

# Challenges of What, Why, and How of Clinical Metadata

## Beginner's Guide to Metadata

d-Wise Technologies

Chris Decker

Life Sciences Director

# Overview

- What is metadata?
- Why do we need metadata?
- How do we collect metadata?
- Case Studies: Innovative and not so innovative metadata projects
- Gaps in Existing Technologies
- Metadata in 2020

# What is Metadata?

- In the past it was limited to variable names and labels
- Always referred to as “Data about Data”
- Encompasses so much more...
  - Information about the process
  - Relationships between data, reports, processes and all clinical artifacts

# Why do we need Metadata?

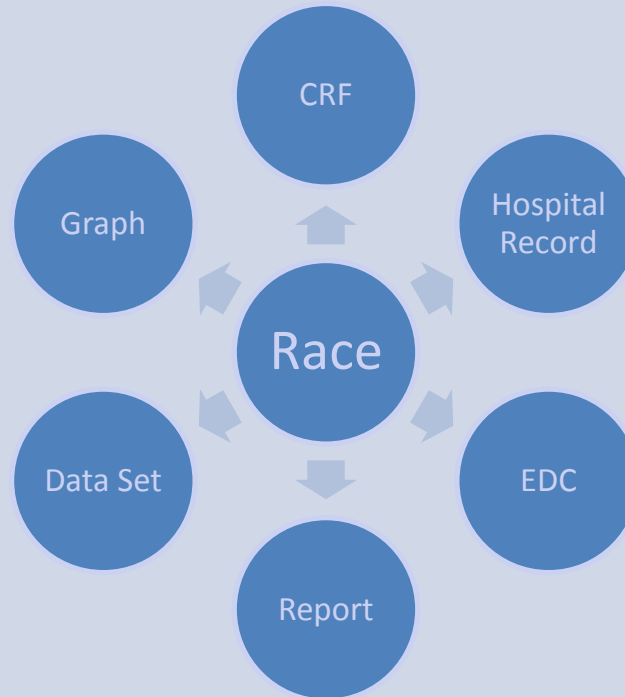
- Metadata can drive and automate processes
- Improve efficiencies across studies
- Support standards and technologies
- However...
  - Metadata must be robust and complete
  - Must go beyond defining just data

# How do we collect metadata?

- Excel is the universal data entry tool (NOT!)
- Need to build a metadata model supporting
  - “data about data”
  - Relationship between data elements
- Standards define “data about data” and some relationships but NO process for applying them
- Need robust tools to store, access, and manage metadata

# How to think about Metadata

- The world is not flat and neither is clinical data
- Need to think about abstract objects



# Requirements of Metadata Management

- Ability to store “data about data”
- Support relationships between data elements
- Support standards and incremental updates to those standards
- Access metadata from a variety of tools
- Capture metadata about the process

So I called this meeting today because the developers say they need our marketing data for their software. Unfortunately, all the data is stuck in an Access database, and I don't know how to use Access, so we have to figure something out. Any ideas?



Access?! If they wanted a database, then why didn't they just use Word?



You idiot! Word is a spreadsheet program. Obviously, we need to somehow get this data into Excel.



I know! I know!  
We could cut and paste!  
I know how to cut and paste. We could do that!



No! No cutting!  
Remember what happened last time?  
Look at my stump!



Ooooooh! I know!  
We could print out the data, scan it in, and then paste the image into Excel!



# Case Study #1: Gaps in an ETL Tool

## Problem:

- Customer implemented an ETL tool for transforming data to SDTM
- Metadata interface did not meet their needs
- Example of a technology not really designed for managing metadata

# Case Study #1: Gaps in an ETL Tool

## Solution:

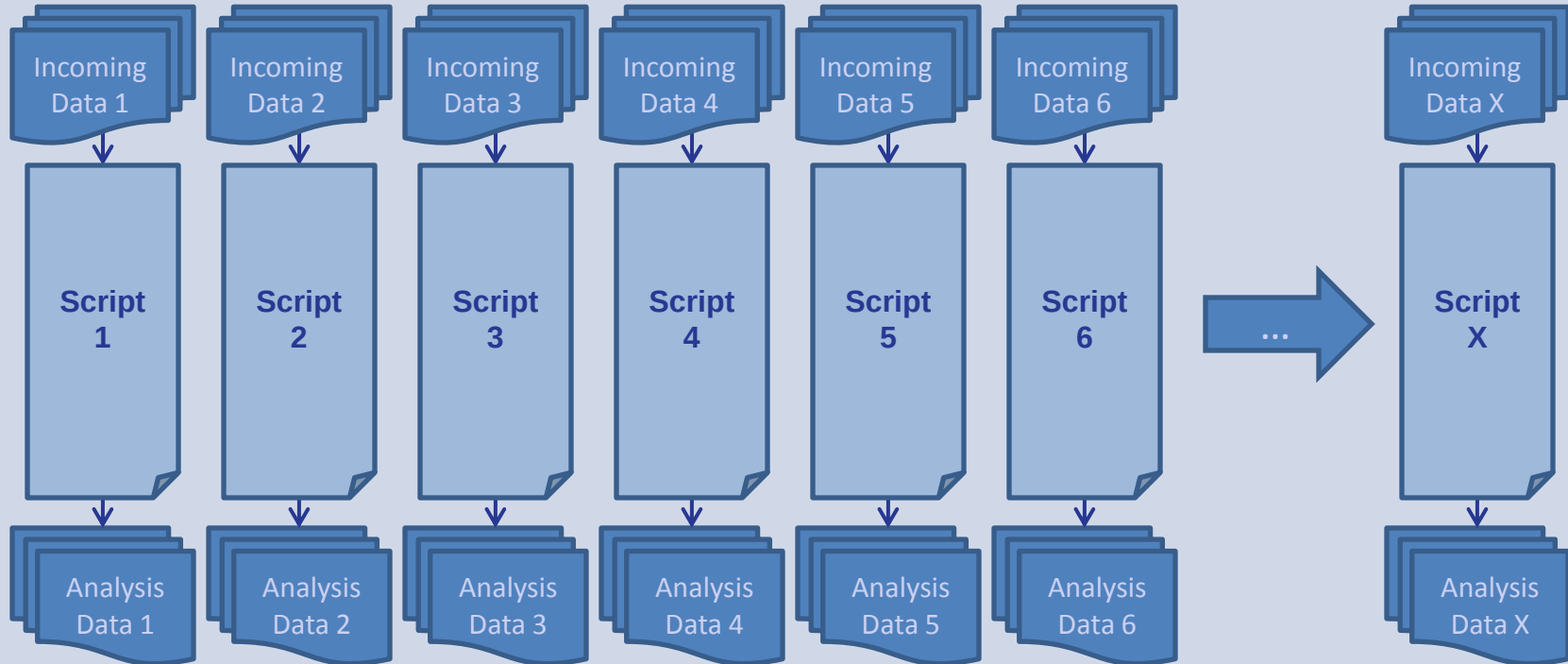
- Customer stuck on using Excel for metadata
- Built integration interface to push Excel metadata in and out of the ETL tool
- Full of Problems
  - Excel not a robust tool
  - Not able to provide rigorous controls
  - Excessive programming to support a non robust tool

# Case Study #2: Automating Data Flow

## Problem:

- Customer outsources 95% of their data
- Received from a variety of vendors through a single portal
- Used scripting language to receive data, run steps, and generate standard data
- Exponentially created many one off scripts
- Poor implementation of an open source technology

# Case Study #2: Process Becomes Unmanageable

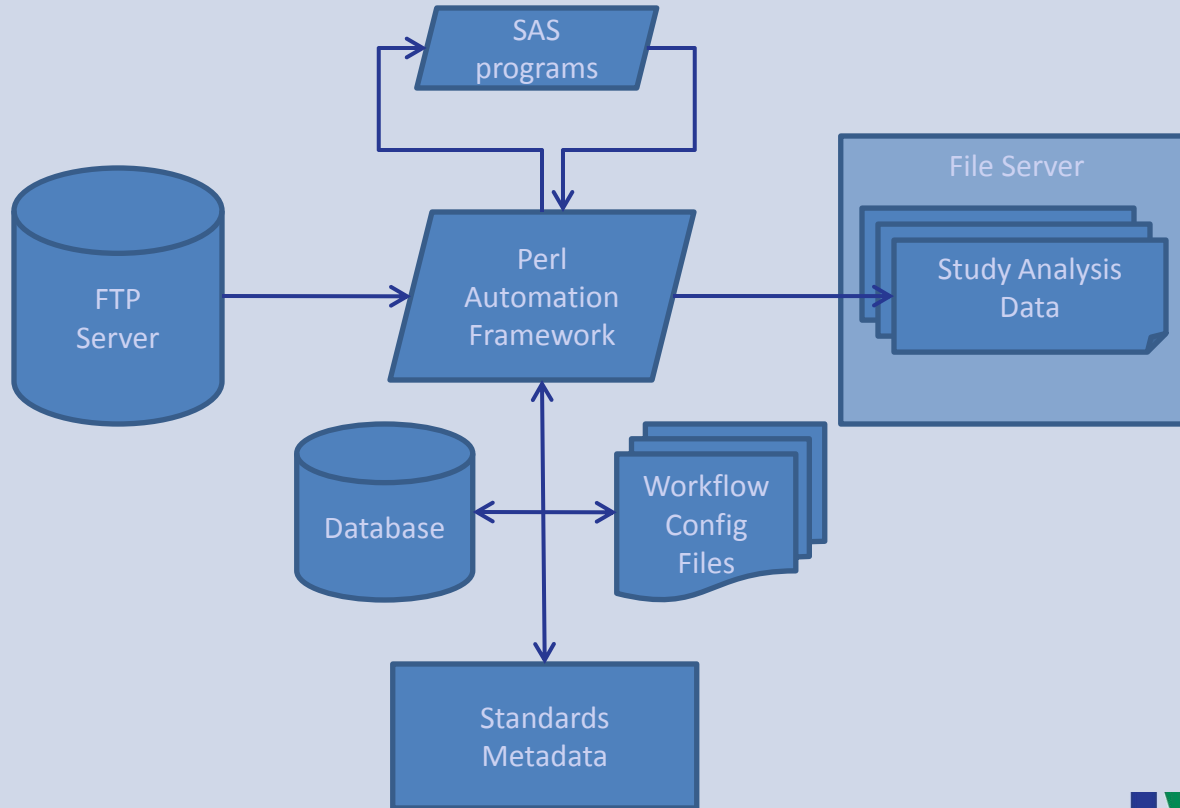


# Case Study #2: Data Flow Requirements

*The system must...*

- support varying input data formats from CROs
- externalize configuration details such as user credentials and file server paths (Process metadata)
- provide a reusable and extensible library of actions from which to build work flows
- Support standards within the library of actions
- provide a “single source of the truth” for understanding how individual data flows are defined

# Case Study #2: Metadata Framework



# Case Study #2: Role of Metadata

- Metadata at the **Data level** – running functions that apply standard conversions
- Metadata at the **Configuration level** – location of data, passwords, files to execute
- Metadata at the **Workflow Level** – triggers to start workflow and rules for workflow direction
- Metadata at the **System Level** – Notification of who, when, and how; information about systems
- **Overall Goals**
  - Proper tradeoff between automation and flexibility
  - Separate programming logic from configuration and metadata logic

# Case Study #3: ODM Drives the Process

## Problem/Challenge:

- Customer collects metadata in ODM
- How to integrate this with the data transformation process
- Proof of Concept to meet objectives:
  - Research/evaluate options to generate standard data from ODM metadata/clinical data

# Case Study #3: ODM Drives the Process

## Questions

- What is the best tool for processing XML?
- Best technology for storing and managing transactional data and associated metadata?
- Optimized language for creating analysis data?
- How do we use metadata to automate the process?

# Case Study #3: ODM Drives the Process

## Solution

- XML Processing Platform
  - Java Parser to process XML data
  - XML database to store metadata/data
  - Views to access 'tabular' data
- Programming Framework
  - Code written has small functional units
  - Manage code units within the metadata framework
- Workflow Wrapper
  - Automate load of ODM data from external systems
  - Study management component to manage metadata about source, data, and programming
  - Triggering mechanism to launch and manage the ODM to data process

# Gaps in Existing Technologies

- Tools manage metadata as an after thought
- ETL tools focus on the transformation process
- Desktop tools are not robust
- ECD vendors focus on data collection
- All solutions are very data centric
- Must start with 'metadata driven' in mind

# Metadata in 2020

- Multi-Dimensional Object Oriented Metadata (M-DOOM – needed a catchy acronym)
- What do I really mean?
  - Need to support hierarchical metadata relationships – ODM?
  - Need to think outside of just the clinical data
- SHARE? HL7? Something else...

# Summary

- Metadata is more than just variables names and labels
- Need a more robust relationship model – ODM? SHARE? HL7?
- Need to think beyond data sets and focus on process
- Technologies exist – defining the metadata model is the hard part

# Questions?