< AnIML >

Getting Started with AnIML 1.0

< / AnIML >

Gary W. Kramer

Biochemical Science Division
National Institute of Standards and Technology

NIST

This presentation will:

- Describe Why AnIML Standards are Needed
- Describe What the AnIML Standards Are
- Describe the Components that Make Up AnIML
- Describe What AnIML 1.0 Supports
- Discuss Where We Are with AnIML 1.0
- Discuss Getting Started with AnIML

NIST

We got TROUBLE, my friends, right here in Science City and that starts with T, and it rhymes with D, and it stands for DATA…

- Can't Move It
  - From Instrument to Instrument
  - From Instrument to Application
  - From Application to Application
- Can't Interconvert It
- Can't Find All its Parts
- Can't Look at It Easily
- Can't Use It Easily with Modern Computing and Networking Technologies

NIST

There is Crisis in Archiving and Retrieving Data

- We're drowning in data, yet increasingly we cannot find our stuff .

- Data mining can increase the value of archived data, but…

- Regulatory agencies are now demanding extraordinarily long data retention

- 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 folks be able to read ours in 100, 20, or even 5 years?

NIST
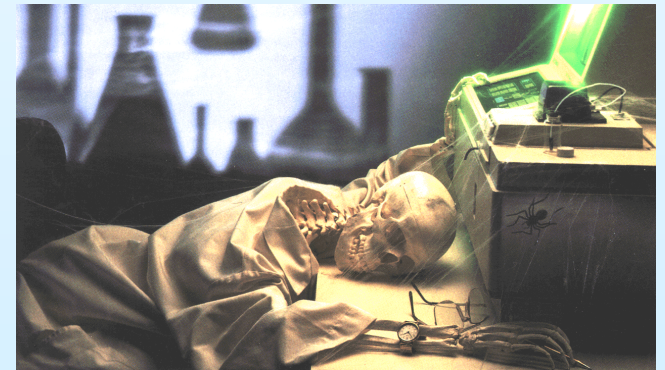
What Are the AnIML Standards?

- The Analytical Information Markup Language (AnIML) is a markup language based on Extensible Markup Language (XML) that provides a structure, syntax, and format for recording, reporting, exporting, importing, and archiving the result data and result metadata outcomes from analytical chemistry experiments.

- The AnIML Standards are projects of ASTM Committee E13 on Spectroscopy and Separation Science, Subcommittee E13.15 on Analytical Data

The Analytical Information Markup Language (AnIML)

❑ Has been Developed Expressly for Analytical Chemistry Result Data and Metadata

❑ Defines a Data Format for Representing All Types of Result Data and Metadata

❑ Provides a System for Identifying Content (Data Elements and Attributes) with Labels (Tags)

❑ Provides a Model and the Means for Defining the Structure, Content, and Semantics of Result Data and Metadata in Documents

❑ Expresses Shared Vocabularies

❑ Allows Machines to Carry Out Rules Made by People

NIST

Some AnIML Files are Simple

Some AnIML Files are Complex

Key AnIML Features

- Extensible – Supports New Vendor and User Requirements

- Verifiable - Supports Digital Signing of Full and Partial Documents to Ensure Data Integrity

- Traceable - Supports Audit Trails to Track Data Alterations and Create a Dataset "Chain of Custody"

- Validatable - Permits Checking of Datasets Against Data Requisites from Multiple Data Authorities
  - XML Syntax and Schema Language
  - AnIML Schemas and Technique Definitions
  - AnIML Naming and Design Rules
  - Semantic Rules
  - Data Requisites from Higher Data Authorities
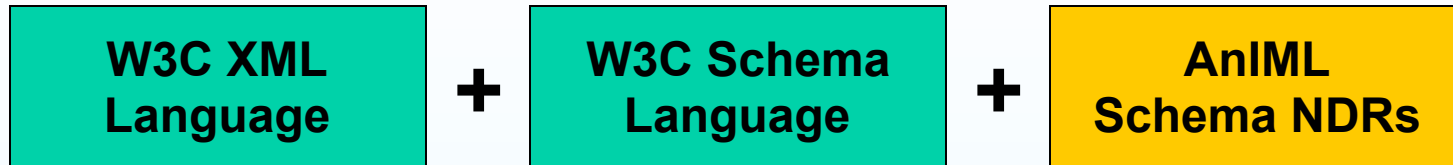
In the Beginning, There was...

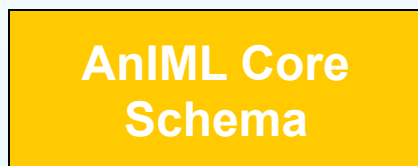|	
W3C XML Language	**W3C Schema Language**

+

To which We Added…

W3C XML Language	+	W3C Schema Language	+	AnIML Schema NDRs

NIST

And this Begat…

W3C XML Language	**+**	W3C Schema Language	**+**	AnIML Schema NDRs

AnIML Core Schema
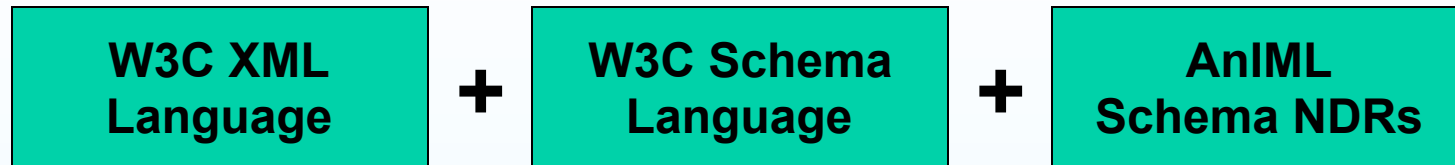
A Generic Data Container
Capable of Holding Any Analytical
Data and Metadata

And it was Good…But Lo,
It was Too Generic
to Support Interoperability

NIST

And so, We Created a Helper for the Core…

W3C XML Language	**+**	W3C Schema Language	**+**	AnIML Schema NDRs

AnIML Core Schema

AnIML Technique Schema
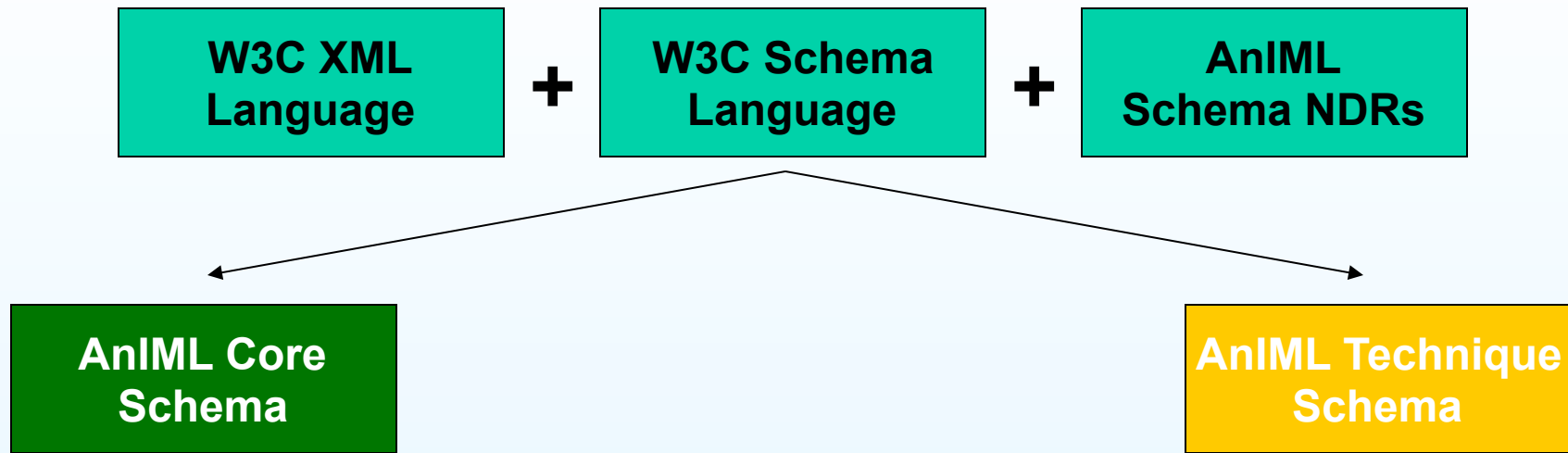
A Mechanism to Constrain the Core And Make the Tags Meaningful to Technique Users

And It was Good…,But, It still Needed Rules

NIST

And Rules Enabling It to Constrain the Core…

W3C XML Language	**+**	W3C Schema Language	**+**	AnIML Schema NDRs

AnIML Core Schema

AnIML Technique Schema
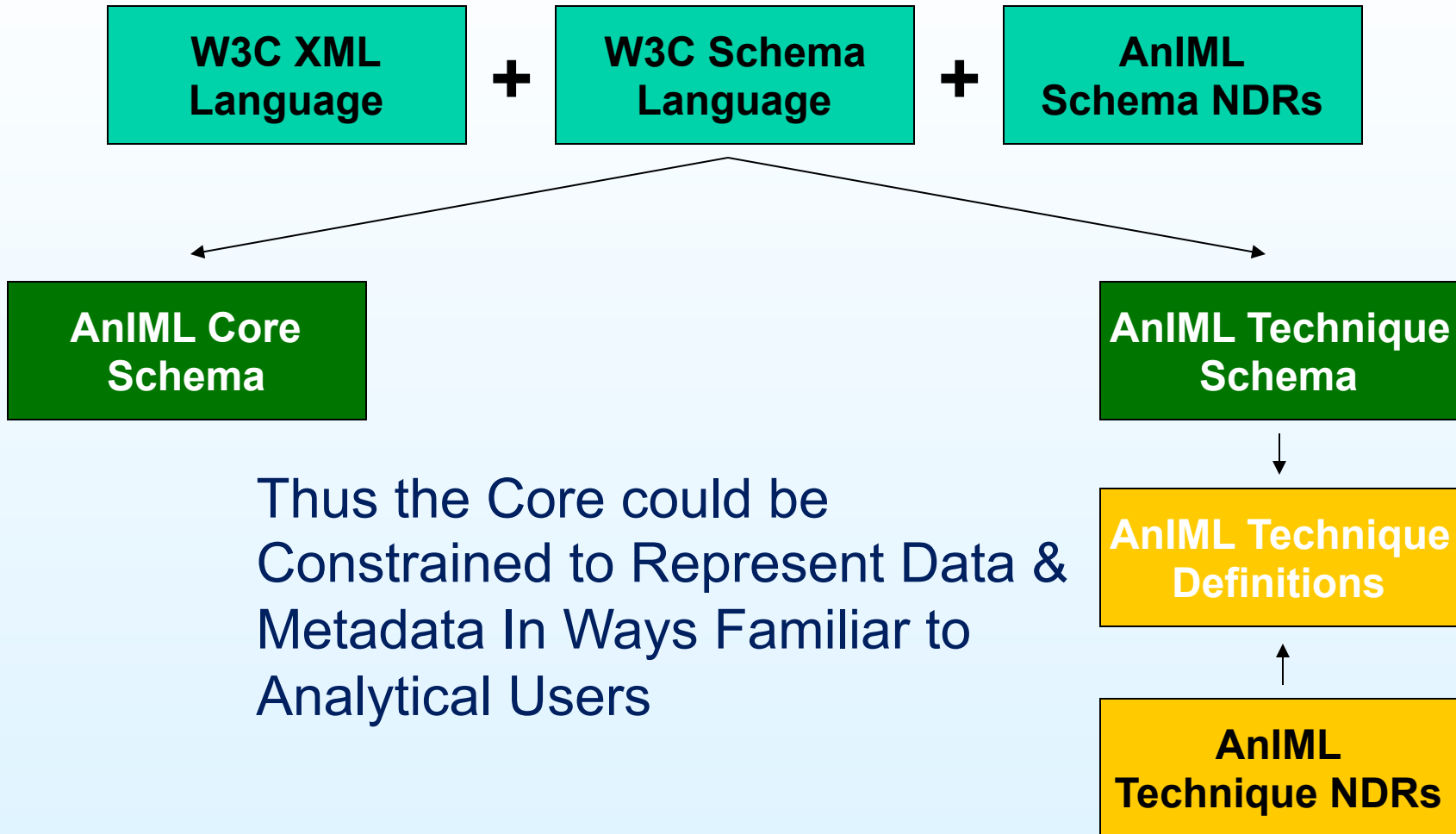
Thus the Core could be Constrained to Represent Data & Metadata In Ways Familiar to Analytical Users

AnIML Technique Definitions

AnIML Technique NDRs

And this was better…, but Vendors & Users Still Needed to Add their Own Tags & Constraints

NIST

Allowing AnIML to Be Extended…

W3C XML Language	**+**	W3C Schema Language	**+**	AnIML Schema NDRs

AnIML Core Schema

AnIML Technique Schema

Extensions can Supplement, but Not Override the Constraints in Technique Definitions & the Core

AnIML Technique Definitions

Vendor Technique Extension

Institutional Technique Extension
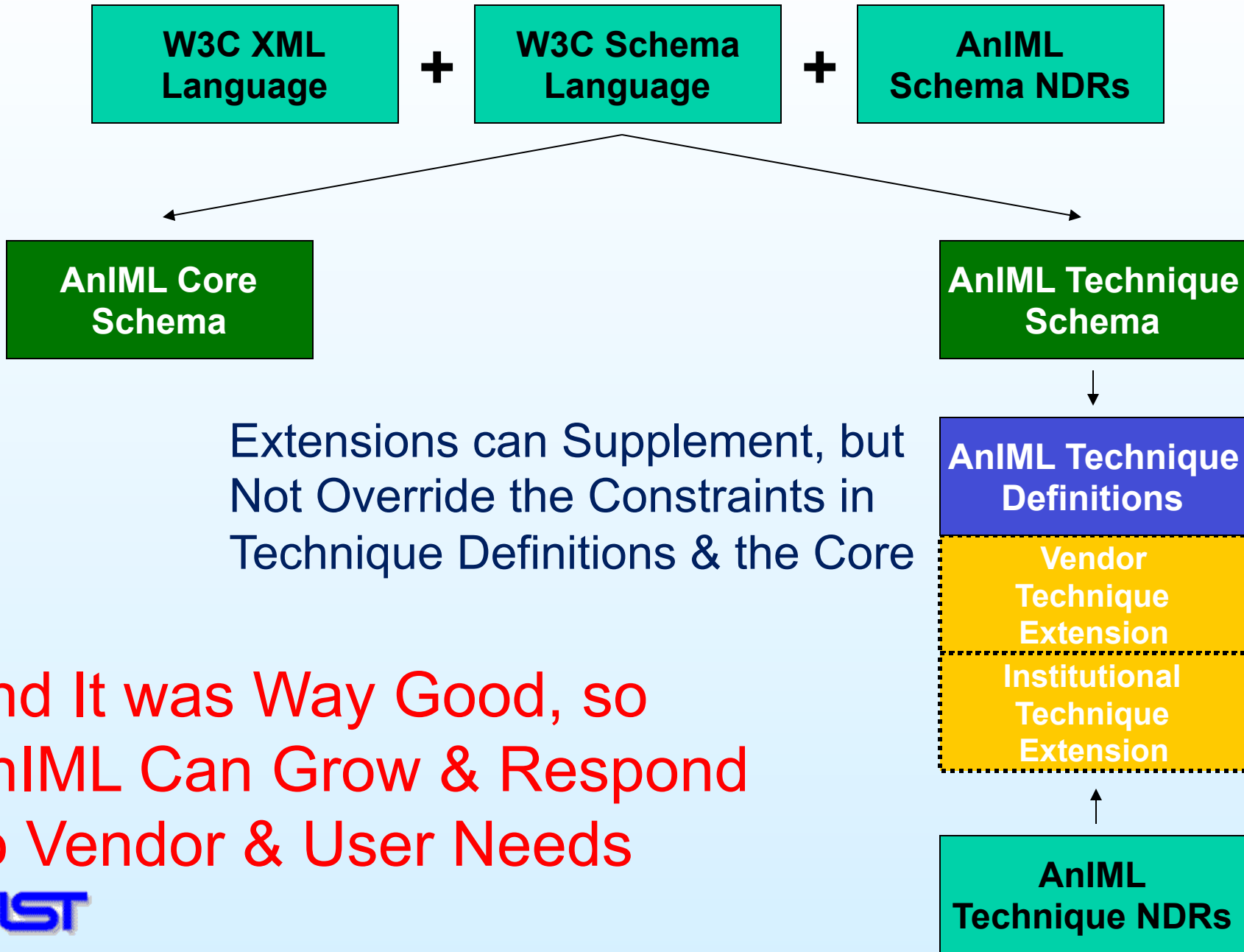
And It was Way Good, so
AnIML Can Grow & Respond
To Vendor & User Needs

AnIML Technique NDRs

NIST

Then Along Came a Programmer…

W3C XML Language	**+**	W3C Schema Language	**+**	AnIML Schema NDRs

AnIML Core Schema

Who Created an App that Put Instrument Data Into the AnIML Framework

AnIML Technique Schema

AnIML Technique Definitions

Instrument Data, Params, Metadata

AnIML Data File Creation App

Vendor Technique Extension

Institutional Technique Extension

And Lo, It was Awesome!

AnIML Data File

AnIML Technique NDRs

NIST
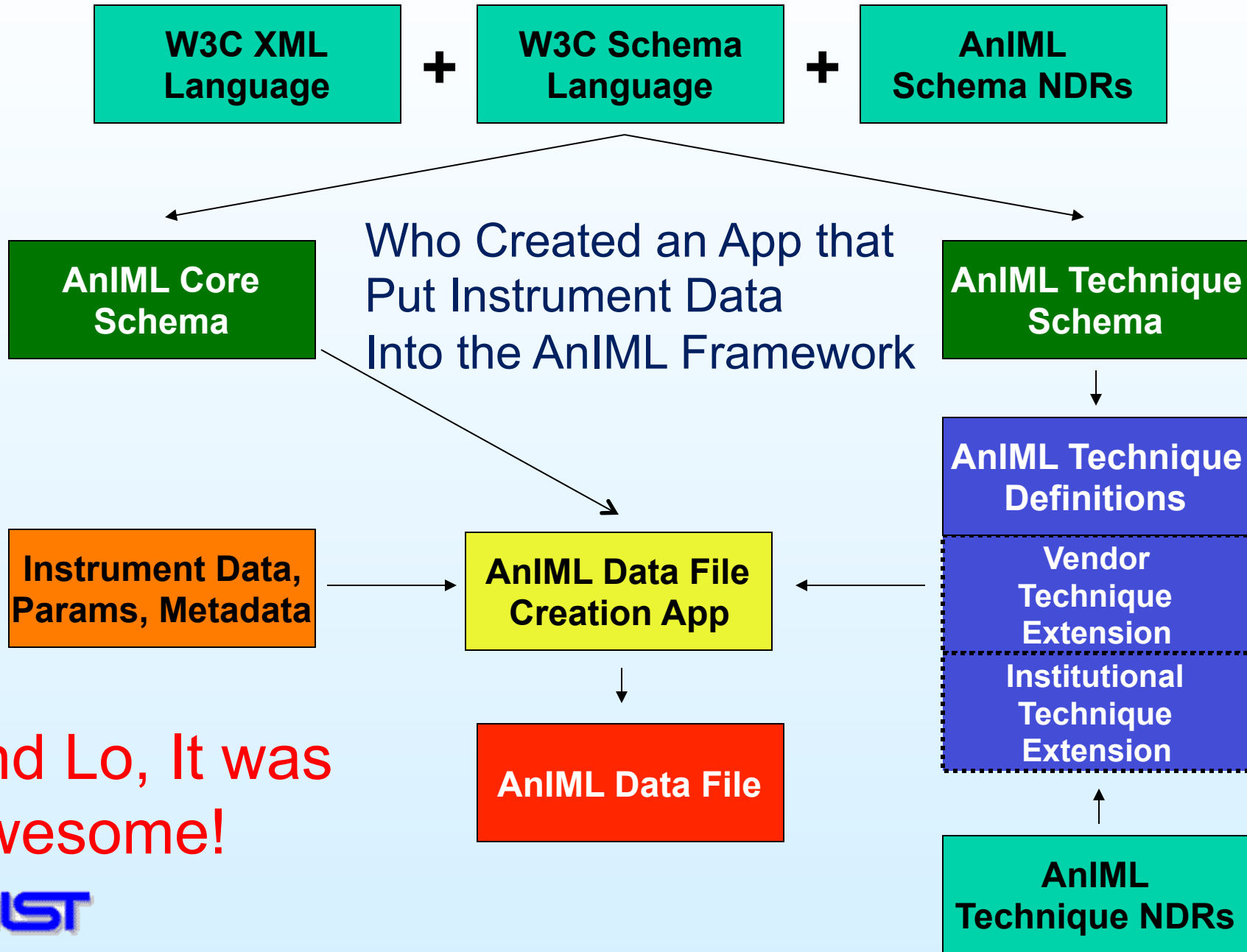
Main AnIML Components

- **AnIML Core Schema**
  - One Schema
  - Maintained by ASTM E13.15

- **AnIML Technique Schema**
  - One Schema
  - Maintained by ASTM E13.15

- **AnIML Technique Definition Documents**
  - One or More Instance Documents per Technique
  - Maintained by ASTM E13 or Appropriate Domain Expertise Organization

- **AnIML Technique Definition Extensions**
  - One or More Instance Documents per Technique
  - Maintained by Vendor, Organization, User, or Whomever Extends the Technique

- **AnIML Result Data Files**

NIST

What Does AnIML 1.0 Support?

- Storage of Raw or Processed Experimental, Simulated, or Annotated Data & Metadata from <u>Selected</u> Analytical Techniques

- Processed Data and Metadata from Trace-Forming/Peak-Finding, Cross-Cutting, Peak Integration, and Peak-Table Creation

- ~~Quantitation~~

NIST

AnIML 1.0 Supported Techniques

- **Chromatography and Other Flow Techniques**
  - GC Point Detectors – FID, TCD, ECD, NPD, FPD
  - LC Point Detectors – RI, ELS

- **UV-Vis**

- **IR**

- **1D NMR**

- **MS**

- **Combinations of the Above Techniques (Hyphenated Techniques)**

- **Indexed Experiments Using the Above Techniques**
  - Kinetic Runs
  - Temperature or pH Profiles

NIST

AnIML: Where We're At for Version 1.0

- **AnIML Core Schema**
  - Complete
  - Frozen
  - Internally Documented
- **AnIML Schema NDRs**
  - Complete
- **AnIML Technique Schema**
  - Complete
  - Frozen
  - Internally Documented
- **AnIML Technique NDRs**
  - In Progress
- **AnIML Technique Exemplar**
  - In Progress

AnIML: Where We're At for Version 1.0

- **AnIML Technique Definitions**
  - UV-Vis Technique Definition – Complete
  - Chromatography Technique Definition – Complete
  - Mass Spec Technique Definition – Started
  - 1D NMR Technique Definition – Started
  - IR Technique Definition – Started
  - Point Detector Technique Definitions
    - GC Detectors (5) – Complete
    - LC Detectors (2) – Started
  - AnIML Misc. Technique Definitions
    - Indexing – Complete
    - Cross-Cutting – Not Started
    - Trace Forming – Not Started
    - Peak Table – Not Started

NIST

AnIML: Where We're At for Version 1.0

- **AnIML Applications and Tools**
  - AnIML Data Viewers – Several Complete
  - AnIML Extensible Validator – Complete
  - Simple AnIML File Writer for Agilent 8453 UV/Vis – Complete
  - Simple AnIML File Writer for PE Lambda 900 UV/Vis - Complete
  - JCAMP-DX to AnIML Data Converters – Several Complete
  - XSLT Documentation Extractor/Stripper for Schemas – Complete
  - Documentation Extractor /Creator for Technique Definitions – Complete

AnIML Documents and Standards

- AnIML Standard Practice Document – Nearly Complete
- AnIML Core and Technique Schema Standard Specification and Guide Document
- AnIML Standard Technique Definition Guide Document
- Individual Technique Standard Specification and Guide Documents
  - AnIML UV-Vis Technique
  - AnIML IR Technique
  - AnIML Chromatography Technique
  - AnIML Mass Spec Technique
  - AnIML 1D NMR
  - AnIML Misc. Techniques
    - Indexing
    - Point Detectors
    - Cross-Cutting
    - Trace Forming/Peak Finding
    - Peak Table

NIST

Expediting the Completion of AnIML 1.0

- Current Volunteer Effort is Too Slow
- There is Still Too Much to Do
- Need Paid Help to Develop Components & Documentation
- AnIML 1.0 Documentation Expedition Project
  - Funded by
    - Agilent Technologies
    - Amgen
    - Glaxo SmithKline
    - PerkinElmer
    - Waters Corporation
  - Administered by ASTM

Getting Started with AnIML?

- Check Out the AnIML Website
  - http://animl.sourceforge.net
  - http://www.animl.org
- Delve into the Internally Documented Core Schema
- Examine the AnIML Schema NDRs
- Check Out Some Technique Definition Docs and Their Documentation
- While There May Be Some Changes Before Vers. 1.0, They Likely Will Be Minor or Organizational

NIST

Want to Help Develop AnIML?

- **Contact**
  - Gary Kramer, NIST
  - +1-301-975-4132
  - gary.kramer@nist.gov

  - Burkhard Schäfer, BSSN
  - +1-888-674-0047
  - +49-6136-75-27-60
  - bschaefer@bssn-software.com

- **Webex Virtual Meetings**
  - Monthly – 10:30 AM EST
  - ASTM-sponsored

NIST

Acknowledgements

Mark Bean – GlaxoSmithKline

Maren Fiege – Waters Informatics GmbH

Burkhard Schäfer – BSSN Software

Stuart Chalk – University of North Florida

Dale O'Neill & Michael Leiber – Agilent Technologies

Adam Patkin – PerkinElmer

Jamie McQuay – Scimatic Software

Peter Linstrom – NIST

Tony Davies – IUPAC Subcommittee on Electronic Data Standards

Reinhold Schäfer – Fachhochschule Wiesbaden

- Alex Rühl
- Martin Peschke
- Aykut Arslan
- Dominick Pötz
- Anh Dao Nguyen
- Alex Roth

- Ronnie Jopp
- Patrick Gleichmann
- Kordian Placzek
- Frank Masur
- Dennis Backhaus

What is a Data Requisite?

- A Data Requisite is the collection of reported result data and result metadata required by a data authority to make a dataset fit for the purpose of an experiment.

- Examples of Data Requisites:
  - MIAME "Format" – Minimum Information About a Microarray Experiment
  - IUPAC Recommendations for the Presentation of NMR Data for Publication in Chemical Journals
  - Dublin Core Metadata Set - resource description for publications and semantic web
  - Experimental Descriptions in National Pharmacopiae
  - USEPA/USACE SEDD – Staged Electronic Data Deliverable

NIST

What is a Data Authority?

- An individual, group, organization, or body—
  - Who decides what result metadata are required to accompany the result data from an experiment in order to make the resulting dataset fit for the purpose of the experiment and
  - Who enforces that decision.

- Examples of Data Authorities:
  - Consensus standards bodies
  - Accreditation agencies
  - Individual researcher policies on recording experimental results
  - Major Professor/Group Leader policies on experimental data
  - QC/QA group requirement policies
  - Journal editors and editorial boards
  - Pharmaceutical company policies on experimental results
  - FDA polices for experimental results supporting NDAs
  - EPA requirements for environmental analysis reporting

NIST

What AnIML Standards Are Not

- AnIML standards are **<u>not</u>** proscriptive and do **<u>not</u>** contain lists of data and metadata elements and attributes that are **<u>required</u>** for recording, reporting, exporting, importing, and archiving the result data and result metadata outcomes from analytical chemistry experiments.

- With very few exceptions, the AnIML standards do **<u>not</u>** require the inclusion of **<u>any</u>** specific data or metadata elements and attributes.

- Instead, AnIML dictates how data and metadata, **<u>if</u> <u>present</u>**, must be represented and located within a dataset.

NIST

What are Result Data?

■ Result Data are the Qualitative and Quantitative Outcomes from an Experiment.

What are Result Metadata?

■ Result Metadata are the Data about the Result Data that Describe Parameters, Attributes, and Conditions Relating to the Result Data, their Collection, and their Manipulation.

What is a Dataset?

■ A Dataset is the Combination of the Reported Result Data and Attending Result Metadata from Carrying Out an Experiment or Experiments.

NIST