Native AnIML LCMS Processing and Viewing

Mark F. Bean, Ph.D.
Analytical Chemistry
Molecular Discovery Research
GlaxoSmithKline, Philadelphia Area

ASTM E13.15 Subcommittee (Analytical Data)

IUPAC Subcommittee on Electronic Data Standards (SEDS)

AnIML – a New Home for Old Standards

ASTM INTERNATIONAL
Standards Worldwide

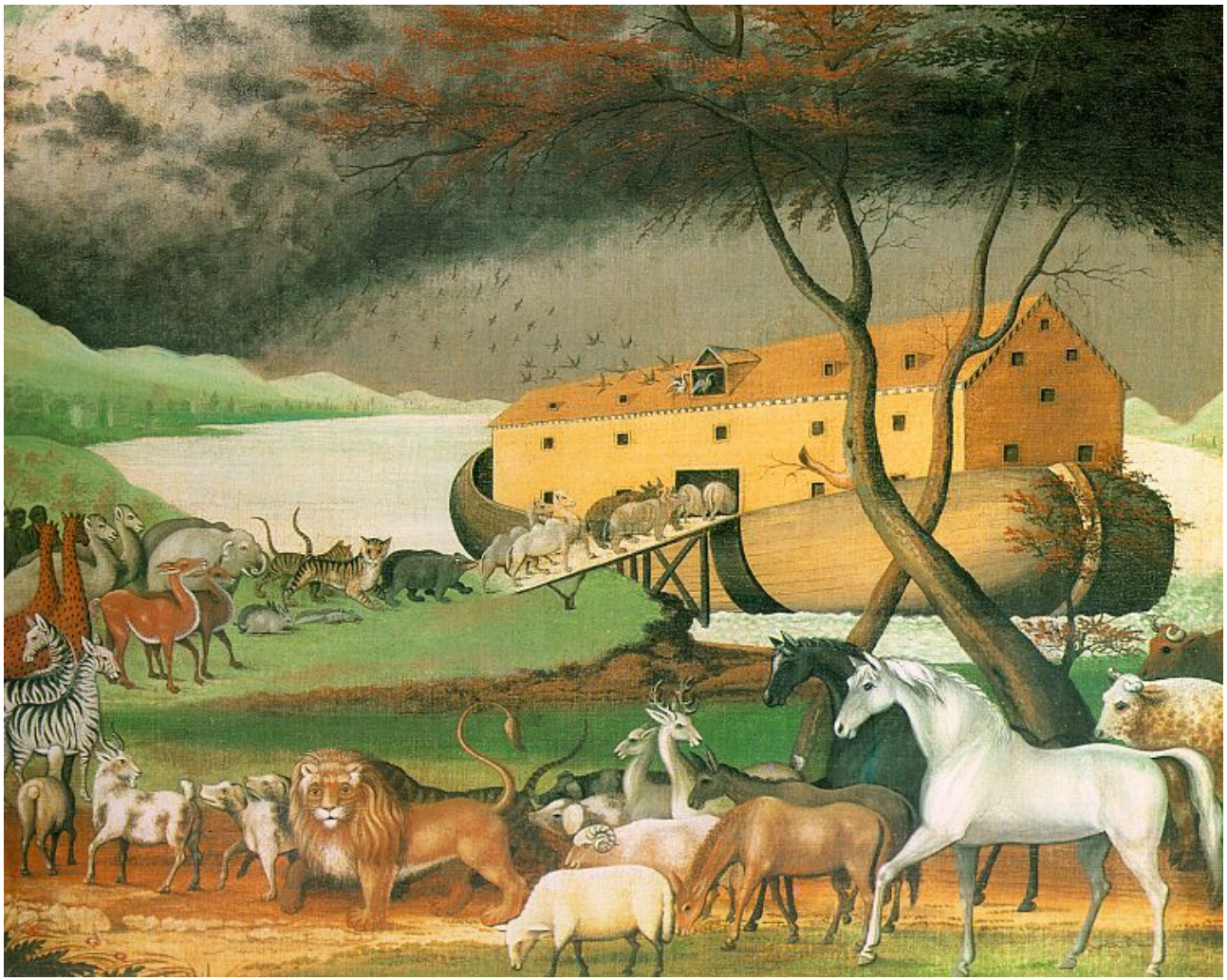
NIST

IUPAC
Committee on Printed & Electronic Publication

Subcommittee on Electronic Data Standards
XML and JCAMP-DX

JCAMP-DX

GAML



AnIML Hybrids (LC-UV-ELS-MS)

Sample Alteration

Separations
chromatography

Chemistry
post-column addition
reaction kinetics

→ *new samples*

Sample Measurement

Detection
spectroscopy
UV, IR, MS, NMR
point-detection
pH, density

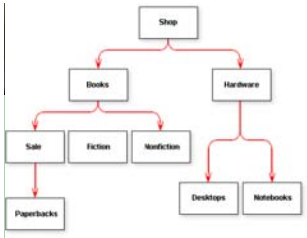
→ *data*

Data Processing

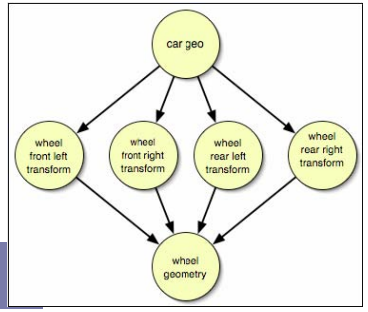
Single Trace
noise filtering
smoothing
peak finding

Multiple Runs
quantitation

→ *new data*



Hierarchical vs. Referenced Relationships



- Hierarchical (one parent only)

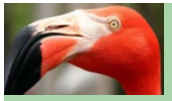
xml

```
<ExperimentStep name="Spectrum">
    <Infrastructure>
        <ParentDataPointReference  seriesID="ix1" startIndex="278"/>
    </Infrastructure>
    <Result name="Spectrum">
        <SeriesSet name="Spectrum" length="138">
            <Series name="Intensity" dependency="dependent" seriesID='
                <EncodedValueSet>AExVRwDQfkcAKpRHAKanRwAqukcANsxHACTbI
            </Series>
        </SeriesSet>
    </Result>
</ExperimentStep>
```

- key\keyref (multiple parents possible)

xsd (schema)

```
<xsd:key name="sampleID">
    <xsd:selector xpath="SampleSet/Sample"/>
    <xsd:field xpath="@sampleID"/>
</xsd:key>
<xsd:keyref name="sampleIDUsage" refer="sampleID">
    <xsd:annotation>
        <xsd:documentation>An ExperimentStep may only r
    </xsd:annotation>
    <xsd:selector xpath=".//ExperimentStep/SampleRefere
    <xsd:field xpath="@sampleID"/>
</xsd:keyref>
```



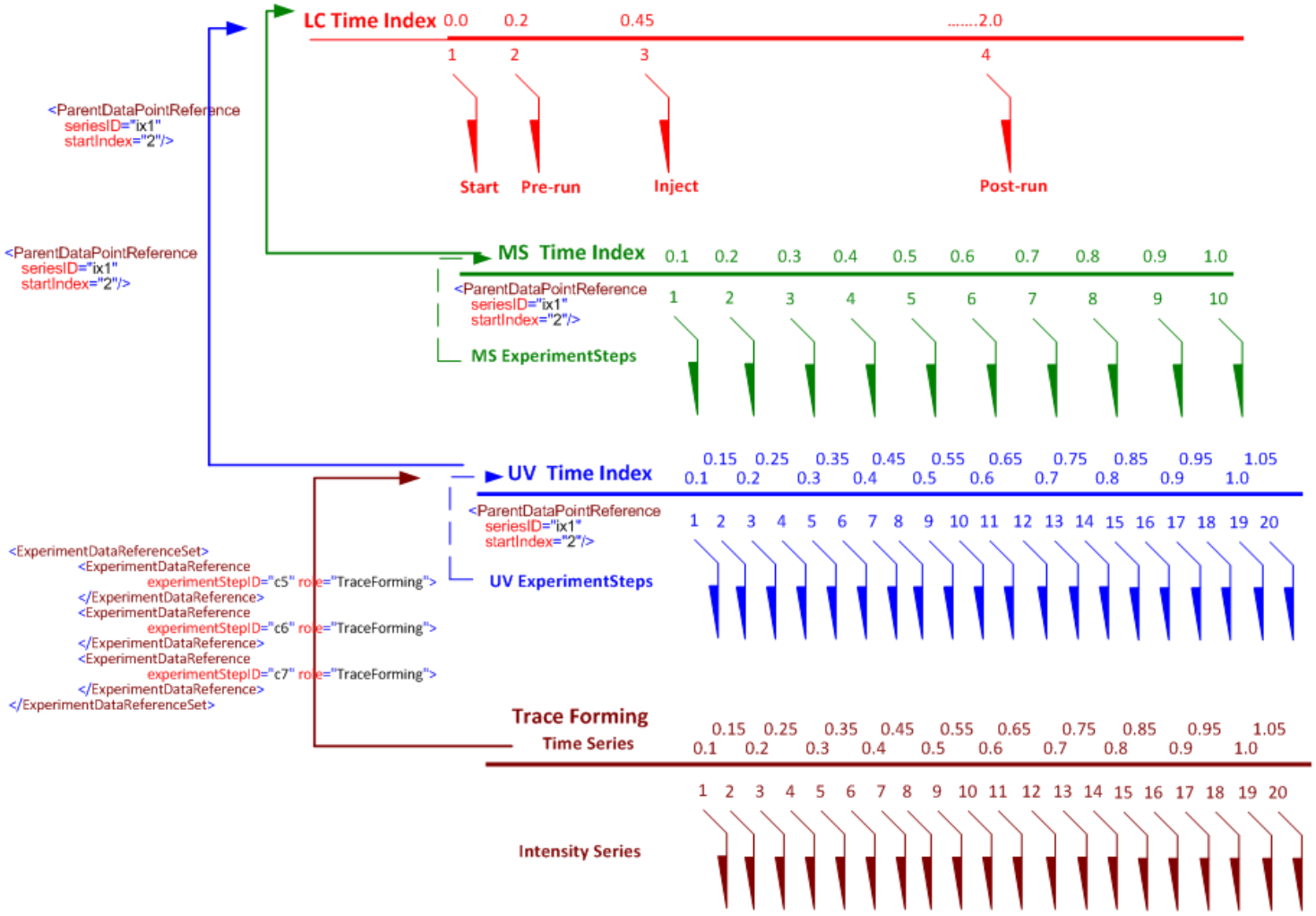
What Do XML Schema Do?

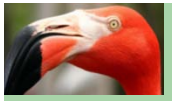
- Controls syntax -- *structure, relationships, constraints,* and *terminology* for XML files
- Used to validate XML documents
- The complex rules governing schema are difficult to understand.
- Hierarchical XML limits ability to model data relationships; a child can have only one parent in a hierarchical tree. This is resolved in schema by introduction of key, keyref (foreign key constraints).

>90% of AnIML Data Looks Like This

[illegible]

Time Synchronization Issues





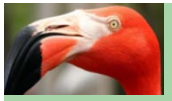
Handling Multiple Dimensions

One of the hardest areas to get right in AnIML

- Are LC-UV or LC-MS truly 2D data sets?
- or are they an array of spectra against time?

- Are chromatograms the result of processing?

- 2D NMR is an array of spectra against parameter changes (T1)



Implementing AnIML

AnIML Reader-Writer
xml

AnIML Core
XML Schema

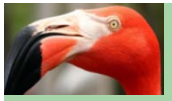
AnIML Technique
XML Schema

AnIML Viewers

Technique Term Writer

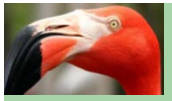
Technique Definition Doc

Vendor File Reader
binary



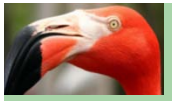
IRSample.AnIML

- Simple IR spectrum with metadata
- (Demo)
  - XML parsing (collaboration with Scimatic)
  - Intelligent Viewer (CANDI)



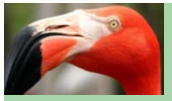
LC-UV-GStd.AnIML

- Complete Array of UV Spectra from LC run
  - Uses Templates
  - Multiple Techniques
    - Chromatography
    - UV-Vis
    - Peak Finding
- (Demo)
  - XML parsing (collaboration with Scimatic)
  - Intelligent Viewer (CANDI)

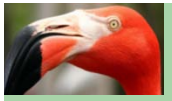


LC-UV-MS-Report.AnIML

- Summed &Subtracted UV and MS Spectra
  - Uses Templates
  - Multiple Techniques
    - Chromatography
    - UV-Vis
    - Peak Finding
    - Summation-Subtraction
- (Demo)
  - XML parsing (collaboration with Scimatic)
  - Intelligent Viewer (CANDI)



AnIML = Working Together



Acknowledgements

- **Burkhard Schaefer**, BSSN Software
- **Maren Fiege**, Waters
- **Dale O'Neill**, Agilent
- **Jamie McQuay**, Scimatic Software
- **Stuart Chalk**, UNF
- Tony Davies
- Bill Browett, GSK
- Lyle Burton, Sciex
- David Martinsen, ACS
- James Duckworth, Thermo
- Randall Julian
- Mark Mullins, Agilent
- Robert J. Lancashire, UWIMONA
- Richard Larsen

NIST

- **Gary Kramer**
- **Dennis Backhaus**
- **Peter Linstrom**
- Alexander Roth
- Anh Dao Nguyen
- Dominik Poetz
- Frank Masur
- Kordian Placzek
- Patrick Gleichmann
- Ronny Jopp