DRAFT
Requirements Specification
for
AnIML - Version 1.0

Prepared by Mark F. Bean, Antony N. Davies, Maren Fiege, Burkhard A. Schaefer, Gary W. Kramer, Dale P. O'Neill, and David P. Martinsen

ASTM E13.15 and IUPAC CPEP SEDS

Revision Date: October 31, 2007

Revision History

Name	Date	Reason For Changes	Revision
Mark Bean	22. Mar 2005	Initial Draft	1
Tony Davies et al.	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 et al.	17 October 2005	Changes requested during E13.15 telephone conference 14 Oct. added	2.02
Tony Davies et al.	11 November 2005	Changes made from E13.15 TC 28 Oct 2005	2.03
Tony Davies, Mark Bean, Richard Larsen, Gary Kramer, Alex Mutin	28 November 2005	Changes made following E13.15 EAS meeting Nov. 13th 2005 Reformatted to add ASTM-conform paragraph numbering.	2.04
Mark Bean	9 March 2006	Wording of some requirements	2.05
Mark Bean	3 August 2007	Added naming and design rules (NDRs). Added new requirements based on discoveries on the AnIML wiki at http://www.animlwiki.com/wiki/Requirements.ashx These need to be reviewed by the committee and have not been approved.	2.06
Tony Davies, Maren Fiege, Dave Martinsen, Robert Lancashire, Peter Lampen, Mohan Cashyap	4. August 2007	Update reference list	2.07
Mark Bean	7 September 2007	Reviewed and tidied recent changes. Made all URLs hyperlinks.	2.08
Mark Bean	18 September, 2007	AnIML data file rather than AnIML file. Req 20 and 21 modified. Analytical data changed to Analytical chemistry data except when naming a committee.	2.09
Burkhard Schaefer, Gary Kramer, Frank Masur, Kordian Placzek, Maren Fiege, Patrick Gleichmann, Mark Bean	25 September 2007	Results of today's teleconference added. Many requirements have been clarified and reworded.	2.10
Maren Fiege, Burkhard Schaefer, Gary Kramer, Frank Masur, Kordian Placzek, Mark Bean, Tony Davies, Dale O'Neill, Stuart Chalk	02 October 2007	Results of today's teleconference added. Many requirements have been clarified and reworded.	2.11
Maren Fiege, Gary Kramer, Frank Masur, Dale O'Neill	16 October 2007	Results of today's teleconference added. Many requirements have been clarified and reworded.	2.12
Gary Kramer et al.	31 October 2007	Results of today's AnIML VM discussion; fix toc; fix ASTM refs;version sent to Tony Davies for publication in Pure & Applied Chemistry	2.13

Table of Contents

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 This document will also define what computing technologies/components are required to work with AnIML to identify impending obsolescence and the need to migrate the format in the future.

1.1.4 This document 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 business/user requirements.

1.1.5 This document can be used where necessary to delimit the AnIML standard.

1.2 Intended Audience and Reading Suggestions

1.2.1 This document was prepared for the ASTM E13.15 Subcommittee (Analytical Data) and the IUPAC Subcommittee on Electronic Data Standards (SEDS).

1.2.2 The intended audience is the ASTM E13.15 Subcommittee, the IUPAC SEDS Subcommittee, instrument vendors, others working on the development of the standard, and anyone seeking an overview of the development goals of the AnIML standardization initiative.

For more detail, the reference list in Section 1.4 should be consulted along with minutes and presentations from AnIML meetings on the AnIML website http://animl.sourceforge.net/ and http://www.animl.org.

1.3 Scope

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

1.3.2 When ratified by the ASTM, the AnIML series of standards will supersede the current ASTM Analytical Data Interchange Protocol (ANDI) standards and guidelines (see 1.3.4 to 1.3.7).

1.3.3 The data dictionaries required will draw on existing ASTM and IUPAC standards (see 1.3.4 to 1.3.16).

References

1.3.4 ASTM E1947–98 (2004) for Standard Specification for Analytical Data Interchange Protocol for Chromatographic Data
1.3.5 ASTM E 1948–98 (2004) Standard Guide for Analytical Data Interchange Protocol for Chromatographic Data
1.3.6 ASTM E 2077–00 (2005) Standard Specification for Analytical Data Interchange Protocol for Mass Spectrometric Data
1.3.7 ASTM E 2078–00 (2005) Standard Guide for Analytical Data Interchange Protocol for Mass Spectrometric Data
1.3.8 JCAMP-DX for IR, Applied Spectroscopy 42(1), 1988, 151-162
1.3.9 JCAMP-DX for Chemical Structures, Applied Spectroscopy 45(1), 1991, 4-11
1.3.10 JCAMP-DX for NMR, Applied Spectroscopy 47(8), 1993, 1093-1099
1.3.11 JCAMP-DX for Mass Spectrometry, Applied Spectroscopy 48(12), 1994, 1545-1552
1.3.12 JCAMP-DX v. 5.01 (IUPAC Recommendations 1999), Pure Appl. Chem. 71(8), 1999, 1549-1556
1.3.13 JCAMP-DX for IMS (IUPAC Recommendations 2001), Pure Appl. Chem. 73(11), 1765-1782, 2001
1.3.14 JCAMP-DX NMR Pulse Sequences (IUPAC Recommendations 2001), Pure Appl. Chem. 73(11), 1749–1764, 2001
1.3.15 JCAMP-DX for Electronic Magnetic Resonance Spectrometry EMR, EPR, ESR, Pure Appl. Chem. 78(3), 613–631, 2006
1.3.16 JCAMP-DX V.6.00 for Chromatography and Mass Spectrometry Hyphenated Methods, Ratified August 2005, Beijing IUPAC General Assembly, Final Review
1.3.17 http://www.jcamp-dx.org
1.3.18 http://animl.sourceforge.net/
1.3.19 SpectroML http://www-1.informatik.fh-wiesbaden.de/~schaefer/JALA-6-6-SpectroML-Paper.pdf and http://animl.sourceforge.net/SpectroML_PittCon_03.pdf
1.3.20 Generalized Analytical Markup Language (GAML) http://www.gaml.org/
1.3.21 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)
1.3.22 XML Namespaces: http://www.w3.org/TR/REC-xml-names
ISO 11179: http://metadata-standards.org/11179/index.html
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

1.3.23 OASIS Universal Business Language NDR Project
http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl
1.3.24 UN/CEFACT NDR Project
http://xml.coverpages.org/UN-CEFACT-ATG-NamingAndDesignRulesV12-20050908.pdf
1.3.25 U.S. Environmental Protection Agency NDR Project
http://www.exchangenetwork.net/dev_schema/index.htm
1.3.26 U.S. Department of the Navy NDR Project
http://xml.coverpages.org/DON-XML-NDR20050127-33942.pdf

Overall Description

1.4 Overview

This section provides an overall description of AnIML; detailed requirements are found in section 3.

1.5 Perspective

This is the first version of a series of standards. When ratified by the ASTM, these AnIML standards will supersede the current ASTM ANDI standards and guidelines E1947–98, E1948–98, E2077–00, and E2078–00. The ASTM AnIML standards will be developed and published through ASTM E13.15. In the long-term, it is hoped that the appropriate ASTM subcommittees will take over jurisdiction over and maintain those AnIML standards that fall into their specific domain of expertise. The generic AnIML standards will remain under the jurisdiction of and be maintained by E13.15.

1.6 Features

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

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

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

1.6.4 AnIML should be designed to accept data converted from the current IUPAC/JCAMP-DX and ASTM ANDI (netCDF-based) formats without loss of information. (REQ 4, 16).

1.6.5 An AnIML data file provides a generic data container that is the same regardless of the particular analytical technique originally used to create the data. This container can be formally constrained to reflect technique-specific terms and conventions, such as infrared spectroscopy or chromatography. An AnIML data file can also store content that is user- or vendor-specific.

1.6.6 Every AnIML data file must contain the agreed minimum technique-specific content to allow parsing and processing. Additional information can be provided, if needed.

1.6.7 In addition to standardized technique-specific content, AnIML shall support amendments to allow for vendor-, instrument-, and end-user-specific content in a prescribed manner. Such amendments shall be formally documented in XML.

1.6.8 Providers of amendments (1.6.7) are responsible for maintenance of the amendment and its associated documentation.

1.6.9 The mechanism by which such amendments (1.6.7) are to be included within an AnIML data file is defined by the AnIML family of standards.

Amendments cannot override content prescribed by the AnIML family of standards. (REQ 1, 2, 4, 17, 20)

1.7 User Classes and Characteristics

The AnIML family of standards should be equally applicable to all fields of analytical chemistry, including regulated environments. Qualitative and quantitative analyses, as well as chemical structure studies, must be supported (REQ 1, 2, 11).

1.8 Operating Environment

AnIML shall be platform independent and usable on computer systems supporting XML and XML Schema.

1.9 Design and Implementation Constraints

1.9.1 The AnIML family of standards must be compliant with current regulatory requirements for electronic records, such as FDA 21 CFR part 11.

1.9.2 The standards shall allow adequate read speed for complex, large files such as those from LC-MS experiments. (REQ 11, 12)

All AnIML standards and amendments (1.6.7) shall conform to naming and design rules, ensuring internal consistency, harmony, and interoperability with other markup languages (for example, 1.3.22 to 1.3.26).

1.10 User Documentation

1.10.1 Sufficient documentation for the standards shall be provided for developers to understand, implement, and deploy AnIML solutions, as well as for users to create amendments (1.6.7). (REQ 3)

1.10.2 ASTM E13.15 will deliver the following documentation as part of the AnIML development project (REQ 17):

- Schema documentation
- Implementation guidelines
- Naming and design rules
- Technique documentation (for each technique defined)
- Users' guides

1.10.3 Example files for each technique as well as all of the necessary files required to parse and validate an AnIML data file must be freely available on the Internet.

1.11 Validation

1.11.1 Any AnIML schema(s) must validate in common software packages handling XML. At the time of publication, this includes Altova XMLSpy (2005 and later), Microsoft Visual Studio (2005 and later), and the current W3C Schema Validator.

1.11.2 More extensive validation, such as semantic validation of AnIML data files, is desirable.

System Requirements

1.12 Flexible Strongly Constrained Standard

1.12.1 **REQ-1:** Sufficiently flexible to represent analytical chemistry data.
AnIML must be sufficiently flexible to represent a wide range of analytical chemistry data, such as, pH meter measurements, alternating positive-negative-ion LC-MS (liquid chromatography-mass spectrometry) with simultaneous PDA (photodiode array spectroscopy) and ELS (evaporative light scattering) detection, 2D NMR (two-dimensional nuclear magnetic resonance spectroscopy), multi-well microtiter plate measurements, multi-dimensional data sets, derived data, simulations, etc.

1.12.2 **REQ-2:** Strongly constrained.
The manner of representing analytical chemistry data shall be strongly constrained for all techniques to reduce representational variability, permitting interoperability, data interchange, and creation of generic data viewers.

1.13 Simple to Understand

REQ-3: Simple to understand.
For AnIML to be a successful standard, it must be relatively easy for developers to understand and implement. Where possible, the arrangement of elements should be similar to that in vendors' data systems. AnIML should avoid creation of new constructs for problems that the analytical chemistry community has not found a need or way to address. Symmetry in element treatments is desirable (peak and method table representation, for example).

1.14 Extensible

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 and should not break compatibility with existing AnIML-enabled applications.

1.15 Long Lived

REQ-5: Longevity.
AnIML data files should be readable for at least 60 years.

1.16 Human Readable

REQ-6: Human readable.
AnIML must be human readable; it shall not be a binary file format. Special software or instructions beyond the AnIML documentation should not be needed to understand its content. However, it is reasonable to adopt ASCII-encoded binary formats as containers for measurement data, if needed.

1.17 Amenable to Validation and Verification

1.17.1 **REQ-7:** Support electronic and digital signatures.
To be human readable, AnIML data files must be text, and text can be altered. To avoid falsification of results, it should be possible to verify that the contents of an AnIML data file have not been altered.

1.17.2 **REQ-8:** Design and documentation.
The AnIML family of standards must be designed and documented so that conformity to the standards can be verified.

1.17.3 **REQ-9:** Audit trail.
To be human readable, AnIML data files must be text, and text can be altered. If altered, there should 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 21 CFR part 11.

1.17.4 **REQ-10:** AnIML data file conformance to the standards must be amenable to validation.
It must be straightforward to demonstrate that an AnIML data file conforms to a standard, either manually or through software. This covers both reading and writing AnIML data files. Validated example files shall be made available on the Internet.

1.18 Viewable

REQ-11: Preserving complex relationships.
Logical relationships between data generated by complex and hyphenated techniques must be preserved for viewing and navigation.

1.19 Database Connectivity

REQ-12: Database-AnIML interchange.
AnIML should be easily parseable into database records and vice versa.

1.20 Supports Analytical Context (Metadata)

1.20.1 **REQ-13:** Track analytical context (metadata).
It must be possible to store metadata that provides the context for the analytical chemistry data stored in the AnIML data file.

1.20.2 **REQ-14:** Sufficient metadata for interpretation of results.
Metadata stored in an AnIML data file should include all parameters and information needed to interpret the results properly.

1.20.3 **REQ-15:** Sufficient method metadata to permit data reprocessing.
It should be possible to store sufficient method information in an AnIML data file to permit proper data reprocessing (e.g., domain transforms, peak integration, etc.).

1.21 Supports Conversion from Prior Standards (ANDI and JCAMP)

REQ-16: Structure adequate to hold ASTM/ANDI and IUPAC/JCAMP-DX Data.
Although the AnIML project itself will not supply converters, the structure of AnIML must be able to hold analytical data held in prior formats without loss.

1.22 Supports Common Analytical Techniques

REQ-17: Supports common analytical techniques in an extensible manner.
To develop an AnIML standard, a comprehensive data dictionary is needed for the technique. These must include information required for complete interpretation of the data sets. The order of development of the standards is contingent on the availability of these data dictionaries:

- Phase 1: IR, NMR, UV-Vis, MS, Chromatography, Ion-Mobility Spectrometry, including combinations of these techniques (hyphenated techniques, such as well plate LC-MS).
- Phase 2: Electron Magnetic Resonance (EPR/ESR), Near IR, Crystallography
- Phase 3: Other Techniques

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

REQ-18: Platform Independence.
AnIML must work on any system that can support XML. This is a high priority requirement.

1.24 Data Must be Distinguishable by their Provenance

REQ-19: Data type records.
The AnIML standards must include metadata to indicate clearly whether the data are as-measured, simulated, subsequently processed, etc.

1.25 Technique-Constrained Software Applications

REQ-20: Technique Partitioning.
An AnIML data file containing data from multiple techniques should be readable by application software that may not understand all of the techniques. This implies clean technique partitioning. In a given AnIML data file, it should be possible to parse data that an application supports in the presence of data that the application does not support.

1.26 Reuse of Commonalities

REQ-21: Reuse of common structures.
Frequently used structures, such as those for organizing data acquired against time, pH, voltage, or other common measures and table structures, should be reusable and generic

in nature. Once definitions have been agreed upon, techniques developed later must use these definitions unless there is a clear, unmet requirement.

2. External Interface Requirements

2.1 User Interfaces

The AnIML specifications do not attempt to set requirements for application software.

2.2 Hardware Interfaces

AnIML does not specify any hardware interface requirements.

2.3 Software Interfaces

AnIML does not specify any software interfaces.

2.4 Communications Interfaces

AnIML does not specify any communications interfaces.

Appendix A: Glossary

2D NMR	Two Dimensional Nuclear Magnetic Resonance Spectrometry experiment
AIA	Analytical Instrument Association
ANDI	Analytical Data Interchange (name of the ASTM / AIA netCDF standards)
AnIML	Analytical Information Markup Language
ASCII	American Standard Code for Information Interchange
ASTM	ASTM International (formerly American Society for Testing and Materials)
cGMP	Current Good Manufacturing Practice
CPEP	IUPAC Committee on Printed and Electronic Publications
CROMERR	Cross-Media Electronic Reporting Rule, issued by EPA
ELS	Evaporative Light Scattering detector
EMR	Electron Magnetic Resonance
EPA	US Environmental Protection Agency
EPR	Electron Paramagnetic Resonance
ESR	Electron Spin Resonance
FDA 21 CFR part 11	Rules from the USA Food and Drug Administration governing the equivalence of electronic and paper records.
GAML	Generalized Analytical Markup Language
GLP	Good Laboratory Practice
IMS	Ion Mobility Spectrometry
IR	Infrared
ISO	International Standards Organization
IUPAC	International Union of Pure and Applied Chemistry
JCAMP	Joint Committee on Atomic and Molecular Physical Data, sponsored and staffed by representatives of the following organizations: • American Chemical Society • American Physical Society • American Society of Mass Spectrometry • American Society for Testing and Materials • Coblentz Society • Optical Society of America • Society for Applied Spectroscopy • Spectroscopy Society of Canada
JCAMP-DX	Joint Committee on Atomic and Molecular Physical Data – Data Exchange
LC-MS	Liquid Chromatography / Mass Spectrometry
MS	Mass Spectrometry
NDR	Naming and Design Rules
netCDF	Network Common Data Format (as opposed to CDF – Common Data Format)
NMR	Nuclear Magnetic Resonance
OASIS	Organization for the Advancement of Structured Information Standards
PDA	Photodiode Array detector
REQ	Used in this document to denote a requirement
RFC	Request for Comments
SEDS	Subcommittee on Electronic Data Standards (within IUPAC, subcommittee of CPEP)

SpectroML	XML for Molecular Spectrometry Data (NIST effort)
UBL	Universal Business Language (OASIS effort)
UN/CEFACT	United Nations Centre for Trade Facilitation and Electronic Business
UV-Vis	Ultra-violet and Visible
XML	Extensible Markup Language
W3C	World Wide Web Consortium