

<  AnIML >

Introduction to the Analytical Information Markup Language

< / AnIML >

Gary W. Kramer

Biochemical Science Division
National Institute of Standards and Technology

This introductory presentation will:

- Detail why we need an markup language for interchanging and archiving analytical chemistry data
- Describe the notion of markup languages in general and extensible markup language (XML) in specifics
- Outline the requirements for developing the Analytical Information Markup Language (AnIML)

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.



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 couple of DVDs.
Know anybody that has a 9-track tape reader interfaced to a PC?



Why should I care about data archiving?

- The FDA wants us to save the data how long?



In terms of retrieving archived data, we were better off with paper forty years ago before the arrival of lab computers....

We can still read the data in Newton's notebooks today; will they be able to read ours in 100, 20, or even 5 years?

Problems with Current Result Data Handling Mechanisms

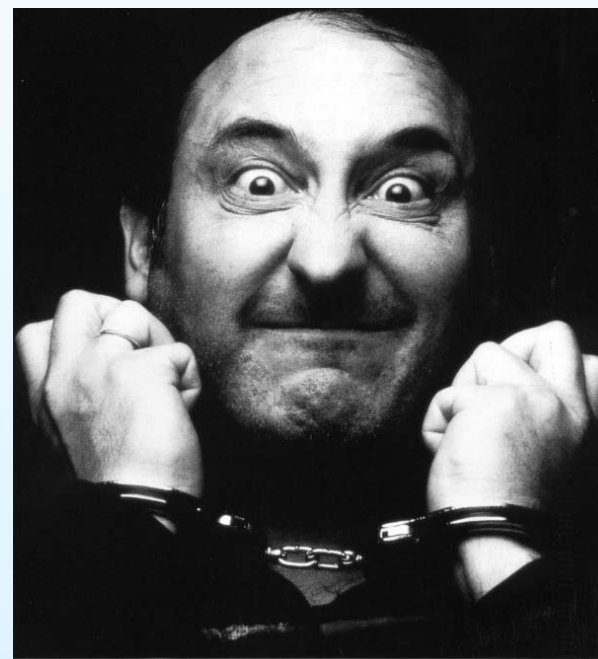
■ Native Data Formats

- 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



AnIML is a series of standards that enable the:

- Interchange
- Archiving
- Interconversion
- Verification
- Validation
- Viewing

of analytical chemistry data.

AnIML uses XML

- AnIML is based on the W3C XML (Extensible Markup Language) Standards
- Useable & Maintainable
 - Platform independent
 - Easy to create, use, adapt, maintain...
 - Readily available tools
- Acceptable
 - Uses standard mechanisms accepted by mainstream computing
 - XML is supported by nearly all major software vendors (Microsoft, IBM, Sun Microsystems, Oracle...)
- Network Friendly

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>
```

So what!

XML Data

Application Routine

Result

```
<?XML Version="1.0" ?>
.
.
<date>2006-03-25-05:00</
date>
.
.
```



March 2006						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
		1	2	3	4	
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

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



AnIML Requirements

- Flexible enough to represent analytical chemistry data
- Strongly constrained to ensure data interchange and interoperability and to enable creation of generic data viewers
- Simple to understand
- Extensible to satisfy current and future needs of vendors, corporate interests, users, and new technologies

AnIML Requirements

- Long-lived
- Human readable
- Support digital signatures
- Verifiable
- Validatable
- Support audit trail
- Viewable
- Support facile interchange with databases

AnIML Requirements

- Track analysis content (metadata)
- Support sufficient metadata for result interpretation
- Support sufficient metadata for reprocessing
- Support conversion of datasets from prior standards (JCAMP-DX and Andi)
- Support common analytical techniques
 - UV/Vis, IR, Chromatography, NMR, MS, IMS
 - Hyphenated techniques
 - Multi-sample techniques such as array-based assays (titer plates), kinetics experiments, analytical mapping

AnIML Requirements

- Platform independent
- Distinguish between raw, processed, re-processed, and simulated data
- Provide sufficient commonality so that technique-constrained software can read technique-specific sections of multi-technique files

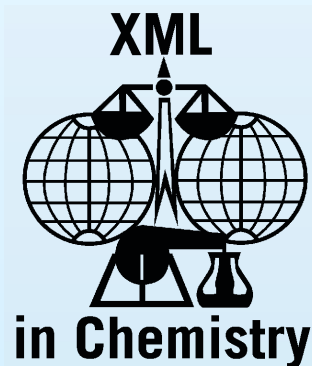
Creating Analytical Information Markup Language (AnIML)

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

AnIML an XML-Based Standard for Analytical Result Data

- 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



More Information

- ANIML

- <http://animl.sourceforge.net>
- <http://www.animl.org>