

Dave Handelsman
Principal Strategist, Clinical R&D, SAS

October 2005

Agenda

- Part I: What is CDISC and Why Standards?
- **Part II: Available CDISC Standards and Models and How SAS Supports These**
- Part III: CDISC Collaborations and Standards in Progress



Clinical Data Interchange Standards Consortium

CDISC is an open, multidisciplinary, non-profit organization committed to the development of worldwide industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trials data and metadata for medical and biopharmaceutical product development.

The mission of CDISC is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.

CDISC Standards

- **Operational Data Model (ODM)**
- **Submission Domain Standard (SDS)**
 - Study Data Tabulation Model (SDTM)
 - Standard for Exchange of Non-clinical Data (SEND)
- **Analysis Data Model (ADaM)**
- **Laboratory Data Model (Lab)**
- Case Report Tabulation Data Definition Specification (CRTDDS - aka define.xml)
- Protocol Representation Standard
- Biomedical Research Integrated Domain Group (BRIDG)



ODM – Operational Data Model

- Written in XML
- Clinical trials data recording / transfer / archive
- Contains both study metadata and clinical data

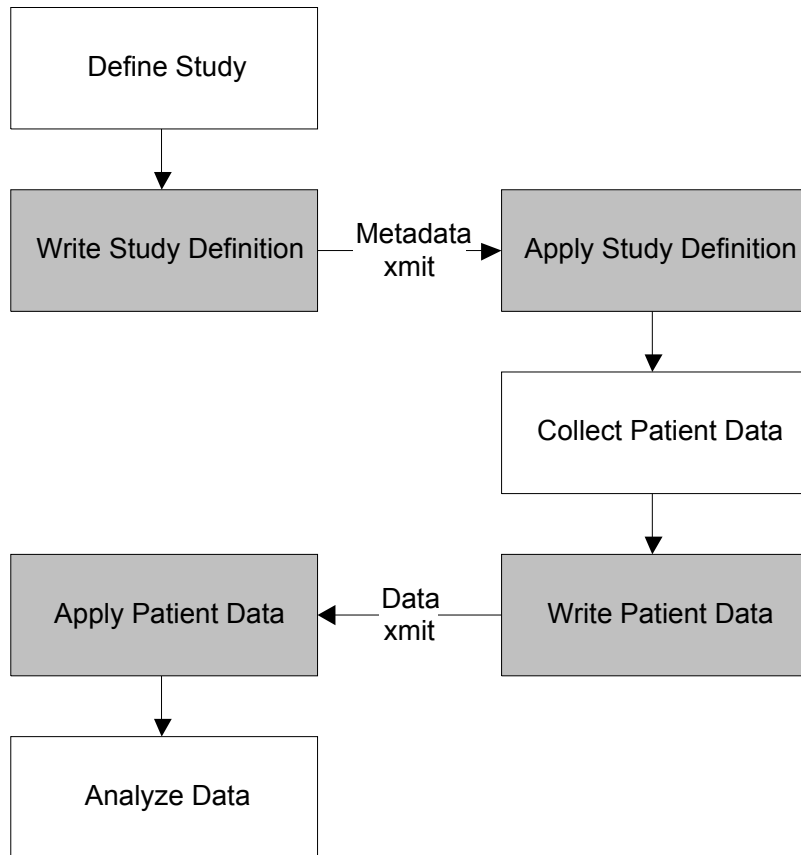
Model Features:

- Framework can be used by any clinical data management (CDM) system or clinical application
- Supports 21 CFR Part 11 audit trail and electronic signatures
- Supports archiving & data retention guidance
- Provides for incremental transfers
- Includes vendor extension capability

Interchange cycle – move to ODM

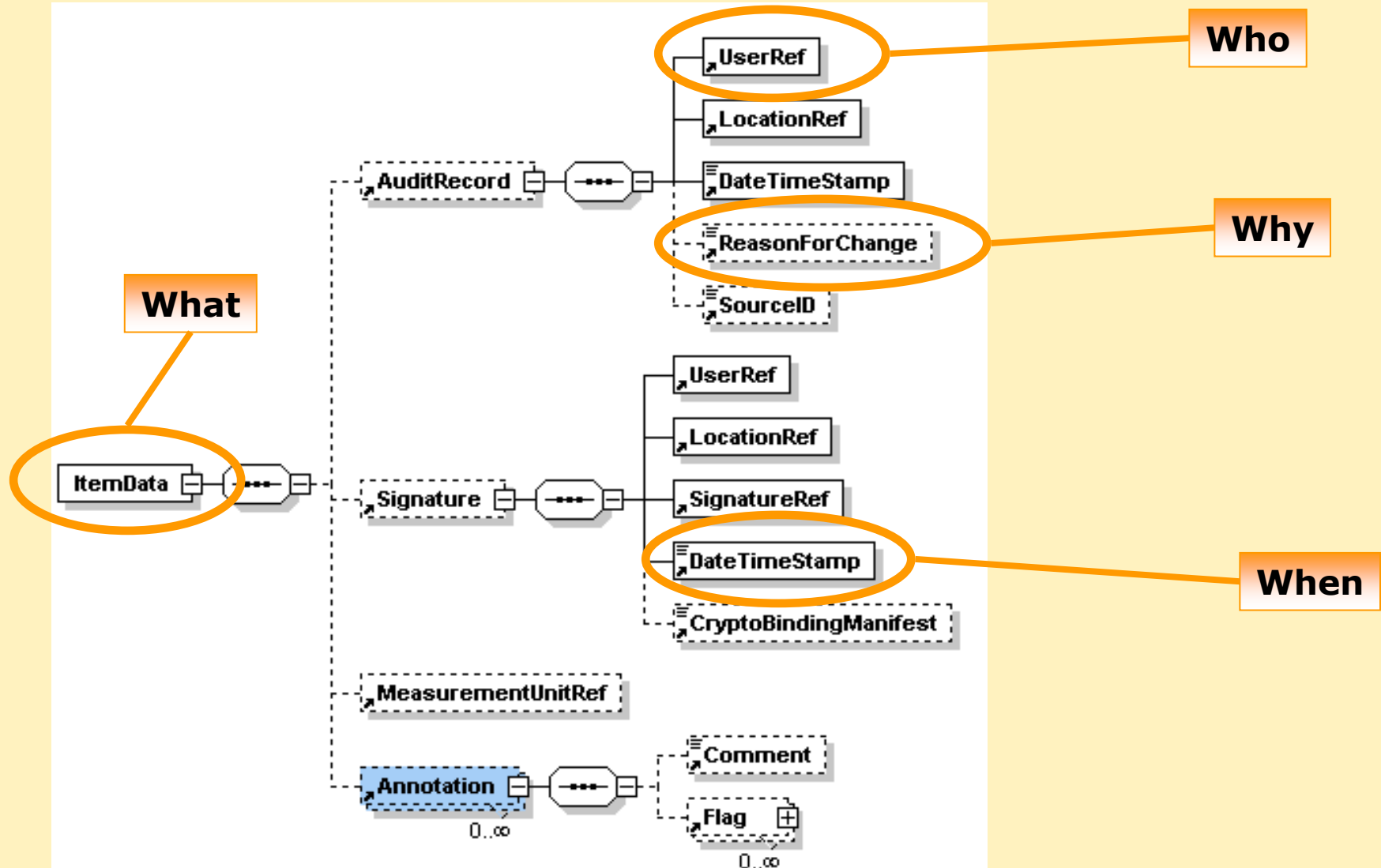
Sponsor

Helper



- Traditional approach
 - MetaData Xmit = Excel spreadsheet
 - Data Xmit = SAS Datasets
- **CDISC approach**
 - **MetaData Xmit = ODM XML**
 - **Data Xmit = ODM XML**
- **CDISC Benefits**
 - **More information exchange**
 - **Increased automation**

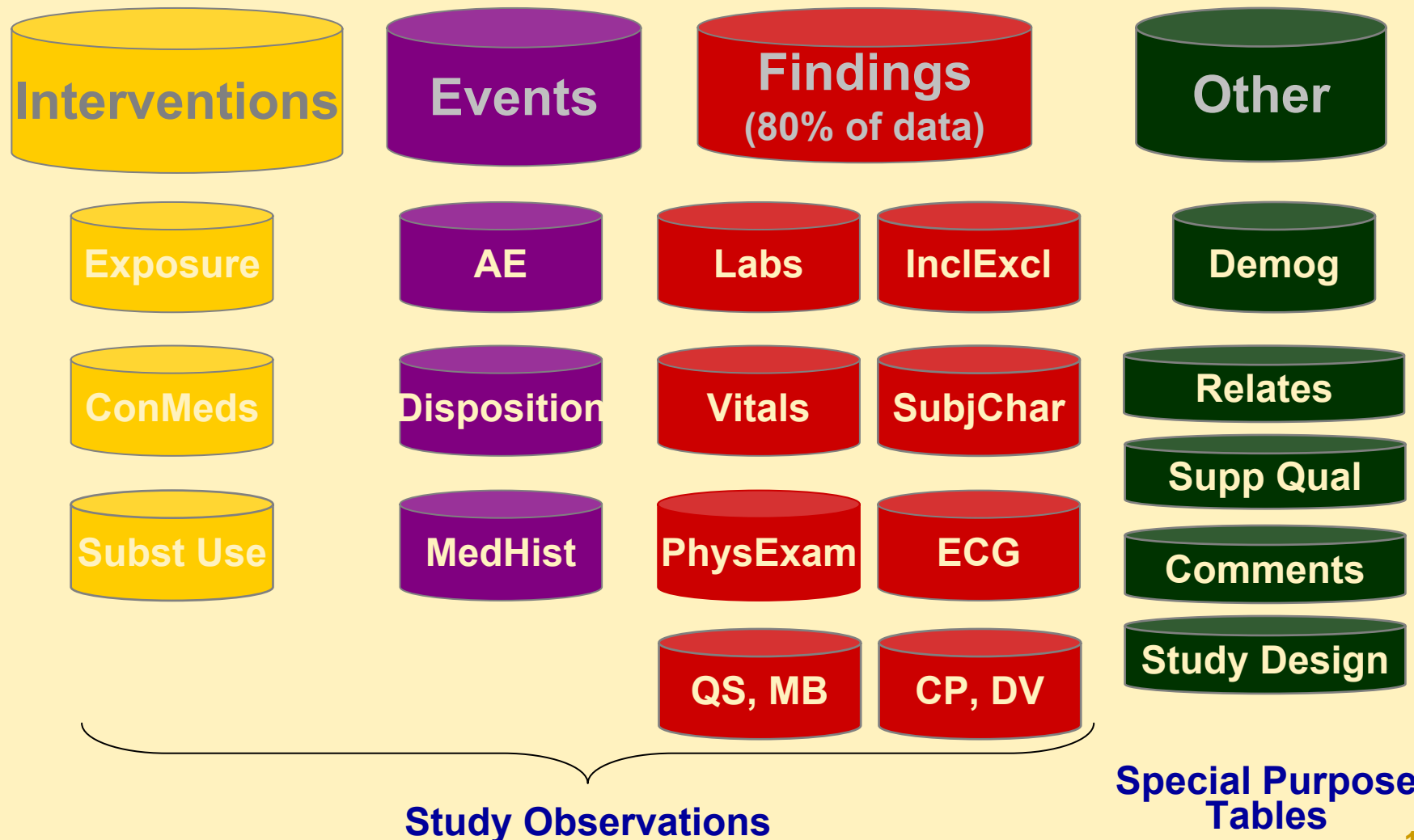
ODM & Audit Trail



SDS, SDTM and SEND – Submission Standards

- **SDS: Submissions Data Standards**
 - Refers to the ***implementation*** of the model for ***clinical submissions*** (e.g. SDS v3.1.1)
- **SDTM: Study Data Tabulation Model**
 - Refers to the **model** that supports ***clinical and preclinical submissions*** data
- **SEND: Standard for Exchange of Nonclinical Data**
 - Defines ***domains and variables*** for ***submitting*** all data generated from **animal toxicity studies**

SDS V3.1.1: CDISC Standard Domain Models Observation Classes



SDTM Basic Concepts

- Utilizes standard variable names, metadata, and data structure
- Each domain has required, expected, and permitted variables
- No redundancy (other than identifying variables)
- SDS team has developed standards, containing specific metadata for >20 domains

SDTM

- 3 general observation classes
 - Interventions
 - Events
 - Findings
- Special domains
 - Demographics
 - Comments
 - Relationship between records in multiple domains
 - Supplemental qualifier domains

Specifications For Organizing The Datasets

<http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.1.pdf>

 [folder name]	Replace with folder name, e.g., m5
 Datasets	
 [study]	Replace with study identifier, e.g., 123-070
 analysis	Contains analysis datasets and associated files
 programs	Contains program files
 ecgs	Contains annotated ECG waveform datasets
 listings	Contains data listing datasets and associated files
 profiles	Contains subject profiles
 tabulations	Contains data tabulation datasets and associated files

SDTM Benefits

- Increase in efficiency
 - Data transfers
 - Data reviews
 - Creation of analysis files
- Flexible
 - Supplemental domains
 - Can accommodate a mixture of raw and derived

ADaM – Analysis Dataset Model

Uses for Analysis Datasets:

- Provide clear communication of the scientific results to regulatory reviewers
- Replicate or verify the sponsor's analyses, results, and conclusions
- Test the validity and robustness of the sponsor's analyses and assumptions (what if...)
- Audit the data for inconsistencies and errors
- Status: Analysis Dataset Models Completed
 - Change from baseline
 - Time to event
 - Categorical analysis
- Working on Additional Dataset Models
 - Linear model
 - Safety Analysis
- Ensuring compatibility with SDS Version 3.1 (SDTM)

Lab – Clinical Laboratory Data Model

- Core model designed to handle “simple lab data”: one test, one result structure
- 10 levels of content includes support of administrative information for transfer of samples, reporting of results
- Production Version 1.0.1 - on CDISC web site
 - Specification, data fields, reference ranges
 - Includes Microbiology extension
- Currently standard approaches are included for
 - Bar delimited ASCII (default representation)
 - SAS
 - XML
 - HL7 RIM message
- More complex tests - handled by extensions
 - Work on pharmacogenomics beginning

Case Report Tabulation Data Definition Specification (CRTDDS aka define.xml)

- Machine readable replacement for define.pdf currently used in submission process
- Organizes and describes submission content
- Extends ODM metadata for submission content
 - Enhances general column descriptive information
 - Adds item-level metadata support
 - Supports linkage to external folders and files

Protocol Representation Standard

- Initiated as an HL7 project to develop standard representation of clinical trial protocol elements
- CDISC team formed to provide domain expertise
- Spreadsheet of elements, with glossary definitions open for public review and comment
- Initial Clinical Document Architecture (CDA) model developed and balloted through HL7

CDISC Biomedical Research Integration Domain Group (BRIDG)

- Follows the HL7 Development Framework
- Initiated in January 2004 by CDISC Board
- Developed through numerous modeling sessions with domain experts and reviewed by CDISC, industry, FDA, NCI/NIH groups; led by HL7 RIM expert
- Purposes and Anticipated Benefits:
 - To help ensure harmonization among CDISC models (present and future)
 - To provide the industry with a standard model to represent the clinical research domain
 - To enable an HL7 implementation of the CDISC ODM
 - To help harmonize the CDISC and HL7 standards
 - To help enable interoperability between clinical research and healthcare systems

7



SDS = Submission Domain Standards

ADaM = Analysis Data Models

ADaM = Analysis Data Models



How SAS Supports CDISC

“At first glance, SAS might appear to be the loser in any standards shift. After all, it owns the format in which clinical data are analyzed....

... But [SAS] declares: *Standards do not scare us. We embrace them.*

... The company is both acutely sensitive to its large user base and attuned to the shifting winds of industry, where the Clinical Data Interchange Standards Consortium (CDISC) has been gaining traction.”

- Mark Uehling, Senior Science Editor, Bio-IT World | eCliniqua
Thursday, Dec 9, 2004

Support for the CDISC Organization

- SAS is a corporate sponsor
- Ed Helton, of SAS, recently elected Chairman of the Board for CDISC
- Dave Handelsman is SAS' representative on the Industry Advisory Board
- SAS has hosted the Industry Advisory Board meeting in the past, and periodically hosts Board of Directors' meetings
- SAS participates in several of the modelling teams

SAS Software Developed for CDISC

- PROC CDISC
 - ODM (v1.2.1): Data conversion between SAS and XML
 - SDTM (v3.1): Data structure verification
- XML Engine ODM Native mode (v1.2.1)
- XML Engine and XMLMap Extensions
- New base SAS formats/informats for ISO-8601
- SAS CDISC Viewer (ODM 1.2.1)

See Tony Friebe's talk at 16:05 today for more details: SAS Dataset Content Conversion to CDISC Data Standards

PROC CDISC Example

```
libname results 'SASEnvironment\SASCode\CDISC Demo\data' ;  
FILENAME XMLINP '\CDISC Demo\xml\ae.xml';
```

PROC CDISC

```
MODEL=ODM                READ=XMLINP  
formatActive=YES         formatNoReplace=NO;
```

```
ODM                       ODMVersion = "1.2"  
                           ODMminimumKeyset=NO;
```

```
CLINICALDATA              OUT = results.AE  
                           SASDATASETNAME = "AE" ;
```

```
RUN;
```

SAS Integration Based on CDISC ODM

- ODM v1.2.1 used to bridge DataTrak EDC with SAS Drug Development
- Incremental data is transferred from EDC server to SAS Drug Development
 - Scheduled, or
 - Ad hoc
- Data is then:
 - Extracted from XML (ODM) into normalized SAS data sets
 - Compared to master files
 - Updated according to ODM transaction types
- Once loaded, data sets are:
 - Versioned
 - Audit trailed
 - Available for further transformation and analysis

SAS ETL Studio

The screenshot displays the SAS ETL Studio - MP2 application window. The interface includes a menu bar (File, Edit, View, Project, Tools, Process, Window, Help), a toolbar, and a left-hand navigation pane with shortcuts for Source Designer, Metadata Importer, Metadata Exporter, Process Designer, and Target Designer. The main workspace is divided into two panes. The left pane shows a hierarchical tree of Repositories, including Foundation, CDISC Data Standards, Clinical Domain Models, Misc, and Projects. The right pane, titled 'Process Designer : Load Nicsah Study 001 Adverse Eve...', displays a data flow diagram. The process starts with an 'ADVERSE' source, followed by a 'SAS Extract' step, then a 'Sort AEs' step, a 'Calculate Sequence Number' step, a 'CDISC Transform' step, a 'Loader' step, and finally a 'Nicsah Study001 Adverse Events - AE' target. The bottom of the window shows a status bar with 'SASMain', 'mp-2\sasadm as Unrestricted', and 'mp-2 : 8561'.

ETL Studio

- Source Designer
- Target Designer
- Process Designer

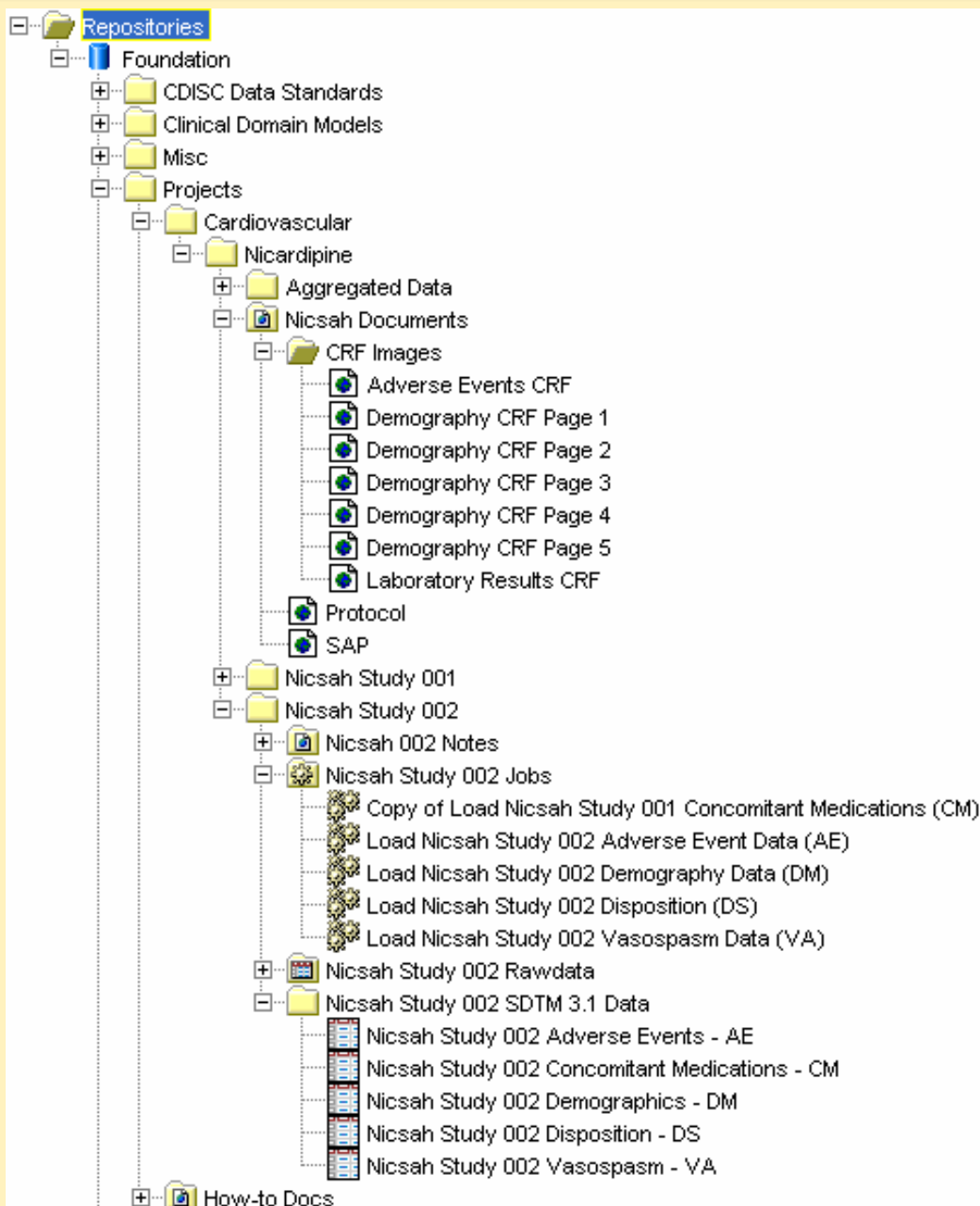
ETL Studio Desktop

■ Customizable Trees

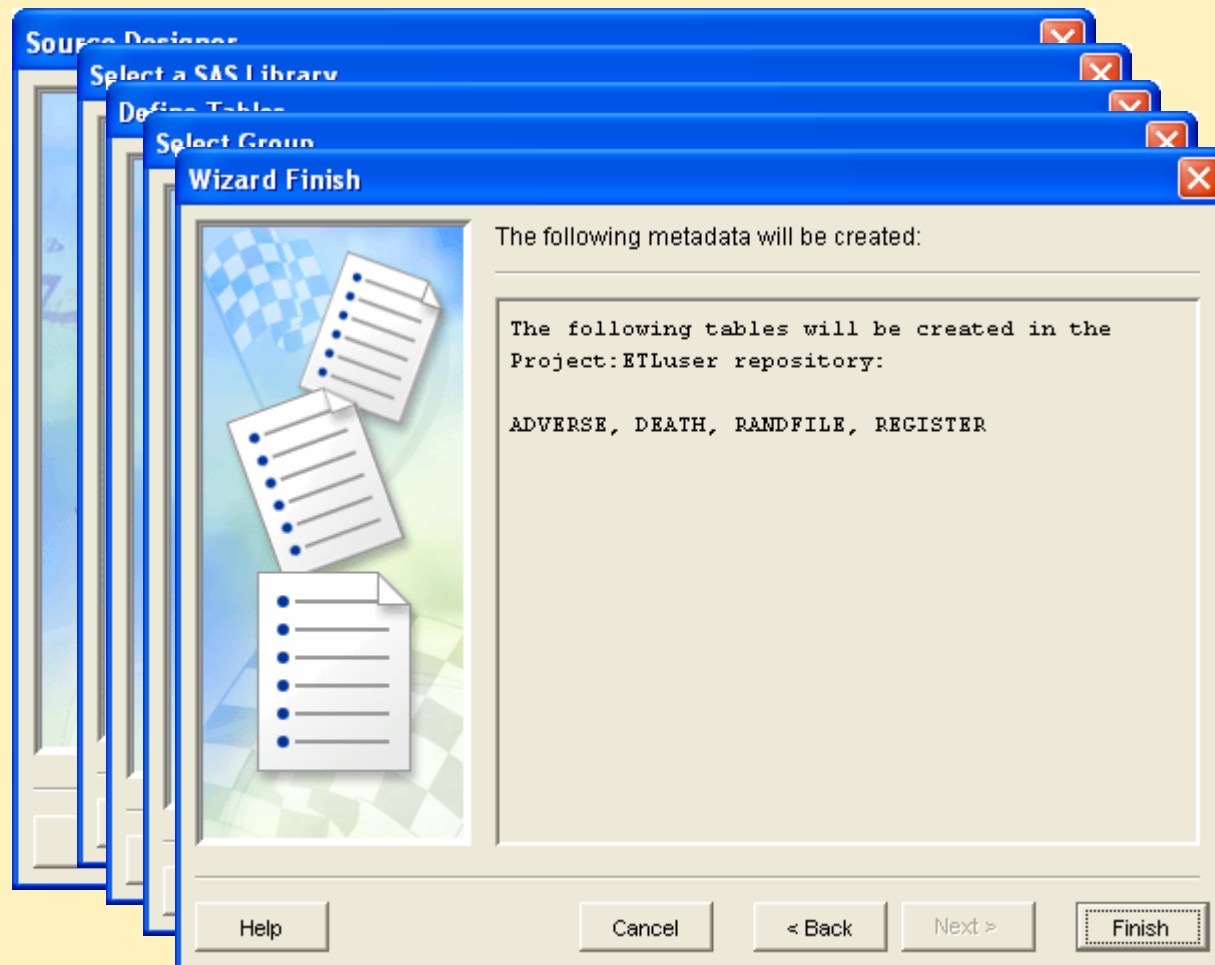
■ Many Object types

- Document
- Source Table
- Target Table
- Job
- External Files
- Groups
- Libraries
- Notes

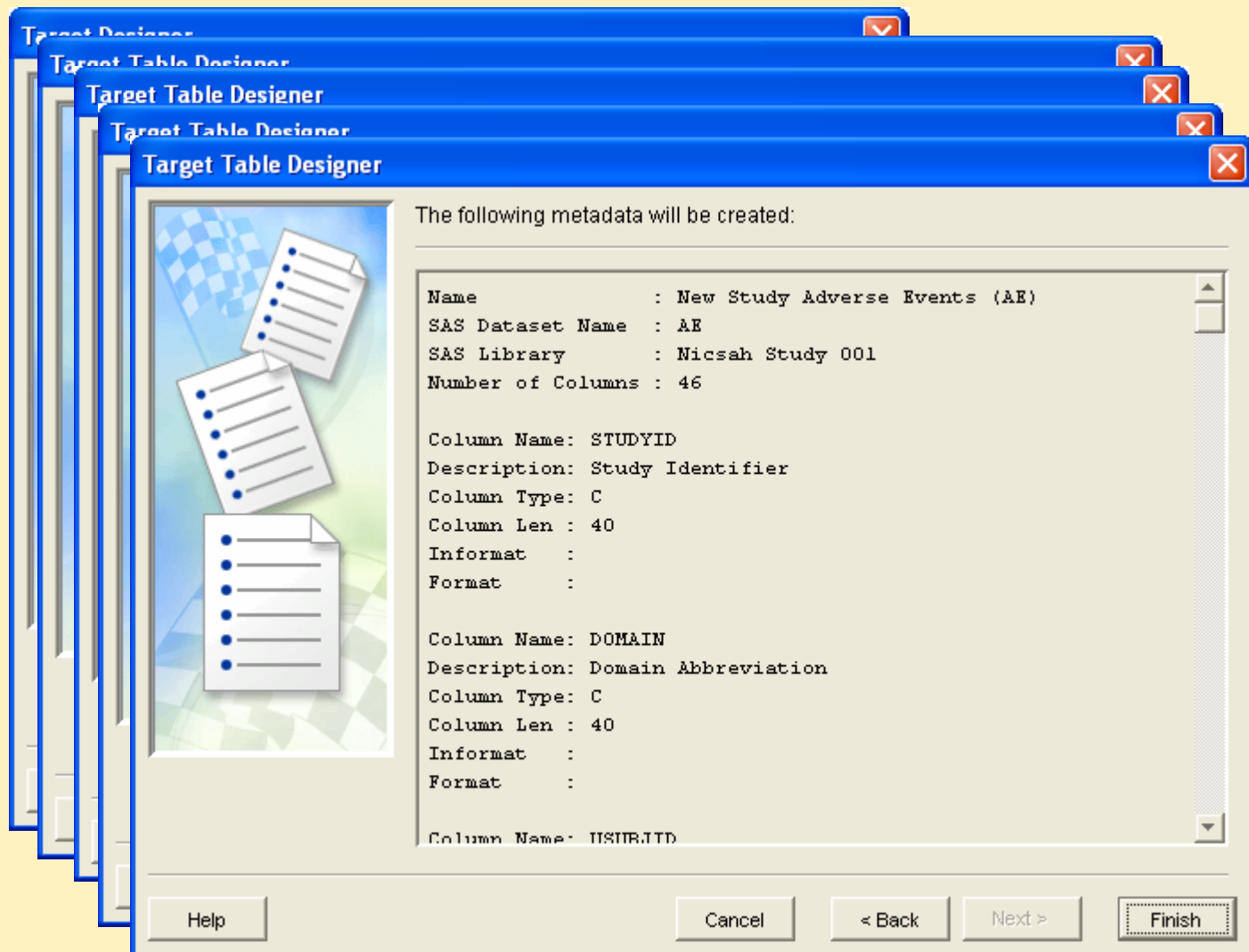
etc



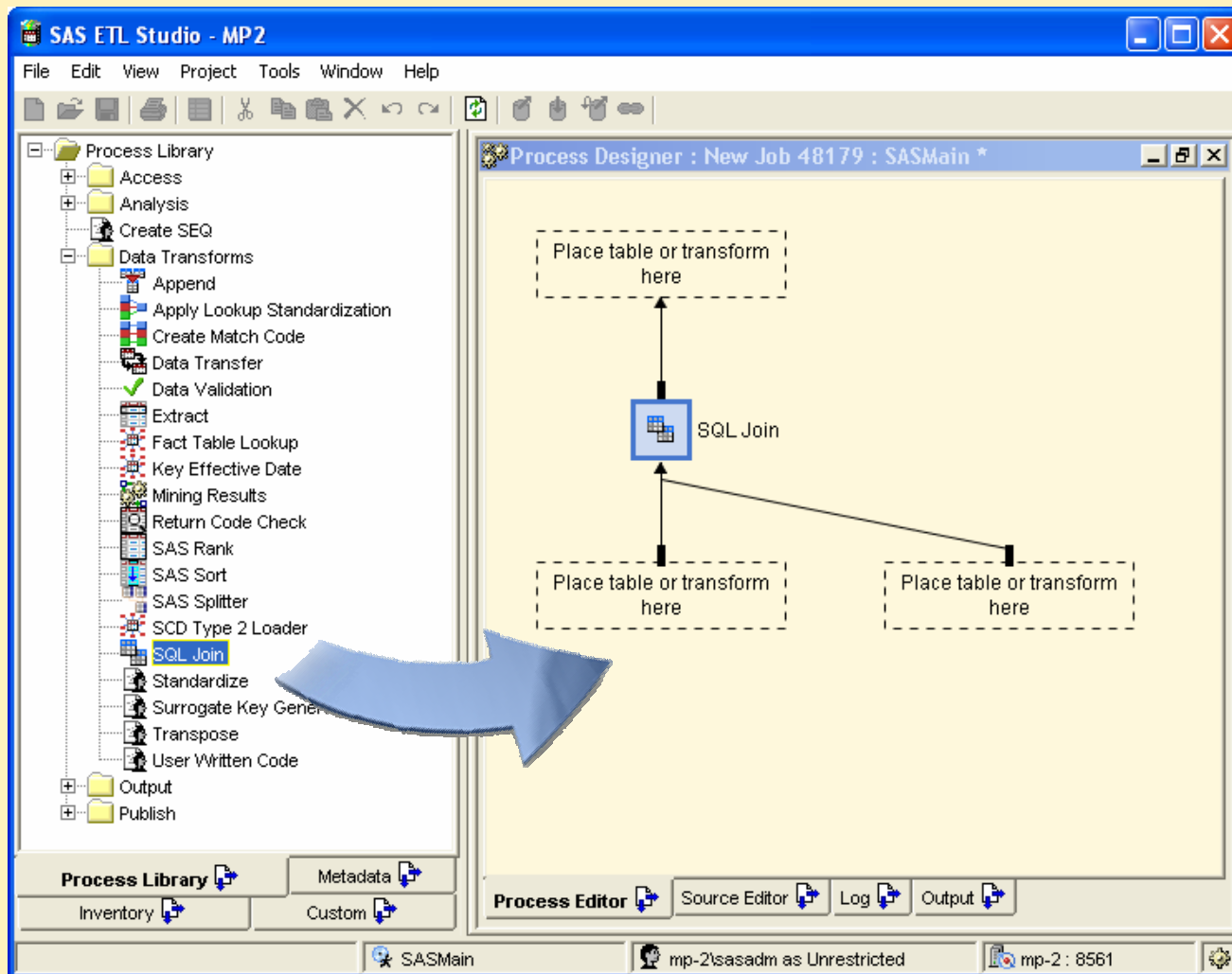
Source Designer



Target Designer



Process Designer



Process Designer

SDTM Transformations Properties

General | Where | Process | Mapping | Options | Notes | Extended Attributes | Advanced

Source table: Demo Target table: W5342FLE

#	Column	Column Desc
1	STUDYID	Study ID
2	SITEID	Site ID
3	INVID	Investigator ID
4	INVNAME	Investigator Name
5	USUBJID	Unique Subject ID
6	SUBJID	Subject ID
7	AGE	Age
8	SEX	Sex
9	RACE	Race
10	TRTCD	Treatment Code
11	TRTGRP	Treatment Group
12	COUNTRY	Country
13	VISITNUM	Visit Number
14	VISITDY	Visit Day
15	VISIT	Visit Name
16	DMACTDY	Actual Study Day
17	DMDT	Date of Visit
18	DMREFDT	Subject Reference
19	DMREFTM	Subject Reference
20	BIRTHDT	Date of Birth
21	WEIGHT	Weight in kg
22	HEIGHT	Height in cm
23	SUBJINIT	Subject Initials

Mapping connections (Source to Target):

- STUDYID (1) to STUDYID (1)
- SITEID (2) to DOMAIN (2)
- INVID (3) to USUBJID (3)
- INVNAME (4) to SUBJID (4)
- USUBJID (5) to RFSTDTC (5)
- SUBJID (6) to RFENDTC (6)
- AGE (7) to SITEID (7)
- SEX (8) to INVID (8)
- RACE (9) to INVNAM (9)
- TRTCD (10) to BRTHDTC (10)
- TRTGRP (11) to AGE (11)
- COUNTRY (12) to AGEU (12)
- VISITNUM (13) to SEX (13)
- VISITDY (14) to RACE (14)
- VISIT (15) to ETHNIC (15)
- DMACTDY (16) to ARMCD (16)
- DMDT (17) to ARM (17)
- DMREFDT (18) to COUNTRY (18)
- DMREFTM (19) to DMDTC (19)
- BIRTHDT (20) to DMDY (20)
- WEIGHT (21) to SUBJINIT (21)
- HEIGHT (22) to SUBJINIT (21)
- SUBJINIT (23) to SUBJINIT (21)

Buttons: OK, Cancel, Apply, Help

Expression Builder

Expression Text:

```
PUT ( DHMS ( DMREFDT , 0 , 0 , DMREFTM ) , IS8601DT. )
```

Operators: +, -, *, /, **, AND, OR, NOT, =, ^=, <, <=, >, >=, =*, ||, ' ', (,)

Functions | Data Sources

Categories:

- All Functions
- Aggregate
- Arithmetic
- Bitwise Logical Operations
- Character
- Character String Matching
- DBCS
- Data Quality
- Date and Time
- Descriptive Statistics
- External File
- Financial

Functions:

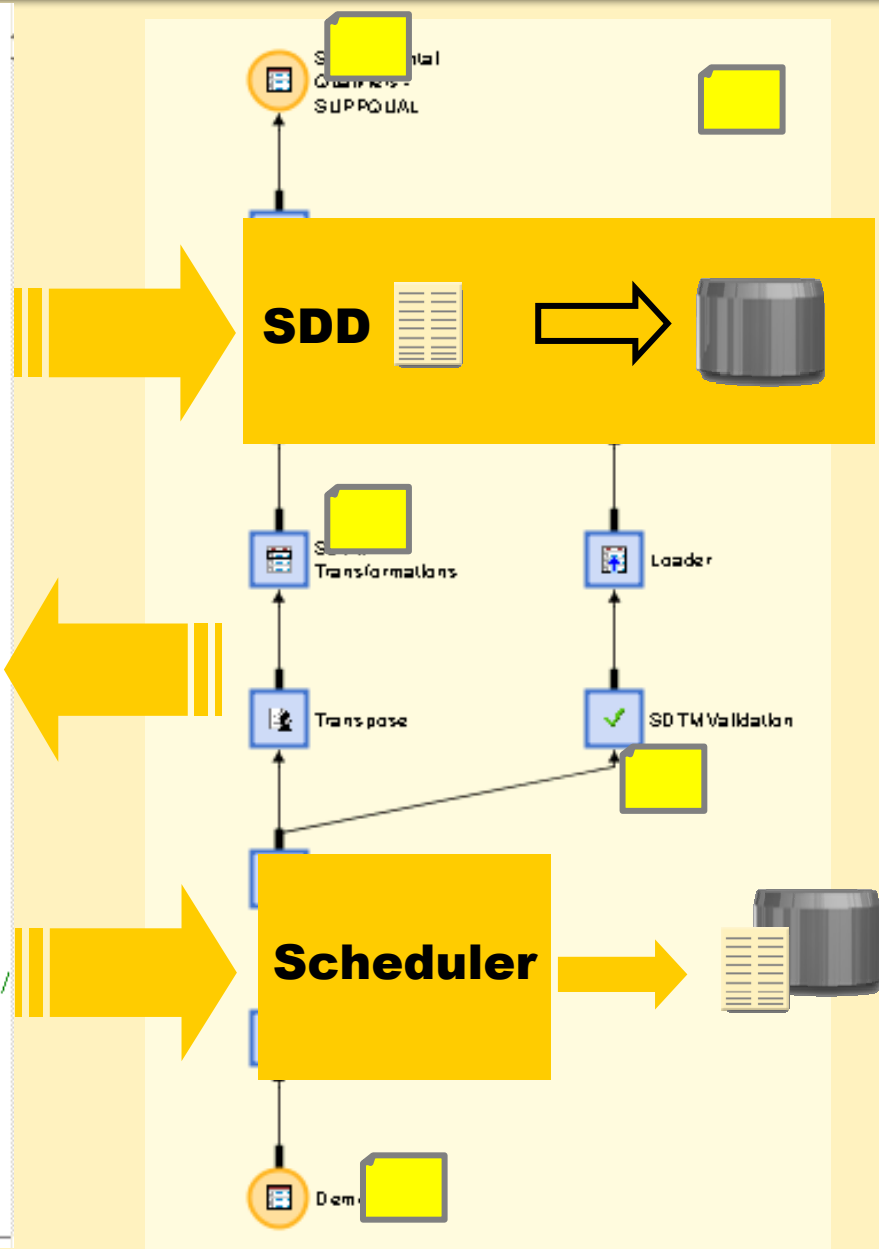
Insert

Buttons: OK, Cancel, Help

Supplemental
Qualifiers -
SUPQUAL

Demo

```
*****  
* Job:          Load Demography          A5853KYA.AN000002 *  
* Description:                                     *  
*                                     *  
* Repository:   Sandbox                   A0000001.A5853KYA *  
* Server:       SASMain                   A500IHYN.AR000001 *  
*                                     *  
* Source Table: Demo - s_pr.Demo          A5853KYA.AG00000RV *  
* Target Tables: Demographics - DM - std_pr.DM A5853KYA.AG00000RT *  
*               Supplemental Qualifiers - SUPPQUAL - A5853KYA.AG00000RZ *  
*               std_pr.SUPPQUAL           *  
*                                     *  
* Generated on: Tue Mar 08 13:38:05 EST 2005 *  
* Generated by: vmxp-l\sasadm              *  
* Version:     9.1.20041201.4014          *  
*****  
  
*****  
  
*****  
  
/* Setup to capture return codes */  
%let job_rc = 0;  
%let trans_rc = 0;  
%let sqlrc = 0;  
%global syserr;
```



CDISC SDTM Data Model

- Implemented in the metadata
- Implements all currently defined models
- Contains table definitions and model metadata
- Facilitates a workflow to define new models
- Provides supplemental tables for validation steps

Demographics - DM Properties

General Columns Indexes Keys Physical Storage Notes **Advanced**

#	Name	Description
1	STUDYID	
2	DOMAIN	Domain Abbreviation
3	USUBJID	Unique Subject Identifier
4	SUBJID	Subject Identifier for the Study
5	RFSTDTC	Subject Reference Start Date/Time
6	RFENDTC	Subject Reference End Date/Time
7	SITEID	Study Site Identifier
8	INVID	Investigator Identifier
9	INYNAM	Investigator Name
10	BRTHDTC	Date/Time of Birth
11	AGE	Age in AGEU at Reference Date/Time
12	AGEU	Age Units
13	SEX	Sex
14	RACE	Race
15	ETHNIC	Ethnicity
16	ARMCD	Planned Arm Code
17	ARM	Description of Planned Arm
18	COUNTRY	Country
19	DMDTC	Date/Time of Collection
20	DMDY	Study Day of Collection

New Import... Delete

OK Cancel Apply Help

Extended Attributes for STUDYID

#	Field Name	Value	Description
1	TERM		Controlled Term or ...
2	ORIGIN	CRF	Origin
3	ROLE	Identifier	Role
4	CORE	Req	Core
5	References	SDTM 2.2.4	References

New Delete

OK Cancel Help

Reference Tables (controlled terminology)

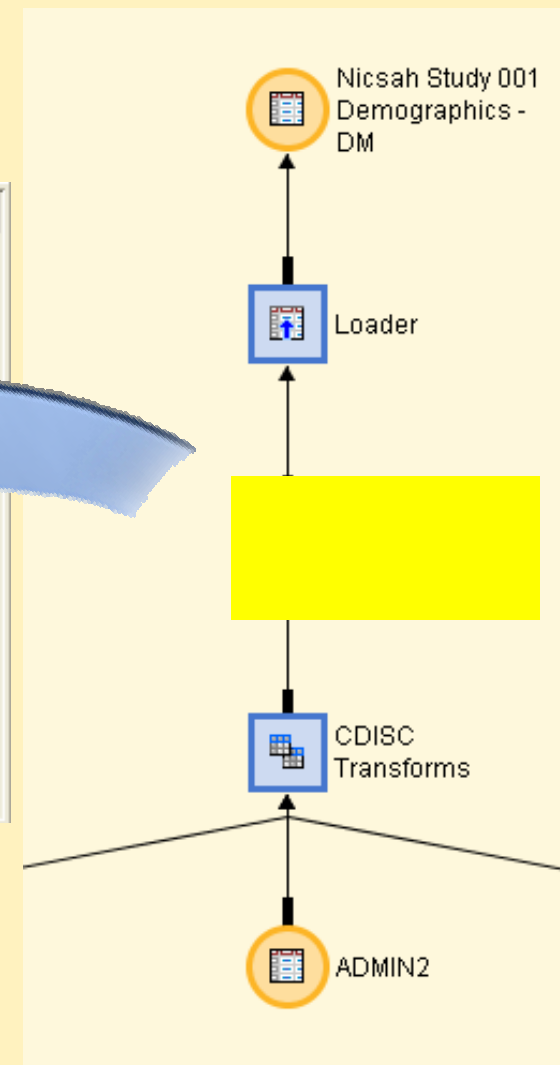
The screenshot shows the SAS Clinical Domain Models interface. On the left, a tree view lists the following categories:

- Clinical Domain Models
 - Custom Domains
 - Reference Tables
 - AGEU
 - DOMAINS
 - EGTESTCD
 - ETHNIC
 - KEYWORDS
 - QNAM
 - SEX
 - VSTESTCD
 - YN
 - SDTM 3.1 Domains
 - Templates

Two arrows point from the 'Reference Tables' folder to the 'View Data: EGTESTCD (rows 1-10)' window. This window displays a table with the following data:

#	ECG TEST Codes	ECG Test
1	FINDING	General ECG Finding
2	HR	Heart Rate
3	INTP	ECG Interpretation
4	PR	PR Interval
5	QRS	QRS Interval
6	QT	QT Interval
7	QTCB	QTc Interval Bazett
8	QTCF	QTc Interval Fridericia
9	RR	RR Interval
10	VRATE	Ventricular Rate

A large blue arrow points from the 'EGTESTCD' table in the left pane to the 'View Data' window. A yellow box is positioned below the table in the window.



SAS and SDTM as the Emerging Standard

**OCTAGON RESEARCH SOLUTIONS IDENTIFIED AS A
PRIMARY TRAINER FOR CLINICAL DATA
INTERCHANGE STANDARDS CONSORTIUM (CDISC)**

**Corporate Sponsor to Provide Training Curriculum on
Emerging Data Standards**

April 4, 2005

**OCTAGON RESEARCH SOLUTIONS IMPLEMENTS SAS
TECHNOLOGY [SAS® Enterprise ETL Server] TO
REVOLUTIONIZE CDISC SDTM LEGACY CONVERSION
OFFERINGS**

April 11, 2005

On the SAS Radar....

- define.xml
- Lab, SEND, ADaM
- Other emerging standards

The Move to Standards...

“Faced with rapid changes, the nation’s healthcare system has **fallen short of its ability to translate information into knowledge** that can be used in practice, and to apply new technology safely and appropriately.

The results are exactly what you would expect. Everyone who uses the current system **constantly confronts large information gaps**, whether it’s at the doctor’s office, on the hospital ward or at government agencies charged with protecting the public health. That goes for the FDA -- we’re no exception.”

*Dr. Mark McClellan, former FDA Commissioner
CDISC Interchange, October 2003*

“Innovation depends upon standardization.”

Dr. Bob O’Neill, Director, Office of Biostatistics, CDER, FDA

Benefits of Standards – Regulatory View

“The importance of a standard for the exchange of clinical trial data cannot be overstated. FDA reviewers spend far too much valuable time simply reorganizing large amounts of data submitted in varying formats. Having the data presented in a standard structure will improve FDA’s ability to evaluate the data and help speed new discoveries to the public.” -*Lester Crawford, Former Commissioner, FDA*

“... But [SAS] declares: *Standards do not scare us. We embrace them.*”

- Mark Uehling, Senior Science Editor, Bio-IT World | eCliniqua
Thursday, Dec 9, 2004



The Power to Know®