and JCAMP)

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

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

3.12 Supports the Following Techniques

3.12.1 **REQ-17:** Supports common analytical techniques in an extensible manner. 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 techniques such as in Multi-Well Plate High-Throughput Experiments
- Phase 2: Electron Magnetic Resonance, Near IR, Crystallography
- Phase 3: Other Techniques

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

3.13.1 **REQ-18:** Platform Independence.
AnIML should work on any system that can work with XML. This is a high priority requirement

3.14 Mechanism for Distinguishing Raw and Processed Data

3.14.1 **REQ-19:** Data type records.
The AnIML standards must include metadata, to clearly indicate whether the data is as-measured or subsequently processed.

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

3.15.1 **REQ-20:** Technique definitions must facilitate the reading of Technique-specific AnIML file sections.

An AnIML file containing data from multiple techniques should be able to be read by software which understands only some of the techniques.

4. External Interface Requirements

4.1 User Interfaces

4.1.1 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

4.2.1 There are no hardware interface requirements in AnIML.

4.3 Software Interfaces

4.3.1 AnIML does not specify software interfaces.

4.4 Communications Interfaces

4.4.1 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">• American Chemical Society• American Physical Society• American Society of Mass Spectrometry• American Society for Testing and Materials• Coblenz Society• Optical Society of America• Society for Applied Spectroscopy• Spectroscopy Society of Canada
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