

<AnIML >

What AnIML is Not

</AnIML >

Gary W. Kramer

Biochemical Science Division
National Institute of Standards and Technology

This presentation will:

- Describe What the AnIML Standards Are
- Describe What the AnIML Standards Are Not
- Define the Concepts of “*Data Authorities*” and “*Data Requisites*”
- Discuss How AnIML Can be Used to Create a “*Data Requisite*”
- Describe How to Enforce “*Data Requisites*” Using AnIML Tools

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.
- AnIML provides markup (or tags) to identify the content of any analytical data.
- AnIML provides data structures to organize the data from any analytical experiment.

Some AnIML Files are Simple



Some AnIML Files are Complex



What AnIML Standards Are Not

- The AnIML standards do not contain lists of data and metadata elements and attributes that are **required** 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 not require the inclusion of any specific data or metadata elements and attributes.
- Instead, AnIML dictates how data and metadata, if present, must be represented and located within a dataset.

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 and their collection

What is a Dataset?

- A dataset is the combination of the reported result data and attending result metadata from carrying out an experiment.

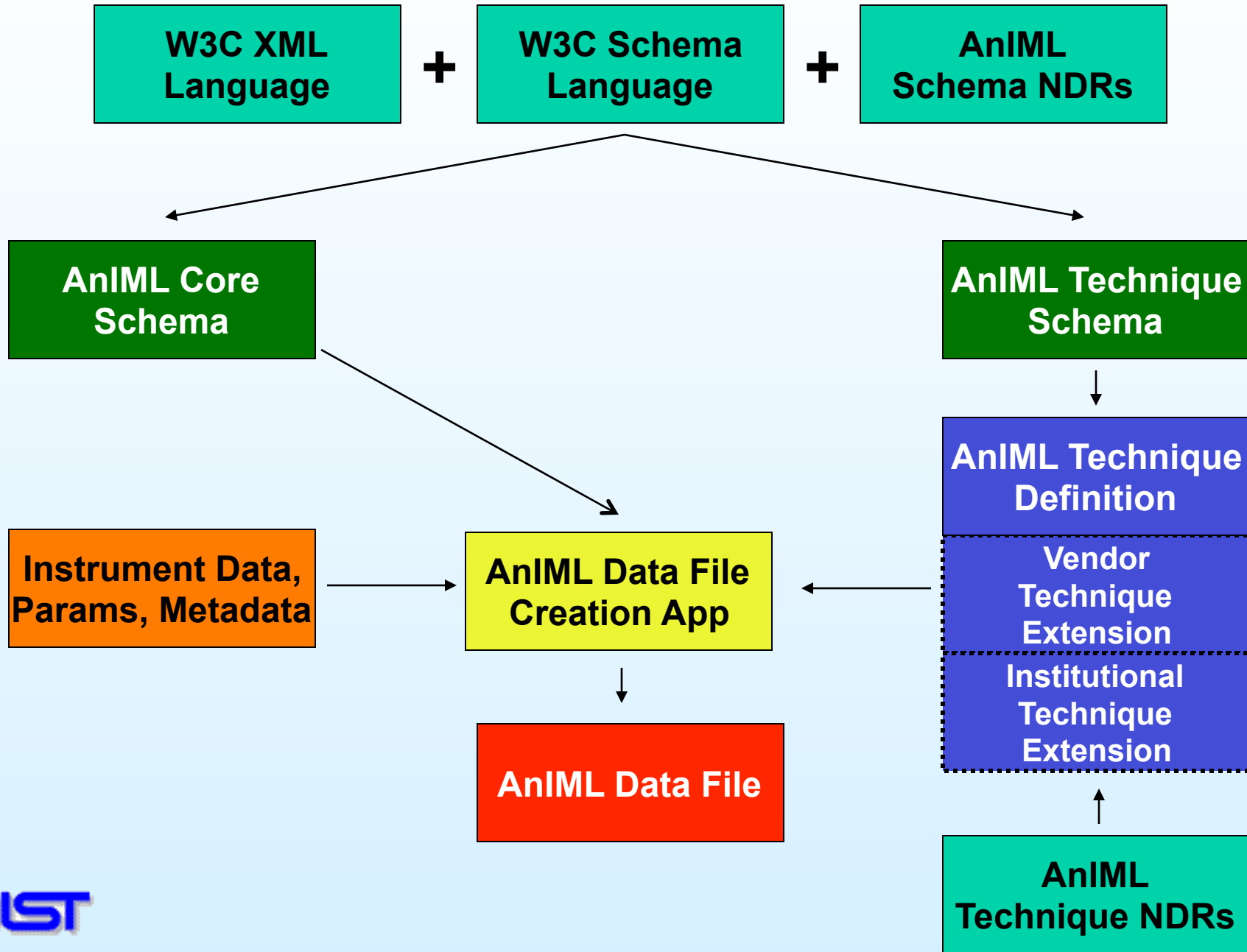
What is a Data Authority?

- 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

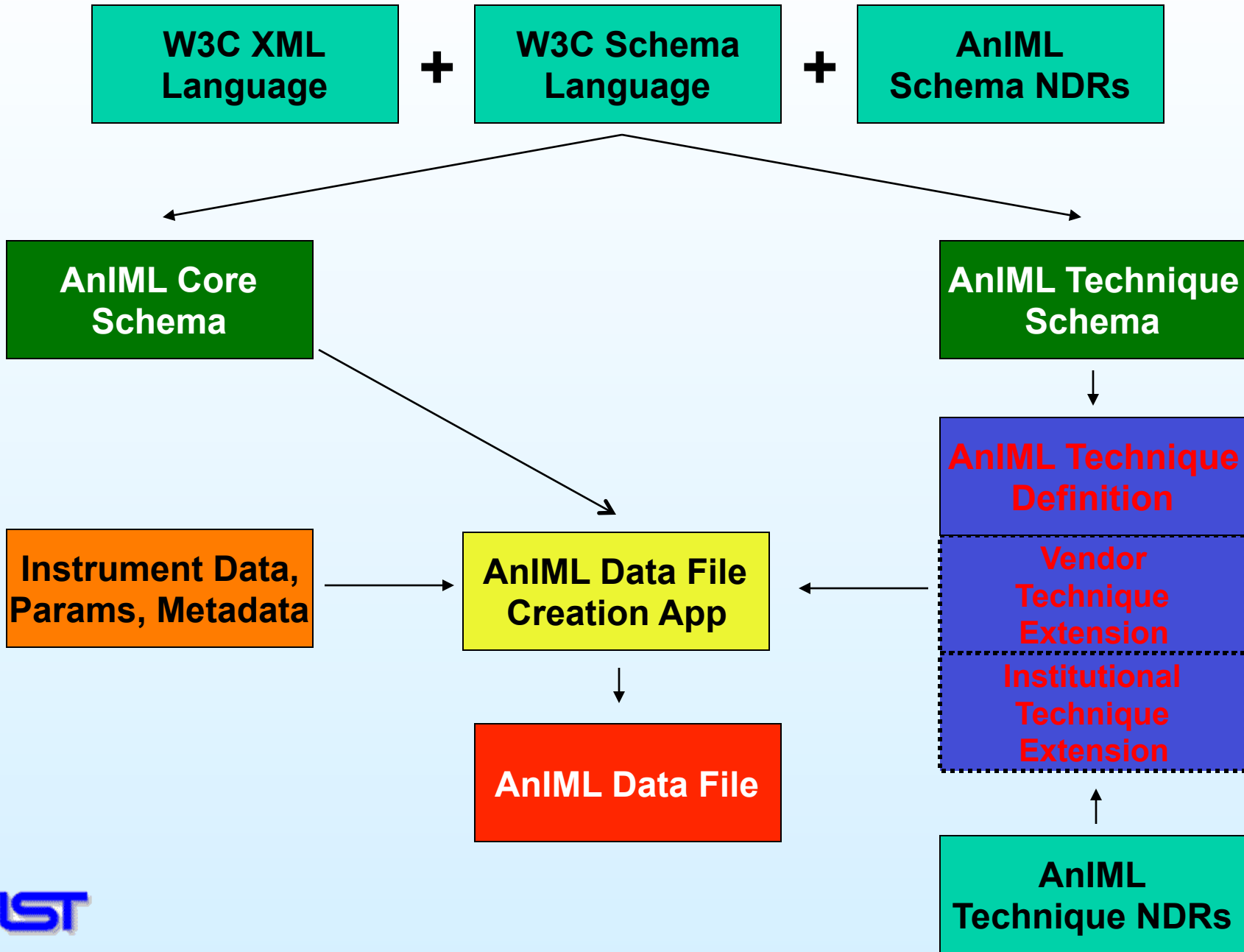
What is a Data Requisite?

- 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

Basis for Creating an AnIML Data File



Creating a Data Requisite with AnIML



Creating a Data Requisite in AnMIL

- Create a Technique Definition Extension
 - Make Optional Components Required
 - Add New Required Components
 - Add New Lists of Enumerated Values
 - ~~● Make Required Components Optional~~
 - ~~● Add to Existing Enumerated Values Lists~~
- Create a Whole New Technique Definition

Technique Definition Extension Scope

- Global Technique Definition Extensions
 - Apply to all AnIML Technique Definitions
 - Components in Global AnIML Technique Definition Namespace
 - Currently Allowable in AnIML

- Local Technique Definition Extensions
 - Apply to Specified AnIML Technique Definitions
 - Must Specify Which Technique Definitions are Affected
 - Components in Specified Local AnIML Technique Definition Namespaces
 - Currently Not Allowed in AnIML

Enforcing Data Requisites with AnIML Tools

- Conformance Testing to determine whether an implemented system fulfills its requirements against a standard or authority
 - XML is more than a data format, more than a system of identifying content (data elements and attributes) with labels (tags)
 - XML provides a means for defining the structure, content and semantics of XML documents.
 - XML expresses shared vocabularies and allows machines to carry out rules made by people

XML Conformance Testing

Well-Formedness Checking & Validation

- Well-Formedness Checking -- Syntax and Language-Rule Checking
 - All XML documents must be well-formed
 - All mal-formedness errors are fatal
 - ◆ When a parser encounters an error, it stops

- Validation -- Checking Against the Content Model
 - With validation, all that is not permitted is forbidden
 - Validation errors and warnings are not necessarily fatal
 - ◆ The outcome of an error depends on the parser's instructions

Types of Validation

For Instance Documents

- Syntactical or Well-Formedness Checking - syntax/grammar/rule checking against W3C XML language rules
- Simple Validation - checking for organization, structure, required elements/attributes, data types, ... against the content model (schema)
- Advanced Compliance Checking - checking against encodable NDRs (used only for Technique Definitions and Extensions)
- Semantic Validation - checking data content against boundary limits, algorithms, and/or rule bases

NIST Validator for AnIML Data Files

- Validates only AnIML Data Files
- Organized into Procedure Blocks (P-blocks) to speed checking by running P-blocks in parallel where possible
- Well-formedness checking using DOM4J
- Validation against AnIML Core Schema using Apache Xerces
- Validation against AnIML Technique Definitions using Java-based P- Blocks
- Limited Semantic Validation using Java-based P-Blocks
- Currently does not validate signatures
- Extensible by creating and linking in new P-Blocks

Requirements for the ASTM AnIML File Validator

- Application tool to validate AnIML Technique Definitions, AnIML Technique Definition Extensions, and AnIML Data Files
- Can not use “standard” XML validation tools
- Check XML well-formedness
- Validate AnIML Technique Definitions and AnIML Technique Definition Extensions against AnIML Technique Schema and encodable AnIML Technique Definition NDRs
- Validate AnIML Data Files against AnIML Core Schema and applicable AnIML Technique Definitions with appropriate AnIML Technique Extensions
- Allow checking with and without audit trailing and signatures
- Some simple semantic validation included
 - Bounds checking
 - Limit checking



How Much Conformance Testing is Enough?

- Depends on business case requirements
- Depends on “customer” for data
- Depends on need for speed verses need for conformance
- As business needs change, AnIML Data Files may need to be re-validated against higher authorities to assure conformance with requirements
- Must realize the limitations of existing data files
 - Legacy data may not meet modern requirements