

# The CDISC-HL7 Standard

## *An FDA Perspective*

**Armando Oliva, M.D.**  
*Deputy Director for Bioinformatics  
Office of Critical Path Programs  
Office of the Commissioner*

*March 11, 2009  
Philadelphia, PA*



# DISCLAIMER

The views expressed in this presentation are those of the speaker and not necessarily those of the Food and Drug Administration ...

# Outline

- Data standards in clinical research
- What does CDISC offer?
- What does HL7 offer?
- How the CDISC-HL7 project will benefit regulatory review of clinical research
- Implementation Issues

# Take Home Messages

- **“The World is Round”**
  - Clinical data are not flat and cannot be exchanged using flat two-dimensional files without significant loss of meaning
- **FDA is transitioning to a “round view of the world” of clinical research**
  - CDISC-HL7 standard will get us there
- **SDTM is here to stay**
  - Will transition from a standard submission format to an standard view of the data in support of simple analyses (e.g. distribution, means, etc.)

# Data Exchange

## Definition:

- Data exchange is the process of sending and receiving data in such a manner that the information content or ***meaning*** assigned to the data is not altered during the transmission.

# The Problem

- Most clinical trials ...
  - don't employ a standard for data exchange
  - don't use standardized analytic tools or techniques
- Result:
  - Analyzing clinical trial data efficiently and systematically is difficult and time consuming, especially across many trials
  - e.g. How many women participate in clinical trials?

### Study #2 – dmj.xpt

ID	GENDER
A1	Male
A2	Male
A3	Female
A4	Female
A5	Male

### Study #1 – demog.xpt

SUBJID	SEX
0001	M
0002	F
0003	F
0004	M
0005	F

### Study #3 – axd222.xpt

USUBID	SEX
00011	0
00012	1
00013	1
00014	0
00015	1

### Study #4 – dmjph.xpt

PTID	GENDER
0001	1
0002	1
0003	2
0004	2
0005	1

# How do you automate counting women?

- No standard file names
  - How does a computer find the file that contains the data?
- No standard variable names
  - How does a computer find the column that contains sex information?
- No standard terminology
  - How does a computer know which code represents which sex?

# NET RESULT: Analysis of study data is . . .



Difficult



Time Consuming



Expensive

# CDISC

**Common Currency**

- Clinical Data Interchange Standards Consortium
- Non-profit standards development organization
- Standards for clinical trial data for submission (SDTM), for collection (CDASH), derived data for analysis (ADaM), and non-clinical data for submission (SEND)
- 
- CDER adopted CDISC SDTM 7/04, now piloting SEND
- CBER evaluating SDTM in a pilot setting
- CVM intends to pilot SEND for animal data

# Example: Women in Clinical Trials

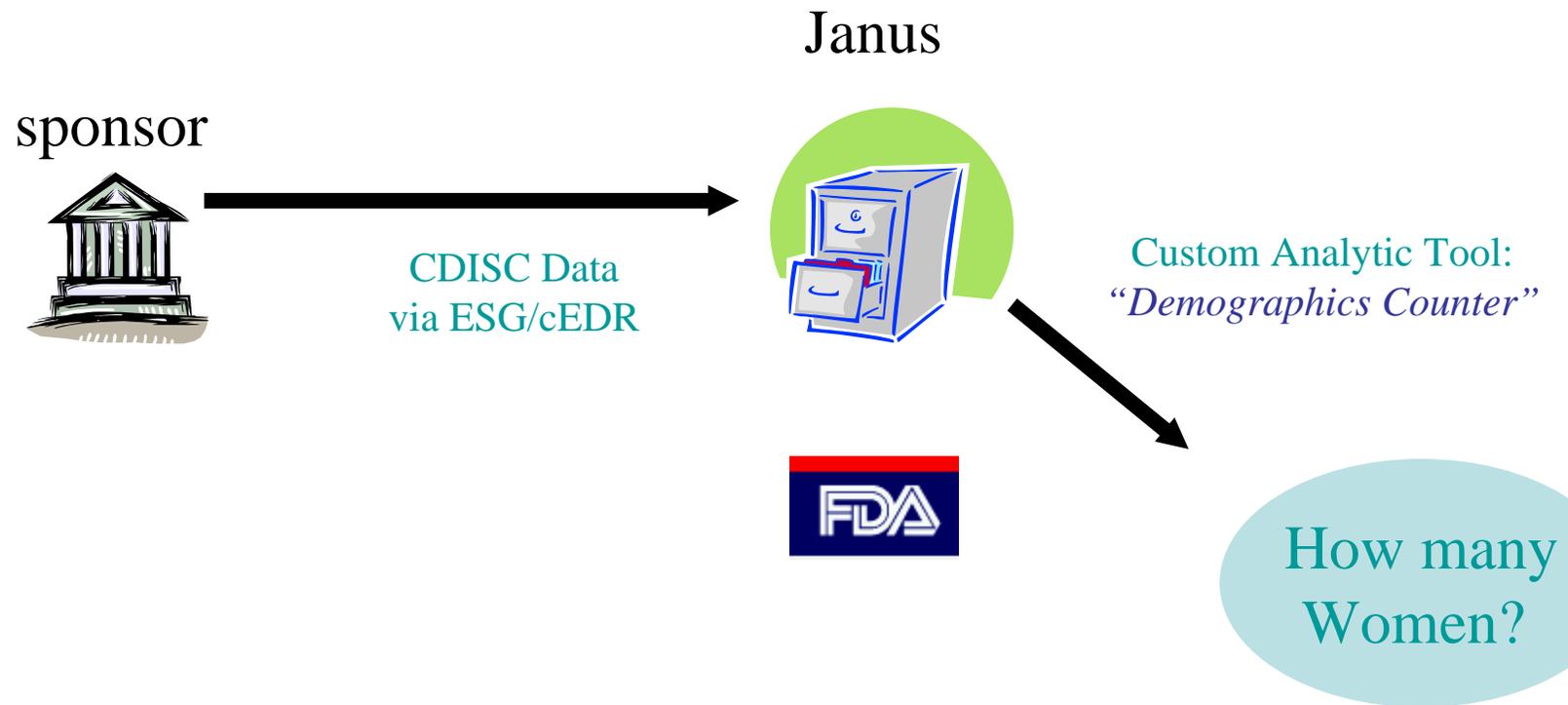
- Demographics File: DM.xpt
- Variable Name: SEX
- Acceptable Codes: M, F, U

## Better Tracking of Women In Clinical Trials

- Applicants submit standardized (CDISC) clinical trial data and are loaded into Janus\*
- New analysis tools (e.g. ‘demographics counter’) can quickly and easily extract demographic information from all trials in the repository and analyze the data.

\*Janus: FDA’s enterprise data warehouse for study data

# Standardized Data Flow:



# Sample CDISC “DM” Dataset

	STUDYID	DOMAIN	USUBJID	AGE	AGEU	SEX	RACE	ARM
1	XY123	DM	XY123-0001-2293	69	YEARS	M	WHITE	Cureall 10 mg
2	XY123	DM	XY123-0001-2294	67	YEARS	F	WHITE	Cureall 5 mg
3	XY123	DM	XY123-0001-2295	70	YEARS	F	WHITE	Cureall 10 mg
4	XY123	DM	XY123-0001-2296	62	YEARS	F	WHITE	Cureall 10 mg
5	XY123	DM	XY123-0001-2297	66	YEARS	F	WHITE	Placebo
6	XY123	DM	XY123-0001-2298	80	YEARS	M	WHITE	Cureall 10 mg
7	XY123	DM	XY123-0001-2473	66	YEARS	M	WHITE	Cureall 5 mg
8	XY123	DM	XY123-0001-379	69	YEARS	M	WHITE	Placebo
9	XY123	DM	XY123-0001-380	63	YEARS	M	WHITE	Cureall 10 mg
10	XY123	DM	XY123-0001-381	70	YEARS	M	WHITE	Placebo
11	XY123	DM	XY123-0001-382	60	YEARS	F	WHITE	Cureall 5 mg
12	XY123	DM	XY123-0001-383	71	YEARS	F	WHITE	Placebo
13	XY123	DM	XY123-0001-384	68	YEARS	M	WHITE	Cureall 5 mg
14	XY123	DM	XY123-0001-895	67	YEARS	F	WHITE	Cureall 5 mg
15	XY123	DM	XY123-0001-896	63	YEARS	M	WHITE	Cureall 5 mg
16	XY123	DM	XY123-0001-897	79	YEARS	M	WHITE	Cureall 10 mg
17	XY123	DM	XY123-0001-898	69	YEARS	F	WHITE	Placebo
18	XY123	DM	XY123-0001-899	70	YEARS	F	WHITE	Cureall 5 mg
19	XY123	DM	XY123-0001-900	66	YEARS	F	WHITE	Cureall 10 mg

# Standard Report on “DM”

SEX	AGE					RACE				
	N	% of Total	Mean	Min	Max	ASIAN	BLACK	HISPANIC	OTHER	WHITE
F	4050	47.29	67.3834568	59	80	737	40	186	76	3011
M	4506	52.71	67.4962273	59	80	833	42	237	67	3327

Created using JMP

## **With the right bioinformatics infrastructure (data standards, data warehouse, analysis tools):**

- Monitoring women in clinical trials is easy!
- Assessing sex differences is almost as easy

# **CDISC is a tremendous advance!!**

- Big improvement from the past
- Easy to find the data
- Easy to understand the data
- Easy to analyze the data

# But...

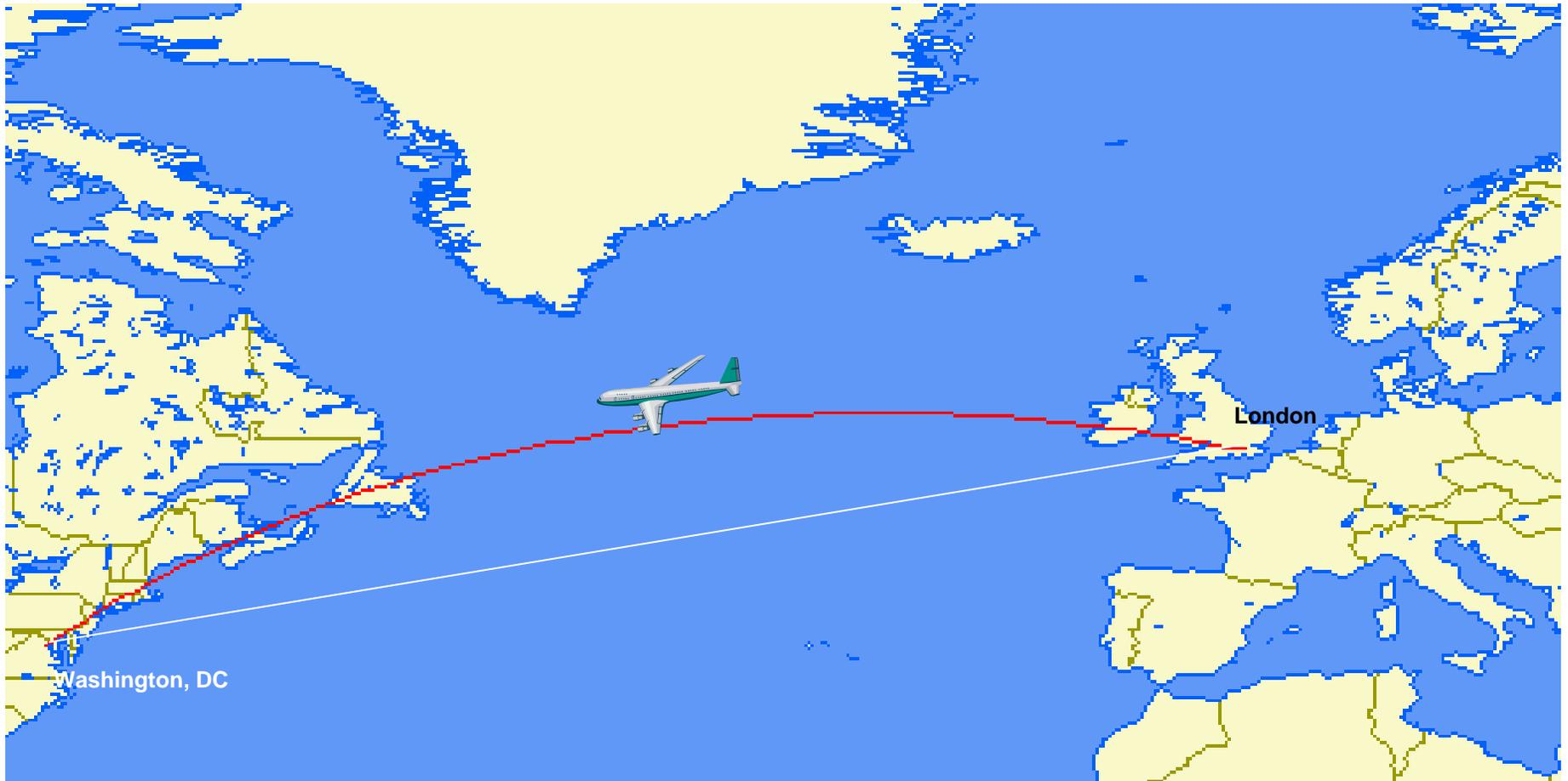
- We've learned some lessons along the way...
- “Flat” 2-dimensional files are not the best way to exchange clinical data
- Some **meaning** is lost when exchanging flat files, making certain analyses difficult or impossible



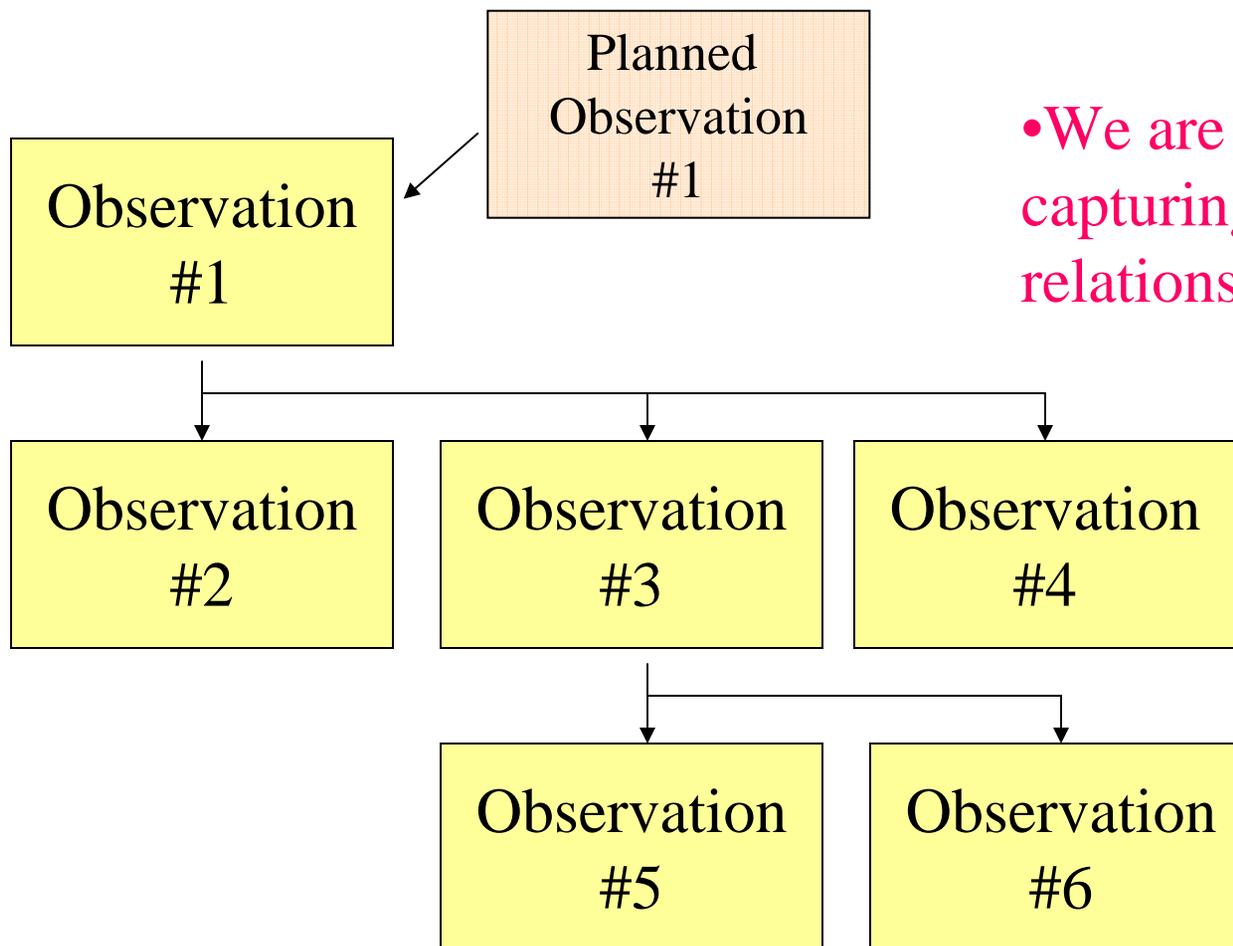
*The World is Round!*

**USA  
ROAD MAP**



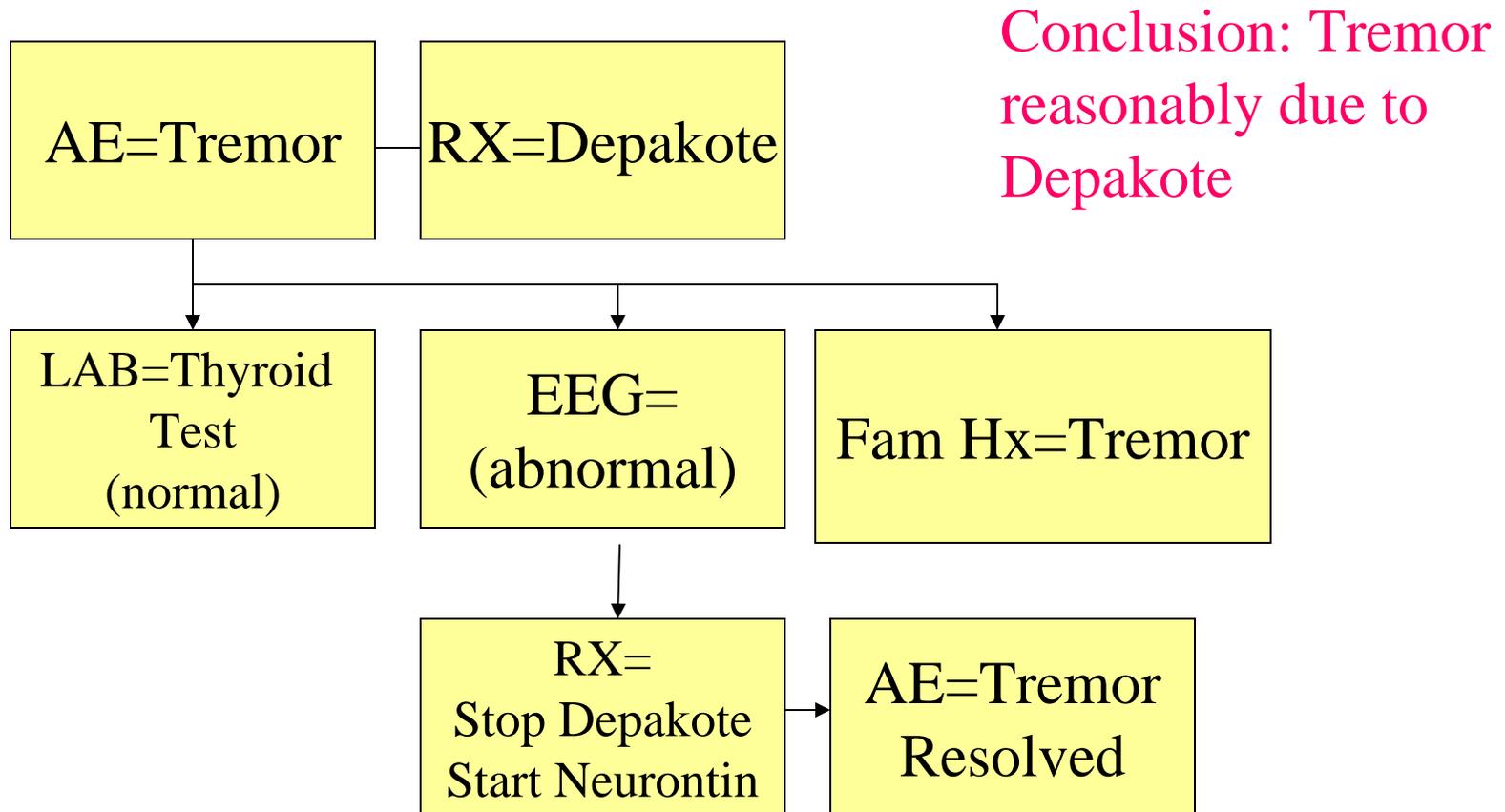


## Clinical Observations: Highly Relational and Hierarchical



•We are currently not capturing these relationships well.

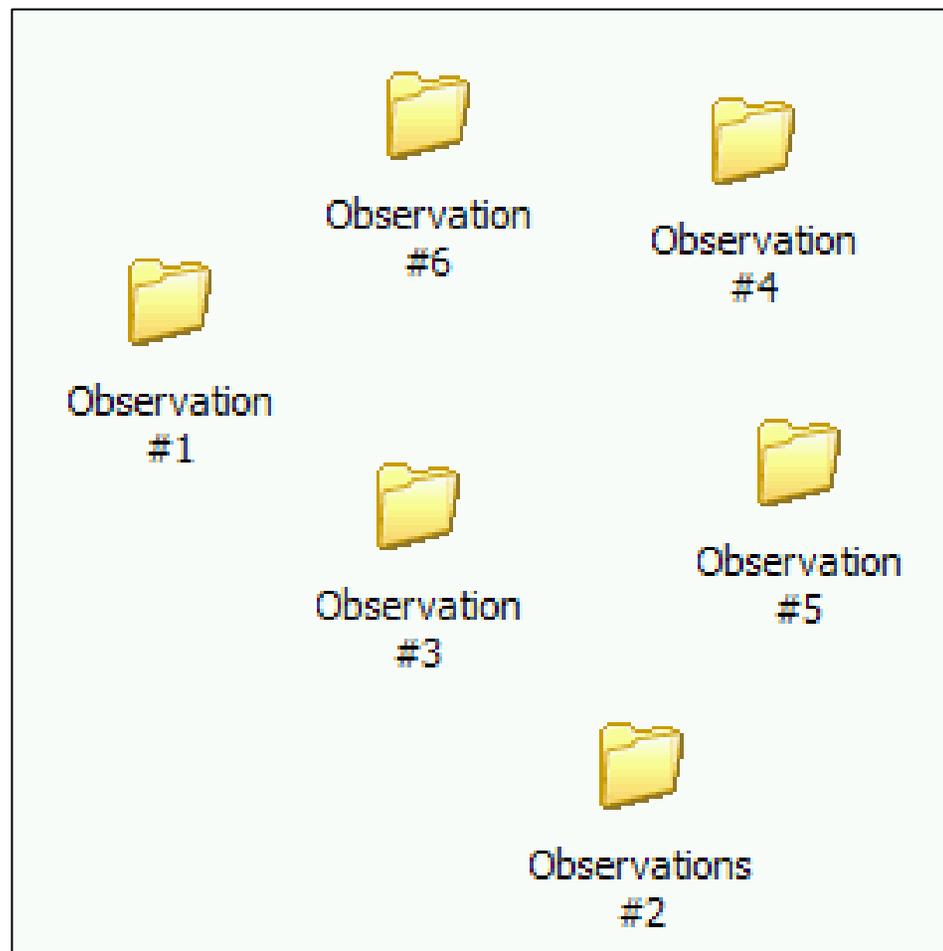
# Example 1 (from my clinic)



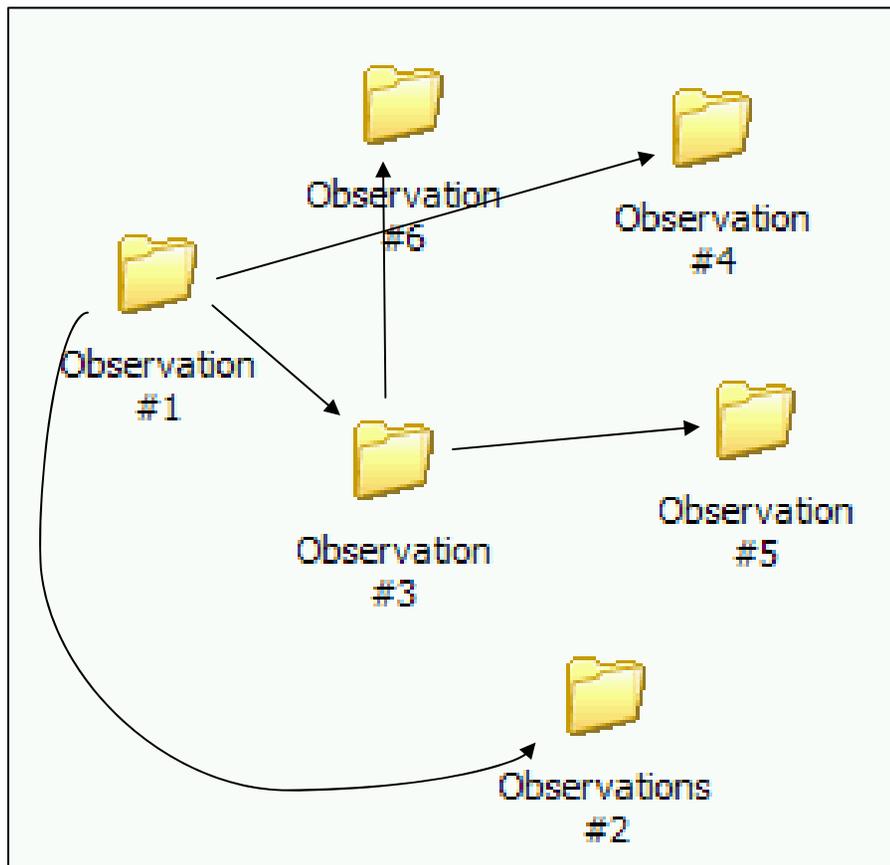
# Think of Clinical Observations as nested folders in a tree structure



**Flat Files Don't  
Inherently Capture  
the Tree Structure,  
which is itself  
important to  
understand the data**



## One approach: capture the relationships separately and add them to the flat file structure



Better approach: data model that inherently captures relationships at the point of collection and can transmit them.

## Example 2

- Study drug: known association with thrombocytopenia in nonclinical studies
- Planned clinical assessment: complete blood count every 3 months

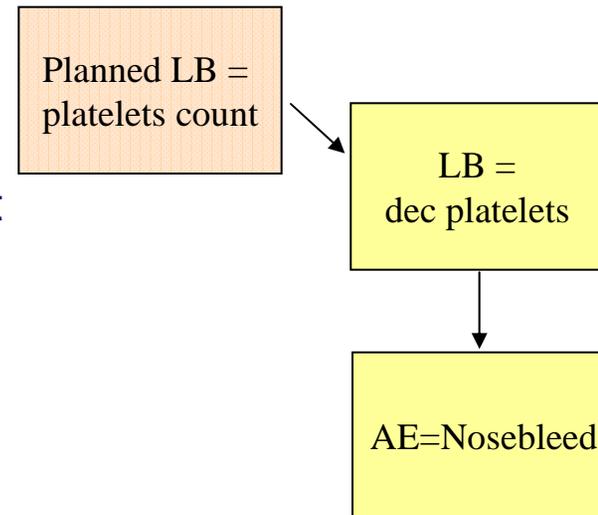
Consider two clinical observations, both observed the same day – reported in flat files as:

- AE: BLEEDING
- LB: PLATELETS = 35K (LOW)

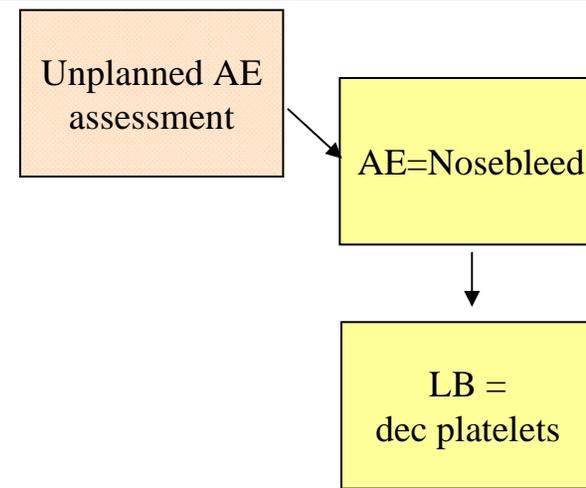
These observations can represent two clinical scenarios.

# Two Scenarios

- #1 - Subject arrives for a planned study visit and laboratory test: complete blood count (CBC). Results show low platelets 35K. Investigator asks about abnormal bleeding episodes. Patient recalls a mild nosebleed earlier that morning.
  - → Hypothesis: Monitoring schedule is adequate (?)



- #2 – Subject calls investigator to report a mild nosebleed. Investigator advises subject to undergo unscheduled complete blood count. CBC shows low platelet count of 35K.
- → Hypothesis: Monitoring schedule is inadequate; more frequent CBC monitoring is needed (?)



# Review Questions?

- How many unplanned CBCs occurred?
- What observations led to unplanned CBCs?
- How often is bleeding associated with unplanned CBC?
- How often is bleeding associated with planned CBC?
- Is monitoring frequency adequate... ?
  - ...For the trial (IND)?
  - ...For labeling (NDA/BLA)?
- Current flat file exchange format makes these analyses difficult.
- A more robust data model will facilitate answering these questions.

# HL7

- Health Level 7 is the world's leading standards development organization for healthcare information exchange.
- ANSI accredited
- Cooperation agreements with ISO and CEN (the European Standards Body)
- HL7 Version 2 is widely implemented in over 20 countries
- HL7 standards are called “messages” because they support healthcare information exchange between systems or organizations

# What is the HL7 RIM?

- HL7 Version 3 is based on the “Reference Information Model” and has been developed for complex health IT systems
- Over 10 years in development
- The fundamental information model from which all HL7 messages are based
- HL7 RIM (V3) has significant international implementation:
  - United Kingdom (National Health Service - NHS)
  - Netherlands
  - Canada
  - Mexico
  - Germany
  - Croatia

# What does HL7 V3 offer?

- More “multi-dimensional” or “multi-relational” representation of the data
- cost-effective method to exchange healthcare information; less manual processing
- In the healthcare domain:
  - Increase patient safety; fewer clinical errors
- Interoperability among loosely coupled systems
- Ability to collect increasing amounts of coded data in EHRs

# CDISC-HL7 Standard



- ***What*** is it?
- ***Why*** do we need a CDISC-HL7 exchange standard?
- ***How*** is the standard being developed?
- ***When*** will the standard be available?
- ...will FDA implement it?

## *What Is It?*

- A project in HL7 to develop HL7 XML exchange standard (“messages”) for CDISC structured study data.
  - Sponsored by FDA and CDISC
- Four messages:
  - Study Design (What will be done?)
  - Study Participation (Who is involved?)
  - Subject Data (What was observed?)
  - HL7 ICSR (expedited AE reporting)

# *Why* **CDISC-HL7?**

- 1. Improved Data Management**
  - Facilitate more analyses
- 2. Harmonize with other HL7 standards for regulated medical product information**
- 3. Better long-term integration with EHRs as they start being used for both Clinical Research and Surveillance**

# *Driver #1 –* Improved Data Management

- Move away from SAS Transport towards XML
- Current exchange standard for CDISC content is SAS Transport file, version 5
- This format has limitations
  - 8 character variable names
  - Flat files → don't inherently capture all of the relationships between study data, or between study design and study data as desired by FDA
- FDA intends to move away from SAS transport file to an HL7 XML-based exchange format for loading into Janus data warehouse
- HL7 is the preferred electronic exchange format for healthcare information
- CDISC is the preferred standard for clinical research data
- CDISC Content in HL7 XML exchange format is ideal

# FDA and CDISC

- CDISC is the preferred standard for clinical research data
- CDER has been accepting SDTM datasets since July 2004
- CBER is evaluating SDTM in a pilot setting
- CVM intends to pilot CDISC for animal data
- CDER Reviewers find SDTM very helpful
  - Successful Phase 1 SDTM pilot (single study)
  - Phase 2 SDTM Pilot underway (cross-study comparisons)
- Phase 2 SEND Pilot underway (for nonclinical data)
- SDTM is here to stay: very useful format
  - Standardized views of the data to support standard analytic strategies, and algorithms with **existing** and **future** review tools (e.g. i/jReview, WebSDM, Patient Profile Viewer)
  - FDA sees SDTM evolving from a submission format to a standard analysis format that Janus will produce
  - Opportunity for SDTM to grow is enhanced once CDISC-HL7 is available

# FDA and HL7

- HL7 XML is the preferred exchange format for health care information
- FDA is committed to harmonize all exchange standards for regulated product structured data with the HL7 Reference Information Model
  - Ensures unambiguous exchange of healthcare information (including clinical research information) between computer systems to promote interoperability

## *Driver #1 – cont'd*

# Improved Data Management

- SAS transport file's flat file structure doesn't inherently capture the complex relationships between clinical observations and between planned and observed observations
  - Adding them *post-facto* (e.g. RELREC in SDTM) is invariably incomplete done inconsistently / time-consuming / inefficient
  - We need a data model that is capable of capturing and conveying these relationships consistently and at the point of data collection
  - HL7 RIM provides a **more robust data model** for clinical observations compared to 'flat' SAS transport files
    - e.g. HL7 ICSR

## *Driver #2*

### Harmonize with Other HL7 Standards

- Harmonization with HL7 ICSR provides a single data model for all pre- and post-marketing observations

# CDISC-HL7 and Janus

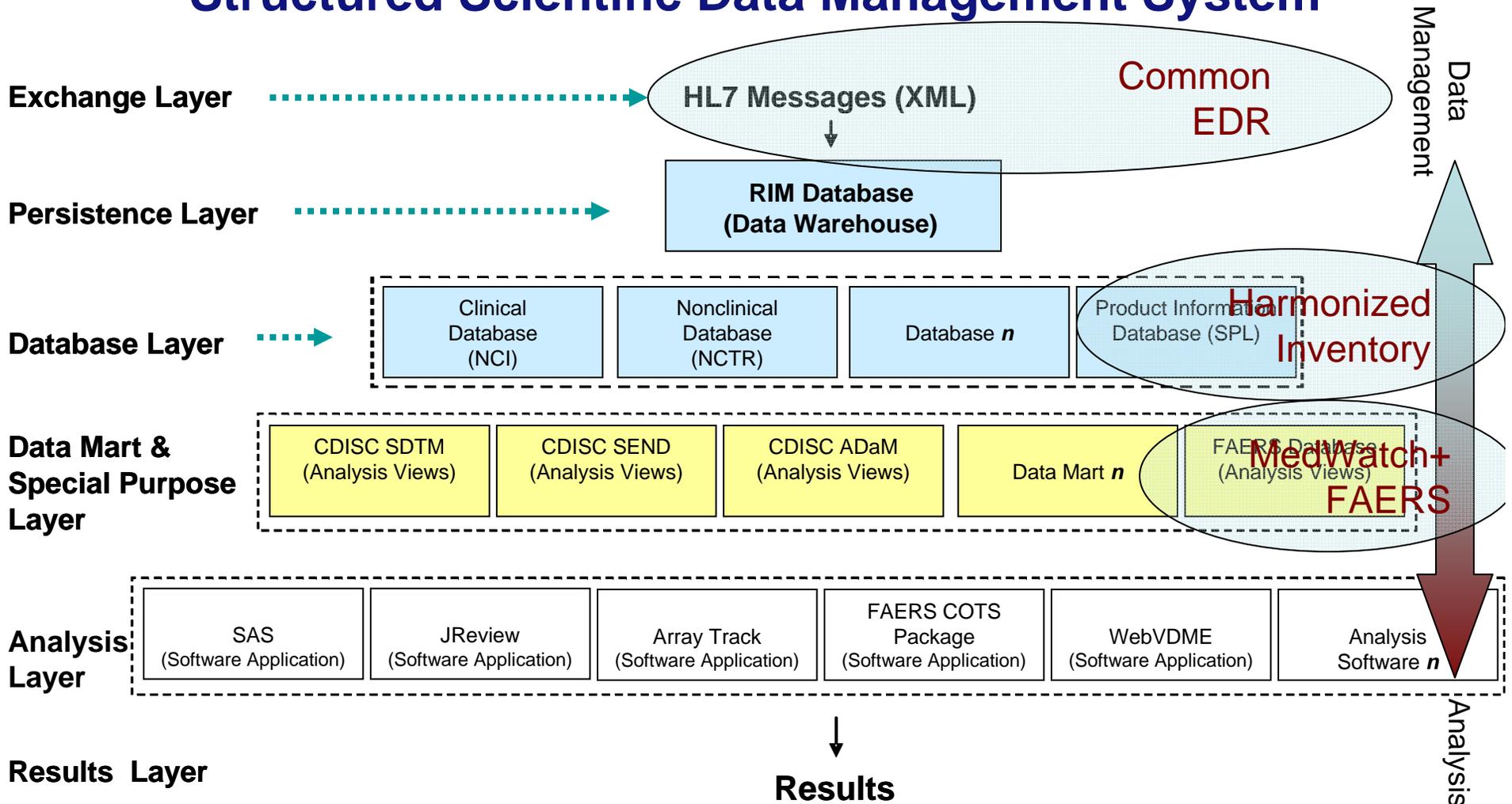
- Janus – FDA’s enterprise logical data model for structured scientific data and physical data warehouses
- CDISC-HL7 XML will facilitate loading study data into the Janus study data warehouse
- HL7 ICSR will be used to load post-marketing observational data into enterprise repositories
- Data checker/loader can be leveraged to work with all HL7 exchange standards
- Harmonization of CDISC-HL7 (pre-market) and HL7 ICSR (post-market) will allow integration of pre- and post-market databases and facilitate analyses across a product’s entire lifecycle.
- → **Better safety assessments**

## Example: Pre- and Post-Market AEs

- AE signal identified post-marketing
- **Question:** How does the AE post-marketing experience compare with clinical trials experience?
  - Current data standards/systems make this analysis extremely difficult/time-consuming to perform
- **Why** did clinical trials miss the signal (if that's the case)?
- **How** can future trials / development programs be improved to detect similar problems earlier?

# Janus Data Pyramid

## Structured Scientific Data Management System



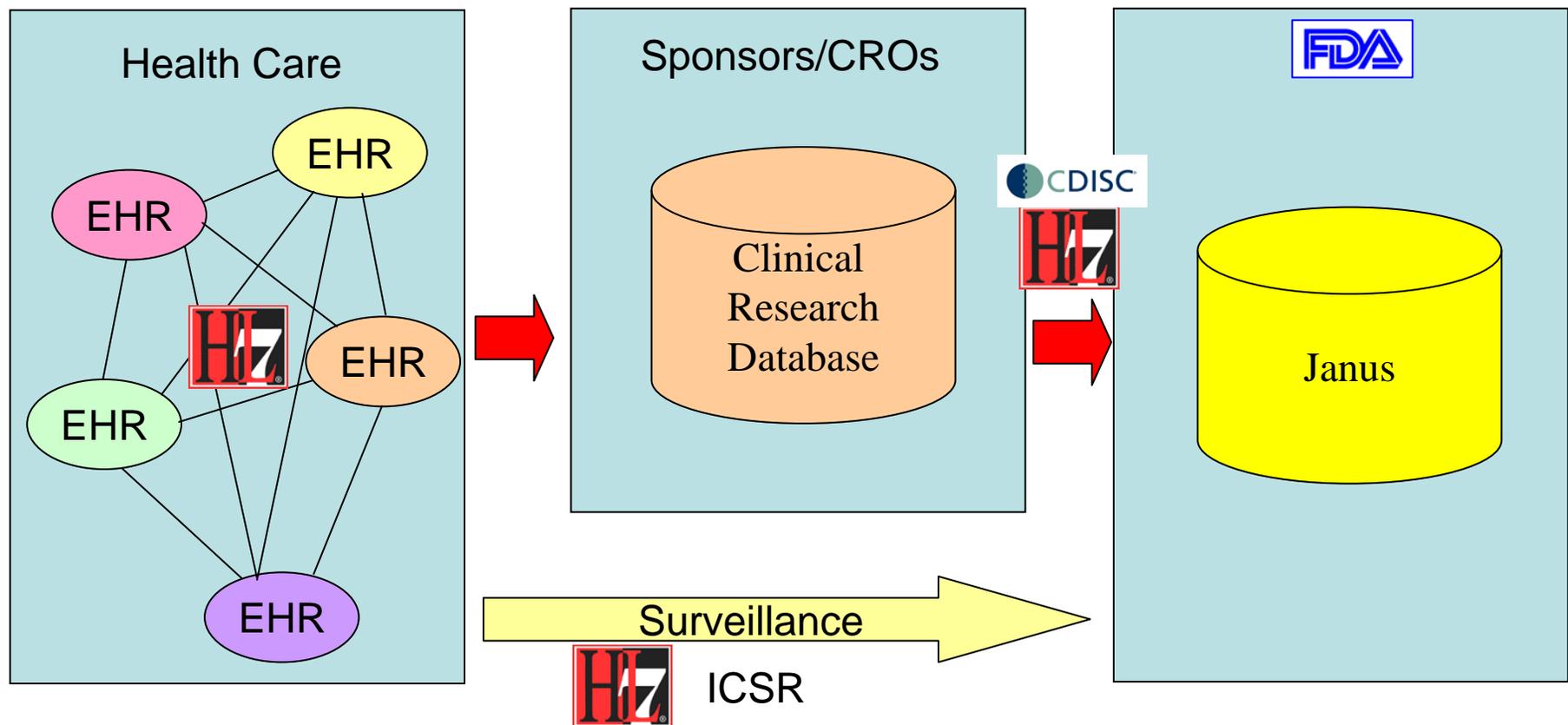
## *Driver #2 – cont'd*

### Harmonize with Other HL7 Standards

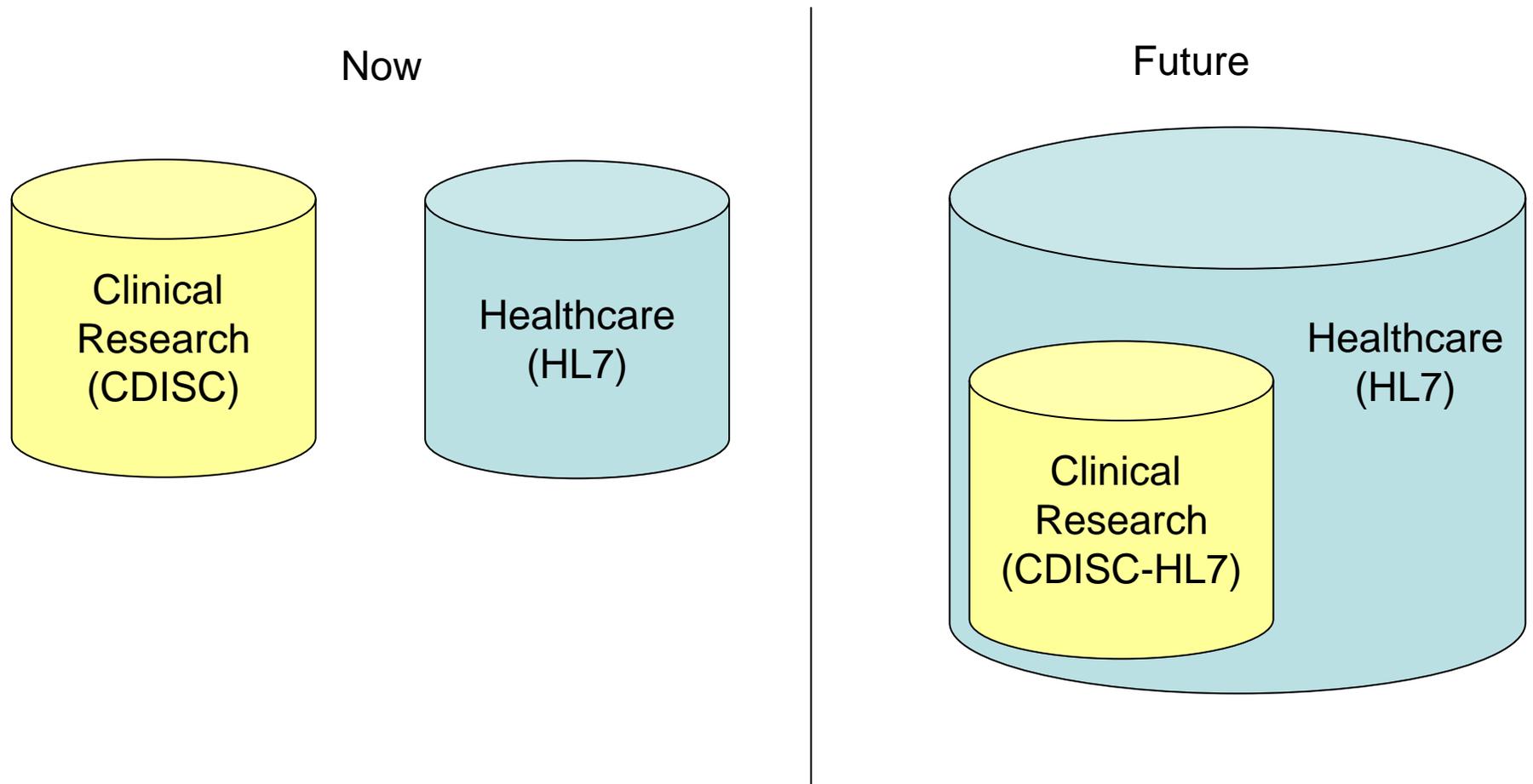
- Harmonization with HL7 SPL provides a better way to associate clinical observations with medical product information
  - This will be particularly important for medical devices, biologics, and drug-device combination products where model #, lot information, other product information is important to interpret the observation

## Driver #3

### Better Long-Term Integration with EHRs to support Clinical Research and Surveillance



# CDISC-HL7 – Integrating Clinical Research with Healthcare



## *How is the standard being developed?*

- Fall 2007 – CDISC-HL7 Project approved by the Regulated Clinical Research Information Management (RCRIM) Working Group within HL7
- Development of 4 messages for CDISC content:
  - study design
  - study participation
  - subject data (Care Record)
  - HL7 ICSR (expedited AE reporting)

# CDISC HL7 Standards Development

- Based on CDISC standards
- Leverages existing HL7 standards (e.g. HL7 ICSR, Clinical Statement CMET\*, SPL)

\*CMET – Core Message Element Type (reusable XML across messages)

# CDISC-HL7 Standard Four Messages

Message	Use
Study Design	•Study Protocol
Study Participation	•Add/changes to participant information
Subject Data (Care Record)	•Study result submission
ICSR	•Expedited AE reporting

*When ...*

**...will the standard be available?**

**...will FDA implement it?**

## **\*DRAFT\* Timeline**

**Actual Timeline will be documented/updated in rolling PDUFA 4 IT Plan**

9/2008*	DSTU Ballot (Draft Standard for Trial Use)
2008 – 2009	Testing (Phase 3 Janus Pilot)
9/2009	Normative Ballot
2009-2012	FDA accepts both XPT and CDISC-HL7 XML
2013	PDUFA V – CDISC-HL7 XML only
2014	EHRs fully deployed nationwide (HHS Goal)
2020 (?)	EHRs fully integrated in clinical research. Full benefit of CDISC-HL7 realized

\*9/08 DSTU ballot failed; re-ballot planned 5/09

# Common Myths

- **SDTM is going away**
  - False: SDTM is a very useful standard view of the collected data to support basic analyses (e.g. means, S.D., distribution)
  - SDTM will transition at FDA from an exchange standard to a standard analysis view that Janus will produce
- **CDISC-HL7 is just for human drug trials**
  - False: CDISC-HL7 is being designed for studies involving any substance (drug, biologic, device, combination) on any organism (human, animal), groups of organisms, part of organisms, and also for studies where the medical product is the subject of the experiment
- **FDA wants study data submitted as flat files**
  - False: Current flat file submissions evolved from paper “tabulations”
  - We now recognize the limitations associated with flat files
  - We intend to move to a more ‘multi-dimensional,’ ‘multi-relational’ model for data submissions
- **CDISC-HL7 is an XML wrapper for SDTM flat files**
  - False: CDISC-HL7 is a new relational model for study data that is able to capture and convey more meaning
  - CDISC-HL7 will support the creation of additional views of the data to support more analyses
  - CDISC-HL7 will enable us to answer more scientific questions and make better regulatory decisions from study data

# Implementation Considerations

- FDA mindful of tremendous resources currently expended in creating SDTM / SAS Transport Files
- Transition plan to CDISC-HL7 needs to consider costs and IT investment lifecycle
- We expect to continue to accept SAS transport files for quite some time (at least thru end of PDUFA4 - 2012?)
- Updates to this Implementation Plan will be made to the PDUFA4 IT Plan
- **Initial implementation** of CDISC-HL7 should be a mapping exercise from SDTM XPT to CDISC-HL7 XML but full benefit of new exchange format will not yet be realized
- Ultimate implementation takes clinical observations from EHRs as input to CDISC-HL7 messages (2014 and beyond)
- **Full benefit** of CDISC-HL7 realized when EHRs fully integrated into Clinical Research as the standard data collection instrument for clinical study data, post-marketing AE data, and exposure data.

# Take Home Messages

- **“The World is Round”**
  - Clinical data are not flat and cannot be exchanged using flat two-dimensional files without significant loss of meaning
- **FDA is transitioning to a “round view of the world” of clinical research**
  - CDISC-HL7 standard will get us there
- **SDTM is here to stay**
  - Will transition from a standard submission format to an standard view of the data in support of simple analyses (e.g. distribution, means, etc.)

# XPT vs. CDISC-HL7 XML

a more robust model of the world of clinical research



SAS Transport XPT



CDISC-HL7 XML