WELCOME

Public Meeting: Solutions for Study Data Exchange Standards November 5, 2012

Mary Ann Slack
CDER Data Standards Program
FDA Center for Drug Evaluation and Research

CDER Computational Science Center

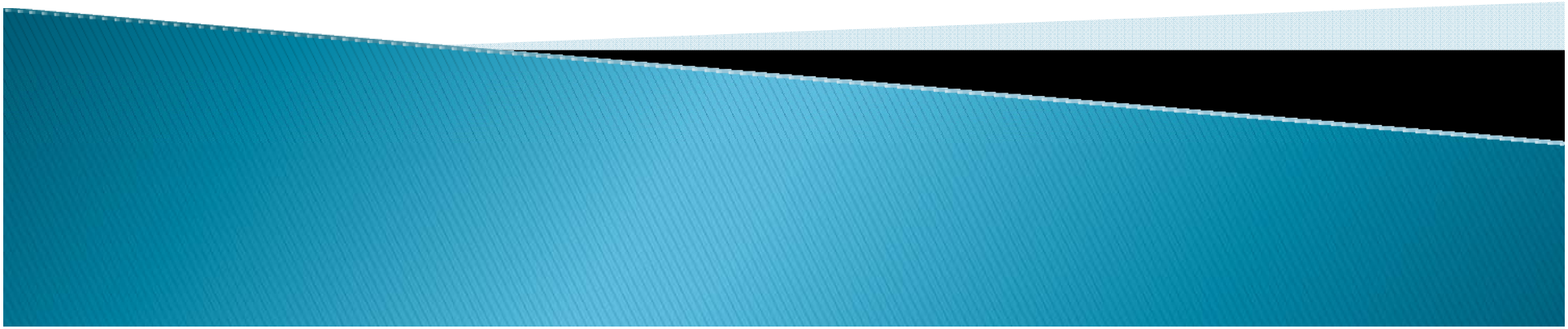
Better Data, Better Tools, Better Decisions

Computational Science Center:

Functional Needs for a Modern Review

Monday November 5, 2012

Chuck Cooper, M.D.
Computational Science Center
Office of Translational Sciences
CDER, FDA

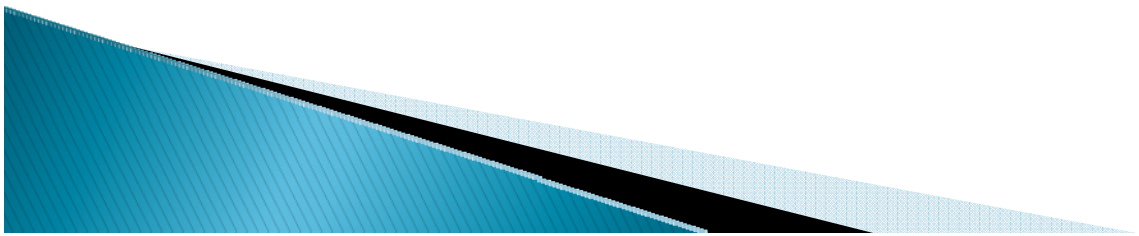


Outline

- Introduction
- Audit trail
- Flexibility
- Integration
- Metadata

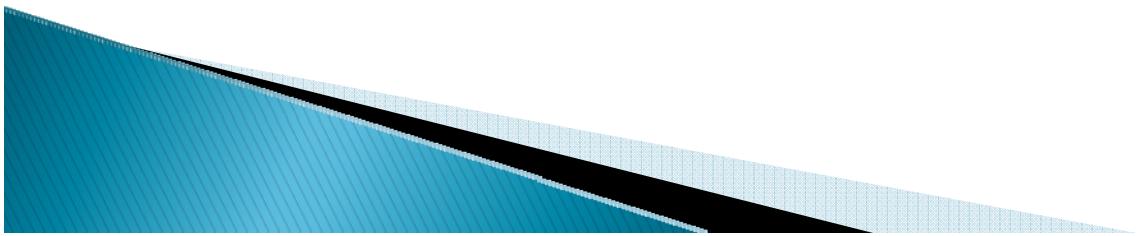
Introduction

- Modern Review Environment
  - Functional considerations
    - Overlap
- Other considerations



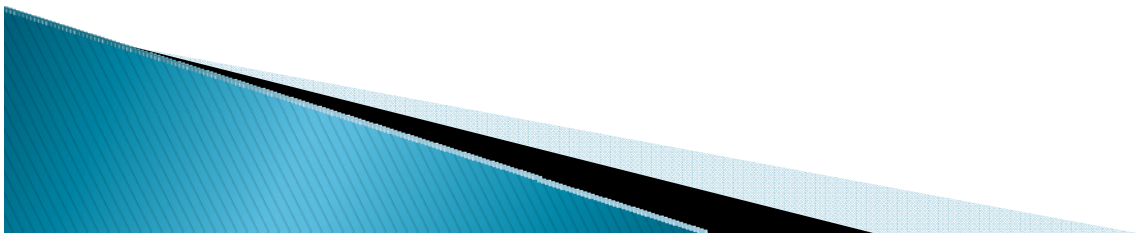
Audit Trail

- Data cleaning/coding/management
  - Reviewers have no window into this
- Analysis specific
  - How a sponsor created their analyses
    - If understood and easily validated, time is saved



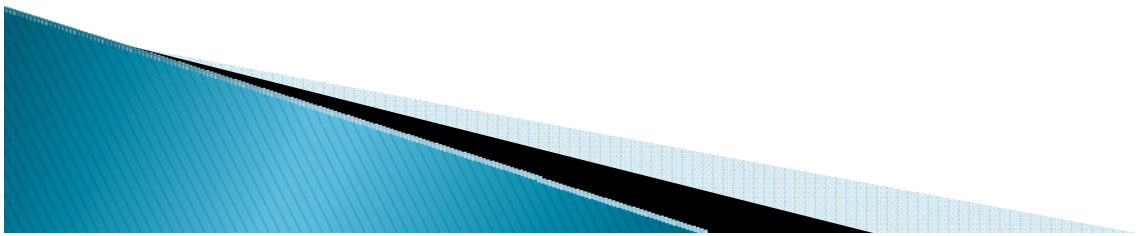
Flexibility

- Rapidly adapt to new, emerging standards
- Accommodate data not accounted for by existing standards
- Semantic clarity
- Usability



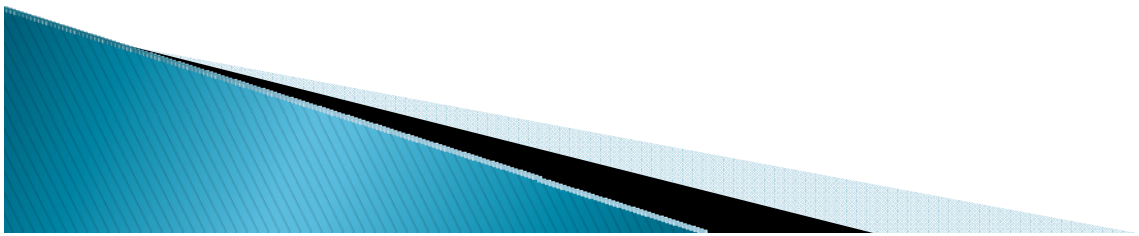
Data Integration

- Reviewers ask questions which involved data scattered across domains
- Alternative analyses
- New analyses
- Tool requirements

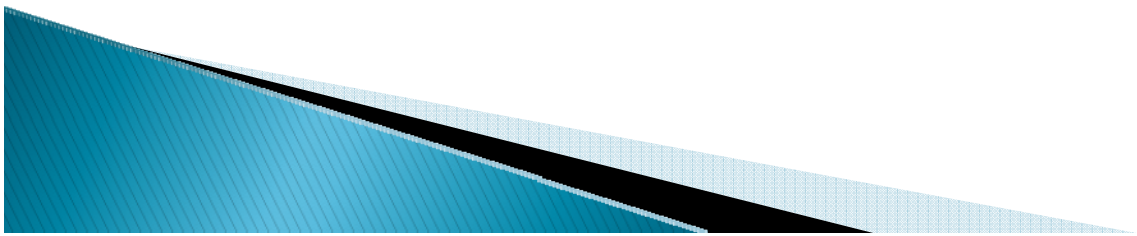


Metadata

- Lack of Robust metadata interferes with review process
  - Reviewers' first step before performing analysis
- Human understandable AND machine readable



END



To All Participants – Thank You For Your Participation And Inputs

Comments/Presentations from:

Industry – 11	Academia – 3	SDO – 7	CRO/Consult/Tech – 7
Roche	Duke	EN13606 Consortium	Formedix
Astellas	University of Arkansas	CDISC XML Team	Quintiles
Novartis	Oxford	CDISC (2)	Assero
Amgen		ASTM Subcommittee	Next Step Clinical Systems LLC
Merck		W3C / MIT	Zifo Technologies
Takeda		HL7	Bioclinica
Seattle Genetics			Octagon Research Solutions
Daiichi Sankyo			Medidata Solutions
GSK			SAS
Sanofi			
Novo Nordisk			
Lilly			

All comments submitted to the docket are available online at:
http://www.regulations.gov/#!docketBrowser;dct=PS;rpp=25;po=0;D=FDA-2012-N-0780

General Information

- All non-FDA guests are limited to this conference area of the campus
  - If you must leave the conference area, please see one of the FDA Support Staff
- The meeting will be recorded for purposes of ensuring accurate meeting summary
  - Recording will be erased when meeting summary is complete
- A concession area will be open during scheduled lunch and breaks
- Restrooms are located in open area near the check-in
  - Look for the signs
- Support Staff are available if you have questions

Meeting Drivers

Today –

- FDA supports the ASCII-based SAS Transport (XPORT) version 5 file format. It has served its purpose well, but…
  - XPORT v5 is not an extensible modern technology
  - Known limitations are causing technical issues; we anticipate these to become increasingly challenging

- Event and data relationships are not currently well captured
  - Information exchange may be part of a solution

Meeting Goal

"to solicit input from industry, technology vendors, and other members of the public regarding current and emerging potential solutions for the exchange of regulated study data."

Pathway to a Solution

- Public input on potential replacement solutions and experiences
  - What solutions are out there?
  - Are they “fit for purpose” today? In 1 year? In 5 years?
  - What should we know before we evaluate?
- Large scale changes cannot be made immediately
- An evaluation will be completed to determine the cost and benefit to both FDA and regulated Industry of any migration to a new exchange format
  - Any solution will need to meet FDA requirements

Example Scenarios

Available on CDER Data Standards website

“Amending an Existing Protocol (Complex)

ACME Pharmaceuticals Inc. has amended the protocol document by increasing the planned enrollment from 400 to 500 and by adding two new eligibility criteria. This is captured as an amendment to the previously submitted protocol. The amendment describes the changes to the protocol and the reason for the changes. The amendment itself may contain multiple sections. The original protocol is unchanged and the amendment is appended to the original document.”

- Associate changes to original documents
- Running “history” of clinical trial

Example Scenarios

“Updated IRB Approval - Protocol Amendment

Acme Pharmaceuticals amends the protocol for study NCT99999999 to extend the duration of experimental treatment by an additional two months. The protocol amendment triggers a review by all the relevant subject protection approving authorities and each grants an updated approval.”

- Associate changes to original documents
- Running “history” of clinical trial
- Trigger actions

Public Comments Received

Question 1: Pressing Challenges

– General

- Lack of high quality data standards across the clinical data lifecycle
- Designing TA specific CRFs
- Timely reporting of data by clinical sites
- Requesting new terminology to have it included in the controlled terminology
- Lack of the universal adoption of CDASH standards leading to disparate CRF structures.

– International Alignment on Submission Standards

- Standards version control
- TA standards to meet our needs
- Lack of standardized industry ontology and terminology

Public Comments Received

Question 1: Pressing Challenges

- **FDA**
  - Expectations to follow the latest standards for each study must be tempered with the need to maintain consistency within a project
  - Conflicting requests from review divisions
  - BIMO data requests
  - Need guidance on managing versions of standards during the lifecycle of a submission.
  - Making changes a few months prior to submission is disruptive.

Public Comments Received

Question 2: How could FDA study data management process more efficient?

- Default agency position is "discuss with your reviewer". Too much variation. Need clear regulatory guidance.
- Differences in approach/expectations across the different FDA therapeutic divisions make it harder for a companies.
- Need guidance on how to manage the existence of multiple versions of the standards through the lifetime of a typical asset
- Continue to share data validation requirements.

Public Comments Received

Question 2: How could FDA study data management process more efficient?

- Synchronized standards and requirements across FDA.
- Need a data standards planning meeting at EOP2.
- Need metrics on what data and data standards are being utilized for decision-making.
- Need much advance notice on standards deprecation.
- Assurance that the specifications of a standard have been tested.
- Need final guidance on lifecycle management of standards.

Public Comments Received

Question 3: What does Industry need to make clinical trials data management more effective and efficient?

- A central repository for the storage, versioning, and dissemination of standards is needed.
- Mature, clearly defined, for purpose, end-to-end standards.
- An agile industry-wide standards development and governance organization.

Public Comments Received

Question 3: What does Industry need to make clinical trials data management more effective and efficient?

- Tools to provide Industry with a pass/fail decision for data format and structure compliance before submitting the data to the FDA.
- EDC vendors need to use the same variable names for the same concepts when building the data capture system.
- We need tools that allow for automatic end-to-end conversion of data from data capture systems to the data exchange standard.

Agenda

- **Agenda was developed to**
  - Ensure that FDA receives input from stakeholders
  - Provide an environment to encourage discussion
  - Hear open feedback from stakeholders on views, concerns, issues and proposals, both short and longer term
- Speakers were scheduled on a “first-come, first-serve” basis and availability of meeting time
- We have allowed approximately 1.5 hours for open discussion in the afternoon

Center for Drug Evaluation and Research

Regulatory New Drug Review: Solutions for Study Data Exchange Standards

Public Meeting Agenda
White Oak Campus
November 5, 2012

Time		
10:00	**Welcome and Introductory Remarks**	Mary Ann Slack Deputy Director, CDER/OPI
10:15	**FDA Drivers for this Meeting Topic**	Mary Ann Slack
10:30	**Discussion – Problems/Challenges Faced Within Current Environment and General Requirements** **Speaker 1**. Doug Warfield (FDA) **Speaker 2**. Chuck Cooper (FDA) **Speaker 3**. Armando Oliva (FDA)	
11:30	**Discussion – Data Exchange Standards and their Advantages and Disadvantages** **Speaker 1**. Bill Gibson (SAS) **Speaker 2.** Peter Mesenbrink (Novartis) **Speaker 3**. Mathias Brochhausen and William Hogan (University of Arkansas)	
12:00	**Lunch**	

Center for Drug Evaluation and Research

Regulatory New Drug Review: Solutions for Study Data Exchange Standards

Public Meeting Agenda
White Oak Campus
November 5, 2012

Time

1:00 **Continue Discussion – Data Exchange Standards and their Advantages and Disadvantages**

Speaker 4. Gary Kramer (ASTM Subcommittee)
Speaker 5. Charlie Mead (W3C)
Speaker 6. Armando Oliva (FDA)*
Speaker 7. Dave Gemzik (Medidata Solutions)
Speaker 8. Wayne Kubick (CDISC)
Speaker 9. Fred Wood (Octagon)
Speaker 10. Diane Wold (GSK)
35 min Open Discussion

2:45 **Break**

3:00 **Continue Open Discussion**

3:50 **Summary & Next Steps** Mary Ann Slack

4:00 **Meeting Adjourned**

*** Armando Oliva (FDA) will present a proposal on behalf of HL7, who were unable to participate.**

Ground Rules

- This is an information-seeking meeting
- Participation is encouraged
- Objective discussion and clarifying questions are welcomed and encouraged
- Please limit side conversations during presentations
- Due to the limited time, we will be adhering to the schedule timeline as closely as possible;
  - Unfinished discussions and questions will be captured for the afternoon discussion, time permitting.

Center for Drug
Evaluation and Research

Regulatory New Drug Review: Solutions for Study Data Exchange Standards

Problems/Challenges Faced Within Current Environment and General Requirements

Douglas Warfield, Ph.D.

Technical Lead/Interdisciplinary Scientist
eData Management Solutions Team
Office of Business Informatics
U.S. Food and Drug Administration
Center for Drug Evaluation and Research

Regulatory New Drug Review: Solutions for Study Data Exchange Standards

Problems/Challenges Faced Within Current Environment

- Limitations of the dataset standard for submission (SAS Transport V5)
- Dataset files size – increasing dramatically

FDA Dataset Transport

Anatomy SAS Transport V5*

- Limitations of Version 5
  - Variable name length limited (8)
  - Variable label length limited (40)
  - Variable size limited
  - Variable size pre-allocated based on length

*FDA specification for submission of datasets - *Study Data Specifications*

http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM312964.pdf

FDA Dataset Transport

*Anatomy SAS Transport V5 (cont.)**

HEADER RECORDS

SAS Symbol | SAS Lib | SAS Ver | SAS OS |…
SAS Symbol | SAS DS Name | SAS Data | SAS Ver |…
ddMMyy:hh:mm:ss….

NAMESTR RECORDS

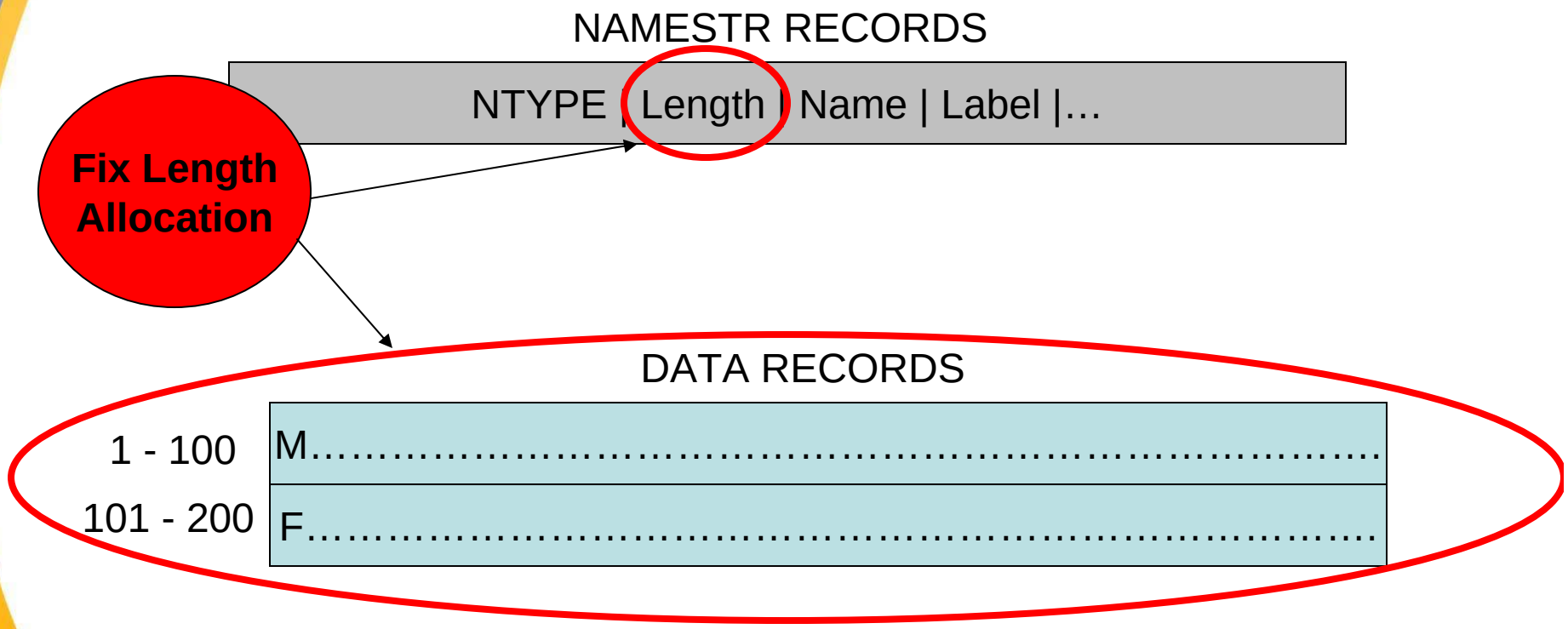
NTYPE | Length | Name | Label |…

NTYPE:1=numeric, 2=char Length: short Name: 8 chars Label: 40 chars

*TS-140 THE RECORD LAYOUT OF A DATA SET IN SAS TRANSPORT
http://support.sas.com/techsup/technote/ts140.pdf

FDA Dataset Transport

Anatomy SAS Transport V5 (cont.)



Example: Type=char Length: 100 Name: Sex Label: Subject Sex

FDA Dataset Transport

*Anatomy SAS Transport V5 (cont)**

- NTYPE = numeric – 8 bytes
- NTYPE = char – bytes based on fixed allocation of length
- CDISC Models use character types extensively for data tabulations

*CDISC - Clinical Data Interchange Standards Consortium

CDISC Submission Dataset Sizes: CDER eData Team Quantitative Research

Introduction

- At Time of Research (March 2011):
  - ~ 650 dataset SDTM* submissions / week
  - ~ 35% All datasets were CDISC/SDTM* formatted study data
- Issues:
  - File size limitations of tools available to conduct regulatory reviews
  - Increased file sizes present challenge to current data management systems/processes

*SDTM – Study Data Tabulation Model

Research: CDISC Submission Dataset Sizes

Research Design and Methodology

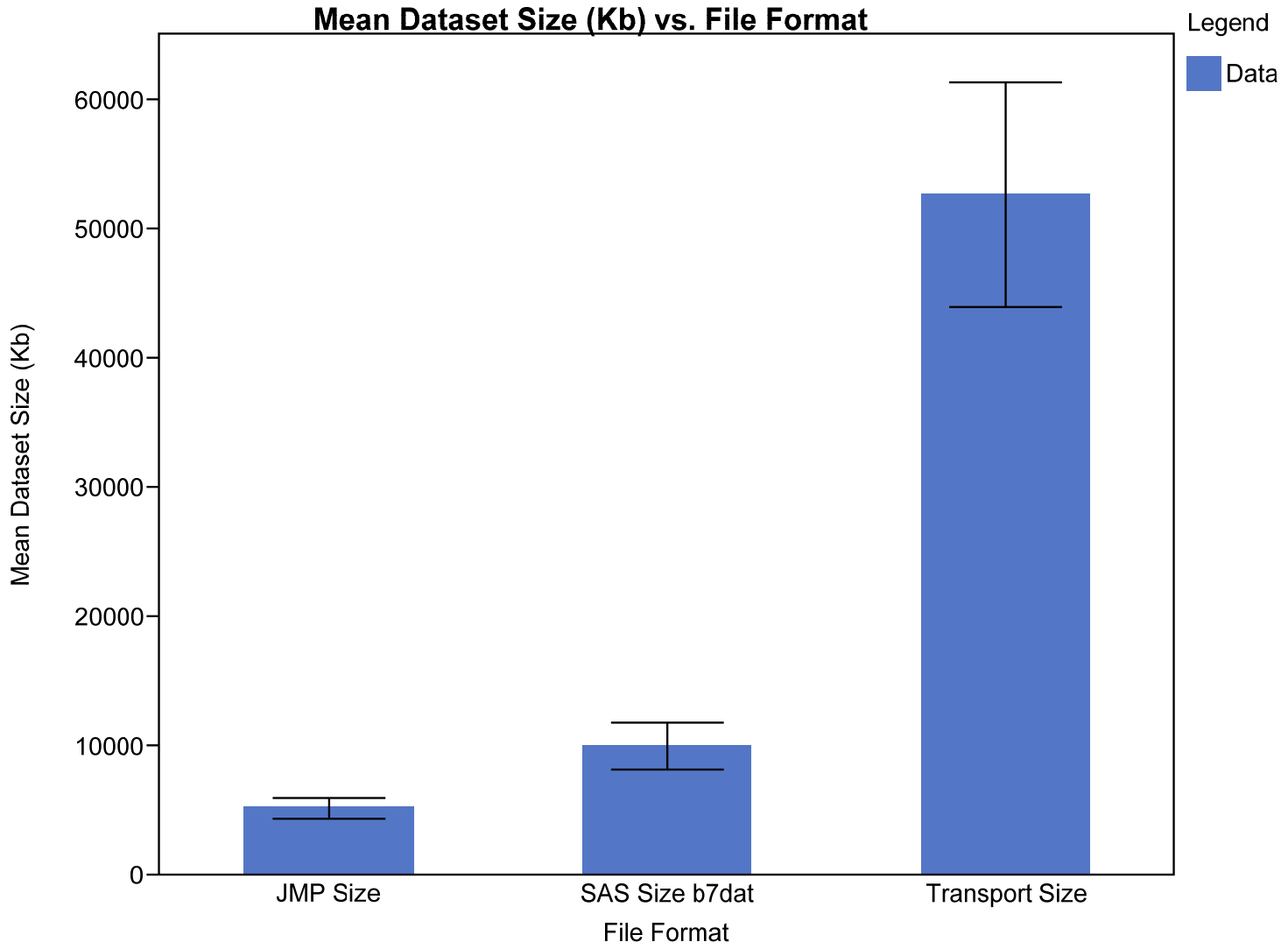
Research study was designed as a *quantitative exploratory* study, reviewing ***20*** randomly selected studies from a list of 565 unique studies tabulated by the eData Team (OBI/CDER) from 2010-2011.

Total of 432 datasets.

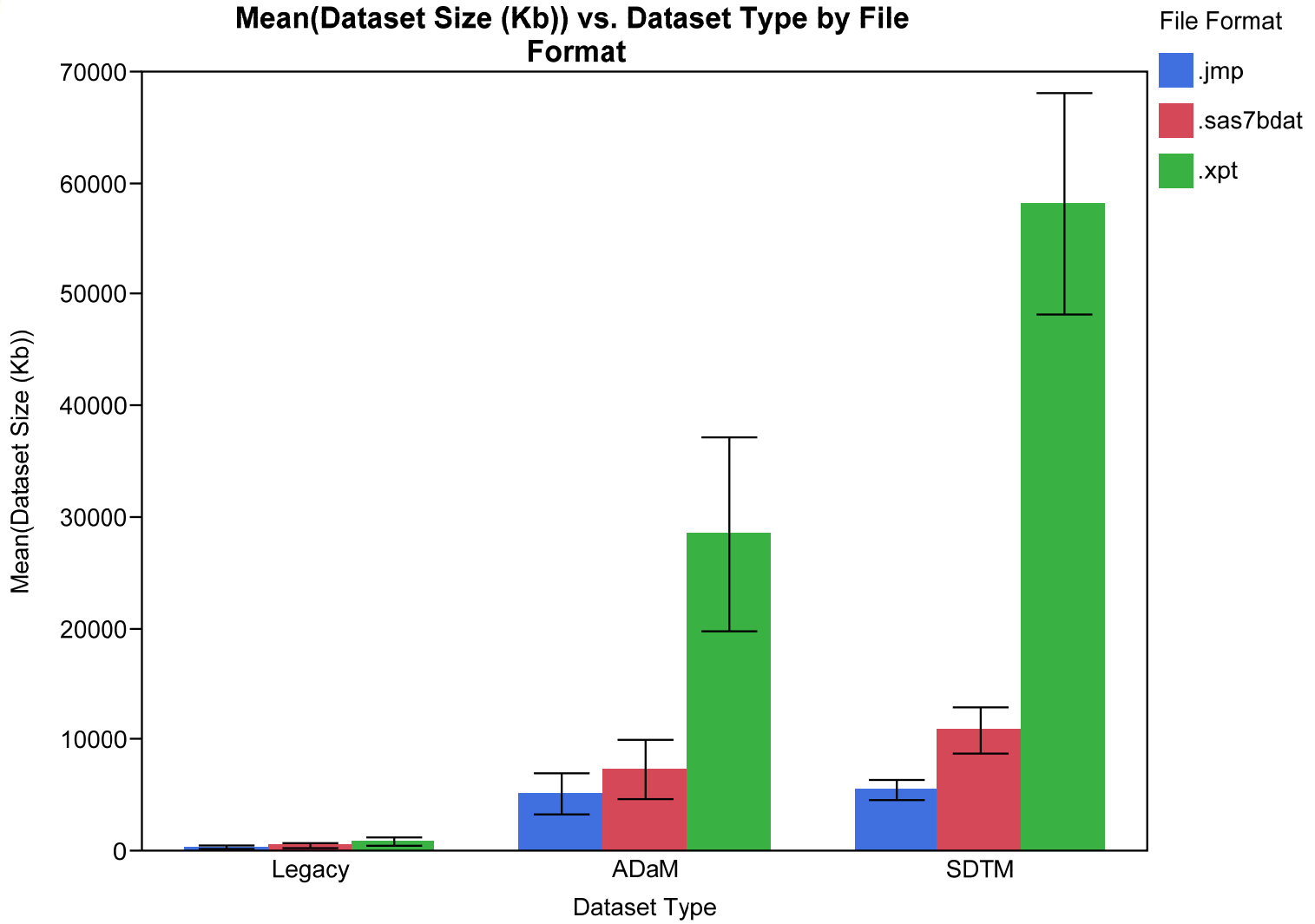
Limitations

- Randomly selected sample size of 20 statistically is too small to significantly reflect the population of CDISC/SDTM clinical trial datasets
- The population sample of 565 studies only reflects the most recent studies from January 2010 to February 2011 and includes only electronic submissions.
- Research focused on XPORT v5 file formats, and only briefly touched upon the .xml file type as another possible format.

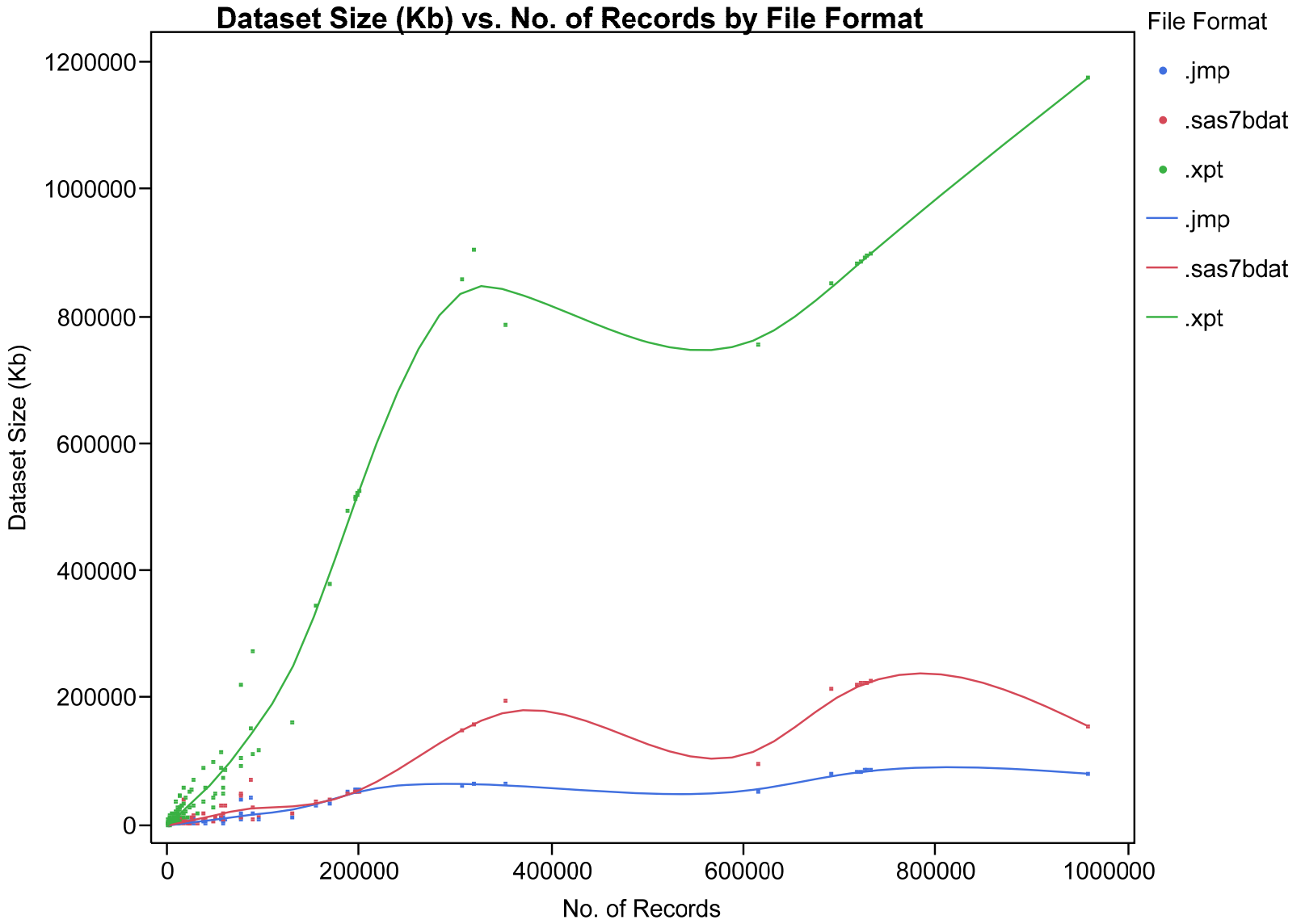
Research: CDISC Submission Dataset Sizes



Research: CDISC Submission Dataset Sizes



Research: CDISC Submission Dataset Sizes



Research: CDISC Submission Dataset Sizes

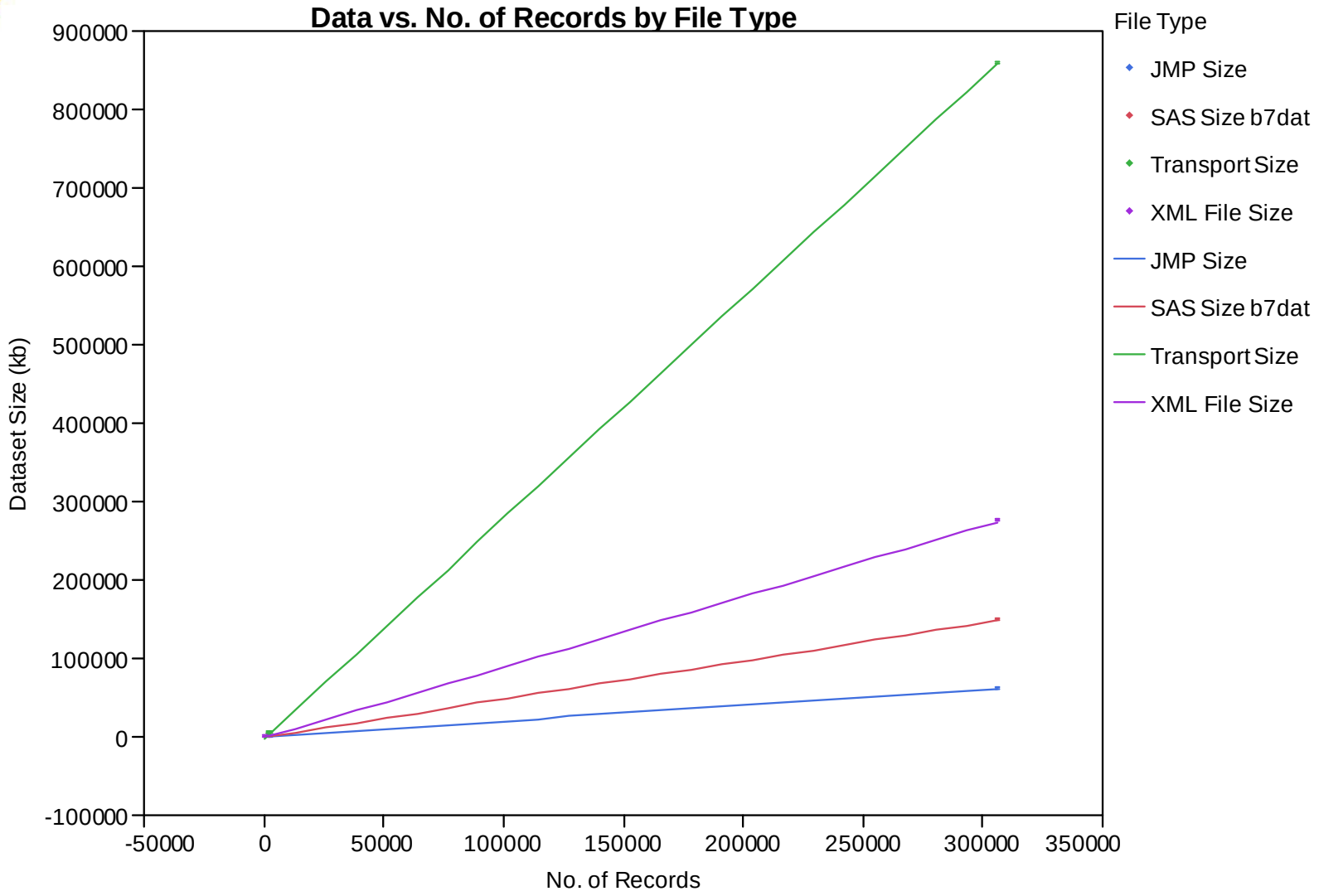
Labs (LB)

Variable Name	Variable Type	Previous Variable Length	Modified Variable Length
DOMAIN	Character	2	2
LBBLFL	Character	2	2
LBCAT	Character	200	20
LBDTC	Character	50	20
LBNRIND	Character	8	8
LBORNRHI	Character	200	10
LBORNRLO	Character	200	10
LBORRES	Character	200	15
LBORRESU	Character	200	10
LBREFID	Character	200	15
LBSEQ	Numeric	8	8
LBSPID	Character	200	5
LBSTAT	Character	8	8
LBSTNRHI	Numeric	8	8
LBSTNRLO	Numeric	8	8
LBSTRESC	Character	200	15
LBSTRESN	Numeric	8	8
LBSTRESU	Character	200	10
LBTEST	Character	200	30
LBTESTCD	Character	8	8
STUDYID	Character	200	10
USUBJID	Character	200	20
VISIT	Character	200	25
VISITNUM	Numeric	8	8
	Total	**2718**	**283**

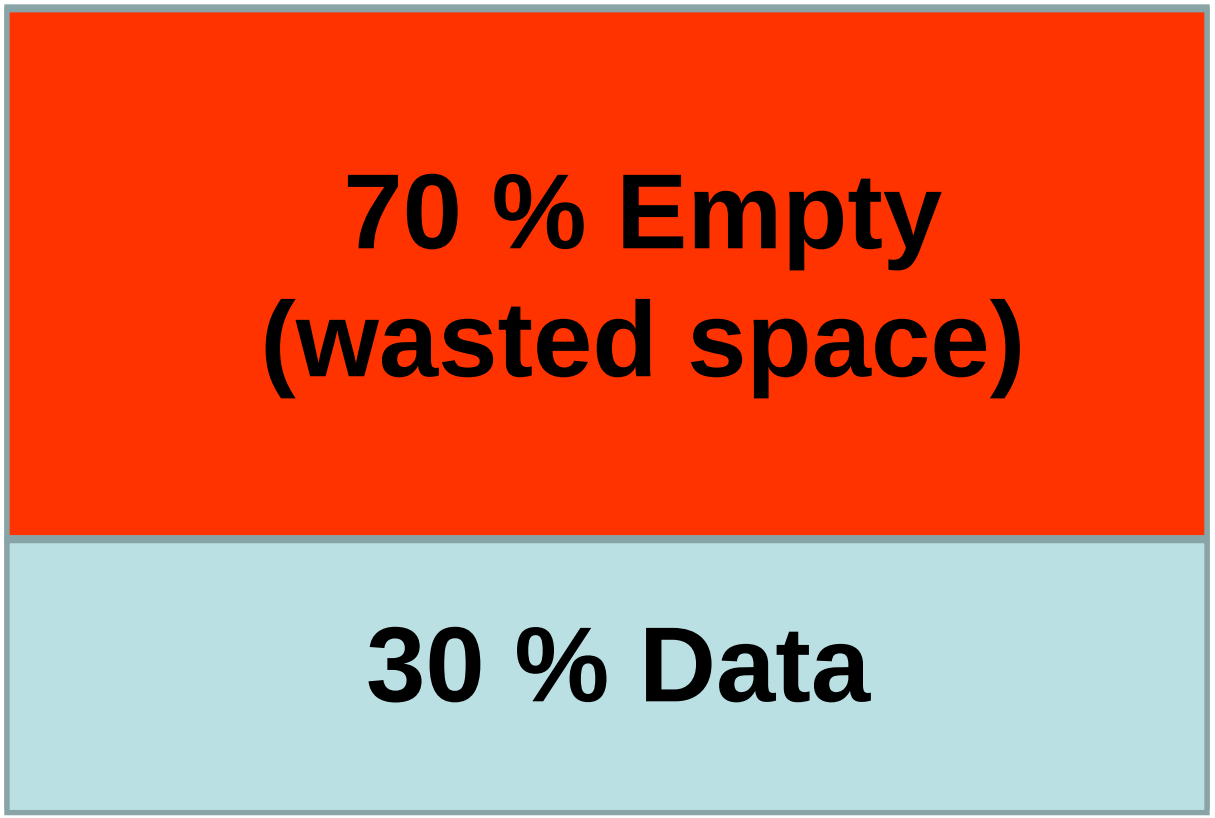
Reduced to the width needed

Totals bytes used ~ 1/10 the size

Research: CDISC Submission Dataset Sizes



Research:
CDISC Submission Dataset Sizes



Submission Datasets

Research: CDISC Submission Dataset Sizes

Research Conclusion

Significant file size differences:

- Legacy and CDISC datasets,
- File types - JMP (.jmp), SAS v7 (.sas7bdat), and Transport v5 (.xpt)
- Wasted Space in character strings

Recommended Immediate Action: *Resize*

Prefer sponsors submit datasets (Transport version 5) using maximum length required (used). Average reduction of **70**%.

Resize Industry Testing

- Industry involvement and testing through PhRMA ERS* group
- Provide 14 participants the following:
  - CDER eData's .sas code used during internal column resizing research
  - Directions on how to select and submit sample study for resizing
  - Template (.xls) to record information and any errors/issues observed

*PhRMA ERS – Pharmaceutical Research and Manufacturers of America Electronic Regulatory Submissions

Research: CDISC Submission Dataset Sizes

Industry Testing Results

Summary (15 studies)

- 15 studies (545 datasets) from 14 participants
- Result - average study size reduction of 68%
  - Range of study size reduction from 27% to 88%
  - Average individual dataset (.xpt) size reduction of 65%
- Average allotted column width reduction by study of 69%
  - Average study column width reduction range from 41% to 90%

FDA Dataset Transport

FDA and Industry Results - Strategy

- Results of FDA research and Industry testing
  - SAS Transport V5 is a poor transport for CDISC standardized data (many character data types)
  - Industry and FDA adversely affected
- Explore options with industry
  - 70% wasted space too costly, and even when eliminated
  - the remaining 30% is still a concern

Center for Drug
Evaluation and Research

Regulatory New Drug Review: Solutions for Study Data Exchange Standards

Problems/Challenges Faced Within Current Environment and General Requirements

Thank You!

Please save your questions until Q & A session.

Challenges of Current Study Information Exchange Format

Armando Oliva, M.D.

CDER Computational Science Center
Food and Drug Administration
armando.oliva@fda.hhs.gov

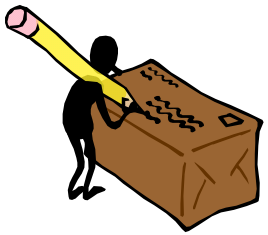
November 5, 2012

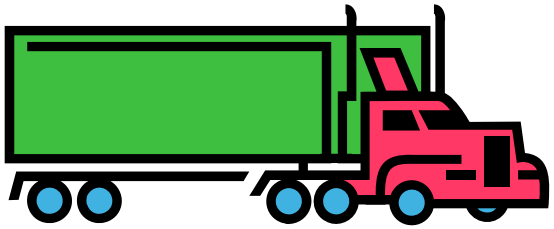
Disclaimer

- This presentation IS about
  - examining the current study information exchange format
  - exploring the future of study information exchange
- This presentation is NOT about
  - study information content standards
  - changing the content standards in current use
- High on Principles; Low on technical details

Exchange vs. Content Stds

- Exchange Standards:

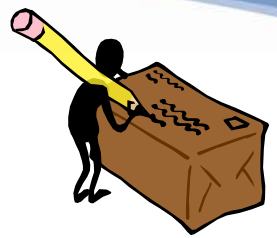




- Content Standards:

The content “requirements” drive the exchange format (“shipping container”)

Limitations of SAS XPT v5



- character limitations for variable names, labels, character fields
  - Easily solved with numerous ‘newer’ exchange formats

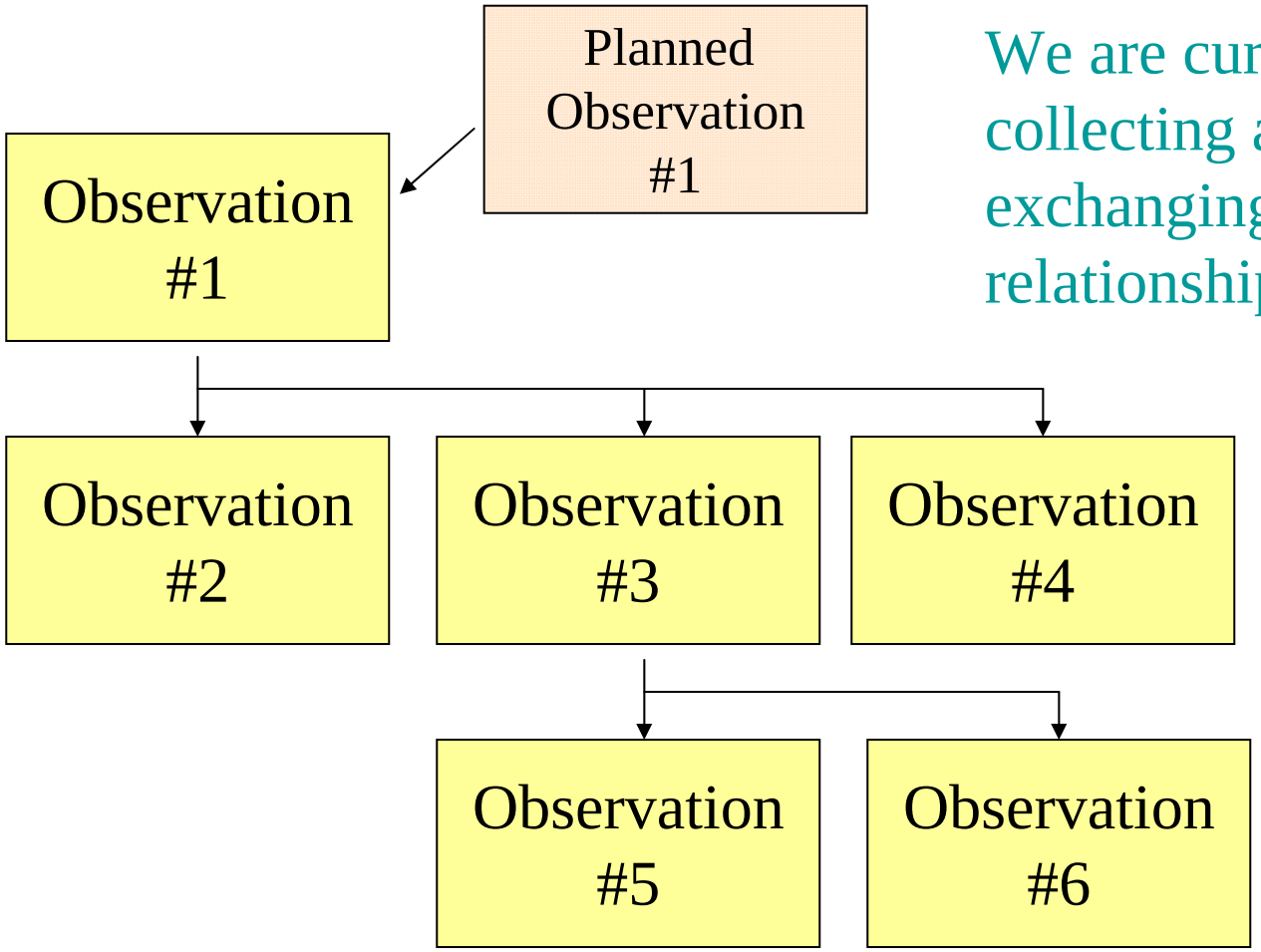
- Flat, two-dimensional data structure for hierarchical, multi-relational data
  - More challenging to solve

Clinical data is Round

- Clinical data is hierarchical and multi-relational – "round"
- Important **meaning** is lost when exchanging 2-dimensional flat files, making some interpretations and analyses difficult or impossible
- i.e. decreased **semantic interoperability**

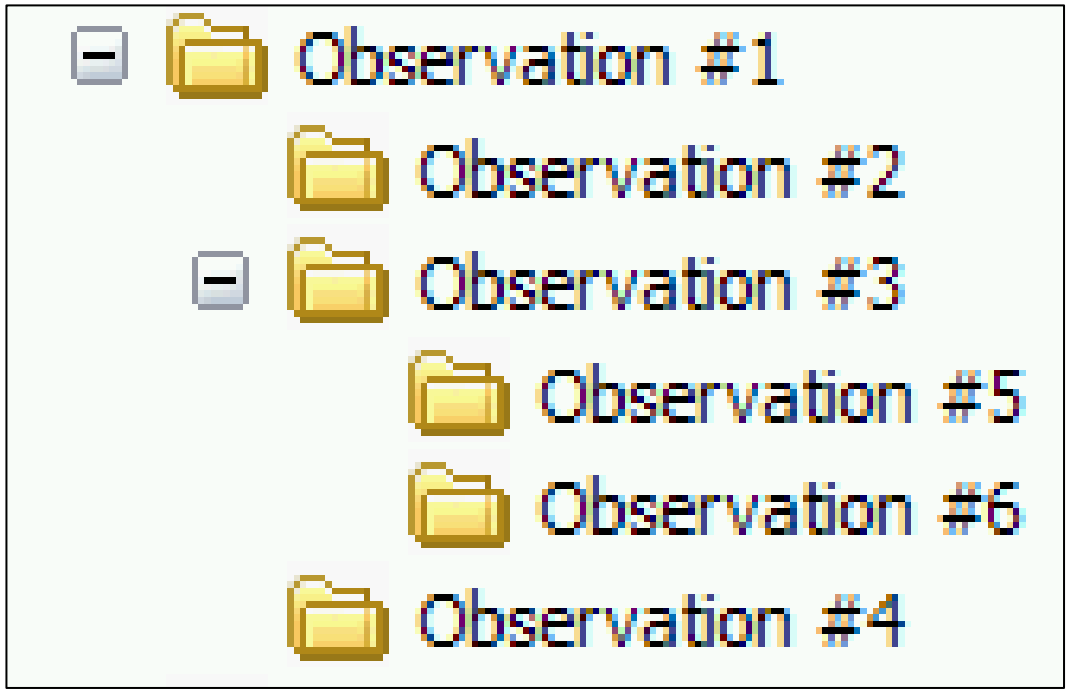
- Just like flat maps are useful for relatively short distances, they are not useful in navigating the globe

Clinical Observations:
Highly Relational and Hierarchical ("Round")



We are currently not collecting and exchanging these relationships well.

Consider Clinical Observations as nested folders in a tree structure



Flat files don't inherently capture the tree or **data structure**, which is itself important to understand and analyze the data

Relational ("Round") Data

- Necessary to create useful relational database(s) and knowledge management systems of study information
- Improve the ability to "slice and dice" the data in many more useful ways
- Support more automated, efficient analytical processes

- CDER has created scenarios to help illustrate the information exchange challenges
  - Scenarios available at: fda.gov (exact link TBD)

- Long-term exchange solution should support the exchange of round, multi-relational data
  - The solution should be based on a robust relational information model that better reflects the real world of clinical data
  - The solution will also necessitate a shift in how data are collected – move away from antiquated, paper CRF-based practices

…and now a shift in gear

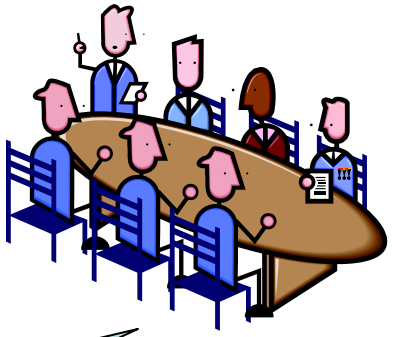
A lighthearted look at study information exchange now and in the future

Now

Hi FDA, I've got this great new drug. It can take you anywhere in the Northeastern U.S. you want to go.

Indication

Sponsor 1

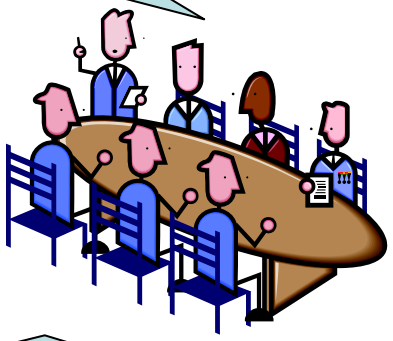


Oh really? Send us your data. We want to check it out and see if it can get us there.

Now

We want to see if it can take us to New York City. Please send us a map from Washington DC to New York City

OK



Sponsor 1

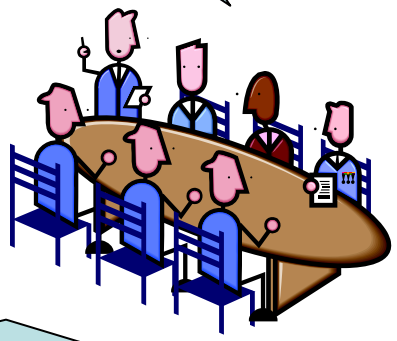
Study Data

FDA

Now

OK, and send us a map from DC to Boston, oh, and also one to Buffalo.

OK

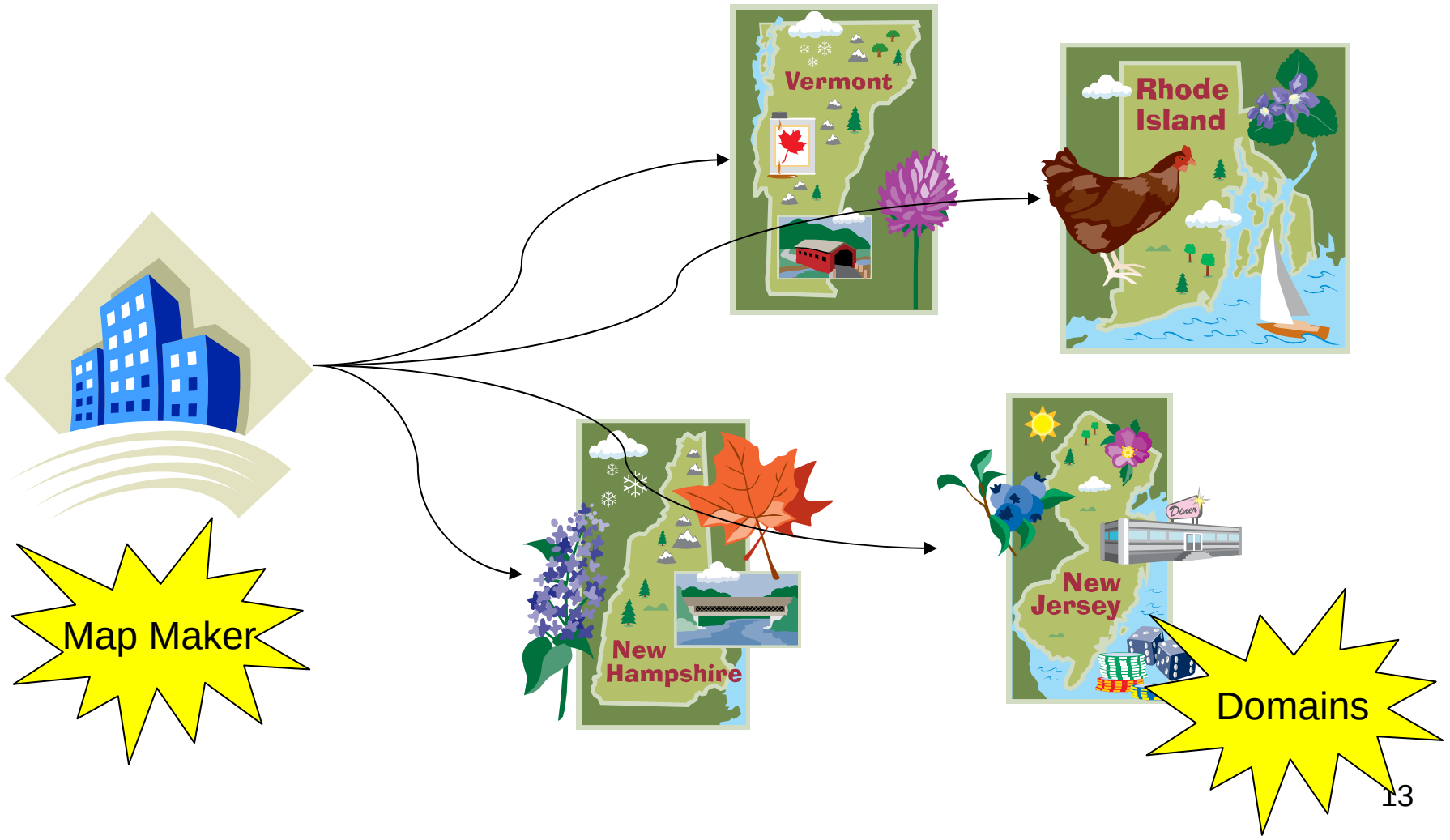


Sponsor 1

Analysis Datasets

FDA

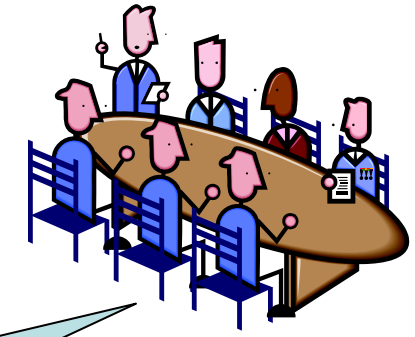
Meanwhile – Standard Maps Emerge



Now

Hi FDA, I've got this great new drug. It can take you anywhere on the East Coast you want to go.

Sponsor 2



OK, send us maps to NYC, Boston, Atlanta, and Miami. We want to see how your drug gets us there.

FDA

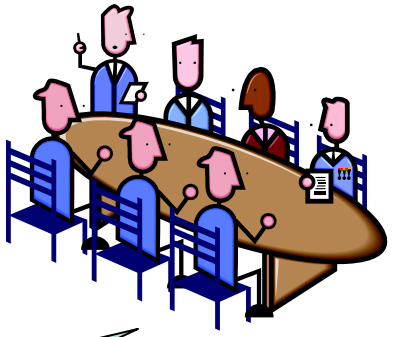
Sponsor 2

- But, standard maps only exist to NYC and Boston.
- Sponsor has to create custom maps to new destinations while standard maps to Atlanta and Miami are created.
  - It takes map makers many months to generate these new maps
  - It takes many months for sponsors to start using the new maps
  - It takes many more months for FDA to start seeing the new maps

Now

Hi FDA, I've got this great new drug. It can take you anywhere in the Continental U.S. you want to go.

Sponsor 3



OK, send us maps to NYC, Atlanta, Chicago, Dallas, and Los Angeles. We want to see how your drug gets us there.

Problem with this strategy

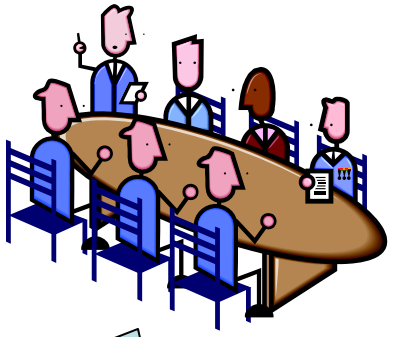
- FDA's never-ending requests for flat maps as the science, data requirements and review needs evolve
- Never-ending request to map makers to generate more and more flat maps to an ever-increasing list of new destinations
- Endless, time consuming, inefficient cycles of map creation, publication, testing, implementation
- This is not sustainable long-term
- Extremely burdensome to sponsors

Future

Hi FDA, I've got this great new drug. It can take you anywhere in North America you want to go.

Sponsor 4

Relational Information Model

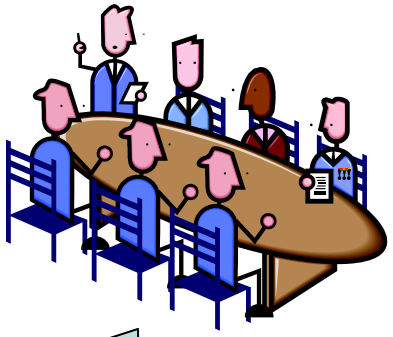


OK, here's a globe. Put all your data on the globe using longitude/latitude coordinates. We'll create all the maps we need and determine for ourselves where your drug can take us.

Future

Hi FDA, I've got this great new drug that you've never seen before. It can take you anywhere in Europe you want to go.

Sponsor 5



Hey, you know what, this globe happens to work for Europe, too. Put all your data on the globe using longitude/latitude coordinates. We'll create all the maps we need.

Advantages of the "Globe"

- The globe is a much more accurate model of the real world
- The globe is also more stable; doesn't change
  - Unless new land gets created
  - Avoids "moving target" submission requirements

- Less burden to sponsors long-term
- More flexibility to FDA to answer all our review questions

Disadvantages of the "Globe"

- Not end-user friendly
  - Can't really take it with you on a road trip

    …or stick one in your back pocket
  - One still has to generate flat maps that reflect where you are going *(e.g. google maps)*
- Data collection practices currently don't capture longitude/latitude
  - Data collection tools and practices need to change to take full advantage of the globe

In Conclusion…

- **"The World is Round"**
  - Clinical data are not flat and the current flat two-dimensional exchange format results in loss of meaning, limiting the ability to support a broad range of analyses of interest to FDA

- **We are transitioning to a "round view of the world" of clinical data**
  - Long-term study information exchange solution should be based on a robust relational information model that can support the complexity of round, multi-relational clinical data

Thank You

CDER Computational Science Center

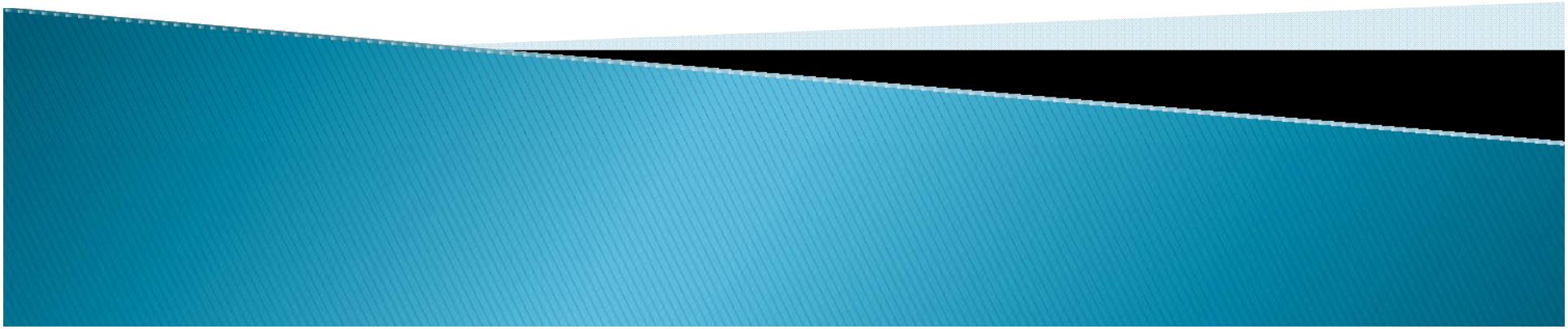
Better Data, Better Tools, Better Decisions

Computational Science Center:

Functional Needs for a Modern Review

Monday November 5, 2012

Chuck Cooper, M.D.
Computational Science Center
Office of Translational Sciences
CDER, FDA

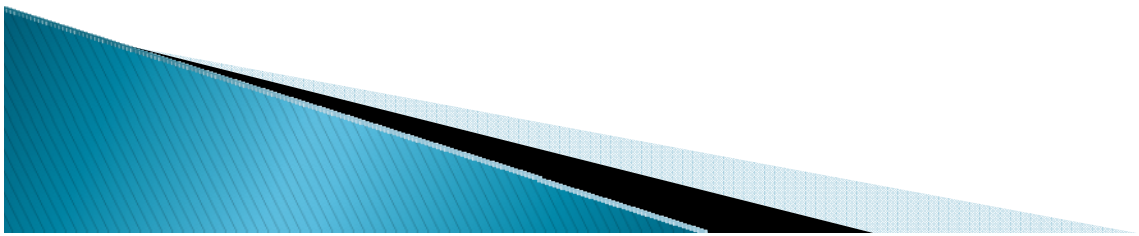


Outline

- Introduction
- Audit trail
- Flexibility
- Integration
- Metadata

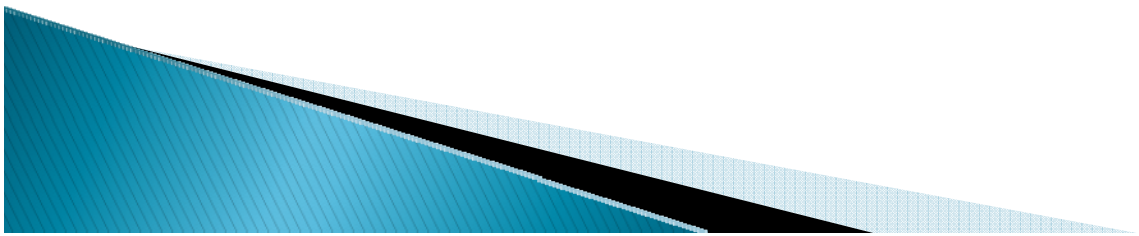
Introduction

- Modern Review Environment
  - Functional considerations
    - Overlap
- Other considerations



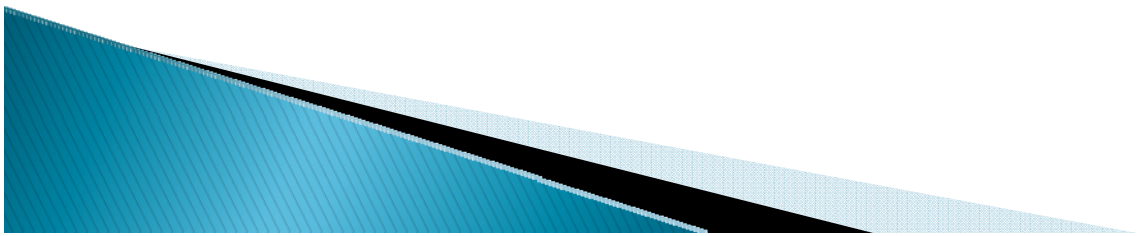
Audit Trail

- Data cleaning/coding/management
  - Reviewers have no window into this
- Analysis specific
  - How a sponsor created their analyses
    - If understood and easily validated, time is saved



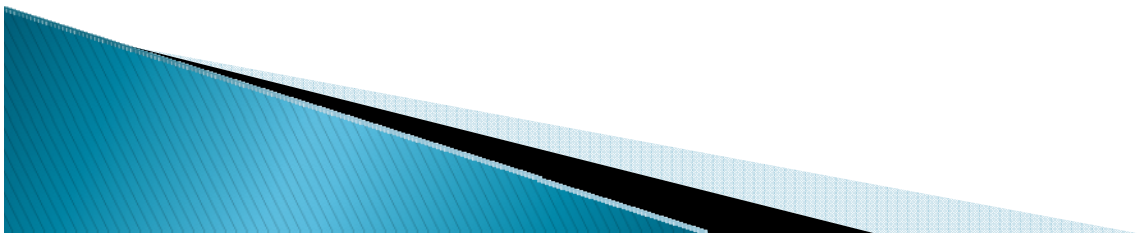
Flexibility

- Rapidly adapt to new, emerging standards
- Accommodate data not accounted for by existing standards
- Semantic clarity
- Usability



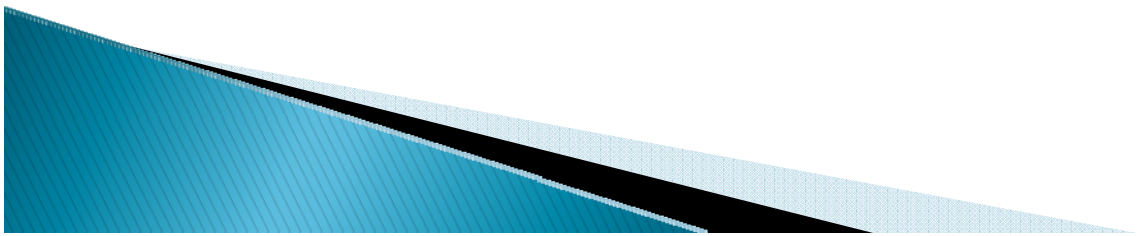
Data Integration

- Reviewers ask questions which involved data scattered across domains
- Alternative analyses
- New analyses
- Tool requirements



Metadata

- Lack of Robust metadata interferes with review process
  - Reviewers' first step before performing analysis
- Human understandable AND machine readable



END