

Putting AnIML to Work

Status of AnIML

March 9, 2009

PittCon 2009 - Chicago, IL, USA

Dr. Maren Fiege

Waters GmbH

ASTM E13.15

Contents

- Why Data Standards?
- Existing Data Standards
 - IUPAC JCAMP-DX
 - ASTM ANDI/AIA
- ASTM AnIML
 - Basics
 - Accomplishments

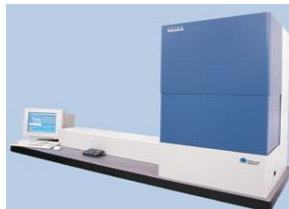
Objectives of data interchange



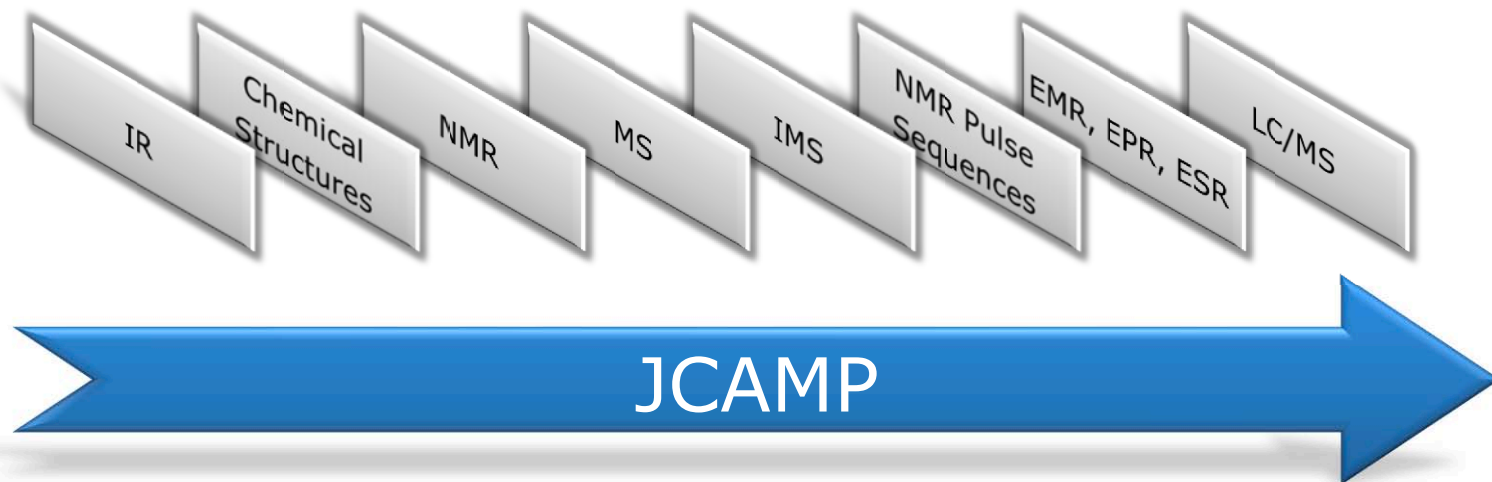
"I speak Spanish to God,
Italian to women,
French to men,
and German to my horse."

Emperor Charles V. (1500-1558)

Why standard formats?



JCAMP-DX Standards



1988

today



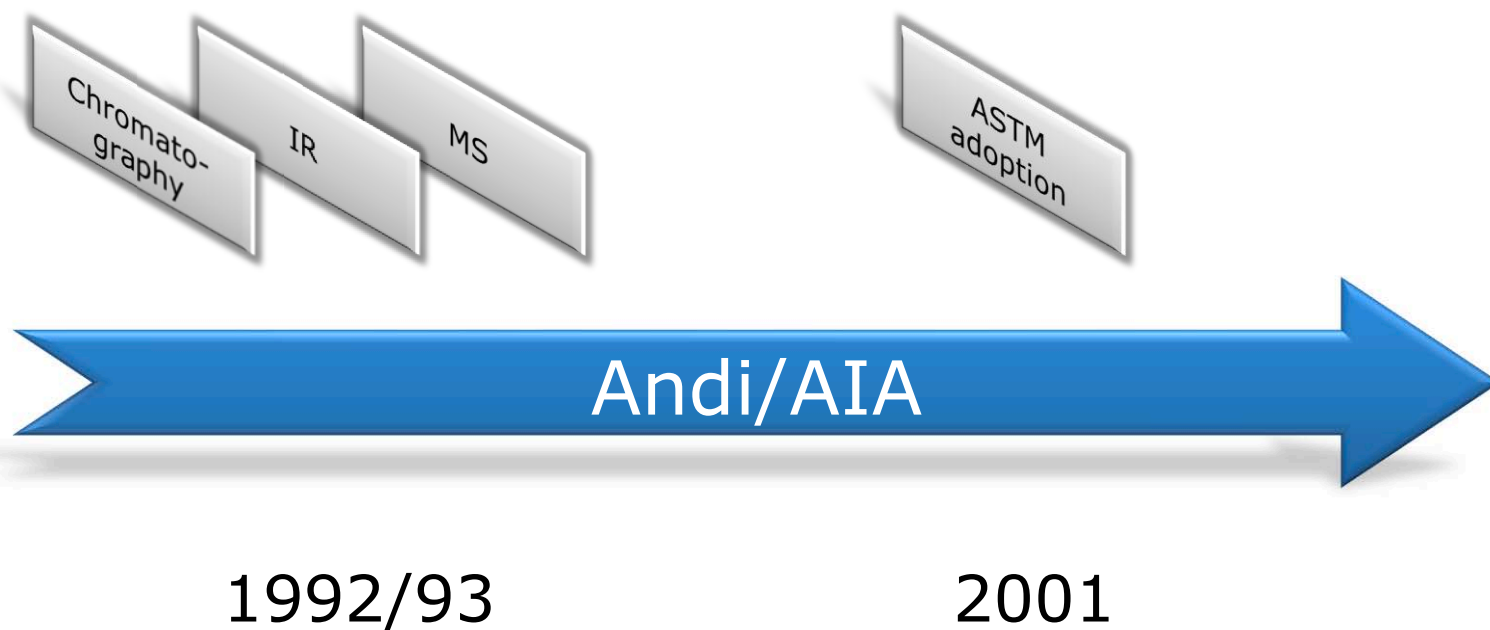
- Open to the public
- Independent from instrument manufacturers and operating systems
- Widely supported
- Human readable
- Combination of spectra and structures
- Extensible without compatibility impacts



- Flexible data format description allows interpretation
- Character set limited to ASCII 30 to 128

ANDI/AIA

- Developed to overcome lack of standards for MS and chromatography
- Based on netCDF, a “platform independent” binary format.





- Independent from instrument manufacturers
- Partially independent of operating systems
- Extensible without compatibility impacts
- Open to the public (upon payment)



- Not human readable
- No combination of data and structures
- Supported only for MS and Chromatography
- Only partially finished
- Not regarded as 21CFR11 compliant
- Will not be re-balloted by ASTM once AnIML is out

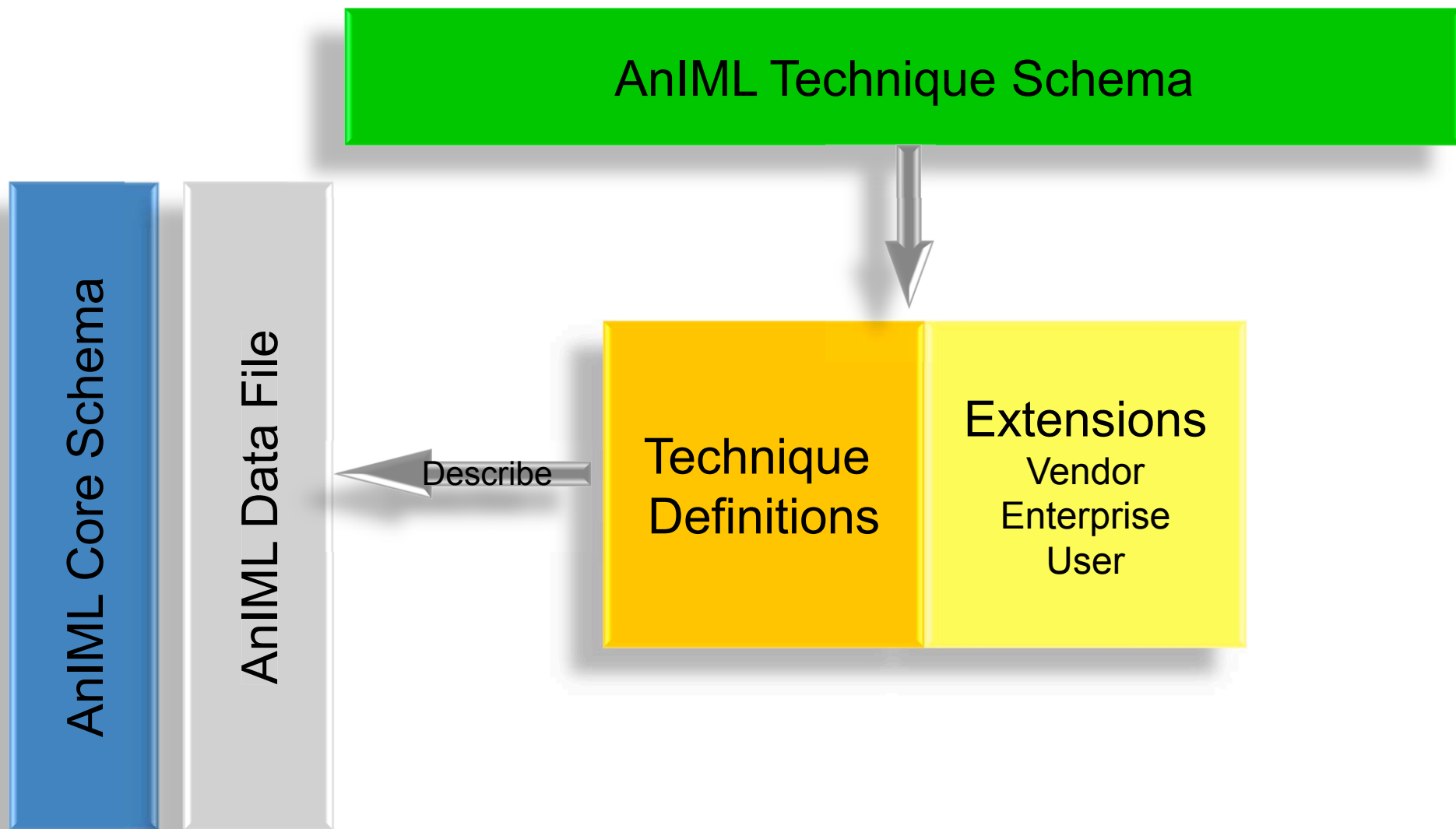
Origins of AnIML

- AnalYtical Information Markup Language
- Necessity for common standard, based on existing standards and experiences
- Proposed AnIML format uses Data Dictionaries from:
 - JCAMP-DX
 - ANDI (netCDF)
 - IUPAC Gold Book
 - ASTM Terminology

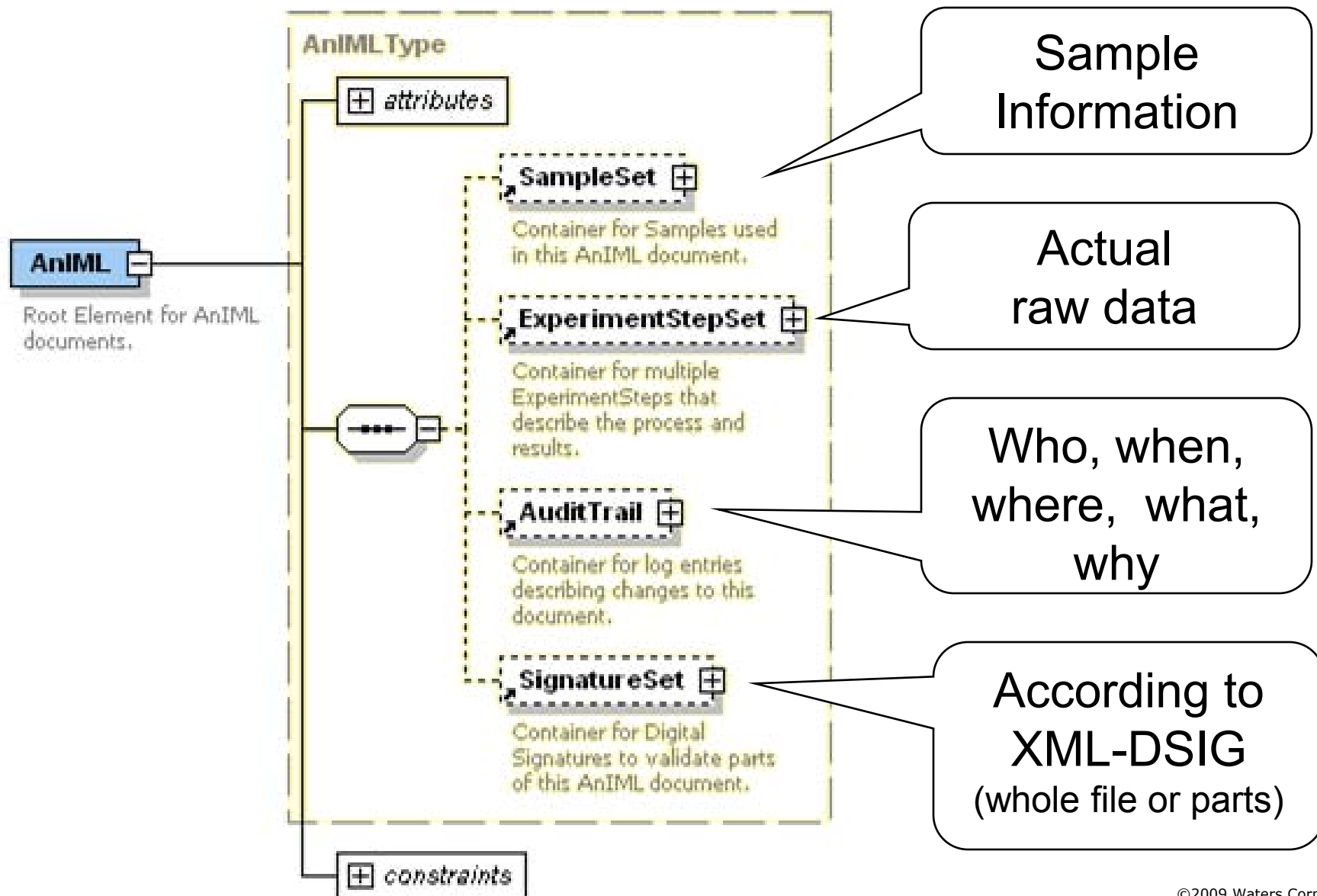
AnIML Design Goals

- XML based format
- 21 CFR 11 compliance; electronic signatures
- Available for all analytical techniques (also hyphenated)
- Long term stable, human readable
- Extensible
- Broad industry support through common data dictionaries

Basics



Core Schema



Core

- Stores:
 - Sample information
 - Method information
 - Processing information
 - Measured data from experiments
 - All data resulting from the application of a technique to a (group of) sample(s)
 - Data representation features:
 - Continuous data
 - Discrete data
 - Sparse data
 - Incomplete data
 - Non-plotted dimensions

Technique Definitions

- XML documents adhering to the AnIML Technique Schema
- Explain how to use the Core for specific techniques
 - Metadata dictionary
 - Describe Sample Information, Axes, Parameters...
- Technique Definitions can be amended using Extensions
- Extensions “overlay” Technique Definitions
- Does not break technique-specific applications

Contents of AnIML 1.0

- Core Schema – 95%
- Technique Schema – 99%
- Technique Definitions
 - IR – 50%
 - UV/Vis – 80%
 - Chromatography – 80%
 - MS – 15%
 - Post-processing Techniques
 - Peak Finding – 80%
 - Trace Forming – 80%

AnIML Documentation

- AnIML Requirements Document - done
- AnIML Naming and Design Rules - done
- AnIML Interagency Report – 20%
- AnIML Core Schema Standard Specification – 80%
- AnIML Technique Schema Standard Specification – 80%
- Standard Guide to AnIML Data Documents – 20%
- Std. Guide to AnIML Technique Definition Documents – 10%
- For each Technique:
 - AnIML Technique Document Standard Specification – 80%
 - Standard Guide to the AnIML Technique – 10%

Sources of Information

“It does not matter how slowly you go as long as you do not stop.” (Confucius, 551-479 BC)

- The AnIML Web Site
 - www.animl.org
- The SourceForge Analytical Information Markup Language Project: Summary Page
 - sourceforge.net/projects/animl





Next:

“Use Cases for the Analytical Information Markup Language (AnIML)”