Agenda

- Who are we?
- Overview of End to End with CDISC
- What Types of Metadata will we talk about?
- What is the ODM and Define.xml?
- End to end uses of ODM and Define.xml
- Typical Clinical Data Transformations
- Metadata driven Extract, Transfer & Load
- Conclusions
Introduction to Formedix

Background

• Involved with CDISC Standards since 2000
• CDISC Software Products
  – Formedix Origin™ - first ODM and define.xml study design tools
  – Formedix Transform™ - EDC and paper integration products
  – Formedix Submit™ – data transformation engine
• Must Work with Existing Processes and Tools
• CDISC Consultancy Services
  – Planning, preparation and implementation
  – Cross trained in all the CDISC metadata & data models
  – Technical Director won 2 awards from ODM team
• Optimize Study Design & Downstream Data Conversions
What Types of Metadata will we talk about?

- **Database Standards**
  - CDISC Operational Data Model (ODM) – Metadata and Data

- **Submission Standards**
  - Define.xml/Case Report Tabulation Data Definition Specification
  - Study Data Tabulation Model – data in SAS transport files

- **Mapping Metadata**
  - Not a CDISC standard
  - Describes how to go from SDTM to Legacy Datasets
  - Describes how to go from SDTM to ODM
  - CDISC please help it become a standard!!!
What is the ODM?

Overview

• The ODM is a model that describes the content and structure of a CRF or database
• Two sections
  – Metadata – design of CRF or database
  – Data – transactional and snapshot
• ODM data is hierarchical and arranged around subjects
• Best Practice ODM Modelling
  – Balance data entry ease vs. downstream transformations
  – Get it wrong here – pay the price later!!
  – A lot of the metadata in the ODM ends up in the SDTM
    • Look at an annotated CRF
    • Category, position etc.
What is Define.xml?

Overview

- Based on the ODM model
- Not to be confused with Operational Data Model which describes the content and structure of a CRF/database
- Define.xml
  - Describes the content & structure of datasets
  - “Mash-up” of ODM with extensions
- SDTM datasets are arranged around findings, interventions and events NOT around subjects
- Define is for metadata
  - No corresponding data model in XML – SAS transport
  - Latterly HL7?
• Formal Specifications do not Exist for the Entire Study Set-up Process
• Raw Data ➔ Proprietary ➔ CDISC SDTM Datasets
• Downstream dataset design/specification not addressed at Study Set-up leading to a specification “gap”
• Study Specification
  – Dynamic Creation of Study Specifications
    • For all Roles
    • For all Systems
• Build
  – Automated build of EDC/CDMS
• Testing
  – “As specified” vs. “As built”
• Downstream
  – Replacement for datasets
  – Easiest way to get to SDTM
Define.xml
End to End Uses

• Study Set-up
  – Design of datasets
  – Establishment of libraries

• Study Execution
  – Automation of dataset conversions
  – Automation of Testing
    • As specified” vs. “As built”

• Submission
  – Lifecycles of define.xml
  – Define.xml for submission in eCTD
Typical Clinical Data Transformations

- 3/2 part & non-ISO 8601 date/time
- Numerically coded text values
- Text & Numeric result values
- Administrative/system variables
- Tables, Panels, Screens, Pages
- Non-standard variable names
- Horizontal data structures
- Repeating-group data structures

- 3/2 part & ISO 8601 date/time
- ISO 8601 date/time
- Decoded text values
- Text result values
- SDTM administrative variables
- SDTM domain & variable names
- Vertical data structures

April 3-5, 2006
Oncology Clinical Trials - Strategies to Accelerate Development and Decrease the Cost of Oncology Clinical Trials
CDISC Metadata Driven
Extract Transfer and Load (ETL) Process

• Define the Source (ODM or CRTDDS)
  – Raw data – use ODM metadata
  – Your Proprietary or Legacy datasets (CRTDDS)

• Define the Destination (CRTDDS)
  – Any destination dataset format is supported (your datasets/SDTM)
  – Destination dataset shell description held within CRTDDS
  – Enables multiple types and combinations of transformations
    • Raw Data ➔ Proprietary/Legacy ➔ SDTM today’s reality
    • Legacy ➔ SDTM legacy conversions
    • Raw Data ➔ SDTM ➔ Proprietary future vision
    • Raw Data ➔ SDTM future vision

• Have Your Cake and Eat It !!!
  • Working with existing “Inward Facing” and “Regulator Friendly CDISC”
CDISC Metadata Driven
Extract Transfer and Load (ETL) Process

• Define the Wiring/Mapping (Mapping Metadata)
  – Needed why? We have a wiring problem ....
  – Extended define.xml or XML language
  – Modelled from destination back to source
  – Destination data comes from
    • ODM, in-house datasets & SDTM data and metadata

• Automatic Data Transformations
  – Today manual raw data ➔ SDTM since no machine readable metadata
  – Metadata drives extract, transfer and load process = no manual coding
  – Engine uses ODM, define.xml, and ODM data

• Convert to proprietary ETL metadata to drive any ETL engine
Ultimate Flexibility
Study by Study Variability with No Coding

• Metadata controls all study to study variability
  – New define.xml and mapping metadata
  – Study to study effort minimized by library re-use

• Transform Engine remains static
  – Structural Transformations
    – One Source ➔ Multiple Destinations
    – Multiple Sources ➔ One Destination
  – Variable Level Transformations
    – Derivations (Function calls)
    – Hard-coded strings
    – Conditional mappings (If, Then, Else)
    – Decodes
    – Codelist mappings and more ….
End to End Clinical Trial Process
With Standards
Conclusions
CDISC Metadata Delivers

• Scalable end to end process
• Define.xml is very important throughout
  – Represent any type of tabular structure
• Populated ODM is better than datasets
• Multiple types of transformation possible
  – Dataset to Dataset
  – Operational Data to Proprietary Dataset to SDTM
  – Have your cake and eat it!
• Gaps still exist
  – Mapping metadata & define for data
Why ODM and Not Datastets?

- Hierarchical standardised and documented structure
- Easy to find all that you want – metadata and data
- Develop generic programs
  - Navigate the tree
  - Metadata driven – get this data from there in the metadata
- Easy to find data
  - Datasets have no relationships – artificially create
- Change SDTM version
  - Datasets approach need reprogramming
  - Metadata approach does not