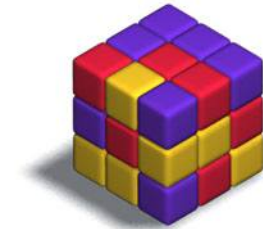## Analytical Instrument Markup Language (AnIML)

## AnIML and Chromatographic Data

## AnIML, TNF, Viewers, and Plenty of Challenges!
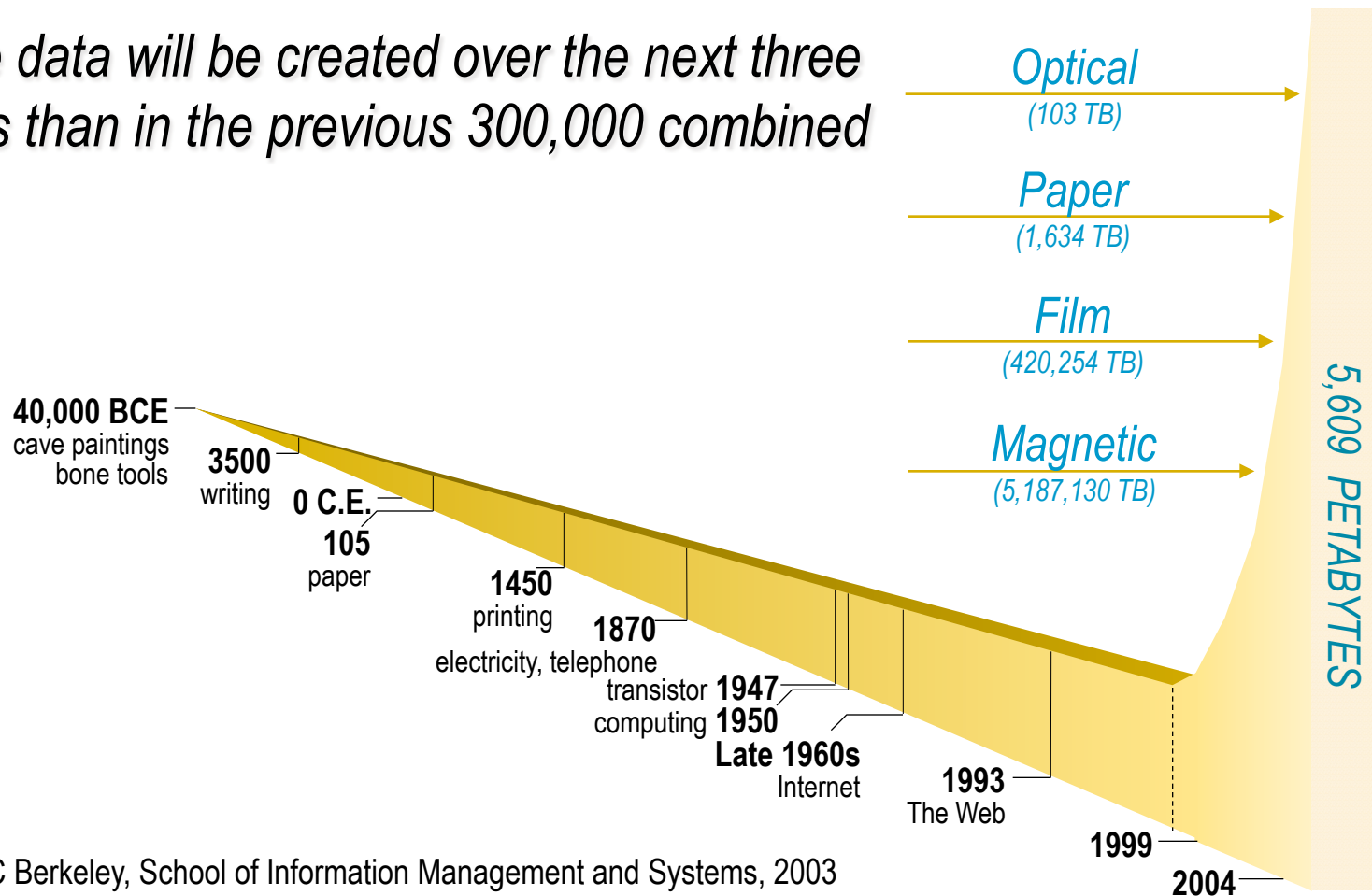
Dale O' Neill
Agilent Technologies

**Agilent Technologies**

# Purpose of AnIML

➢ Sharing of Data

➢ Data Preservation

**Agilent Technologies**

# More and more data…

*More data will be created over the next three years than in the previous 300,000 combined*

*Optical*
*(103 TB)*

*Paper*
*(1,634 TB)*

*Film*
*(420,254 TB)*

*Magnetic*
*(5,187,130 TB)*

*5,609 PETABYTES*

**40,000 BCE**
cave paintings
bone tools

**3500**
writing

**0 C.E.**

**105**
paper

**1450**
printing

**1870**
electricity, telephone

transistor **1947**
computing **1950**

**Late 1960s**
Internet

**1993**
The Web

**1999**

**2004**

Source: UC Berkeley, School of Information Management and Systems, 2003

**Agilent Technologies**

# Different sources and types of data…

**Files**
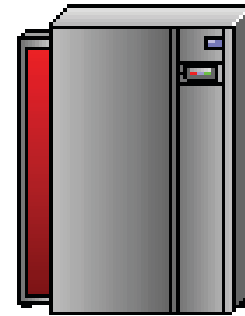
**Databases**
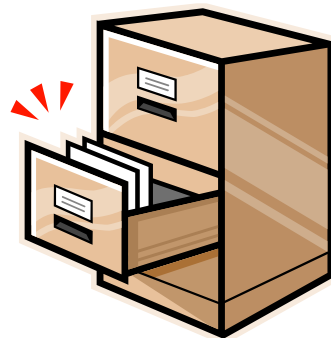
**Structured Data**

**Unstructured Data**

Agilent Technologies

# Retention periods…

- **Regulations**
  - 10, 20, 30 years
- **SOPs**
  - 40, 50… sometimes upwards of 100 years!

Agilent Technologies

# The need for Technology Neutral File (TNF) formats

Critical data must:

- **Be preserved in its entirety**
- **Be OS independent**
- **Outlive the creating application**
- **Must be human readable (not binary or proprietary formats)**
- **Must be usable today (viewing and analysis)**

**Agilent Technologies**

# The problems with multiple TNF formats

- **Little or no interoperability**
- **Must create multiple viewing and analysis tools**
- **Proliferation of more formats**
- **Maintenance and versioning nightmare for developers**
- **New applications must support all previous formats**
- **"Our format is best" syndrome**
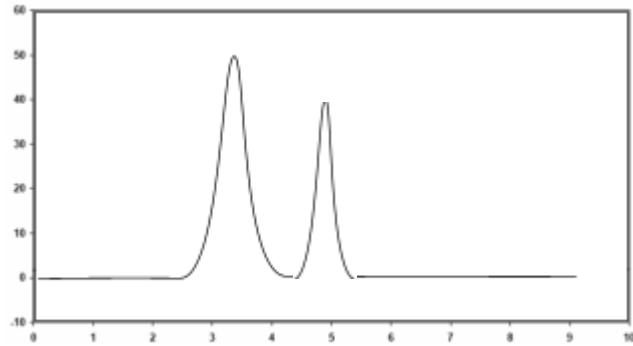
**Agilent Technologies**
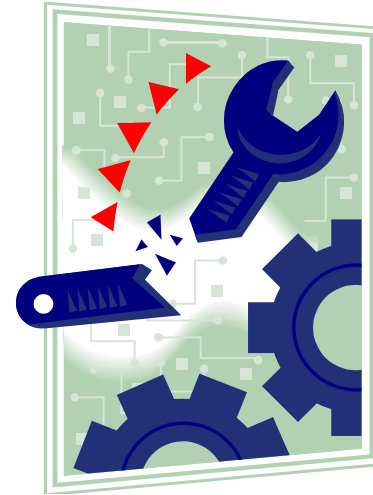
# The advantages of a standardized format

- **Easy exchange of data between applications**
- **Consistent and well known architecture**
- **Tools can be designed to work across versions**
- **Generic tools can be developed and shared**
- **Shared vendor support for standard format**
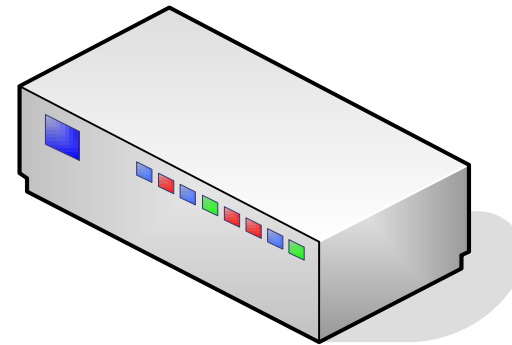- **Format will be maintained and supported, even if vendors come and go**

**Agilent Technologies**

# Chromatography Data System

Separation

Detection

Peak Finding

Analysis

| Text | Text | Text | Text | Text | Text |
|------|------|------|------|------|------|

Title

**Agilent Technologies**

# Chromatography Flow

Agilent Technologies

# Chromatography Flow

**Agilent Technologies**

# Chromatography Flow

Agilent Technologies

**Templates**

- Method Parameters Separation
- Method Parameters Detection
- Method Parameters Integration

ExperimentStep
Name = Sample Introduction
Input = Sample
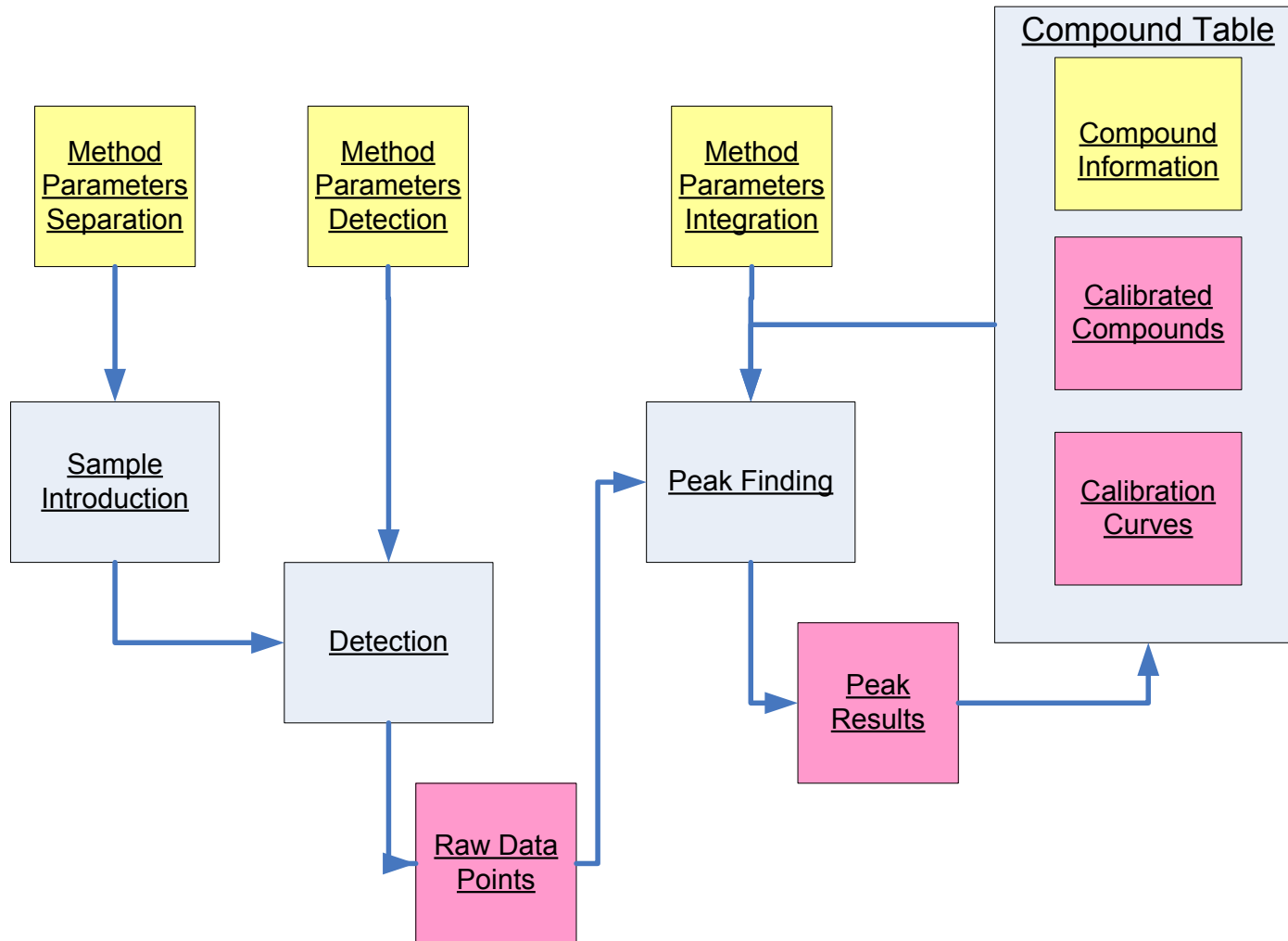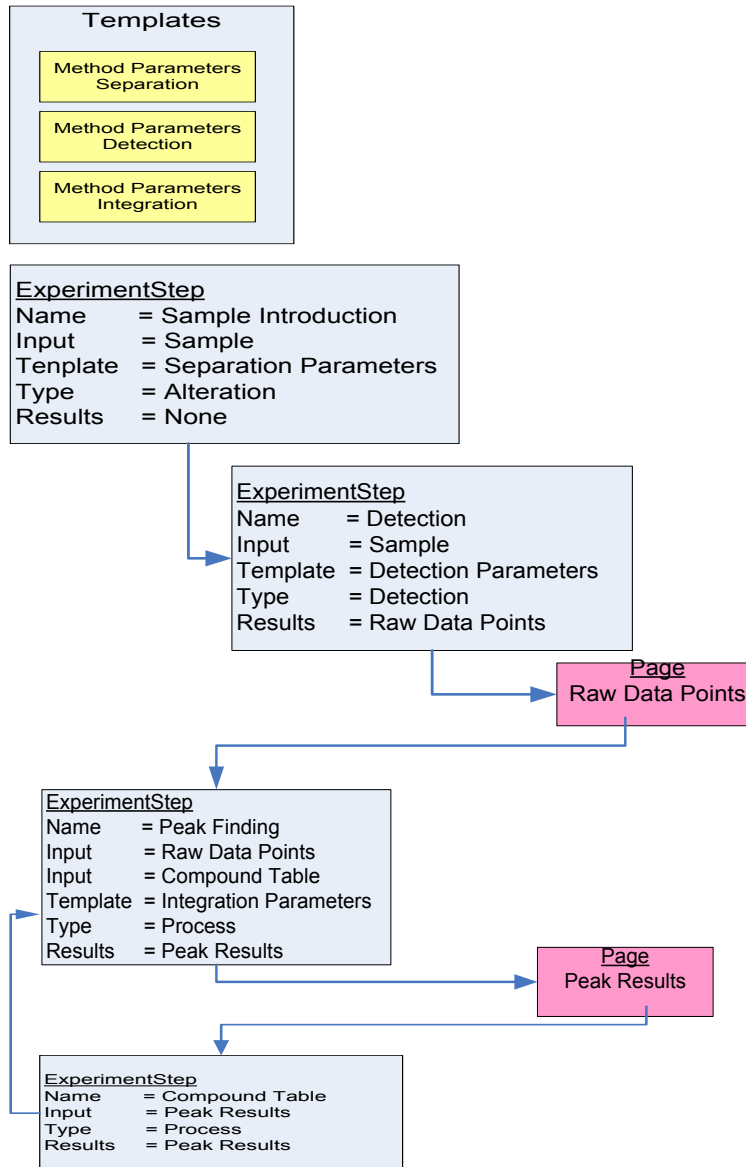Tenplate = Separation Parameters
Type = Alteration
Results = None

ExperimentStep
Name = Detection
Input = Sample
Template = Detection Parameters
Type = Detection
Results = Raw Data Points

Page
Raw Data Points

ExperimentStep
Name = Peak Finding
Input = Raw Data Points
Input = Compound Table
Template = Integration Parameters
Type = Process
Results = Peak Results

Page
Peak Results

ExperimentStep
Name = Compound Table
Input = Peak Results
Type = Process
Results = Peak Results

**Agilent Technologies**

# Example of 3 injections

ExperimentStep (Sample Introduction)

ExperimentStep (Detection)

Raw Data Points

ExperimentStep (Sample Introduction)

ExperimentStep (Detection)

Raw Data Points

ExperimentStep (Sample Introduction)

ExperimentStep (Detection)

Raw Data Points

ExperimentStep Peak Finding
Rev 1

Peak Results

Peak Results

Peak Results

ExperimentStep Peak Finding
Rev 2

Peak Results

Peak Results

Peak Results

**Agilent Technologies**

ExperimentStep (Sample Introduction)

ExperimentStep (Detection)

Raw Data Points

ExperimentStep (Sample Introduction)

ExperimentStep (Detection)

Raw Data Points

ExperimentStep (Sample Introduction)

ExperimentStep (Detection)

Raw Data Points

ExperimentStep Peak Finding
Rev 1

Peak Results

Peak Results

Peak Results

ExperimentStep Peak Finding
Rev 2

Peak Results

Peak Results

Peak Results

**Reference to the
Raw Data for
which results are
being calculated**

Agilent Technologies

```xml
<?xml version="1.0" encoding="utf-8"?>
  <ExperimentStepSet id="ID000008">
    <Template name="SeparationParameters"  templateID="SeparationMethod1" id="ID000001">
    <Template name="DetectionParameters"   templateID="DetectionMethod2" id="ID000002">.
    <Template name="IntegrationParameters" templateID="method3" id="ID000003">...

    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="PeakFinding">...
    <ExperimentStep name="PeakFinding">...
  </ExperimentStepSet>
```
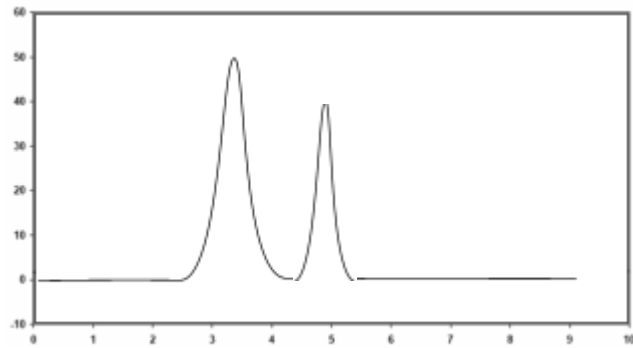
Agilent Technologies

# Detection ExperimentStep

```xml
<?xml version="1.0" encoding="utf-8"?>
  <ExperimentStepSet id="ID000008">
    <Template name="SeparationParameters"  templateID="SeparationMethod1" id="ID000001">.
    <Template name="DetectionParameters"   templateID="DetectionMethod2" id="ID000002">...
    <Template name="IntegrationParameters" templateID="method3" id="ID000003">...

    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">
      <ExperimentStep name="Detection" templateUsed="DetectionMethod2">...
    </ExperimentStep>
    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="PeakFinding">...
    <ExperimentStep name="PeakFinding">...
  </ExperimentStepSet>
```
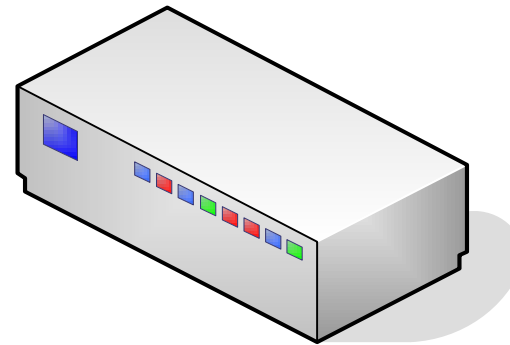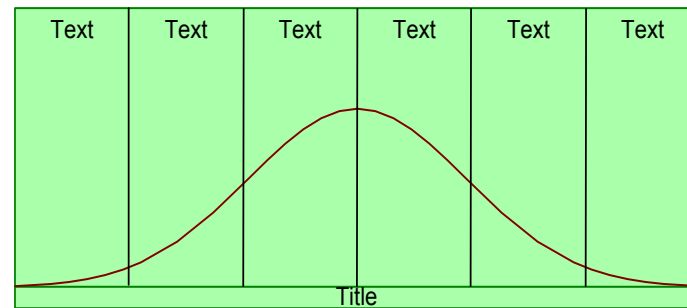
**Agilent Technologies**

# Peak Results

```xml
<?xml version="1.0" encoding="utf-8"?>
  <ExperimentStepSet id="ID000008">
    <Template name="SeparationParameters"  templateID="SeparationMethod1" id="ID000001">.
    <Template name="DetectionParameters"   templateID="DetectionMethod2" id="ID000002">..
    <Template name="IntegrationParameters" templateID="method3" id="ID000003">...

    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="SampleIntroduction" templateUsed="SeparationMethod1">...
    <ExperimentStep name="PeakFinding">
      <Technique name="token" uri="http://www.MyCompany.com" id="ID000068"/>
      <PageSet>
        <Page name="PeakResults"></Page>
        <Page name="PeakResults"></Page>
        <Page name="PeakResults"></Page>
      </PageSet>
    </ExperimentStep>
    <ExperimentStep name="PeakFinding">...
  </ExperimentStepSet>
```

**Agilent Technologies**

# Techniques

Separation

Detection

Peak Finding

Analysis

| Text | Text | Text | Text | Text | Text |
|------|------|------|------|------|------|

Title

**Agilent Technologies**

# Detector techniques

LC Detection

➤UV

➤Mass Spec

➤Flourecence

➤Refractive index

➤PDA

GC Detection

➤TCD

➤FID

➤NPD

➤ECD

➤PFPD

**Agilent Technologies**

# Mapping data to AnIML

Application developers can begin to map analytical data into AnIML by educating themselves on the following topics:

- **AnIML Core Schema**
  - This schema is the heart of AnIML, and ultimately defines the structure for all data in AnIML XML files
- **AnIML Technique Documents**
  - These schemas define the rules for your structured data, given a particular analytical technique

**Agilent Technologies**

# Mapping data to AnIML

Example

- **Mapping Position of Peak and Height of Peak into the AnIML schema**
- **Without a technique document, where do we put these items, and what are they called?**

**Agilent Technologies**
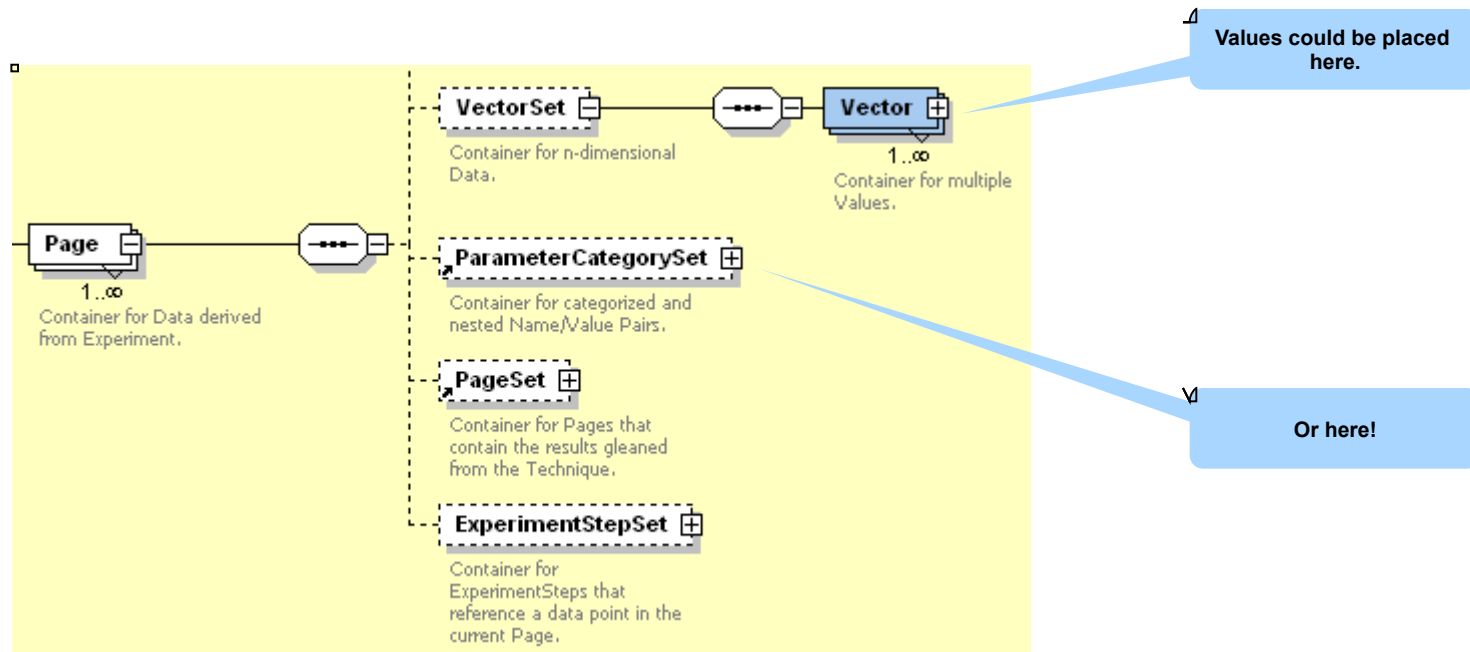
# Mapping data to AnIML
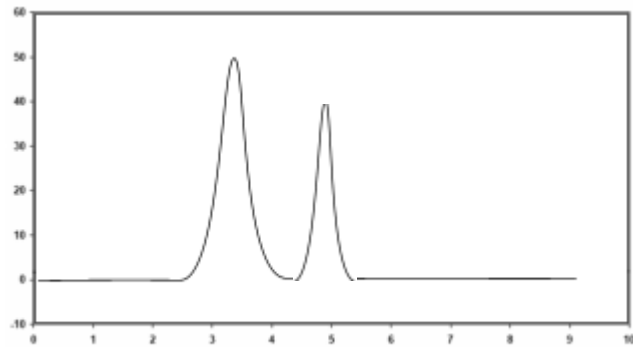
Example

- **The technique document tells us to put these items inside of a Vector, and call them PeakPosition and PeakHeight, respectively**
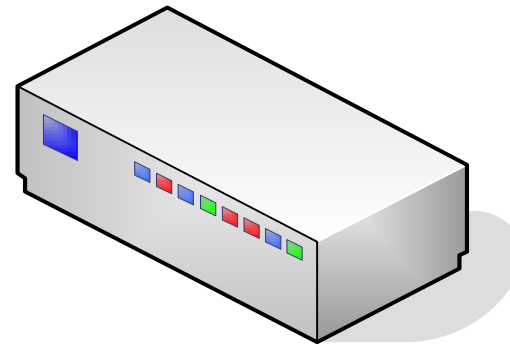
Agilent Technologies
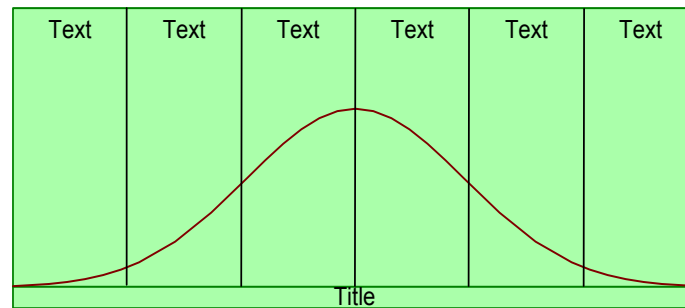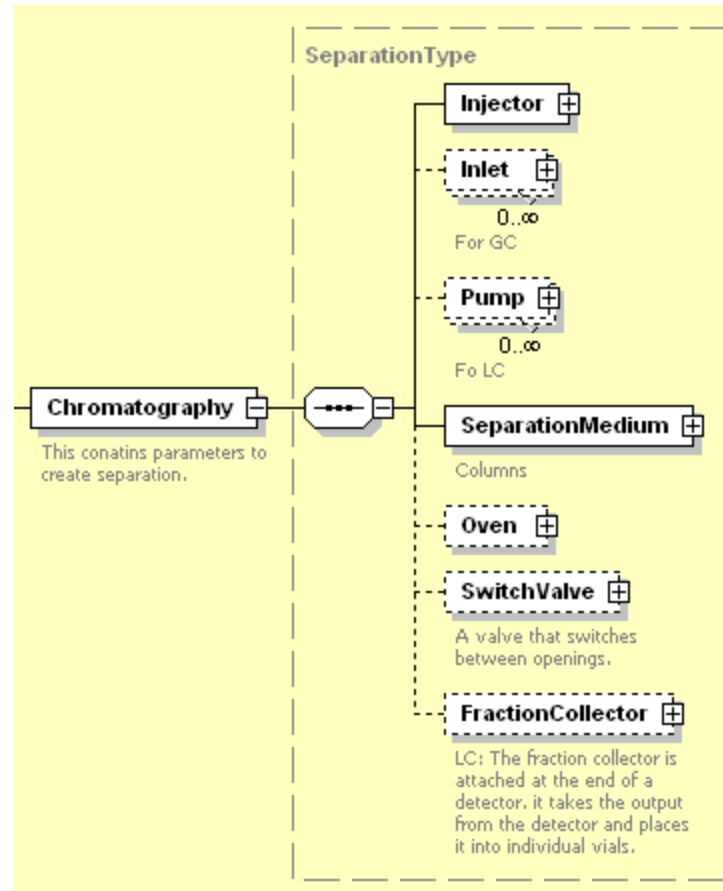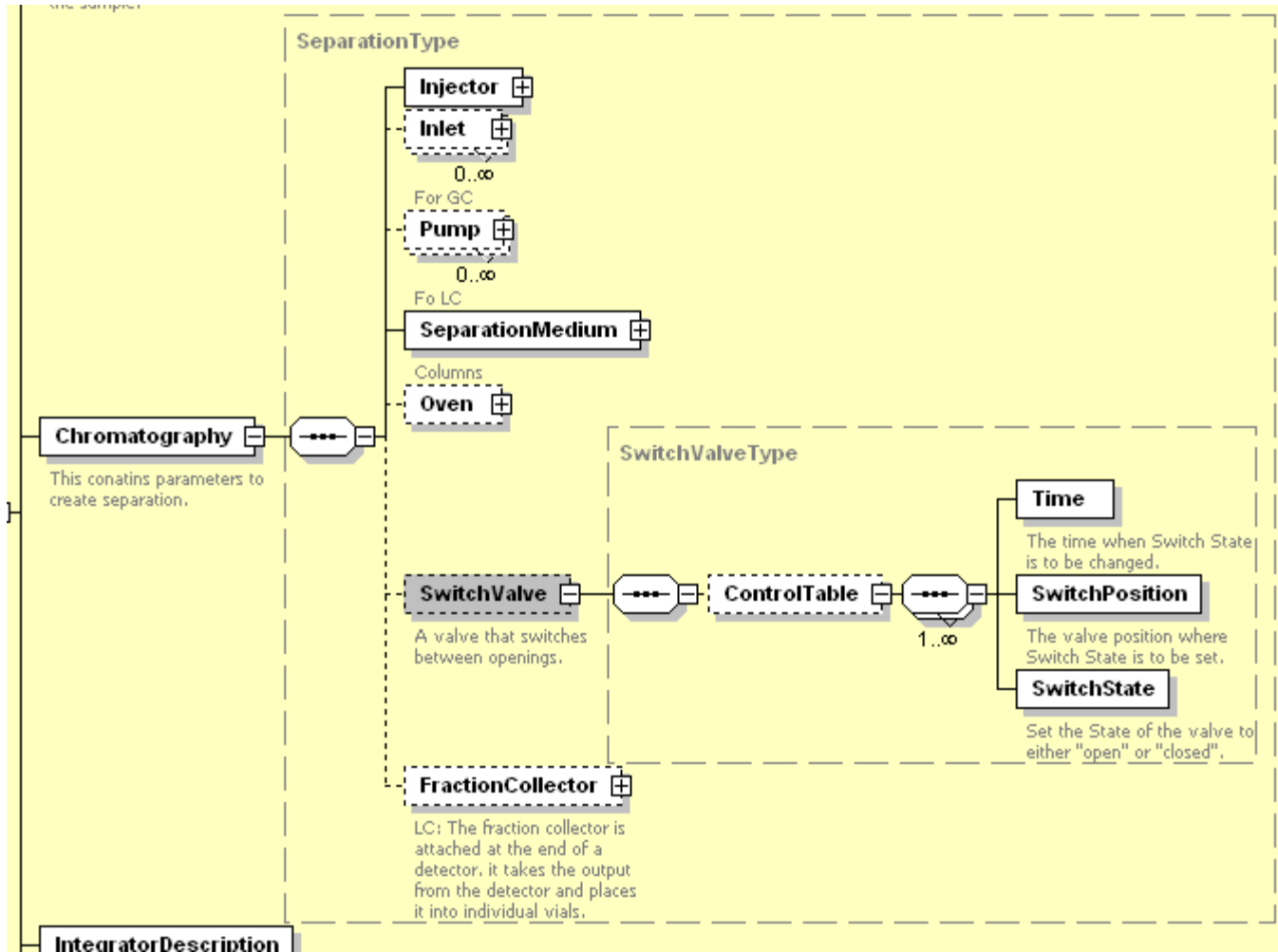
# Techniques

Separation

Detection

Peak Finding

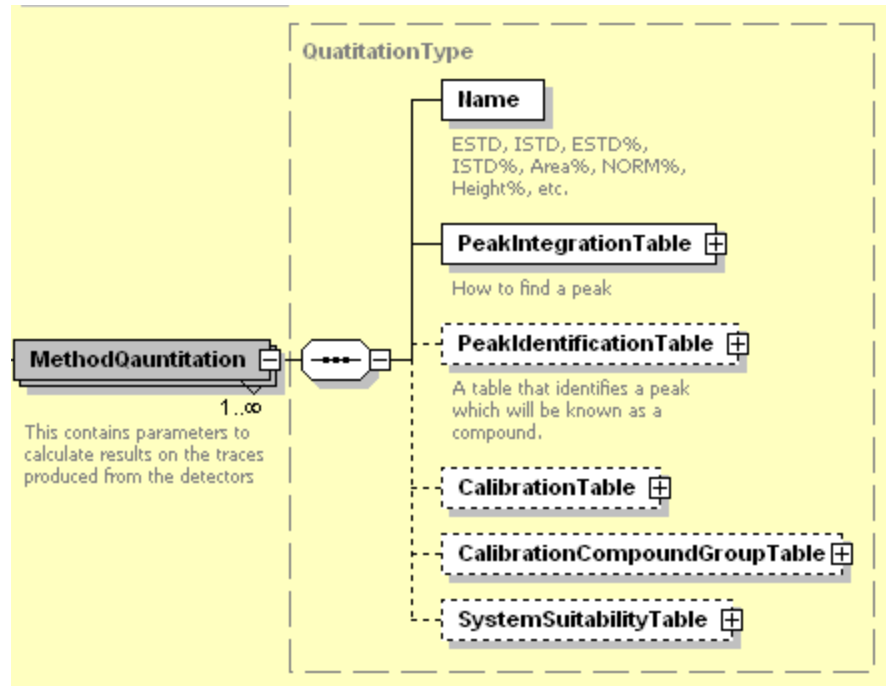Analysis

| Text | Text | Text | Text | Text | Text |

Title
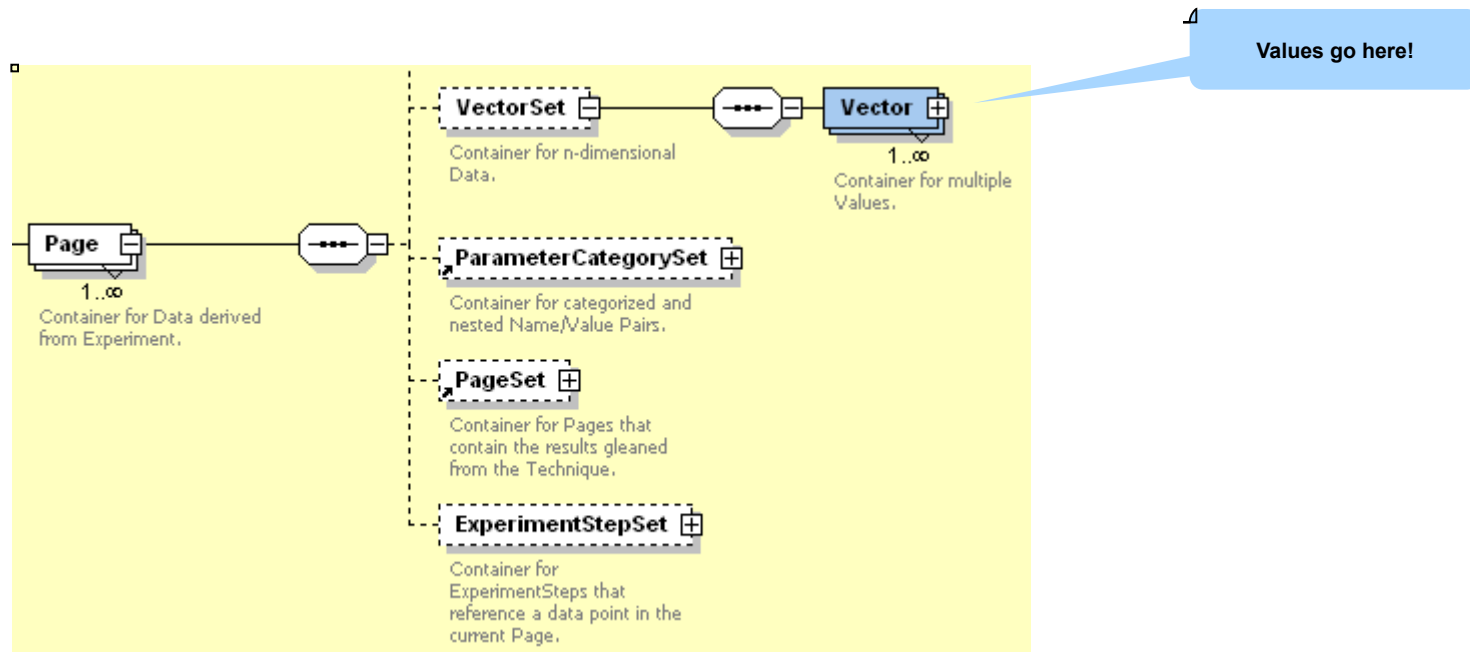
Agilent Technologies

# Chromatography - Separation

Agilent Technologies

# Switch Valve

**Agilent Technologies**

# Method Quantitation

Agilent Technologies

# Peak Identification Table

Agilent Technologies
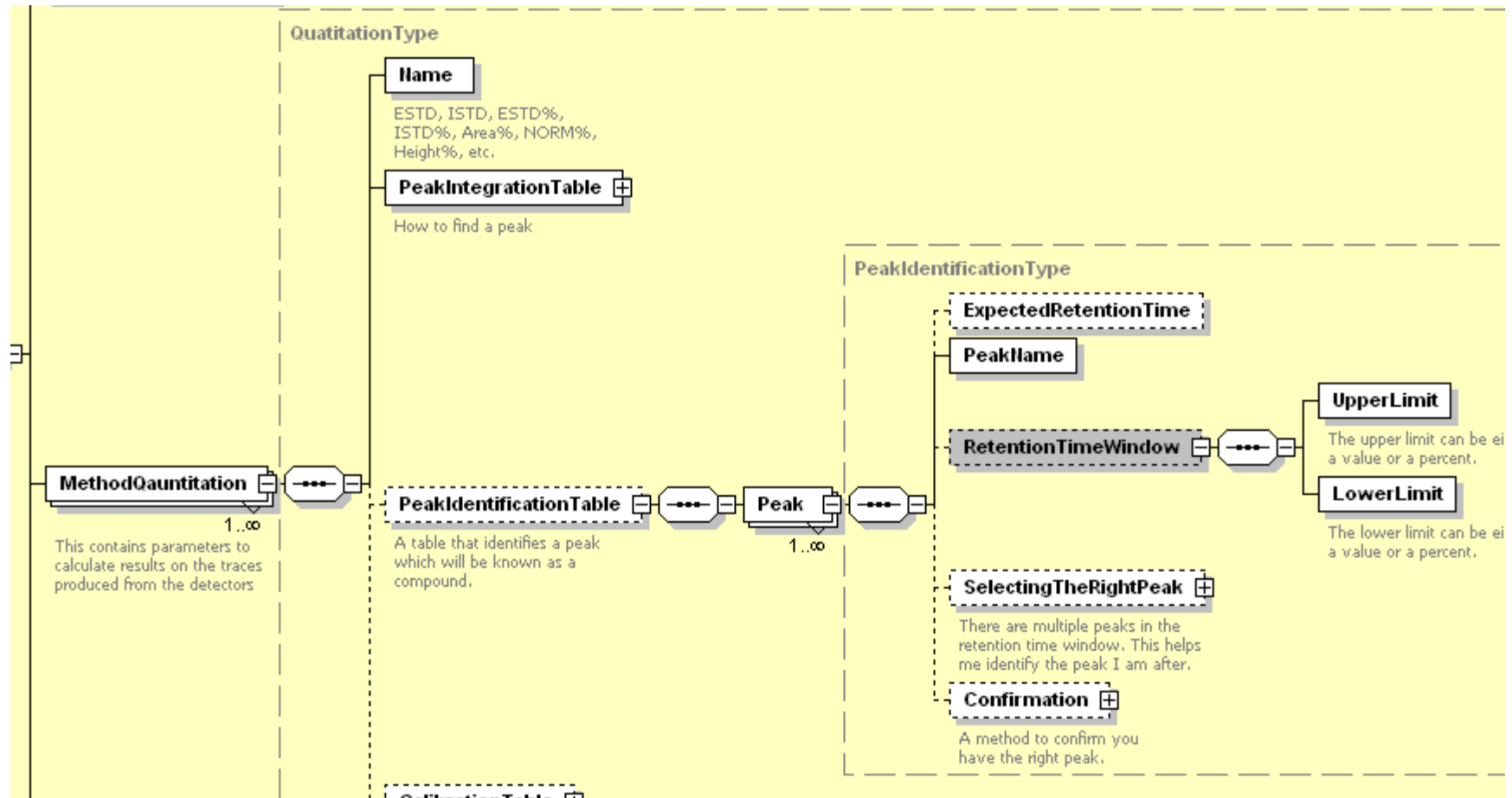
# Issues with Legacy Data

- **Understanding the data structure and organization of the target application**
- **Terminology differences between applications**
- **Finding people with knowledge of the older application**
- **Successive CDS revisions may create different formats.**
- **The original software might not be available to view old data.**
- **Each CDS system has its own data model with a long development history**
- **Documentation of the data model is incomplete in most cases.**
- **API's to data is sometimes incomplete.**
- **DateTime issues – what date is 2/5/79. There is no locale**

Agilent Technologies

# Demo

- **View real AnIML XML file**
- **View same AnIML file in Agilent's AnIML File Viewer**

**Agilent Technologies**

# Summary

- **Massive amounts of data are being generated**
- **Much of this data must be kept for 30+ years**
- **Applications retire, but the data must live on, in a TNF format**
- **AnIML is being created by the ASTM subcommittee E13.15, and is <u>the</u> standard for TNF representations of analytical data**
- **AnIML is a highly structured, but flexible file format**
- **Tools, applications, and viewers are already being generated around AnIML**

**Agilent Technologies**

# Questions

Agilent Technologies