Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

AnIML:
An Analytical Data Standard for the Lab and the Enterprise

Converting Legacy JCAMP-DX and ANDI Data to AnIML

March 1, 2010
PittCon 2010 - Orlando, FL, USA

Dr. Maren Fiege
Waters GmbH
ASTM E13.15

Contents

- The Need for Standard Formats
- Existing Standards
- Moving Data Forward
- From JCAMP-DX and ANDI to AnIML

Regulatory Requirements

- Copies of Electronic Records for Inspection
  - Accessibility for Inspection
  - Integrity of Content and Meaning
  - Human Readable Form
  - Standard Portable Formats
- Retention and Maintenance of Records
  - GxP demands archival of records for extended time
  - Reprocessibility

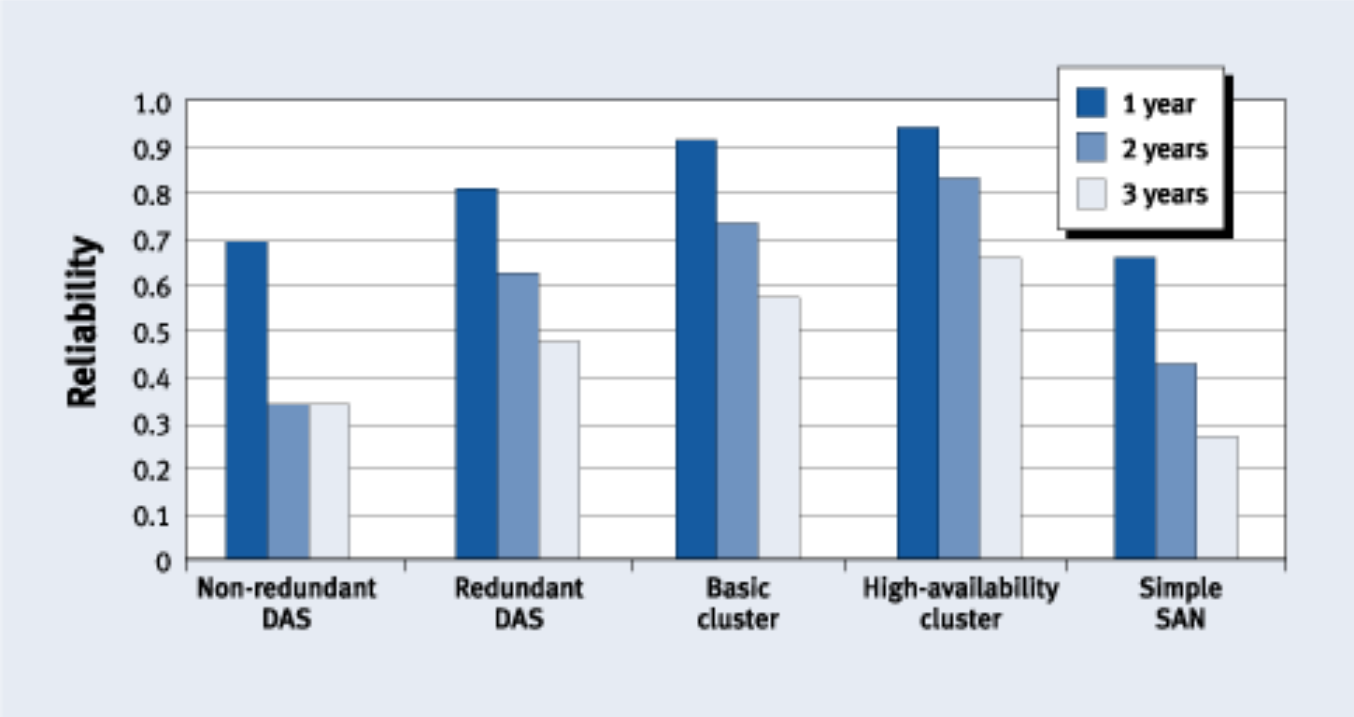
Record Retention Requirements

Industry Segment, Regulator and Type of Record	Typical Retention Period
Pharma: Good Laboratory Practices; records related to a new drug application (NDAs)	Date of submission plus 5 years
Health Care: medical records	Life of patient plus “n” years
Drug/device study records	Marketing application plus 2-3 years
Government records	20-50 years, or permanent
Copyright records (all organizations)	Life of copyright = 95 years or as business needs dictate
Patent records and supporting data	Application plus 17 years

Lifespans

Software lifetime: approx. 9 years*

Hardware depreciation: 3-5 years



Comparison of reliability over three years

*Software lifetime and its evolution process over generations, Tamai, T., Torimitsu, Y., Conference on Software Maintenance, 1992, Proceedings

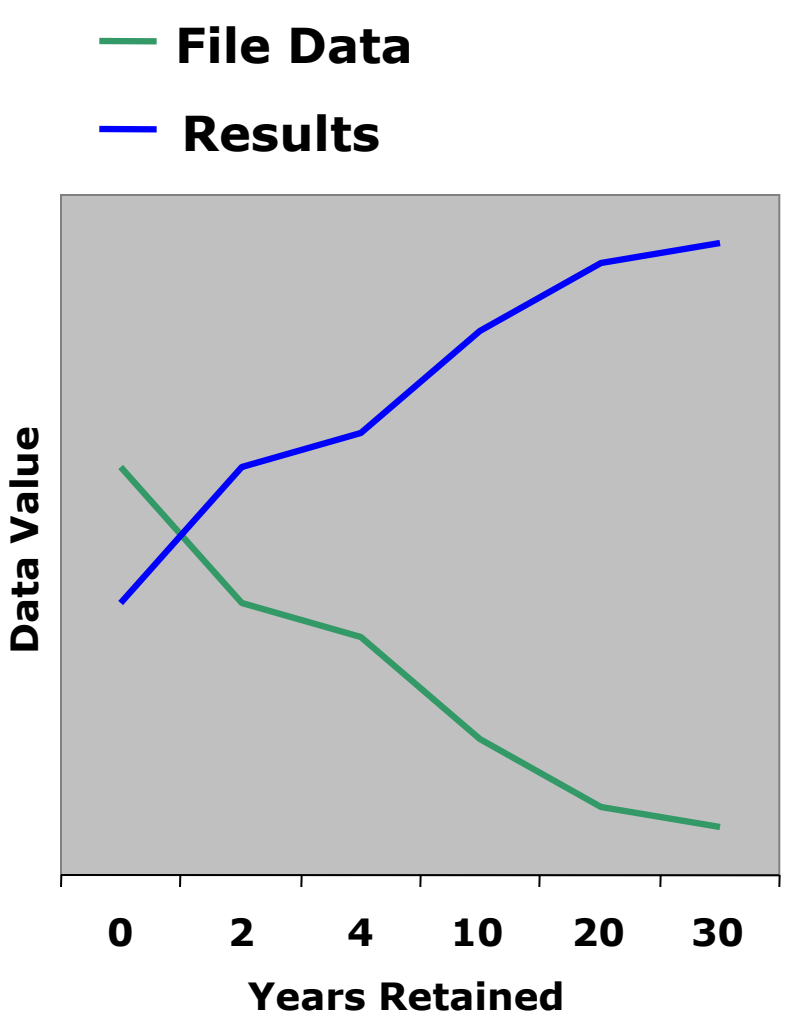
Other Considerations

- Cost and effort of repeat analysis
  - May not even be possible!
- Litigation
- Instrument data not retrievable if laboratory or manufacturer goes out of business
- Other government requirements

Common Strategies

- Keep old hardware
  - Hardware failures
  - Lack of spare parts/maintenance
  - Lack of knowledge
- Virtualize old software
  - Lack of knowledge
  - No updates
- Convert data
  - Either updated proprietary formats
  - Or platform independent standards
  - Cave: Algorithms!

What Is the Relative Value of Data?



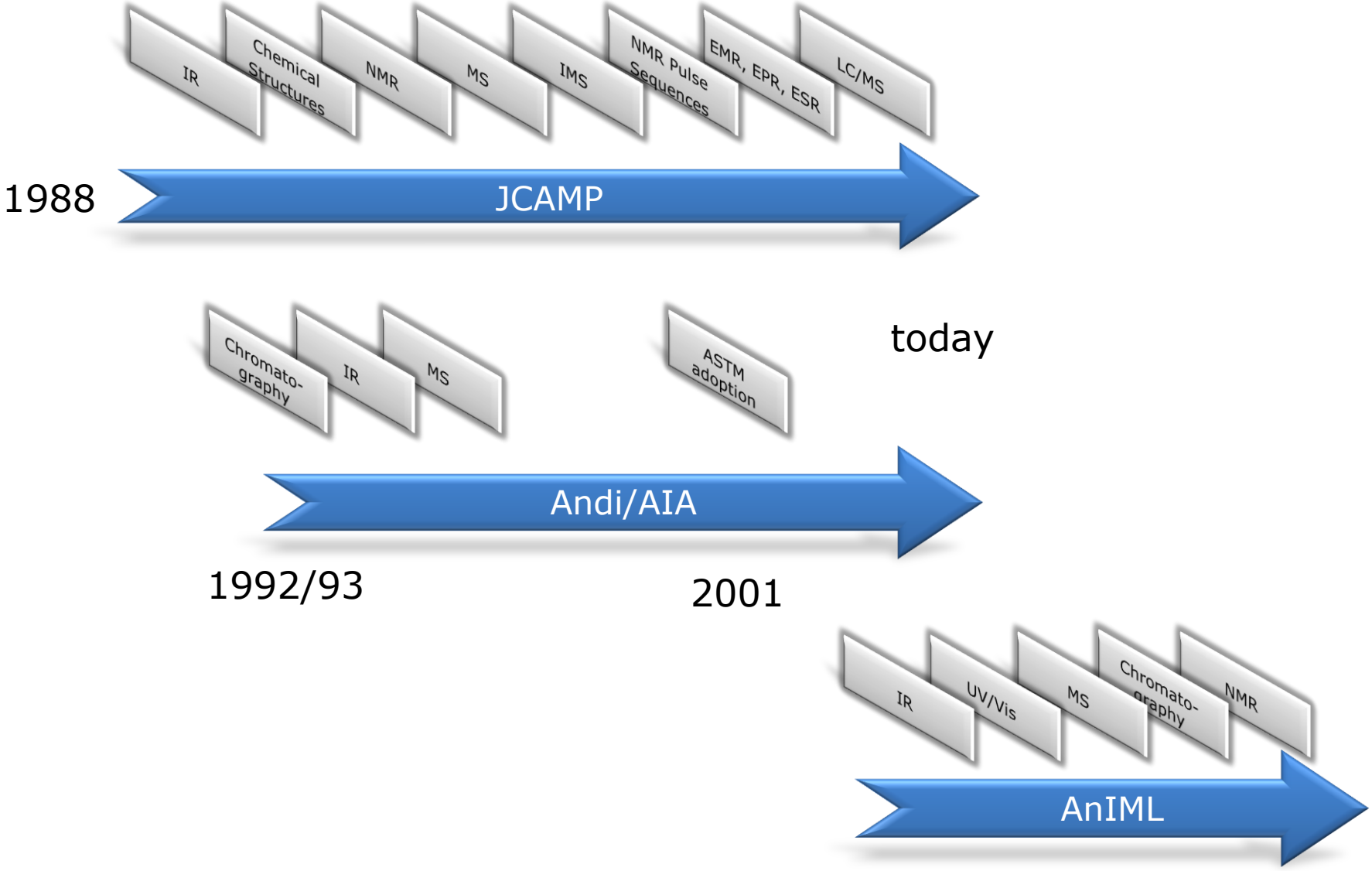
- ISPE white paper (risk-based approach to 21 CFR 11 1/7/2003) suggests **risk-based analysis**
- Raw file data most critical early in process due to need to reprocess
- Need to reprocess decreases with time, but may still be present!
- Result data value increases as data becomes older
- Decisions are based on results - in the future, this is the critical information

Standards Development Requirements

A Data Standard must...

- Fulfill 21 CFR Part 11 Requirements (complete and accurate representation)
- Be Human Readable
- Provide Normalized Terminology
- Provide Flexibility
- Provide Rigidity for Interchangeability

Current Data Standards



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

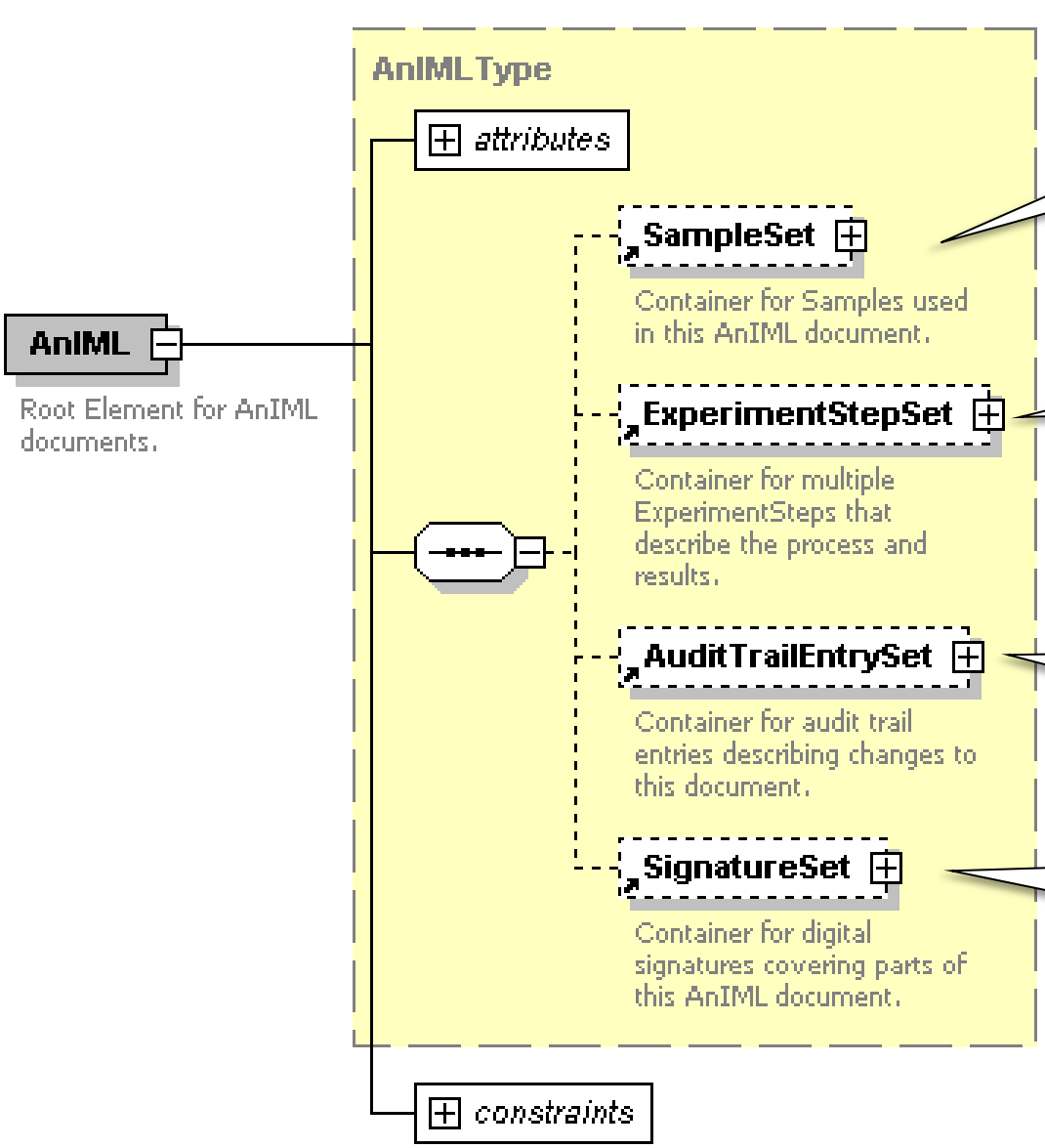
New Standards Development Requirements

- Data dictionaries from previous standards must be...
  - Mapped to the new standard
    - Faithful representation - Integrity of Content and Meaning
    - Allowed values
  - Updated to reflect current technology
    - Additions only!
- Consider mandatory/optional information
- Unify terminology between standards
- Provide clear documentation

User perspective

- Capture and secure reports
- Capture and secure original file data
- It helps to centralize and catalog data
  - Share/find/view
- Create copies in public standard formats such as JCAMP, ANDI, AnIML
  - Before the original software becomes obsolete!
  - Keep the originals!
- Migrate to new standards as they become available
- Validate conversion tool on examples – no need to validate all

AnIML Core Schema



Sample Information

Actual raw data

Who, when, where, what, why

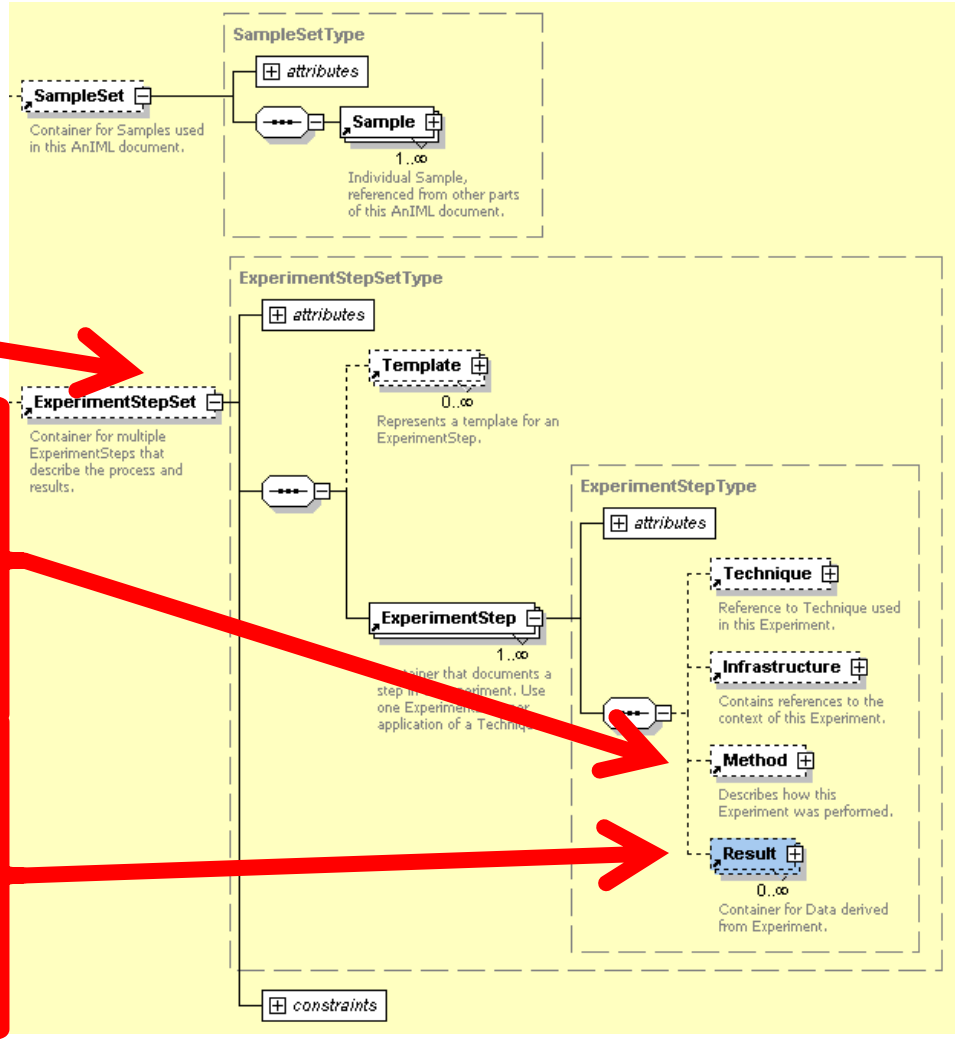
According to XML-DSIG (whole file or parts)

From JCAMP to AnIML

JCAMP-DX

- SAMPLE INFORMATION
  - ##MW
- HEADER INFORMATION
  - ##TITLE=
  - ##DATA TYPE=
- EQUIPMENT
  - ##SPECTROMETER/DATASYSTEM
- SAMPLING INFORMATION
  - ##PATHLENGTH
- REQUIRED SPECTRAL PARAMETERS
  - ##XUNITS, ##YUNITS
- OPTIONAL SPECTRAL PARAMETERS
- TABULAR SPECTRAL DATA

AnIML

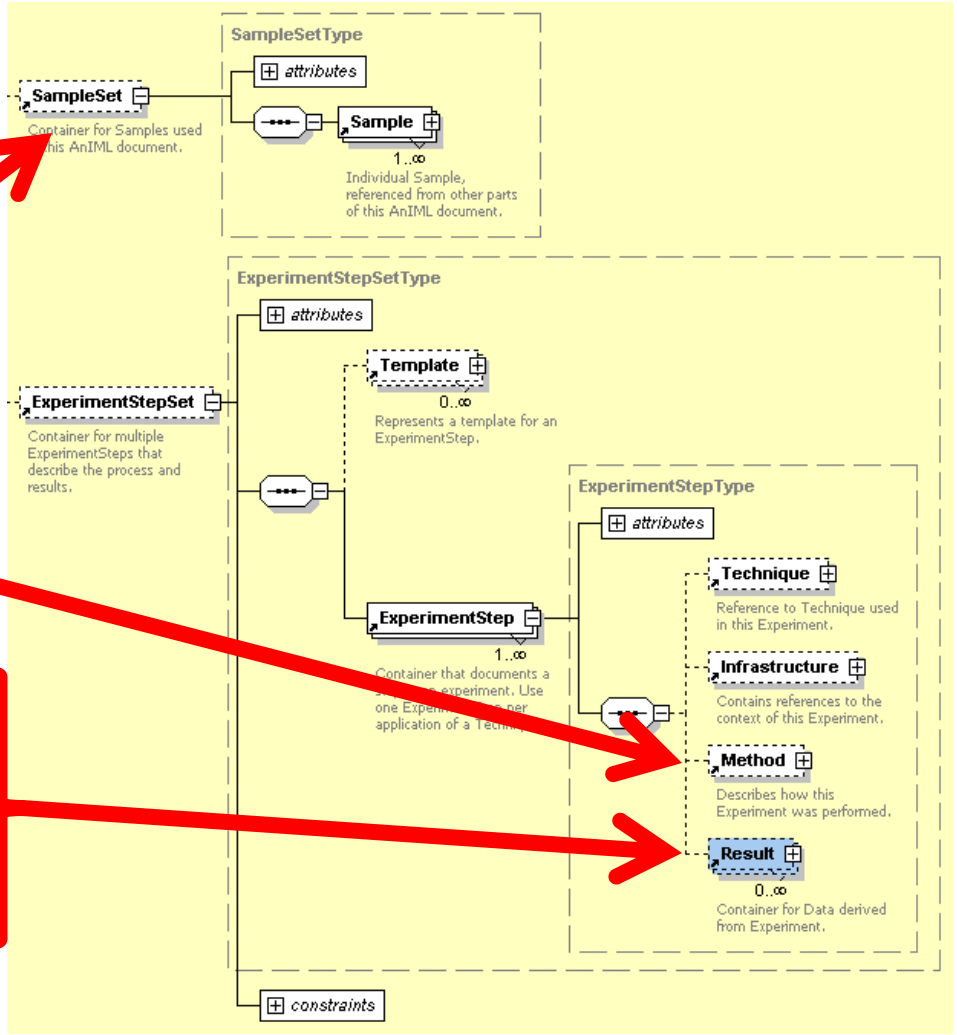


From ANDI to AnIML

ANDI (Chrom)

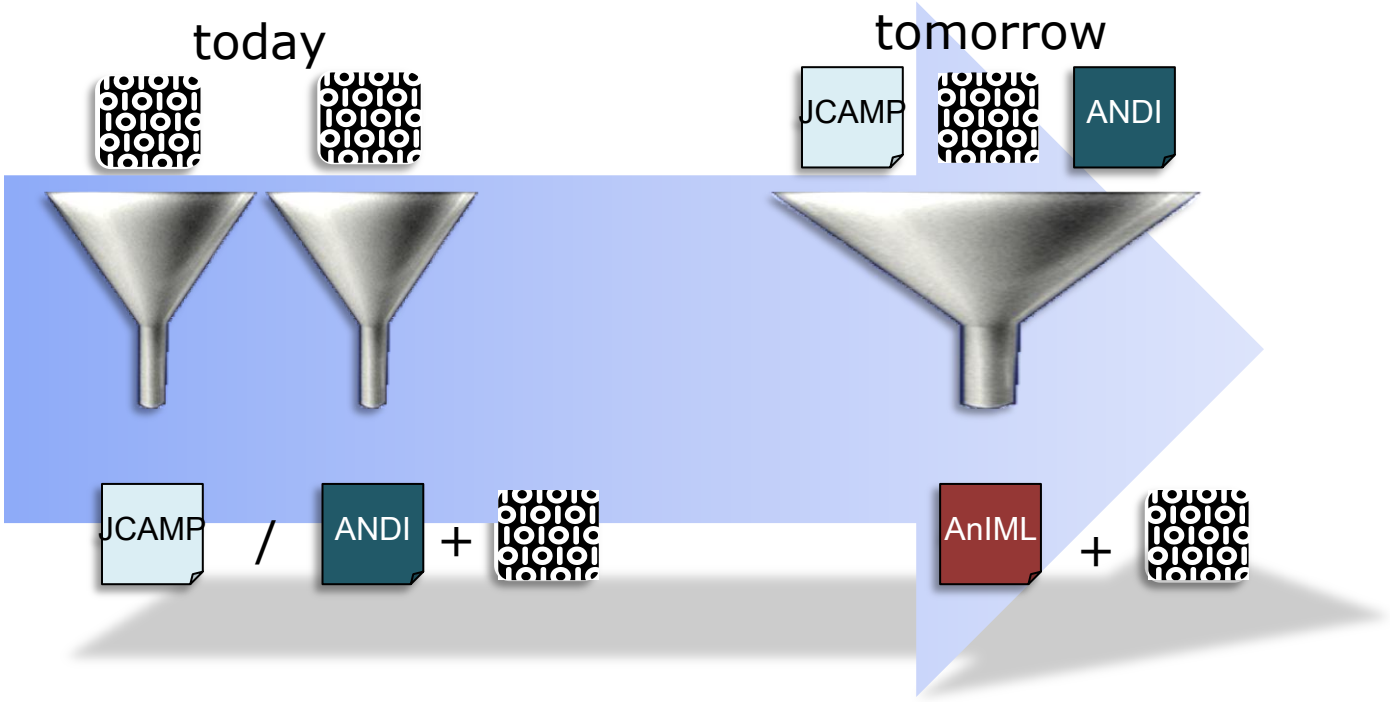
- Administrative Information Class
- Sample-Description Information Class
- Detection-Method Information Class
- Raw-Data Information Class
- Peak-Processing-Results Information Class

AnIML



Summary

- Keep original (binary) data
- Create copies in public standards
- Migrate to new standards



Sources of Information

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

Q&A

Next:

“SEDD - A Path Forward...”

