Playing with the AnIMLs: Demonstrations of AnIML Generic Viewers

Gary W. Kramer

Biochemical Science Division
National Institute of Standards and Technology

NIST

AnIML

- AnIML Provides Structures to Organize Result Data and Metadata from Simple to Complex Analytical Chemistry Experiments
- AnIML is the Only Analytical Data Structure to Support Fully Multi-detector, Multi-dimensional, Multi-sample, Multi-step, Multi-sequence Analytical Procedures in a Single File Format
- AnIML 1.0 Supports Sample Data, Raw Analytical Data, Data-Acquisition Parameters, Processed Analytical Data, Data-Processing Parameters, Trace Forming, and Peak Tables

What Does AnIML Provide?

- AnIML Provides Data Structures to Organize the Data from any Analytical Experiment
- AnIML Provides Markup (or Tags) to Identify the Content of any Analytical Data

Some AnIML Files are Simple

Some AnIML Files are Complex

General AnIML File Organization

- AnIML File Header
- Samples
  - Sample ID, Container, Location, & Properties
- Experiment Steps
  - Technique Used
  - Infrastructure
  - Method
    - Author
    - Instrument Properties
    - “Antecedent” Instrument Parameters
  - Result
    - Independent
    - Dependent
    - “Measured” Instrument Parameters
- Audit Trail
  - Log Entry
- Signatures

How is an AnIML File Organized?

- The Answer is that it Depends to Some Extent on how the Writer of the AnIML File Thinks about the Experiment and the Data
- There are Wrong Ways, and Validator Programs should Find the Errors
- But, there is No Single "Right" Way
- Consider how Two Spectroscopists "See" an ICP-MS Instrument:
  - The Atomic Spectroscopist Thinks of it as an Atomic Emission Spectrometer with a Fancy Detector
  - The Mass Spectroscopist Thinks of it as a Mass Spectrometer with a Fancy Inlet System

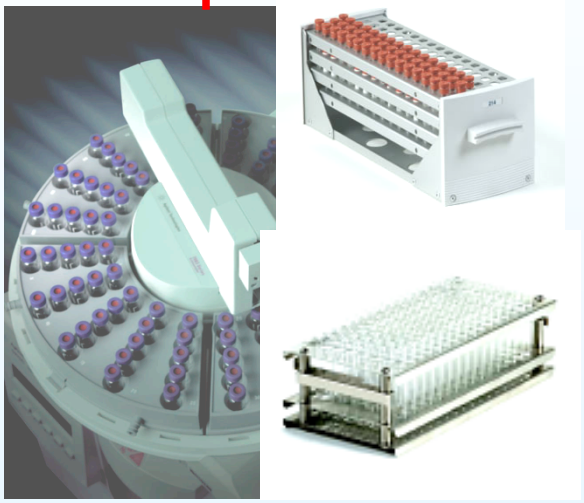
Eliminating Redundancy with Templates

- Components of Multi-step, Multi-sequence or Multi-dimensional Datasets often Contain Redundant Information
- For Example, Each Spectrum in a Set of Spectra may Share Common Wavelength Axis Parameters or Samples may Share Common Property Descriptions
- AnIML Utilizes a Data Structure Called a Template to Record such Common Parameters
- Templates may be Referenced Subsequently in lieu of Repetitively Recording the Redundant Data
- In an AnIML File Templates must be Declared and Populated before being Referenced

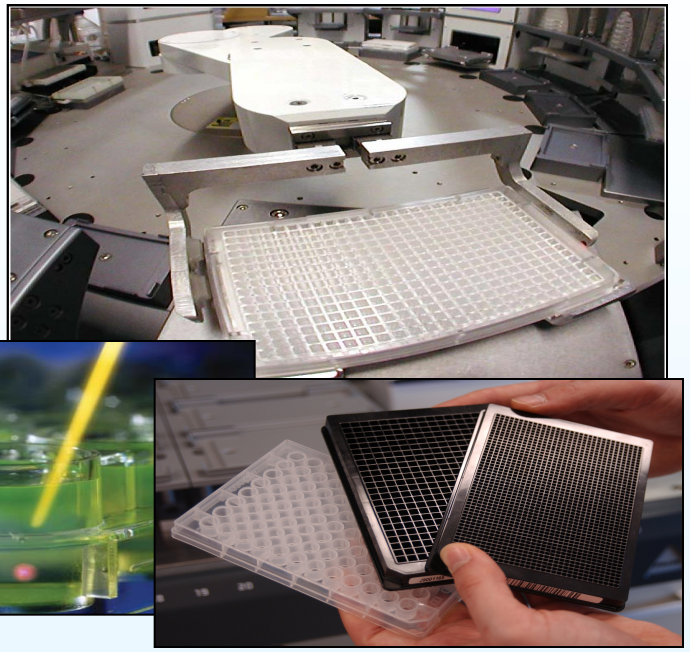
What is a Sample?

Today's Samples Can be Complex

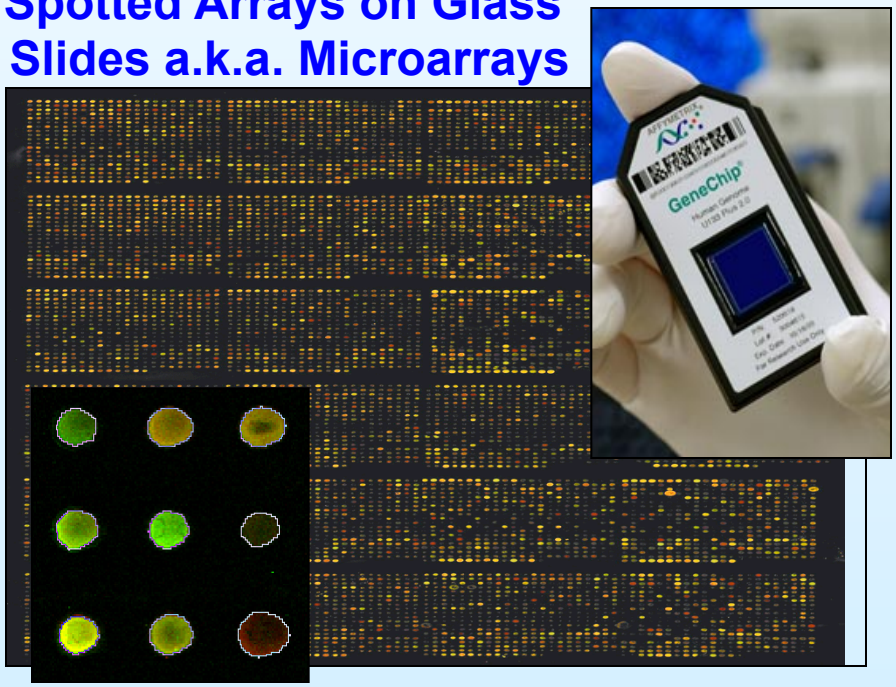
Racks and Arrays of Individual Vials



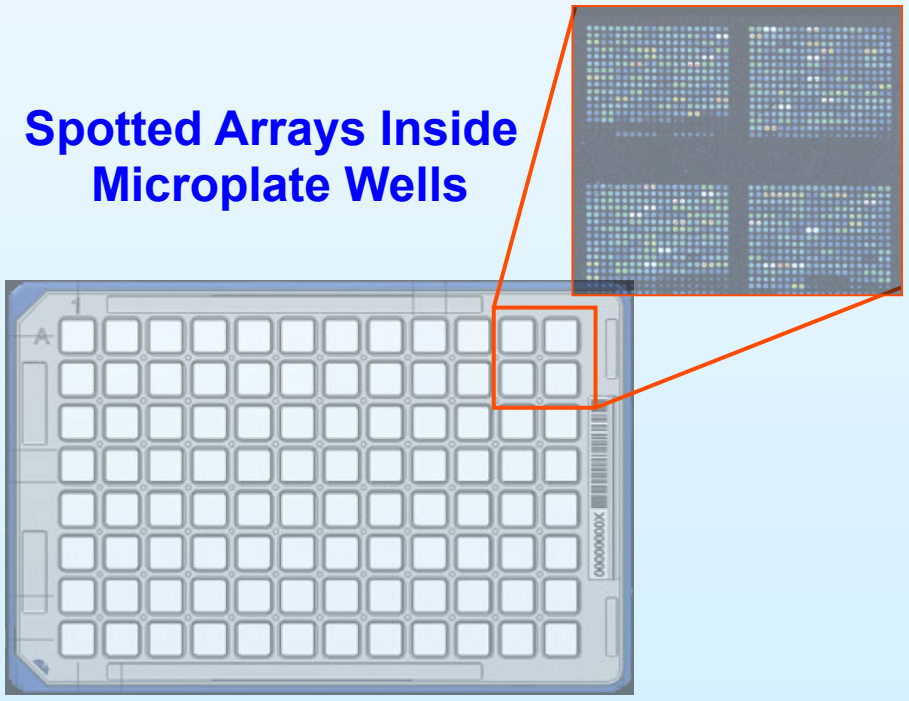
Molded Arrays of Vials a.k.a Microplates



Spotted Arrays on Glass Slides a.k.a. Microarrays

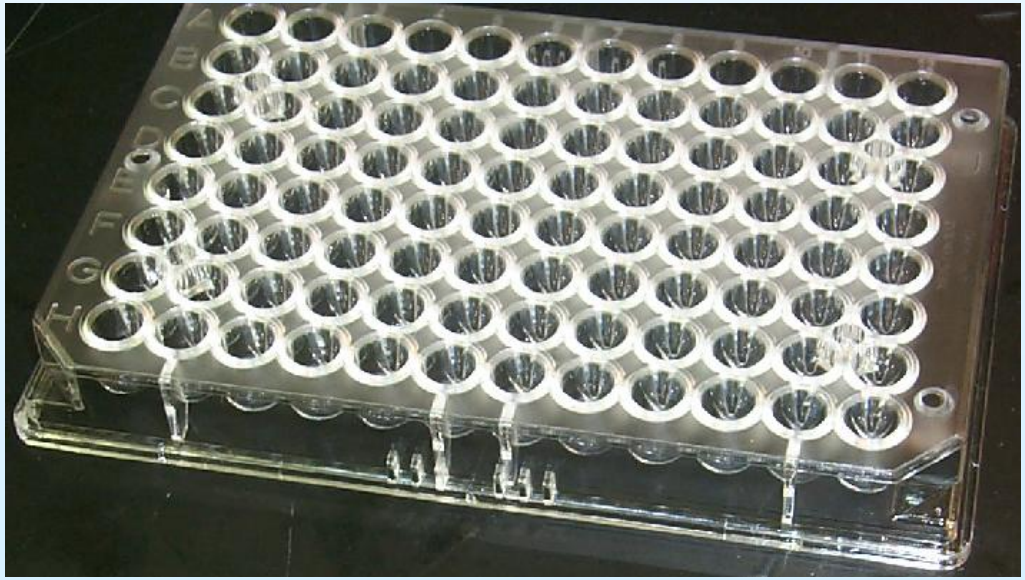


Spotted Arrays Inside Microplate Wells



Consider a Microplate…

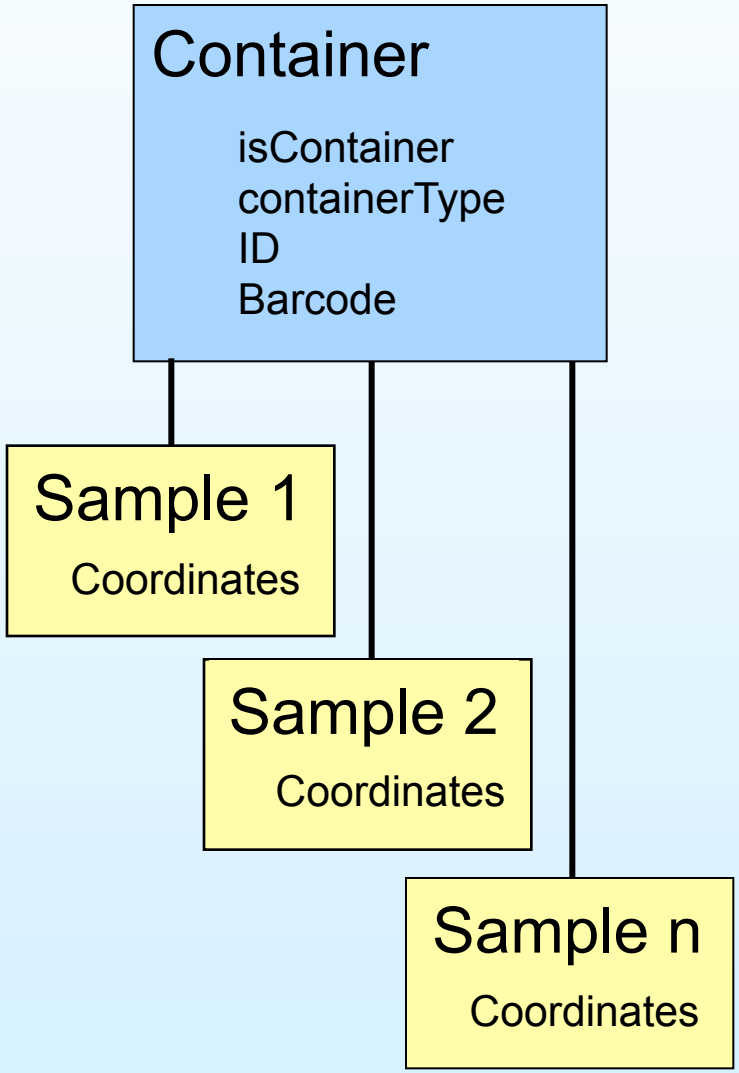
- How Many Samples Does a Microplate Represent?
  - 96 Different Samples?
  - A Single Sample?
  - Something in Between?



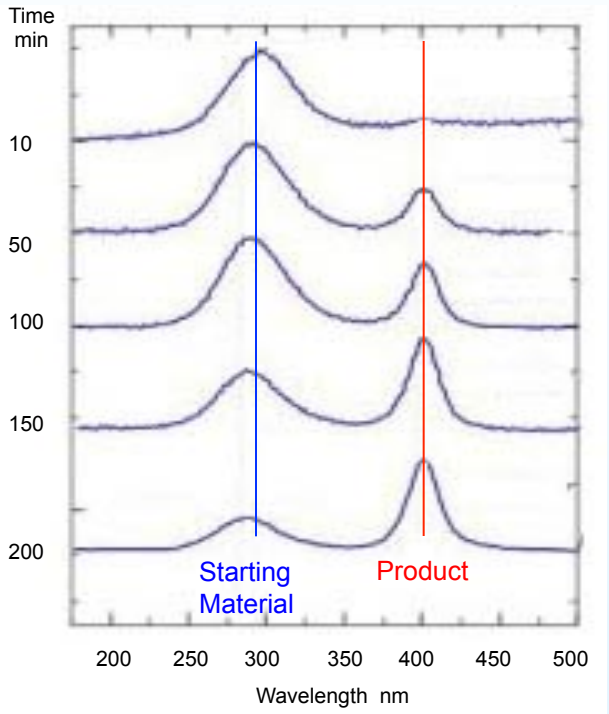
- AnIML does NOT Force the Answer to this Question
- The Answer is Up to the Writer of the AnIML File
- It Depends on How Scientists "See" their Data

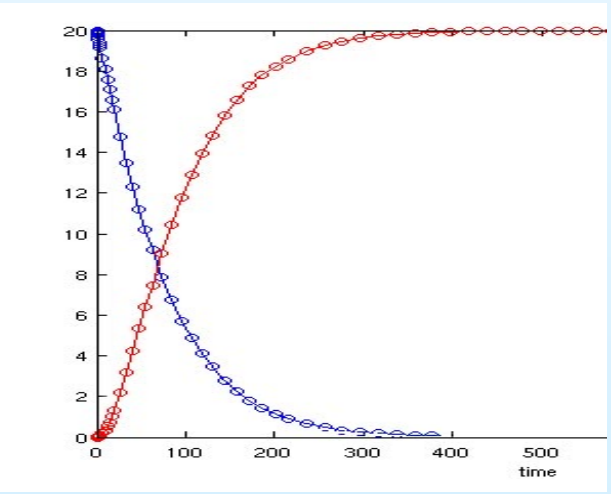
Dealing with Samples

- Samples may be:
  - Individual (Single Vial, Single Bottle)
  - Containerized (Racks, Microtiter Plates, Arrays)
- Containerized Samples may be:
  - Simple
    - Vials in Rack
    - Wells in Microtiter Plates
    - Spots on Arrays
  - Compound
    - Discrete with Prescribed Coordinates
      - Racks
      - Microtiter Plates
      - Spots on Arrays
    - Non-discrete with Relative Coordinates
      - Bands in Gels
      - Points on Surfaces
      - Spots on Arrays
- Experiment Steps may Act on:
  - Individual Samples
  - Individual Samples within a Container
  - Containerized Samples as a Whole



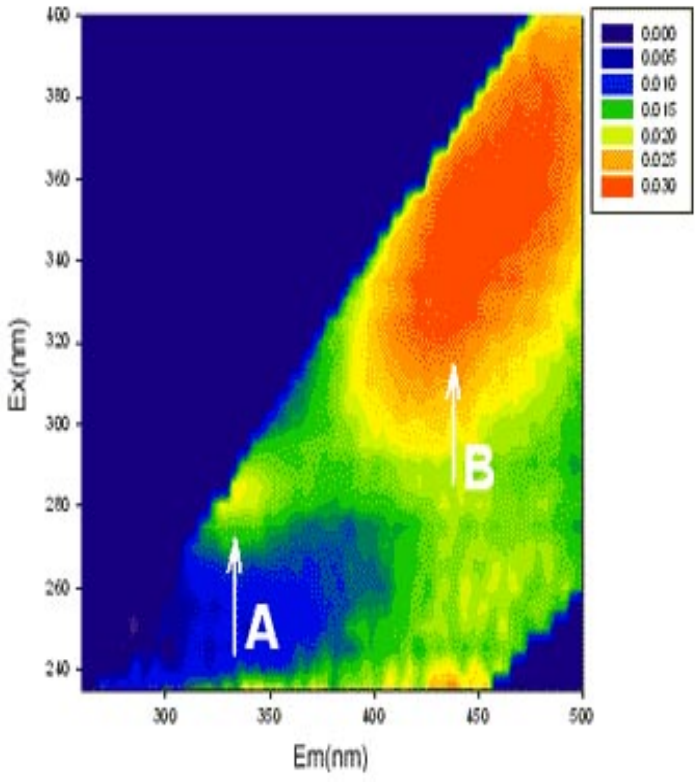
Analytical Data Sequences: Kinetics





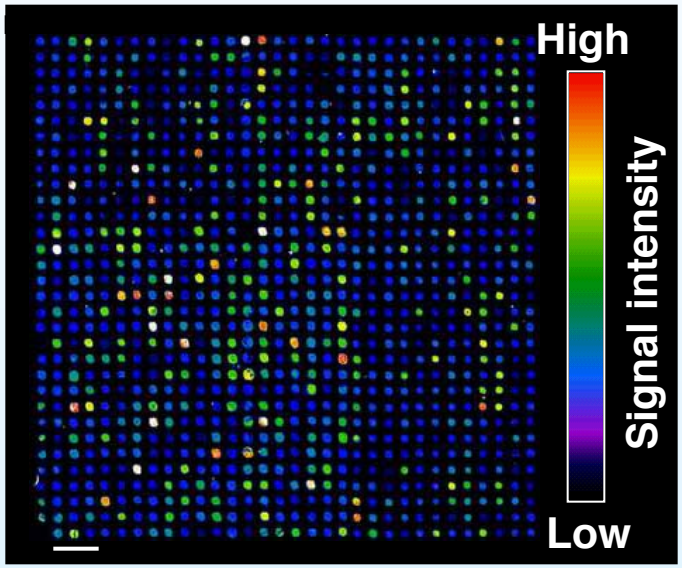
- AnIML Handles Data Sequences Using a Concept Called Indexing
- An Index may be a Sequence Number, a Variable Quantity (such as Time, Pressure, Spatial Location, or pH), or an Analytical Technique
- In this Example:
  - Index Variable (Time) is the Primary Independent Axis
  - Wavelength Variable is the Secondary Independent Axis
  - Intensity is the Dependent Axis
  - The Data can be Considered to be an Array of UV-Vis Spectra Indexed by Time
  - To Generate the Kinetics Plot, the Array is Cross-Cut at the Absorbance Peak Wavelengths to Determine the Data Relating to the Starting Material and Product Concentrations

Multi-Dimensional Data: Fluorescence EEM



- AnIML Handles Data Sequences Using an Indexing Concept
- An Index may be a Sequence Number, a Variable Quantity (such as Time, Pressure, Spatial Location, or pH), or an Analytical Technique
- In this Example:
  - Index Variable (Excitation Wavelength) is the Primary Independent Axis
  - Emission Wavelength Variable is the Secondary Independent Axis
  - Fluorescence Intensity is the Dependent Axis
  - The Data can be Considered to be an Array of Fluorescence Emission Spectra Indexed by Excitation Wavelength
  - To Generate an Excitation Spectrum, the Array is Cross-Cut at an Emission Wavelength

Multi-Dimensional Data: Spatial Fluorescence

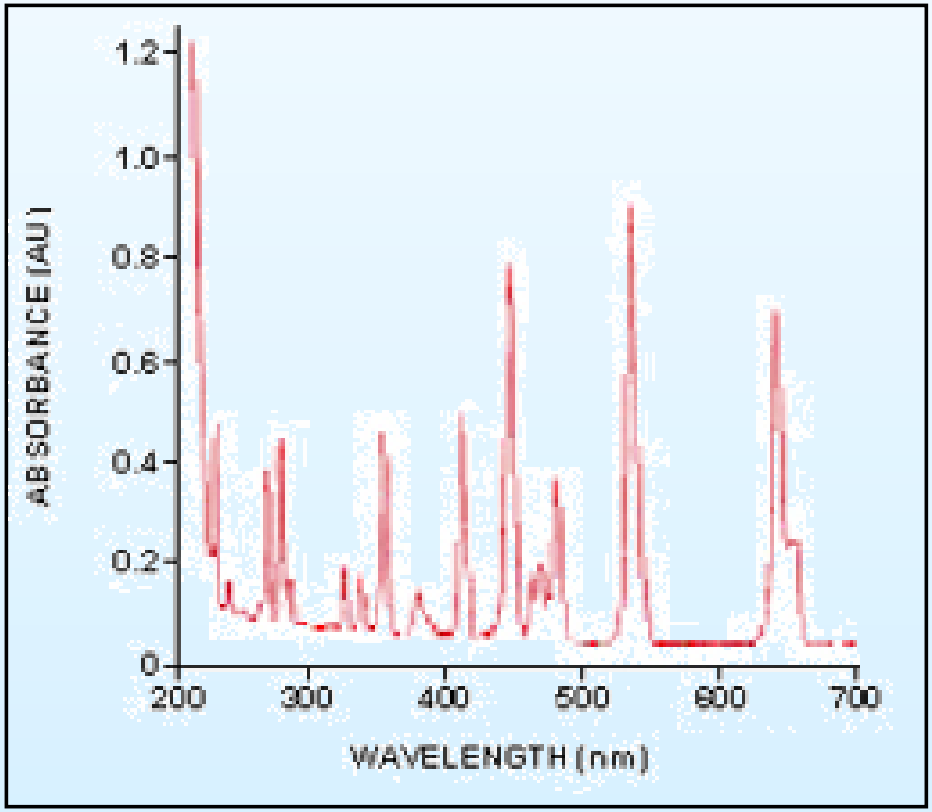


- AnIML Handles Data Sequences Using an Indexing Concept
- An Index may be a Sequence Number, a Variable Quantity (such as Time, Pressure, Spatial Location, or pH), or an Analytical Technique
- In this Example:
  - Index Variable (X-position) is the Primary Independent Axis
  - Index Variable (Y-position) is the Secondary Independent Axis
  - Fluorescence Intensity is the Dependent Axis
  - The Data can be Considered to be a Spatial Array of Fluorescence Emission Intensities

Mark Schena

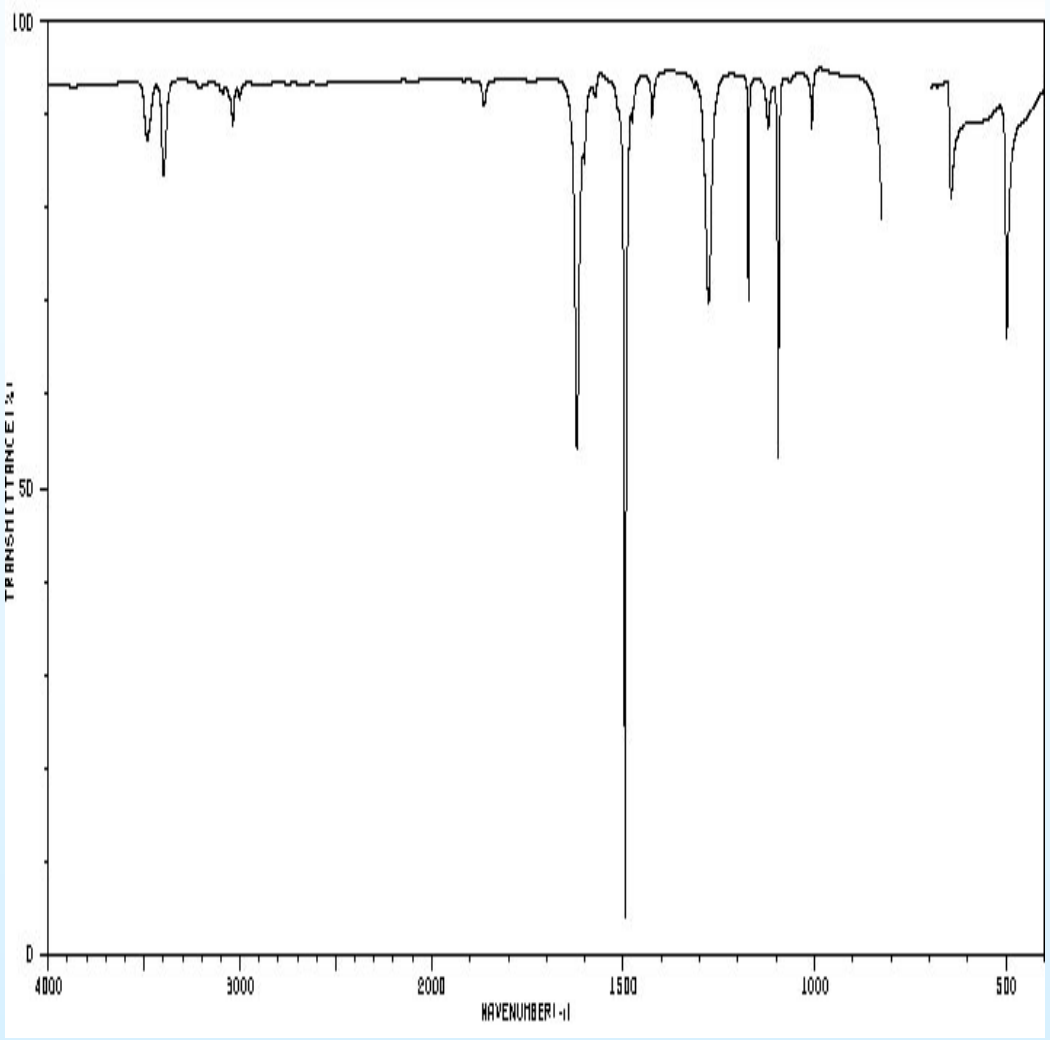
Holmium Oxide UV-Vis Spectrum AnIML File

- AnIML File Header
- Sample Set
  - Sample & Sample Properties
  - Reference & Reference Properties
- Experiment Step Set
  - UV-Vis Technique
  - Infrastructure
    - Sample ID
  - Method
    - Author
    - Instrument Properties
  - Result
    - Independent - Wavelength
      - Auto Increment Value Set
      - Start 700 nm by -1 nm
    - Dependent - Absorbance
      - Encoded Value Set
- Audit Trail
  - Log Entry



p-Chloroaniline IR AnIML File

- AnIML File Header
- Sample Set
  - Sample & Sample Properties
- Experiment Step Set
  - IR Technique
  - Infrastructure
    - Sample ID
  - Method
    - Author
    - Instrument Properties
  - Result
    - Dependent - Transmittance
      - Individual Value Set
    - Independent - Wavelength
      - Auto Increment Set
      - Start 400 cm^{-1} by 2 cm^{-1}
- Audit Trail
  - Log Entry
- Signature Set
  - Signatures



LC-UV (PDA) AnIML file

- AnIML File Header
- Sample Set
  - Sample & Sample Properties
- Experiment Step Set
  - Template (UV-Vis Time Course)
    - Independent - Time
      - Individual Value Set
  - UV-Vis Time Course
    - Dependent – Intensity
      - Encoded Value Set
  - Chromatography Technique
    - Infrastructure
    - Method
    - Result
      - Individual Value Set
        Time = 0.0
      - Experiment Step Set
        Template
        UV-Vis Technique
        ***Infrastructure
        Method
        Result
      - Independent - Wavelength
        Encoded Value Set

- UV-Vis
  - Infrastructure
    - ParentDatapointReference Value = **0.01**
  - Result
    - Dependent - Intensity
      Encoded Value Set

……. Encoded Value Set

- UV-Vis
  - Infrastructure
    - ParentDatapointReference Value = **2.00**
  - Result
    - Dependent - Intensity
      Encoded Value Set