ETHOSOFT

Legacy Data Migration to a New LIMS

Bill Pingpank, *VP Client Services*
Ethosoft, Inc.

Presentation Overview

1) To Migrate or Not to Migrate…That is the Question
   - Decision factors
2) What Data to Migrate
   - What stays and what goes?
   - Details matter
3) Data Formatting
   - How organized is your data?
4) Data Validation After the Fact

Decision to Migrate

- Identifying Reasons for Migration Requirement
  - May be technical in nature
  - Operational needs

Reasons To Migrate

- Technical Considerations for Migration
  - Current systems are no longer supported
    - Hardware/Software
  - Data becoming corrupt and getting worse

- Operational Considerations
  - Seamless historical trending moving forward
  - Reports ability to access both new and old results
  - Regulatory requirements

Reasons Not to Migrate

- Isolate and Keep Legacy System On-Line
  - Virtualization makes this easier
  - Supportable and stable current technical environment
  - Data currently organized and accessible
  - Need to keep historical reporting available
  - Regulatory requirements
    - 5 year data on-line and available

What to Migrate

- Results … *of Course!*
  - *What Constitutes a Result?*
    - *Where is it from? Location centric or naming conventions*
    - *Analyte name*
    - *Numeric or text result value*
    - *Units of measure*
    - *Qualifier(s)*
    - *Date of analysis*
    - *Analysis by*
    - *Method references*
    - *Dilution factors*

What to Migrate (cont.)

- Quality Control vs. Non QC Results
  - Complexity consumes time
  - Creating necessary traceability is complex
    - Mapping to LIMS production structure

      OR
    - Mapping to LIMS historical specific structure

Where Data Originates

- Identifying Historical Data Repository(s)
  - Why It Matters:
    - Historical data formatting variation influences cost
  - Most Common:
    - Legacy LIMS system
    - Data warehouse
    - Spreadsheets
    - Various file types (COAs, Instrument Outputs…)

Where Data Originates (cont.)

- Data Consistency
  - Identification information consistent?
    - Sample numbering
    - Sample location naming constructs
    - Client naming
    - Test / method / analyte naming
    - Cross referencing becomes easy

Legacy LIMS

- Centralized or Relational Structure
  - Database
    - Oracle
    - SQL Server
    - Access
  - Complex structure = more cost
  - Familiarize yourself with your legacy data
    - Work with your IT staff

Data Warehouse

- Combination of Lab and Other Data
  - All or some of the warehouse information
  - Non-laboratory data
    - SCADA data
    - Operational data

Spread Sheets

- Identify The Required Formats
  - Number / datatypes of format columns
  - Single format
    - Most cost effective to import
    - Easiest to validate
  - Multiple Formats
    - Each different format requires unique approach
    - Cost grows by number
    - Increases validation work

Data Files

- Least Preferred Method
  - Data likely disparate and in multiple formats
    - Formats / file types change over time
    - Examples
      - COAs
      - Instrument output
  - Requires more thorough approach
    - Costs increase significantly
    - Validation efforts grow

Determine Migration Specifics

- Devil is in the Details
  - Identify the range of data
    - Date ranges
    - Specific locations
  - Identify the data attributes
  - Reduce the number of sources
  - Document the requirements

Data Migration Validation

- Validate Each Source/Data Set
  - Spot check using reports / queries defined prior to migration
  - Confirm all elements of spot checked data
  - Identify source of any conflicts
  - Re-migrate source data
  - Validate and repeat until successful

Keys to Success

- Prior to Beginning:
  - Identify all data sources
  - Specify data elements
  - Review data sources
  - Document requirements
- Once Underway:
  - Minimize deviations
  - Validate the migrated data

Questions and Answers

- Q&A