 AnIML >

A Brief Introduction to XML and AnIML

< /  AnIML >

Gary W. Kramer

Biochemical Science Division
National Institute of Standards and Technology

NIST

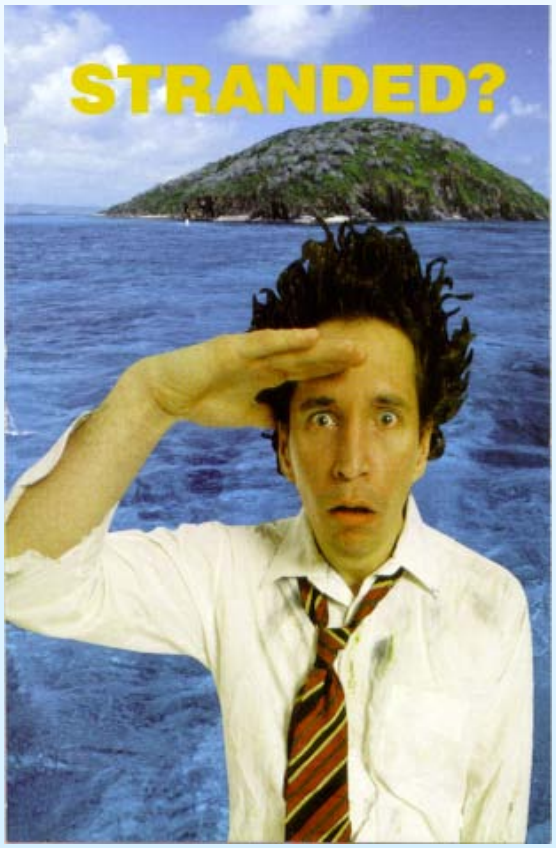
Why should I care about data interchange?

- I just want to paste a chromatogram from our data system into a Word document. But the @#$%& thing seems to be stored in some kind of in binary format.
- I just want to copy a peak table into a spreadsheet. What happened? It looks like the transporter beam scrambled my data.
- I'd like to be able to look at, expand, plot,…that xyzoscopy data that my colleague at EJU sent me. What do you mean I have to buy the vendor's $5K software package? I just want to look at the @#$%& data.

NIST

Why should I care about data archiving?

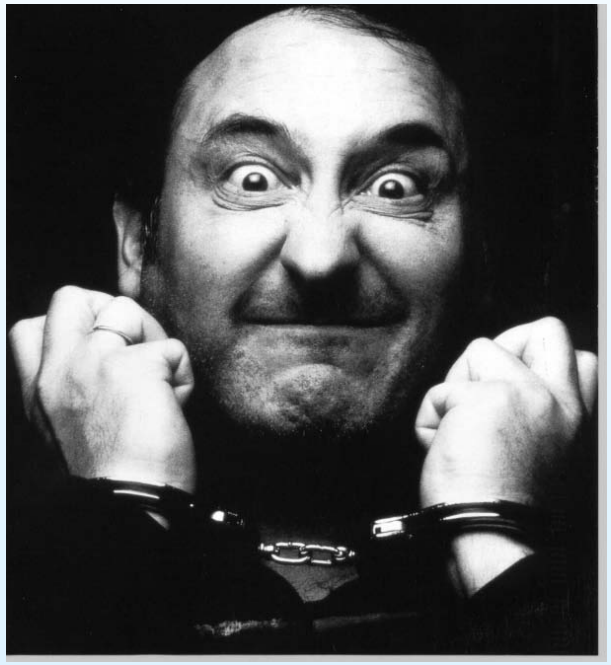
- We did that study before. The data are all here…
  on these DEC 8-inch floppy disks…
- If I could just open & read this file, I wouldn't have to…. But it's a Visicalc file that requires the CP/M operating system.
- This old NMR data "might" be valuable, but the tapes on this shelf take up too much room. I'll bet all their contents would fit onto a a couple of DVDs. Know anybody that has a 9-track tape reader interfaced to a PC?



NIST

Problems with Current Result Data Handling Mechanisms

- Native Data Formats
  - Proprietary Formats
  - "Metadata" Separated from Result Data
    - Metadata & Data in Multiple Files
    - Metadata Not Available in Electronically
    - No Way to Link Metadata with Result Data
- Interchange Data Formats
  - Available for Only a Few Important Techniques
    - ANDI — GC, LC, MS
    - JCAMP-DX — IR/FTIR, NMR, UV/Vis, IMS
  - Fixed Order, Fixed Syntax, Immutable Formats
  - Content Limitations
  - Inconsistent Implementations
- Formats Incompatible with Modern Network Technologies



NIST

In the beginning...

- Consortium on Automated Analytical Laboratory Systems (CAALS)
  - CAALS-I Communication Protocol
  - High-Level Communications Protocol
  - Common Command Set
  - Device Capability Dataset
- Laboratory Equipment Control Interface Specification (LECIS) - ASTM E1989-1998.
- System Capability Dataset
- How to Deal with Result Data?
- NIST In-House Data Interchange/Archiving Needs

NIST

Goals for New Result Data Handing

- Extensible
  - Easy to Add New Elements without Breaking Existing Applications
- Flexible
  - Useful for Diverse Needs: Interchange, Interconversion, Archiving...
- Useable & Maintainable
  - Easy to Create, Use, Adapt, Maintain...
  - Readily Available Tools
- Acceptable
  - Use Standard Mechanisms Accepted by Mainstream Computing
- Network Friendly

☞ Extensible Markup Language

NIST

What is a markup language?

- Markup is a way using delimiting tags to tell a computer program something about a data element.
- Hypertext Markup Language (HTML) is used for describing text on a page.

```
<head>
    <title>What is a markup language?</title>
</head>
```

- Extensible Markup Language (XML) is used for describing data.

```
<date>2006-03-25-05:00</date>
```

NIST

Intro to XML

- Markup
  - Encloses parts of a document within tags
  - Structures tags hierarchically
  - Has attributes to specify tag
  - ID attributes allow unique identification & reference from other elements
  - XML files are fully tagged text files, starting from a root tag

```
<title>analysis</title>
```

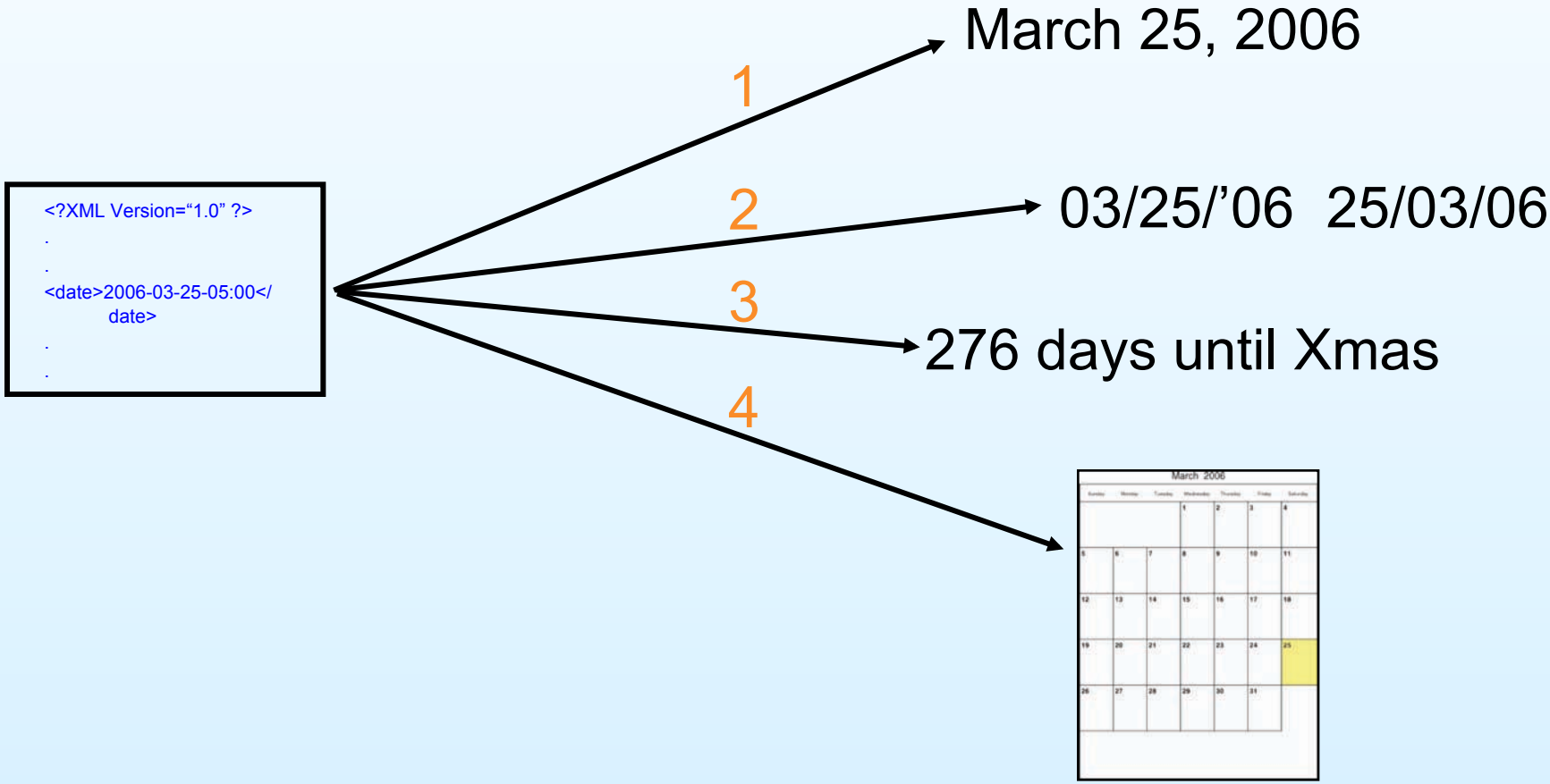
```
<sample>
   <number>1</number>
   <name>water</name>
</sample>
```

```
<person no="1">
   Alex
</person>
```

```
<SpectroML version="1.0">
        .
        .
        .
        .
</SpectroML>
```

So what!

XML Data | Application Routine | Result



XML in 10 Points

- XML is for structuring data
- XML looks like HTML
- XML is text, but is not meant to be read
- XML is designed to be verbose
- XML is a family of technologies
- XML is new, but it has a history & a heritage
- XML turns HTML into XHTML
- XML is modular
- XML is license-free, platform-independent, and well-supported
- XML is a standard maintained by the W3C

www.w3.org/XML/1999/XML-in-10-points

Intro to XML

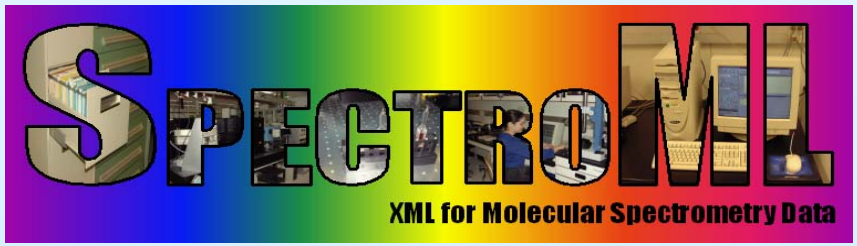
- Stylesheet
  - A stylesheet processor transfers one XML document into another
  - Allows conversion between data formats and visualization of XML data
  - Example: HTML output in browser
- Digital Signature
  - Signs an XML document with a unique value based on its content

- Permits secure and unaltered transmission and content validation

NIST

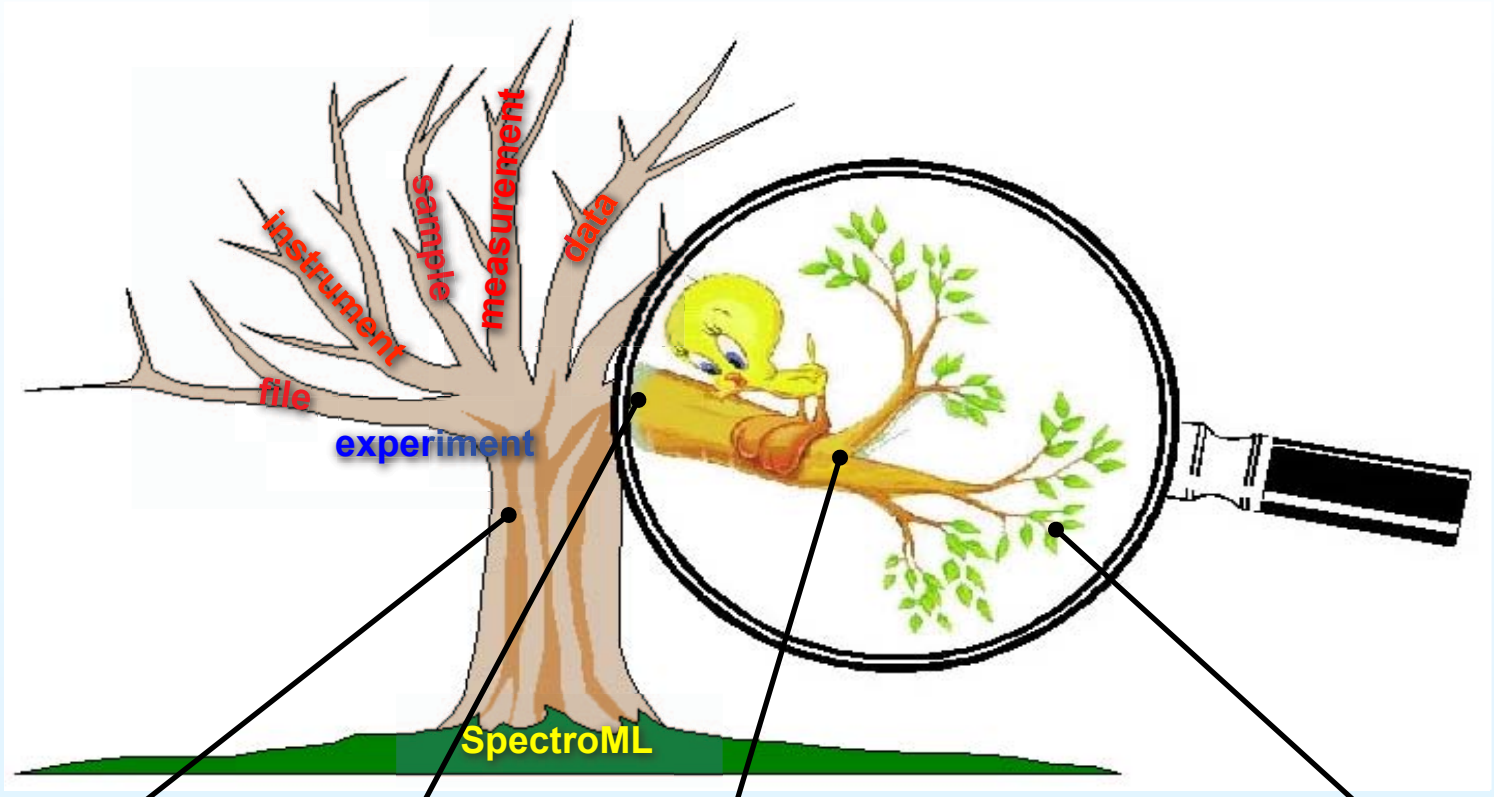
SpectroML

- SpectroML is a markup language for spectroscopy data based upon:
  - XML
  - JCAMP-DX (IUPAC)
  - ANDI/NetCDF (ASTM)
  - Thermo-Galactic GRAMS and SPC file format
  - Data definitions from instrument manufacturers
  - ASTM Definitions
- SpectroML is defined for UV-Visible spectral data
- www.xml.org→registry→ chemistry→SpectroML

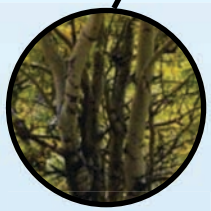


NIST

SpectroML Structure




experiment

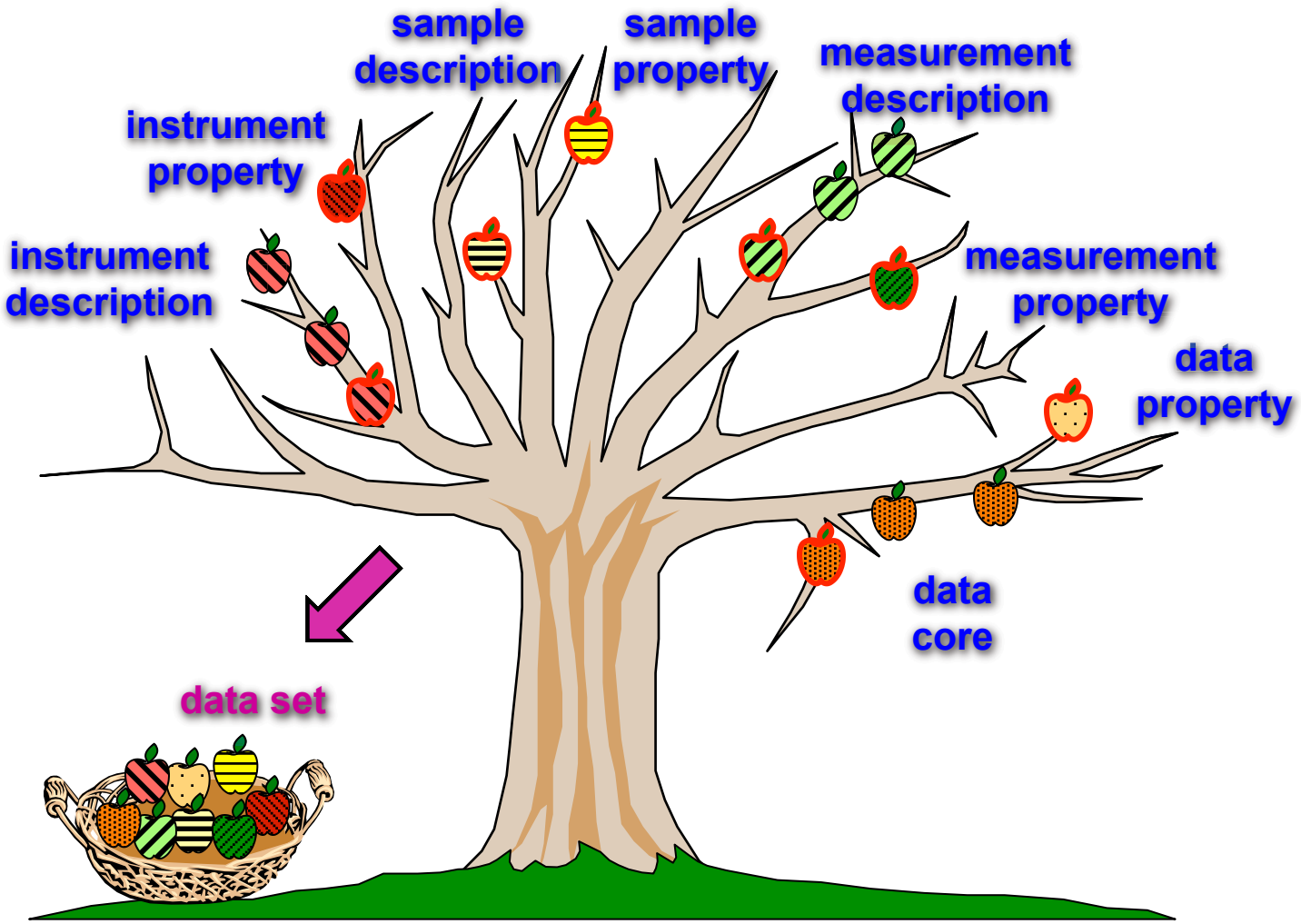

groups


blocks


sections


elements

SpectroML Dataset



SpectroML Elements

Instrument

- instrumentDesignation
  - identifier
  - manufacturer
  - model
  - owner
  - location
- instrumentApplication
  - software
  - version
  - operatingSystem
  - firmware
  - operator
- instrumentSetting
  - resolution
  - linearDispersion
  - spectralBandWidthRange
  - wavelengthRange
  - absorbanceRange
  - detectorTypes
  - sourceTypes
- instrumentParameter
  - slitWidth
  - spectralSlitWidth
  - beamChannel
  - sampleHolder
  - samplePosition
  - scanSpeed
  - pointSeparation

Sample

- sampleDesignation
  - identifier
  - owner
  - location
  - casNumber
  - formula
  - storageMethod
  - disposalMethod
- samplePreparation
  - procedureMethod
  - timeStamp
  - operator
  - supplier
  - preparationDescription
- sampleAttribute
  - molecularWeight
  - meltingPoint
  - boilingPoint
  - density
  - refractiveIndex
- sampleParameter
  - state
  - pathLength
  - amount
  - pressure
  - temperature

Measurement

- measurementDesignation
  - identifier
  - title
  - owner
  - laboratoryReference
- measurementExecution
  - project
  - timeStamp
  - operator
- measurementParameter
  - measurementType
  - scanMode
  - referenceSample
  - filter
  - signalNoise
  - scanNumbers
  - scanDuration
- measurementCorrection
  - qualificationTimeStamp
  - qualificationReference
  - proficiencyTimeStamp
  - proficiencyReference
  - transmittanceTimeStamp
  - transmittanceReference
  - wavelengthTimeStamp
  - wavelengthReference

Data

- dataParameter
  - axisLabel
  - axisUnit
  - minimumValue
  - maximumValue
- dataCalculation
  - scaleFactor
  - numberPoints
  - pointIncrement
  - startValue

Data Values

- Single data points
- A single spectrum
- Multiple spectra
- Multi-dimensional data

NIST

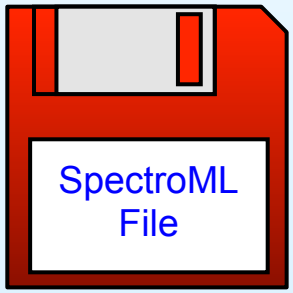
Example SpectroML File

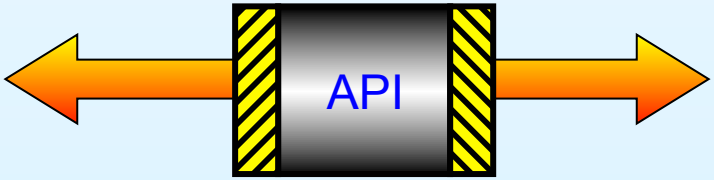
```
<SpectroML version="1.0">
  <experiment type="UV/Vis" language="en-us">
    <file>
      <title>sample experiment</title>
      <timeStamp>
        <date>2002-11-07</date>
        <time>10:12:43</time>
      </timeStamp>
      <path pathId="p0"
        ...
        dataPropertyLink="dp0"
        dataCoreLink="dc0"/>
    </file>
    <instrument>
      ...
    </instrument>
    <sample>
      ...
      <sampleProperty samplePropertyId="sp0">
        <sampleParameter>
          <state>liquid</state>
          <amount unit="ml">5</amount>
        </sampleParameter>
      </sampleProperty>
    </sample>
    <measurement>
      ...
    </measurement>
    <data>
      <dataProperty dataPropertyId="dp0">
        <dataParameter>
          <axisLabel>
            <axis dim="x">Wavelength</axis>
            <axis dim="y">Transmittance</axis>
          </axisLabel>
          <axisUnit>
            <axis dim="x">nm</axis>
            <axis dim="y">%T</axis>
          </axisUnit>
        </dataParameter>
      </dataProperty>
      <dataCore dataCoreId="dc0">
        <values dim = "x">270 576</values>
        <values dim = "y">10 23</values>
      </dataCore>
    </data>
  </experiment>
</SpectroML>
```

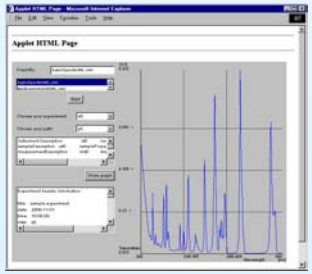
NIST

Applying SpectroML

- DTD, Schema, & Sample File for SpectroML
- Stylesheet for Textual Output in Web Browser
- SpectroML API for Java and C++
- Demo Applications & Applets Using APIs



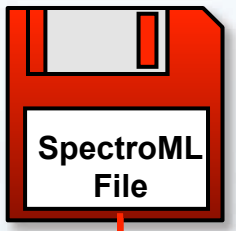


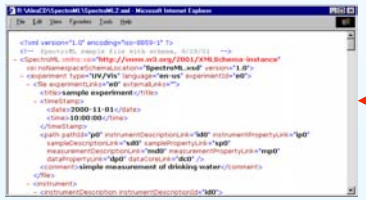


API

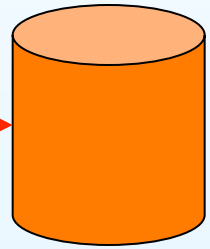
- Creating
- Selecting
- Editing
- Storing
- Searching
- Validating

Applying SpectroML



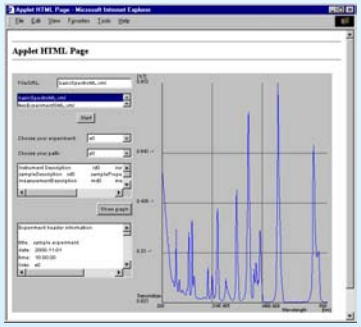


Web Browser

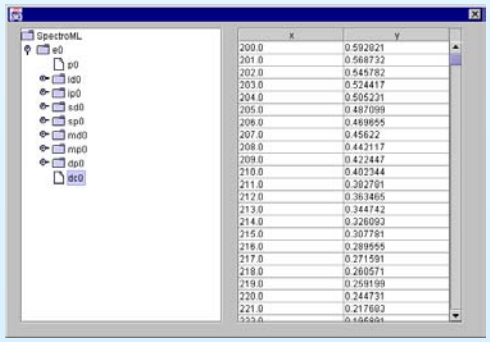


Database

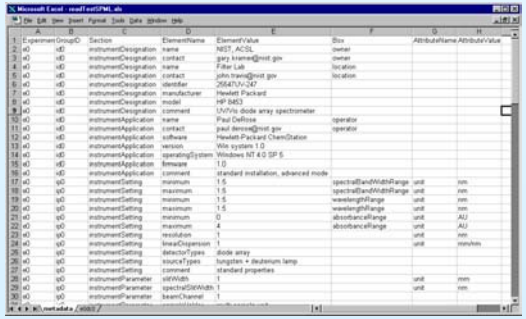
Java Applet



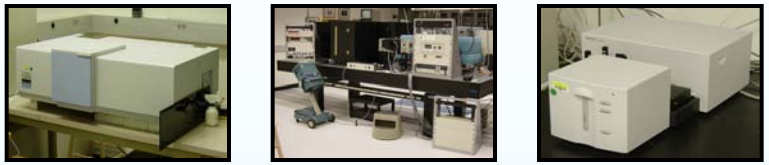
Java Program



Converting to Another Format

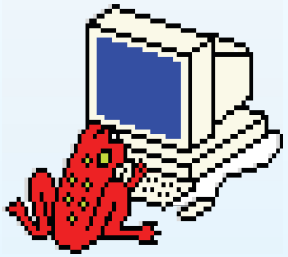


Applying SpectroML at NIST

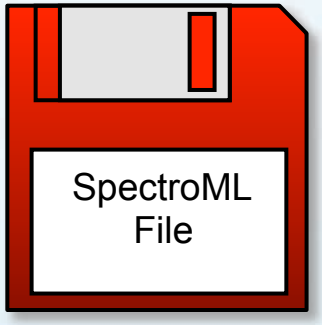


Result Data In Device-Dependent Formats

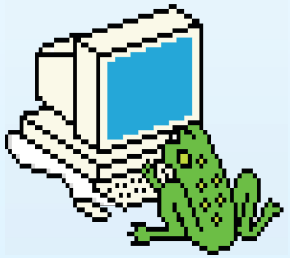
Device-Dependent Converter Software



[User Interface]



Result Data In SpectroML



User Interface

Result Data In Database

NIST

Generalized Analytical Markup Language (GAML)

- Represent Analytical Data from Multiple Spectroscopy & Chromatography Techniques
- Compact, Simple Dictionary & Hierarchy (Schema)
- Use XML Datatypes & Hierarchical Structure to Mimic Relationships in Data Sources
- Avoid Parameter "Mapping" Problem
- Minimize the Need for Complex Dictionaries
- Permit Future Expansion
- Keep File Sizes Small
- www.gaml.org
- www.xml.org→registry→chemistry→GAML

NIST

James Duckworth, Thermo Electron Corporation

An XML-Based Standard for Molecular Spectrometry and Chromatography Result Data

- Possibility of XML-Based Approach Raised
  - ASTM E01.25 and E13.02 Meetings PittCon '00 Atlanta
- Demo of SpectroML and Applications
  - ASTM E13.01 Meeting PittCon '01 New Orleans
- First Organizational Meeting Held
  - ASTM E13.01 Meeting EAS '01 Atlantic City
- Task Group Organizational Meeting
  - ASTM E13.01.03 Meeting PittCon '02 New Orleans
- Task Group Working Meeting
  - ASTM E13.01.03 Meeting Shimadzu, Inc. 9/'02
- New Subcommittee Meeting
  - ASTM E13.15 Meeting PittCon 3/'03 Orlando
- E13.15 Meets at EAS and PittCon
  - AnIML Working Group Meets "Virtually" ~biweekly

NIST

ASTM INTERNATIONAL
Standards Worldwide

AnIML an XML-Based Standard for Analytical Result Data

- SpectroML and GAML serve as starting points for an XML-based standard interchange format for molecular spectrometry and chromatography.

- Instrument manufacturers, data system & LIMS developers, software developers, end-users, consensus standards organizations, regulatory agencies, and other interested parties are invited to participate in this effort.

NIST

XML in Chemistry

ASTM INTERNATIONAL Standards Worldwide

Creating Analytical Information Markup Language (AnIML)

- Creating AnIML does NOT mean "starting over."
- AnIML should be built on existing ASTM, IUPAC, instrument vendor, and LIMS-developer efforts to define common data dictionaries.
- Once the schemas for AnIML are in place, straight-forward translators can be written to bridge current datasets to the new standard.
- AnIML should be developed in a way that makes it extensible to multiple techniques, yet avoids duplication of effort and dictionary entries.

NIST

More Information

- XML
  - http://www.w3c.org
  - http://www.xml.org
  - http://www.xml.com
  - http://www.xmlfiles.com
- SpectroML
  - ftp://caals.nist.gov/pub/download/spectroml
  - http://www.xml.org→registry→chemistry→SpectroML
- GAML
  - www.gaml.org
  - http://www.xml.org→registry→chemistry→GAML
- ANIML
  - http://animl.sourceforge.net
  - http://www.iupac.org/standing/cpep.html

NIST

Acknowledgements

- NIST Systems Integration for Manufacturing Applications (SIMA) Program
- SpectroML
  - Professor Reinhold Schäfer - Fachhochschule Wiesbaden
  - Alexander Rühl - Definition, DTD, & Schema
  - Martin Peschke - Applications, Applets, C++ & Java APIs
  - Aykut Arslan - Instrument-to-SpectroML Applications
  - Anh Dao Nguyen - SpectroML-to-Database Application
- GAML
  - James Duckworth - Thermo Electron Corporation

NIST

Acknowledgements

- AnIML
  - Burkhard Schäfer - BSS&N - AnIML Core & Technique Schemas
  - Dominik Pötz - SAP, AG - AnIML Core & Technique Schemas, AnIML Validator, & Technique Creator
  - Maren Fiege - Waters Informatics - AnIML Technique Definition Base Documents
  - Tony Davies - Waters Informatics - Chair, IUPAC SubCommittee on Electronic Data Standards
  - David Martinsen - American Chemical Society - E13.15 Secretary
  - Anh Dao Nguyen - AnIML Example Data Files, Technique Base Documents for UV/Vis and IR, & Generic AnIML Viewer
  - Peter Linstrom - NIST
  - Mark Bean - GlaxoSmithKline
  - Bob McDonald - JCAMP-DX
  - Ronny Jopp & Alexander Roth - Incorporating UnitsML & NDRs for AnIML

NIST