DATA DELIVERY STRATEGY FOR INDUSTRY AND FDA

TRANSITIONING TO CDISC VIA THE NEW SAS TRANSPORT FILE EXTENSIONS

5-NOVEMBER-2012

BACKGROUND

- The SAS V5 Transport file format is an open standard developed by SAS to support data transfers between systems, especially those running different operating systems
- The SAS V5 Transport file is used across multiple industries
- FDA adopted the SAS V5 Transport File in 1999 as the expected format for receiving data from industry as part of a new drug application
- The ability to create and load SAS V5 Transport files is part of SAS/Base (SAS' entry-level offering)
- Any software vendor can develop a tool to create or load files that conform to this open standard

SAS V5 TRANSPORT FILE

Top of the line computer specs in 1999

- **10GB hard drive**
- **3.5 MB floppy drive**
- **CD-ROM**
- **56k modem**

KNOWN LIMITATIONS

- There have been no changes in the SAS V5 Transport file format since 1999
  - Field names are limited to 8 characters
  - Field names are limited to traditional alphanumeric characters
  - Field labels are limited to 40 characters
  - Character variables are limited to 200 bytes

NEW CAPABILITIES

- SAS has released a new format specification along with new SAS macros that extend and modernize the SAS V5 Transport file
- The macros have been tested for all SAS releases dating back to SAS 8.2
- Existing SAS V5 Transport file limitations are eliminated when the extensions are applied
- The new format specification is open, just like the existing SAS V5 Transport file specification

	V5 Transport File	Transport File w/Extensions
Field name size	8 characters	32 characters
Field name characters	Traditional alphanumeric	Any character other than null (‘00’x)
Field label size	40 characters	256 characters
Character value size	200 bytes	23,767 bytes

WHAT DOES THIS MEAN FOR INDUSTRY AND FDA?

- FDA and biopharmaceutical companies can continue to use the existing SAS V5 Transport files with no change in their current processes
- As of 19-October-2012, organizations can download the new macros from support.sas.com, along with installation and use instructions (http://support.sas.com/kb/46/944.html)
- Value of the new capabilities will be limited until
  - Industry and FDA integrate these new macros into their business processes
  - Other software products (JMP, JMP Clinical, jReview) integrate these new macros
- Similar limitations will apply for any new delivery structure, but these limitations are easily addressed for the Extended SAS Transport file format given the systems already in place

EXTENDED SAS TRANSPORT FILE AND CDISC OPERATIONAL DATA MODEL (ODM)

- The CDISC Operational Data Model (ODM) and related technologies strongly align with existing investments by industry and FDA, and represents the best long-term option for delivering data to FDA
- The new SAS Transport file extensions can bridge the gap between
  - Using the older SAS V5 Transport format with its recognized limitations
  - Newer alternative such as the CDISC ODM structure

BILL GIBSON
BILLJ.GIBSON@SAS.COM
919 531 0996

Solutions for Study Data Exchange Standards

Melissa Binz MS, Global Head Clinical Data & Reporting Standards

Peter Mesenbrink PhD, Director, Statistical Scientist

Silver Spring, Maryland

:012

NOVARTIS

Outline of discussion topics

- Controlled terminology management
- Development and management of disease-level clinical data standards in collaboration with the FDA
- Communication plan across Review Divisions to ensure alignment on data standards requirements
- Frequency of communications/meetings between FDA and Sponsors on data collection/exchange standards during Drug Development Life Cycle
- Expectations for data conversion during the 5-year transitional period

Process for controlled terminology management critical

- Over time, the controlled terminology used in clinical trials will continue to evolve as the number of disease-level clinical data standards available increase
- Disease-level clinical data standards will also be updated over time as they are used in clinical trials across different compounds from different sponsors
- Evolution of controlled terminology needs to be managed with impact assessments to prior terms and data pools. This might include use of synonyms in NCI Enterprise Vocabulary Services (EVS) or the creation of mapping expectations from old to new terminology.
- Clear communication on the version to be used should also be readily available

Development and management of disease-level clinical data standards in collaboration with the FDA

- With clinical data standards released in a staggered manner, not all diseases will have clinical data standards that have been defined by CDISC or the FDA.
- Flexibility is requested from the FDA to work with Sponsors as clinical data standards emerge and evolve
  - This will be critical in therapeutic areas where there are large number of Orphan Diseases
  - Heightened collaboration with Industry is requested to allow for the development of flexible clinical data standards that will allow for the rapid development and approval of novel drugs and expanded diversity of research into new disease targets to save lives and improve quality of life

Development and management of disease-level clinical data standards in collaboration with the FDA

- As new versions of data standards become available
  - Announcements of public review of new and revised Data Standards should be made through the FDA Data Standards website
  - A clear timeline for the expectations for transitioning to a new version of a data standard
    - 12 months if the data standard is not currently in use
    - 24 months if another version of the data standard is currently in use.
  - Any new versions of data standards should be backward compatible if it is necessary to analyze data in an early version of the standard
- Metrics should be developed that will show what data is being used by Medical and Statistical Reviewers so that future versions of the data standard will focus on the data necessary to be collected to demonstrate that a drug is safe and effective.

Communication across Review Divisions critical to ensure alignment on data standards requirements

- As the number of biological compounds in drug development ↑, the number of diseases evaluated with a compound has also ↑
- Decisions on Data Standards implementation by the FDA need to be collaborative effort between Industry and Sponsors
  - Clear and transparent public review by Industry should occur before implementation
  - Careful consideration of the impact of a data standard that may be in use across Review divisions for diseases that are often evaluated in parallel (e.g. plaque psoriasis and rheumatoid arthritis for many monoclonal antibodies)
- Once a decision is made, the communication of this decision and timeframe of implementation should be made clear through the FDA Data Standards website.

Open communications between Sponsors and FDA on data standards during drug development will be key

- The following meetings would be helpful as data standards are developed and evolved during the 5-year transition period
  - Review of the initial proposed package of data standards to be used in human studies and the data standards used in non-clinical studies to determine if any updates are required
  - Any updates to the data collection standards package to be used in Phase III studies and samples of the structure of SDTM and ADaM data sets used to date at the time of the End-of-Phase II meeting
  - At the time of Pre-NDA/BLA meeting a walk through of the data submission package along with Define.xml to confirm acceptability and allow for any updates to be made prior to submission
- These meetings should be separate from the other regulatory meetings at these time points since there will be differences in meeting attendees.

Alignment on data conversion will be key during transition

- During the transition to a common data exchange standard, the expectations for conversion of legacy studies should be clearly communicated.
- Any conversions from non-CDISC to CDISC standards should be discussed in advance and determination made if both legacy and converted versions of the data need to be included in the dossier
- Only those clinical and non-clinical studies that provide data supporting regulatory approval should require conversion while other data should submitted in its original format.
- Meta-analyses can be considered if there is need to evaluate safety risks across all the subjects exposed to a particular compound.
- Any conversions across CDISC-compatible data standards versions will need to be backward compatible.

Thank you for the opportunity to speak on behalf of Novartis

UAMS

UNIVERSITY OF ARKANSAS
FOR MEDICAL SCIENCES

Study Data Exchange and Open Source Ontologies

M. Brochhausen, W.R. Hogan

Division of Biomedical Informatics

Problems with RxNorm

- does not keep history of National Drug Codes

  ⇒creates problems with legacy data

- historically version control has been poor; has been optimized recently
- use-mention confusion

  ⇒Vicodin: trade name vs. drug product

Problems with NDF-RT

- fundamental scientific mistakes
  - wrongly indicates oral vancomycin for bacterial endocarditis and intravenous vancomycin for pseudomembranous colitis
  - Why? NDF-RT wrongly attributes to individual molecules the therapeutic properties of drug products
- errors in logic
  - wrongly attributes relations that do not always hold (may_treat)

Proposal

The problems highlighted previously need to be solved.

Therefore, we propose that the FDA encourage and foster the development of logically robust and scientifically exact resources to support or optimize pre-existing standards.

This will help overcome existing problems and enable more extensive data integration.

Specifically, we suggest to consider supporting the use of **open source ontologies**.

Semantically rich ontologies

- representation of types of things and relations in a given domain
- specify and unify semantics
- optimally encoded in Web Ontology Language 2 (OWL2)
- logical models

Semantically rich ontologies

⇒ allow semantically richer queries

⇒ foster integration of heterogeneous data

- ⇒ key for translational science due to granularity issues

⇒ enable reasoning over data

- ⇒ checking consistency
- ⇒ discovering implicit knowledge

Open source ontologies

- Benefits:
  - iteratively improved through peer review
  - incorporate state of the art
  - low barriers for adoption, no cost
- Challenges:
  - need to encourage development of ontology-driven systems

Open source ontologies

- Best practices
  - open
  - use of URIs
  - versioning
  - delineated content
  - textual definitions
  - well-defined relations
  - collaborative development

(http://www.obofoundry.org/crit.shtml)

Ontology-based forms

- Subsequent annotation of biomedical data has been proven to create obstacles to data integration.
- Biomedical data, e.g. in clinical trials or drug development, should be captured in a way that immediately captures the data in accordance with the ontology.
- To achieve that we propose to research and evaluate ontology-based biomedical forms.

Ontology-based forms

- Examples (2-4 are work in progress):
  1. European Union: ObTiMA (Ontology-based questionnaires for clinical trials)
  2. UAMS: Ontology- and RT-based demographics for clinical data
  3. UAMS: Ontology-based patient questionnaires for translational studies
  4. UAMS: Ontology-based entry form for biomaterial repository

In conclusion

We advocate supporting development of data exchange standards based on existing open source ontologies.

We strongly suggest to desist from premature commitment to a restricted set of standards to allow scientific competition.

<  AnIML >

Analytical Information Markup Language

< /  AnIML >

Gary W. Kramer

We got TROUBLE, my friends, right here in Science City and that starts with T, and it rhymes with D, and it stands for DATA…

- Can't Easily Move It
  - From Instrument to Instrument
  - From Instrument to Application
  - From Application to Application
- Can't Interconvert It
- Can't Find All its Parts
- Can't Look at It Easily
- Can't Use It Easily with Modern Computing and Networking Technologies

ASTM INTERNATIONAL
Standards Worldwide

There is Crisis in Archiving and Retrieving Data

- We’re drowning in data, yet increasingly we cannot find our stuff.
- Data mining increases value of archived data.
- Regulatory agencies are demanding increasingly long data retention times.
- In terms of retrieving archived data, we were better off with paper forty years ago before the arrival of lab computers…
- We can still read the data in Newton's notebooks today. Will folks be able to read ours in 100, 20, or even 5 years?

What are the AnIML Standards?

- The AnIML Standards are projects of ASTM Committee E13 on Spectroscopy and Separation Science, Subcommittee E13.15 on Analytical Data.
- The Analytical Information Markup Language (AnIML) is a markup language based on Extensible Markup Language (XML) that provides a structure, syntax, and format for recording, reporting, exporting, importing, and archiving the result data and result metadata outcomes from analytical chemistry experiments.

What are the AnIML Standards?

- AnIML Provides Structures to Organize Result Data and Metadata from Simple to Complex Analytical Chemistry Experiments.

- AnIML Supports Fully Multi-detector, Multi-dimensional, Multi-sample, Multi-step, Multi-sequence Analytical Procedures in a Single File Format.

- AnIML 1.0 Supports Sample Data, Raw Analytical Data, Data-Acquisition Parameters, Processed Analytical Data, Data-Processing Parameters, Trace Forming, and Peak Tables.

Some AnIML Files are Simple

ASTM INTERNATIONAL
Standards Worldwide

Some AnIML Files are Complex

Key AnIML Features

- Extensible – Supports New Vendor and User Requirements
- Verifiable – Supports Digital Signing of Full and Partial Documents to Ensure Data Integrity
- Traceable – Supports Audit Trails to Track Data Alterations and Create a Dataset "Chain of Custody"
- Validatable – Permits Checking of Datasets Against Data Requisites from Multiple Data Authorities
  - XML Syntax and Schema Language
  - AnIML Schemas and Technique Definitions
  - AnIML Naming and Design Rules
  - Semantic Rules
  - Data Requisites from Higher Data Authorities

AnIML 1.0 Supported Techniques

- Chromatography and Other Flow Techniques
  - GC Point Detectors – FID, TCD, ECD, NPD, FPD
  - LC Point Detectors – RI, ELS
- UV-Vis
- IR
- 1D NMR
- MS
- Combinations of the Above Techniques (Hyphenated Techniques)
- Indexed Experiments Using the Above Techniques
  - Kinetic Runs
  - Temperature or pH Profiles
- Processing Techniques
  - Trace Forming/Peak Finding
  - Cross-Cutting
  - Peak Table Formation

AnIML Components

- AnIML Core Schema
  - One Schema
  - Maintained by ASTM E13.15
- AnIML Technique Schema
  - One Schema
  - Maintained by ASTM E13.15
- AnIML Technique Definition Documents
  - One or More Instance Documents per Technique Created from the Appropriate Base Technique Definition
  - Maintained by ASTM E13 or Appropriate Domain-Expert Organization
- AnIML Technique Definition Extensions
  - One or More Instance Documents per Technique
  - Maintained by Vendor, Organization, User, or Whomever Extends the Technique
- AnIML Result Data Files

AnIML: Where We're At for Version 1.0

- AnIML Core Schema
  - Complete
  - Frozen
  - Internally Documented
- AnIML Schema NDRs
  - Complete
- AnIML Technique Schema
  - Complete
  - Frozen
  - Internally Documented
- AnIML Technique NDRs
  - In Progress
- AnIML Technique Exemplar
  - In Progress

AnIML: Where We're At for Version 1.0

AnIML Technique Definitions

- UV-Vis Technique Definition – Complete
- Chromatography Technique Definition – Complete
- Mass Spec Technique Definition – Started
- 1D NMR Technique Definition – Started
- IR Technique Definition – Started
- Point Detector Technique Definitions – Mostly Complete
- AnIML Miscellaneous Technique Definitions
  - Indexing – Complete
  - Cross-Cutting – Not Started
  - Trace Forming – Not Started
  - Peak Table – Not Started
- SEDD 4 Demo – Complete for GC-MS Data

AnIML: Where We're At for Version 1.0

AnIML Applications and Tools

- AnIML Data Viewers – Several Complete
- AnIML Extensible Validator – Complete
- Simple AnIML File Writer for Agilent 8453 UV/Vis – Complete
- Simple AnIML File Writer for PE Lambda xxx UV/Vis – Complete
- JCAMP-DX to AnIML Data Converters – Several Complete
- XSLT Documentation Extractor for Schemas – Complete
- Documentation Extractor for Technique Definitions – Complete

AnIML: Where We're At for Version 1.0

AnIML Documentation & Draft Standards

- AnIML Standard Practice – Complete
- Guide to the AnIML Schemas – Under Contract
- Guide to the AnIML Technique Definitions – Under Contract
- Guides for Individual Techniques – Not Started

Getting Started with AnIML Now

- Check Out the AnIML Website
  - http://animl.sourceforge.net
  - http://www.animl.org
- Delve into the Internally Documented Schemas
- Examine the AnIML Schema NDRs
- Check Out the Technique Definition Docs and their Documentation
- While There May Be Some Changes Before Vers. 1.0, They Likely Will Be Minor or Organizational

AnIML Players & Sponsors

- Agilent Technologies USA
- Agilent Technologies GmbH
- Waters Incorporated USA
- Waters Informatics GmbH
- PerkinElmer USA
- PerkinElmer UK
- GlaxoSmithKline
- Amgen
- Bristol Meyers Squib
- BSSN Software
- Scimatic Software
- U.S. Environmental Protection Agency
- U.S. Army Corp of Engineers
- National Institute of Standards and Technology
- ASTM International
- International Union of Pure and Applied Chemistry

ASTM INTERNATIONAL
Standards Worldwide

Acknowledgements

Mark Bean – GlaxoSmithKline
Maren Fiege – Waters Informatics GmbH
Burkhard Schäfer – BSSN Software
Stuart Chalk – University of North Florida
Dale O'Neill & Michael Leiber – Agilent Technologies
Adam Patkin – PerkinElmer
Jamie McQuay – Scimatic Software
Robert Lancashire - The University of the West Indies, Mona
Peter Linstrom – NIST
Tony Davies – IUPAC Subcommittee on Electronic Data Standards
Reinhold Schäfer – Fachhochschule Wiesbaden

- Alex Rühl
- Martin Peschke
- Aykut Arslan
- Dominick Pötz
- Anh Dao Nguyen
- Alex Roth
- Ronnie Jopp
- Patrick Gleichmann
- Kordian Placzek
- Frank Masur
- Dennis Backhaus

Solutions for Study Data Exchange Standards

White Oak Campus

November 5, 2012

Charlie Mead
Co-Chair, W3C Healthcare Life Sciences Working Group
Director, Healthcare IT, Octo Consulting Group
Eric Prud'hommeaux
Semantic Web Engineer, W3C
Research Scientist, Massachusetts Institute of Technology

Goals of data/meta-data exchange

- Understandable by humans
- Understandable by machines
- Analyzable in a single-study context
- Analyzable across multiple studies
- Re-purposeable beyond collection context

What is required of “Standards”?

- Domain/Application Relevance

- Semantic expressiveness
  - Concepts
  - Relationships/context
  - Layered representation
    - Information models → meta-data, relationships, context
    - Data or terminology models→ instance-level bindings
    - Data type models → complex semantics beyond “standard” data types

How are “relevant standards” developed?

- Bottom-up development ensures concrete instantiation and situational relevance
  - Can lead to development of multiple silos
- Top-down development ensures maximum breadth and coherence
  - Usually very time-consuming
  - Often results in broader scope then required in many contexts
  - Often requires inordinate amount of time on “edge cases”
- Synthesis of the two approaches provides maximum practicality and reuse
  - Operationally difficult from methodology/tooling perspective

Barriers to cross-standard interoperability:

- Technology
  - Differences in implementation technologies, e.g. proprietary databases
- Syntactic/Lexical
  - Differences in transport serialization, e.g. difficulty of combining XML schemas
- Semantic
  - Difficulty in computationally detecting “semantic equivalences”

Semantic Web Value Proposition

- Common representation between “design-time” and “run-time” (*graph patterns*)
  - Eliminates “impedance mismatch” that normally occurs between these two contexts
- Emphasis on “pure semantics”
  - Reduces technology barriers
    - SPARQL end-points (a broadly-applicable standard)
  - Reduces syntax and lexical barriers
    - Simplified model and syntax → higher level of abstraction → immunity to serialization brittleness

Semantic Web Value Proposition (2)

- Tools and infrastructure designed to promote reuse, discovery, amalgamation, harmonization of data and meta-data from multiple contexts
  - "Point-wise" semantic harmonization
  - Decentralized extensibility
    - Reuse made easy: "Interoperability emerges from laziness…"
- Significant amount of life sciences data already available
  - Linked Clinical Trials (http://linkedct.org)
  - Open Linked Data (http://linkeddata.org)
  - Linked Structured Product Labels (http://purl.org/net/nlprepository/linkedSPLs)

HL7 and CDISC

- Both organizations have developed standards that attempt to support data/meta-data
  - collection
  - submission
  - analysis
  - re-purposing
    - cross-study exchange

HL7 and CDISC Standards

- However, both HL7 V3 messages or CDA documents and CDISC standards (SDTM or ODM) fall short of achieving the goals of standards for data exchange and re-purposing
  - At minimum, these standards currently have issues with XML serialization brittleness →
  - limited cross-study utility →
  - need for “top-down” governance to explicitly specify representation and encoding choices that →
  - limit innovation and inherent “new study” variability

W3C Proposal

- A general approach to the use of *multiple standards* that *collectively support standardized data/meta-data exchange and re-purposing*
  - Standards-as-is*
  - Standards-in-context*
  - Interoperability across standards and the data collected using them*

-- *3-phase project plan of CDISC2RDF project

1. Standards-as-is

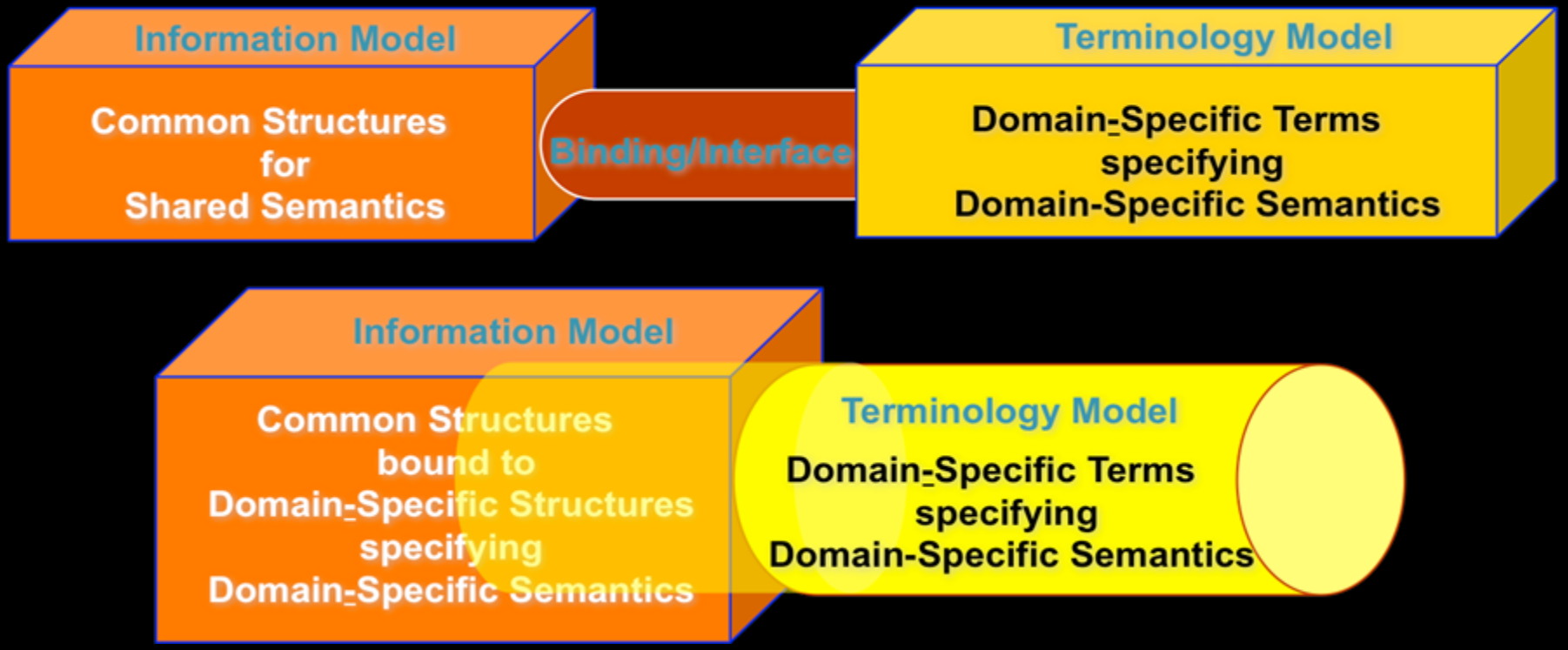
- Use Semantic Web standards to define standards for meta-data/data exchange
  - URIs for all elements in a given standard
  - Expressed using RDF and SKOS
  - Standards become addressable and queryable
- A substantial amount of the work has been done…
  - HL7 “O-RIM”
  - HL7 FHIR
  - BRIDG 3.2
  - SNOMED-CT
  - NCIt
  - CDISC2RDF

2. Standards-in-context

- Link Semantic Web expressions of one standard to those of another
  - Maximize cross-standard utilization based on semantic equivalences
  - Focus on cross-standard semantic incompatibilities as targets for semantic harmonization

3. Interoperability-across-standards

- Example: a “TermInfo” example
  http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/TermInfo



Solutions for FDA Data Exchange Standards

- W3C Proposal: Semantic Web-enabled standards and submission format → cross-trial focus on semantic equivalence → enable re-purposing

- Example: CDISC2RDF
  - RDF representations of all CDISC standards
    - schemas, models, and transform processes for CDISC content standards
    - completion by Q4 2012
    - ODM not yet included but sh/could be

- Example: HL7 FHIR
  - REST interface (default)
  - RDF expression of RIM bindings→ SPARQL endpoints → data interchange with other Semantic Web-enabled standards → CDISC SHARE, ODM, etc.

Summary

- A Semantic Web-based approach to standards development
  - All standards represented using RDF and/or SKOS
    - Supports collaborative (top-down) development *and*
    - Supports context-grounded (bottom-up) development
- A Semantic Web-based approach to standards use means:
  - Data/meta-data (including relationships) transmitted using Semantic Web standards
- Getting started
  - Meta-data represented using Semantic Web standards
    - e.g. represent ODM in RDF
  - Submit first study in RDF
    - The Easy Path has started…"Interoperability emerges from laziness…"

HL7 Study Data Standards Project

Armando Oliva, M.D.

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

November 5, 2012

Limitations of SAS XPT v5

- character limitations for variable names, labels, character fields
- Flat, two-dimensional data structure for hierarchical, multi-relational (“round”) data
- HL7 v3 provides a promising alternative for
  - transporting CDISC content and
  - addressing current information exchange limitations

HL7 – Increasing Use in the U.S. Federal Government for Clinical Information Exchange

HHS/ONC/CMS: HL7 standards specified in meaningful use regulations

- HL7 CDA (Clinical Document Architecture)
  - A v3 standard
- HL7 v2
- Certified EHRs – certifying criteria based on HL7 EHR-S FM (Electronic Health Record-System Functional Model)

HL7 Use (Cont'd)

FDA: SPL; ICSR; RPS; ECG Waveform; eStability

CDC: HL7 CDA in use for National Health Safety Network (NHSN), …

NCI: Use of CDA in building CRFs for oncology trials

NLM: CTR&R (Clinical Trials Registration and Reporting)

DoD and **VA**: v3 Services

AHRQ: CDA

HL7 Study Data Standards Development Project

- An FDA-sponsored project within HL7 to **harmonize** CDISC study data content standards with HL7 v3 exchange standards
- FDA continues to hear from stakeholders that harmonization of standards across SDOs is important
  - This is one example

HL7 Study Data Standards Project

- Three related standards:
  - Study Design Structured Document (Protocol; what will be done?)
  - Study Participation (Who is involved?)
  - Subject Data Implementation Guide (I.G.) for CDA Release 2 (What was observed?)
- All three designed to transport existing CDISC content and additional high-level FDA study information exchange requirements (captured as storyboards or scenarios)
- All three support exchange of “round” hierarchical, multi-relational data

CDER's Position on the HL7 Study Data Standards Project

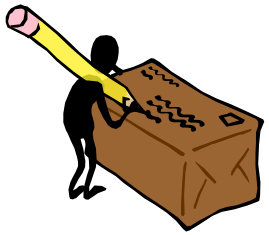
- This is a Data Standards Research and Development (**R&D**) Activity
- **No** policy decisions have been made regarding adoption or implementation
- Position statement posted publicly on fda.gov
  - http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm269946.htm

CDER Position

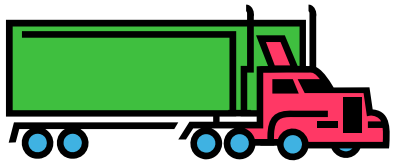
- Ongoing and increased CDER support for CDISC standards for study data content
  - 2004: SDTM
  - 2011: SEND
- Currently, CDER supports the submission of CDISC content using SAS Transport File (XPT) v5 exchange format
  - Limitations of this file format, and desire to move away from this format
- HL7 v3 XML exchange standard is a promising alternative to replace XPT v5.

CDISC Content **AND** HL7 Exchange

Now

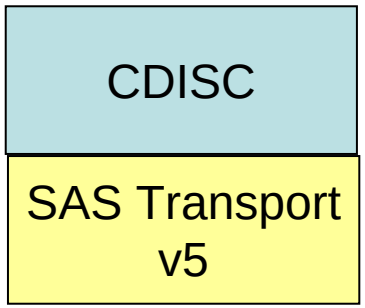


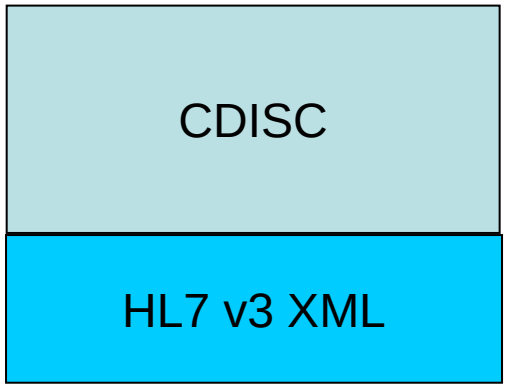
CDISC
+
HL7



Content

Exchange



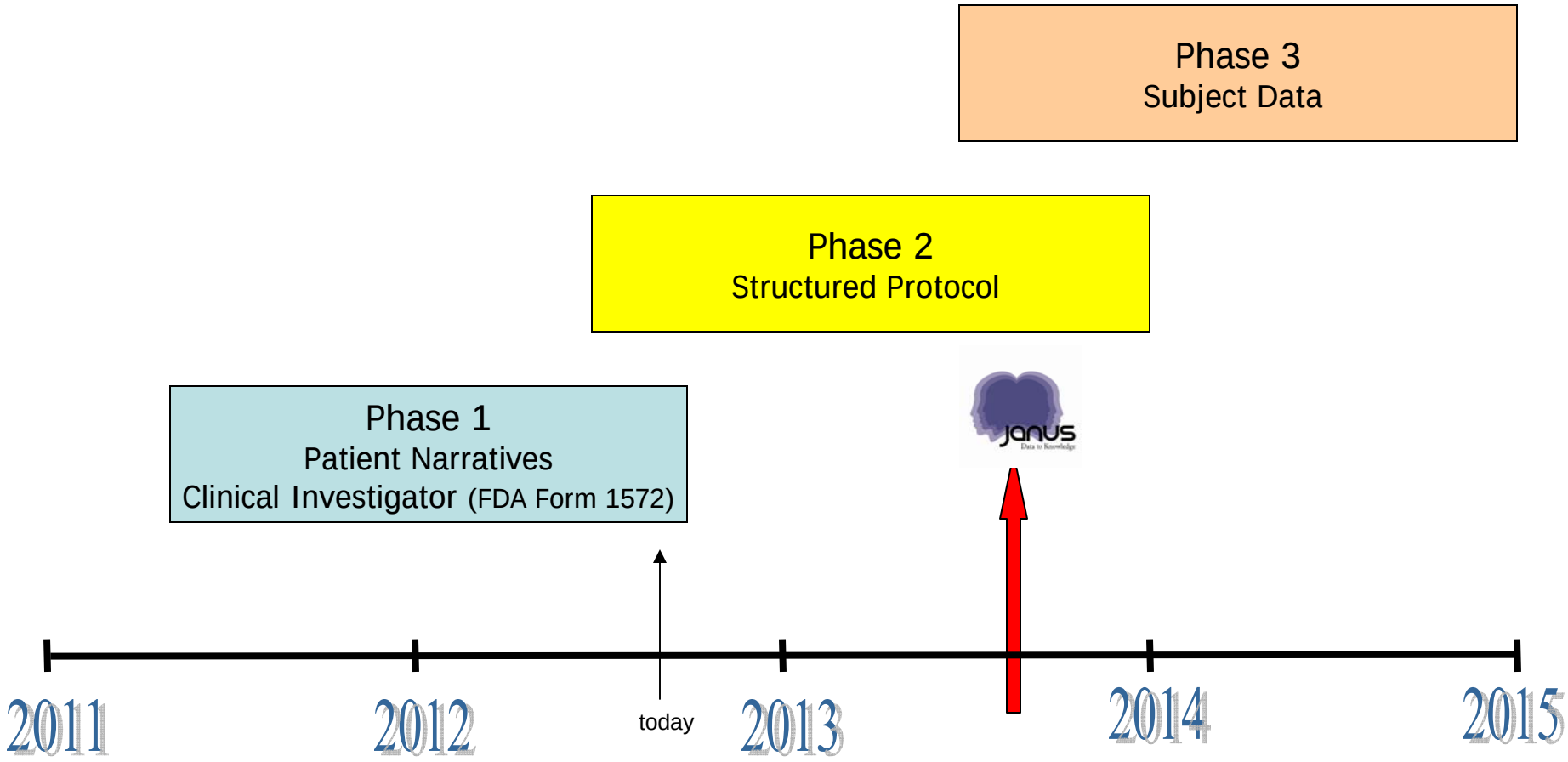


HL7 v3 for data exchange supports more content (i.e. larger container), improved semantic interoperability; resulting in better, more efficient data interpretations and analyses by the recipient

Proof of Concept

- CDER conducting a proof of concept to test HL7 Study Data Standards (R&D activity)
  - In parallel with ongoing robust CDISC implementation efforts and Therapeutic Area (TA) standards development
- Incremental Testing:
  - Phase 1: Patient Narrative / Clinical Investigator Info (1572)
  - Phase 2: Structured Protocol
  - Phase 3: Subject Data
- Testing will include opportunity for interested stakeholders to participate
  - Open and transparent
  - Through HL7/RCRIM “Stage 2” project team
    - Bi-weekly meetings, Wednesday 11am ET
    - Crystal Allard, PM (crystal.allard@fda.hhs.gov)

HL7 Study Data: Proof of Concept



Phase 1: Current Status

- FDA Form 1572 (Clinical Investigator Information) and related information
  - Testing Complete
  - Test Report being written
- Patient Narrative
  - Testing underway
  - Expect completion Q4 2012

Phase 2: Current Status

- Study Protocol content, as defined by ICH E6 guideline and CDISC SDTM Trial Design domains
- Testing initiated Sept 2012
  - Draft Study Design S.D. Implementation Guides complete for clinical and nonclinical protocols; to be shared Nov 2012 with Stage 2 team for comment
  - Currently we are developing data entry tool, data visualization tool for testing
  - Expect completion Q3 2013

Phase 3: Current Status

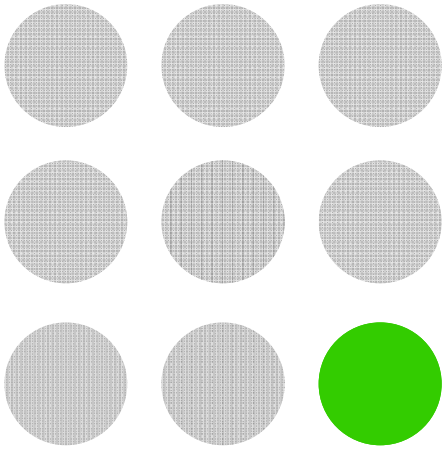
- Subject Data using CDA R2
- Testing not yet started
- Draft CDA Implementation Guides for clinical (SDTM) and nonclinical (SEND) data complete; to be shared Nov 2012 with Stage 2 team for comment
  - Support for “round” data
  - “Forward compatibility” – I.G. can accommodate new SDTM content as it is developed
  - I.G.s written using VA’s Model Driven Health Tool (MDHT): Conformance rules part of the I.G.

Conclusions

- CDISC standards describe the content for study data submissions
- SAS Transport File format v5 is used for transport
- We need a better transport format
- HL7 v3 XML is a promising alternative
- CDER is testing HL7 v3 XML while supporting current submission format
- No policy decisions regarding its adoption and implementation have been made

Medidata and Open Industry Standards

CDISC & CDISC ODM

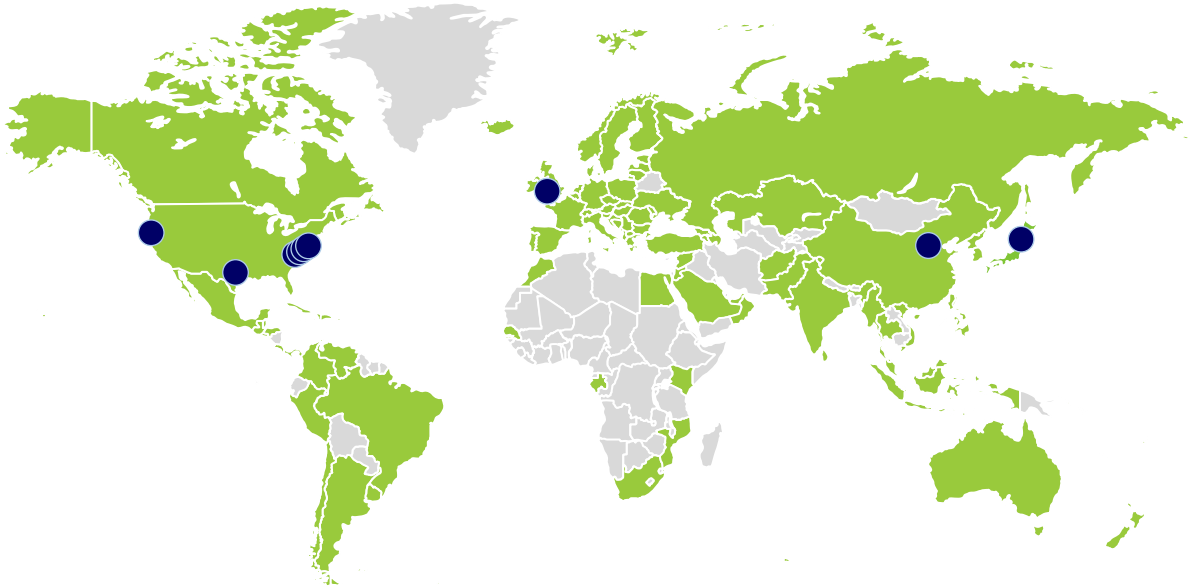


David Gemzik | Vice President, Implementation Services

medidata
Medidata Solutions Worldwide

Optimizing Clinical Trials:
Concept to Conclusion™

Global Footprint and Market Leadership



315+ customers

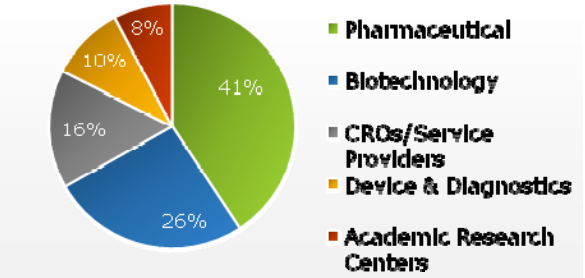
9 offices worldwide

890 headcount

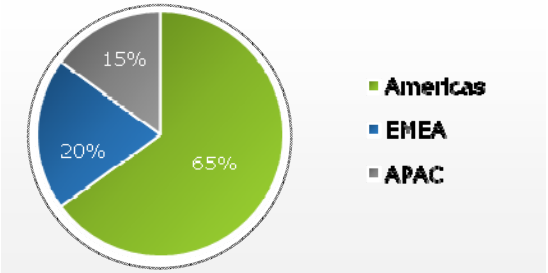
Live data infrastructure for over 3,000 studies, with over 2,500,000 subjects

Medidata Technology Used in Over 115 Countries

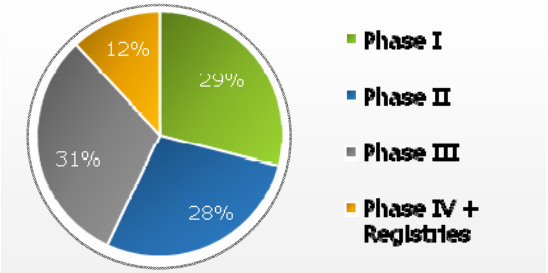
Customers by Industry Segment



Customers by Geography



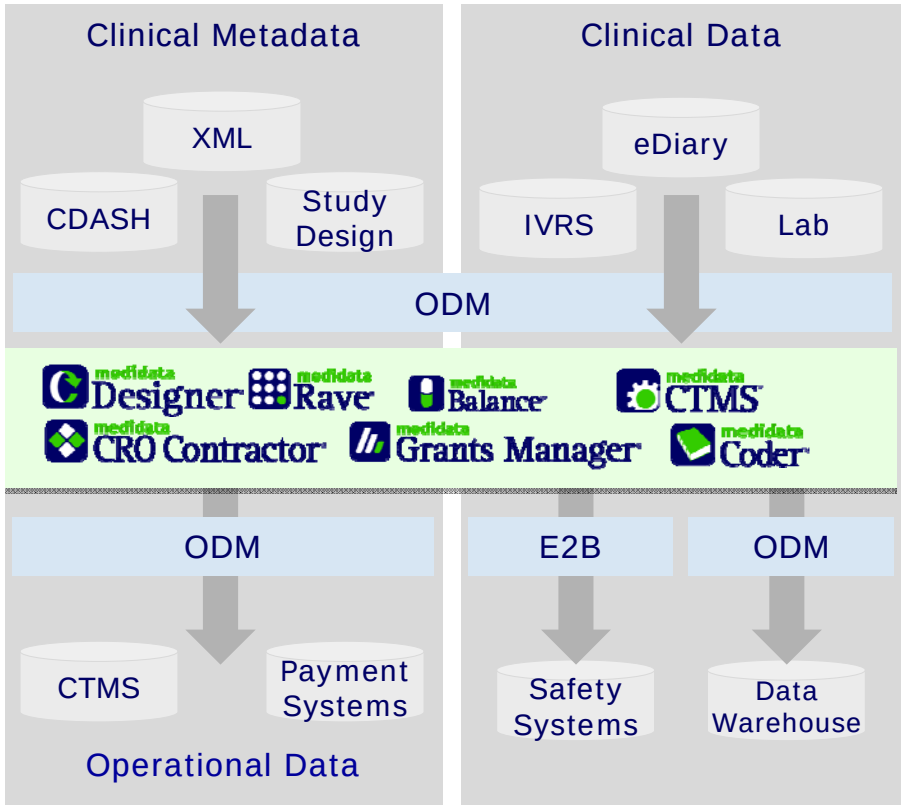
Studies by Trial Phase



Interoperability

Supporting Best-of-Breed Technology

150+ members on Developer Central



- CDISC Corporate Sponsor
- Support of all CDISC standards
- Robust integration toolkit for sponsors and CRO's
- Developer Central & technology partners program
- Experience integrating with a wide range of systems

CDISC Certified on all ODM Use Cases, CDISC PRM, CDASH

medidata
Medidata Solutions Worldwide

Medidata Standards Accelerator

medidata
Medidata Solutions Worldwide

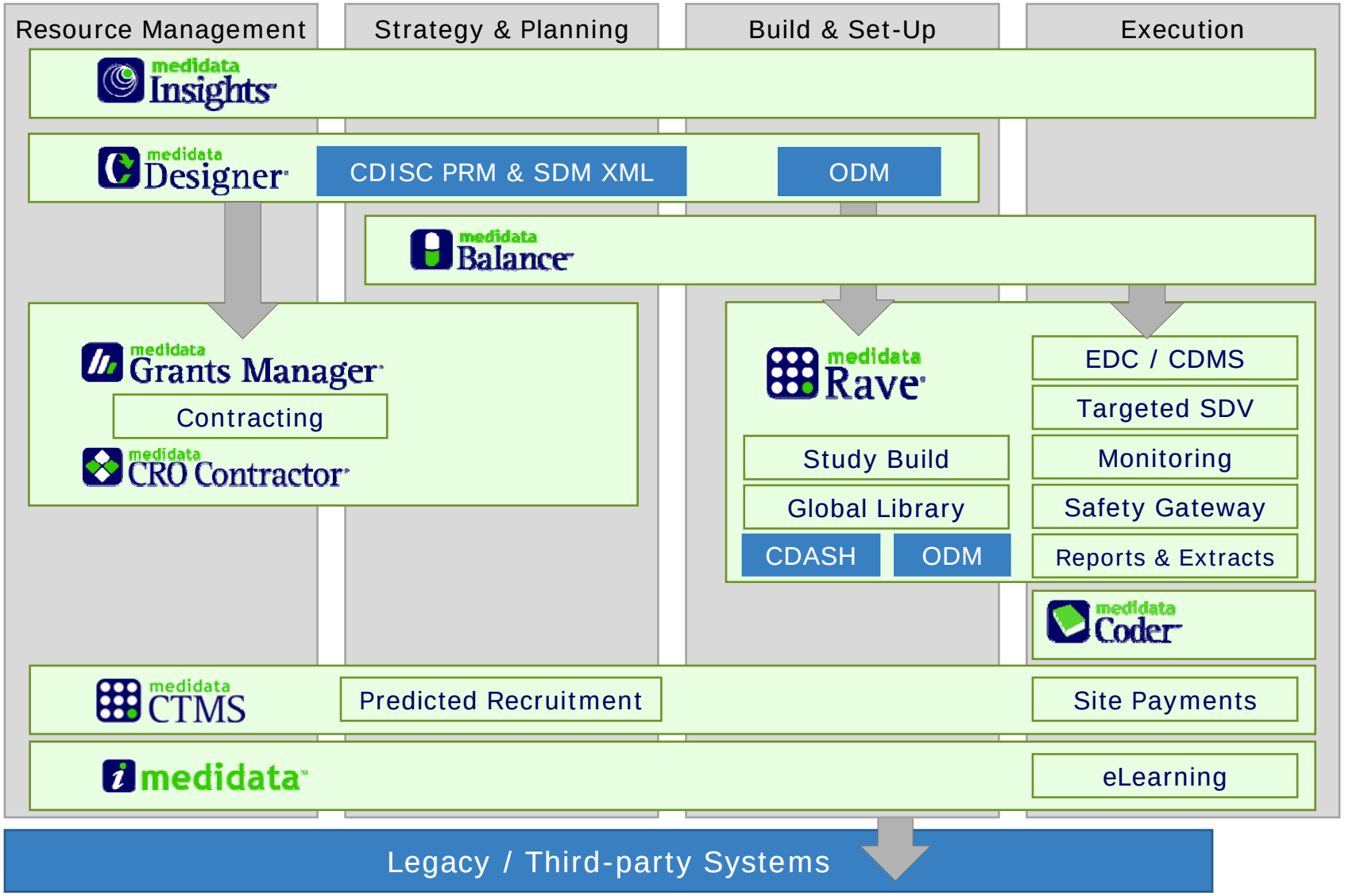
Reduced time to generate SDTM data	Faster study builds	Improved data quality	Simplified, expedited output to SAS
• Support for CDISC Controlled Terminology • CDISC Validation Services	• Repeatable processes drive continued improvement in subsequent trials	• Better data discipline • Fewer data entry errors	• 17 SDTM domains • SDTM validation • Define.xml

Web Services and CDISC ODM

- We provide our clients with validated tools needed to retrieve that data without dependency on Medidata.

- Rave Web Service (RWS) Integrations are enabled by a catalog of RESTful Web Service APIs that expose useful data to be consumed by a 3rd party, be it a user or another application. Medidata's platform API's are based around the CDISC Operational Data Model for both import and export of data, and E2B for Pharmacovigilance data. The API is real-time, synchronous, and fully leverages Medidata's robust role-based-access security model.

Medidata Clinical Cloud

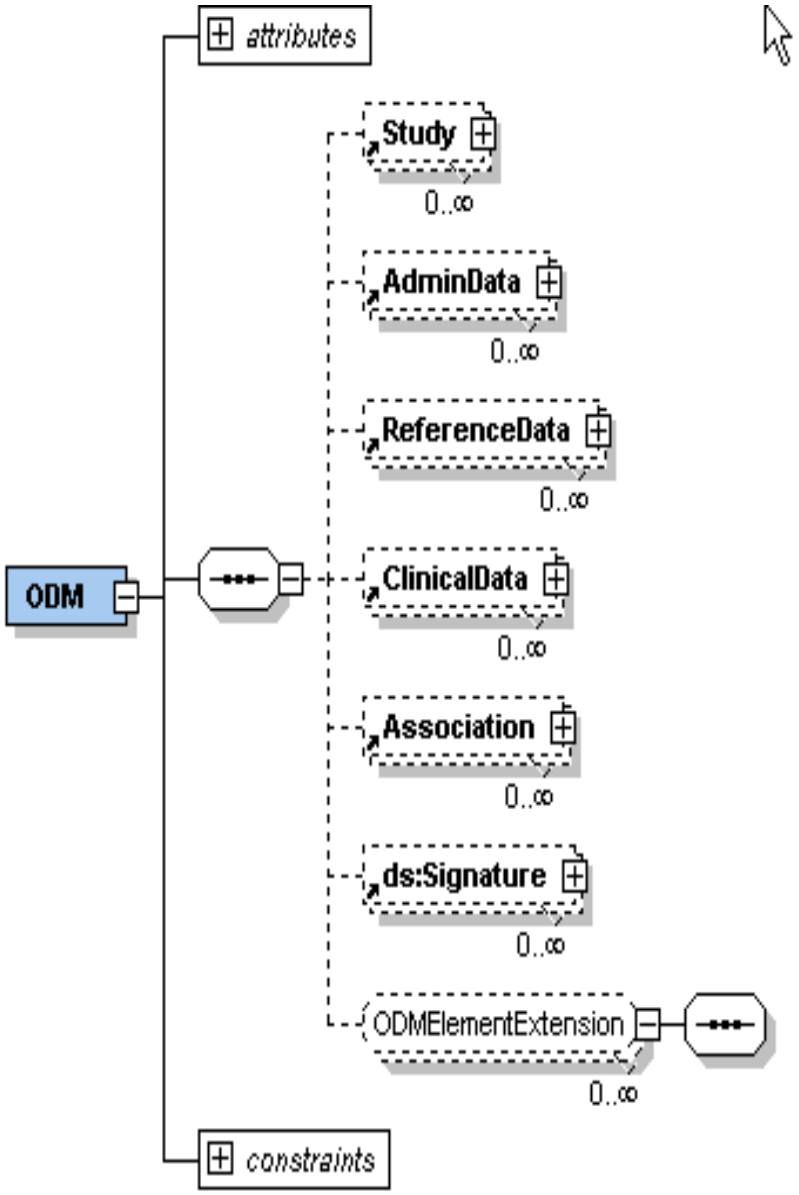


Using CDISC ODM for Study Data Exchange

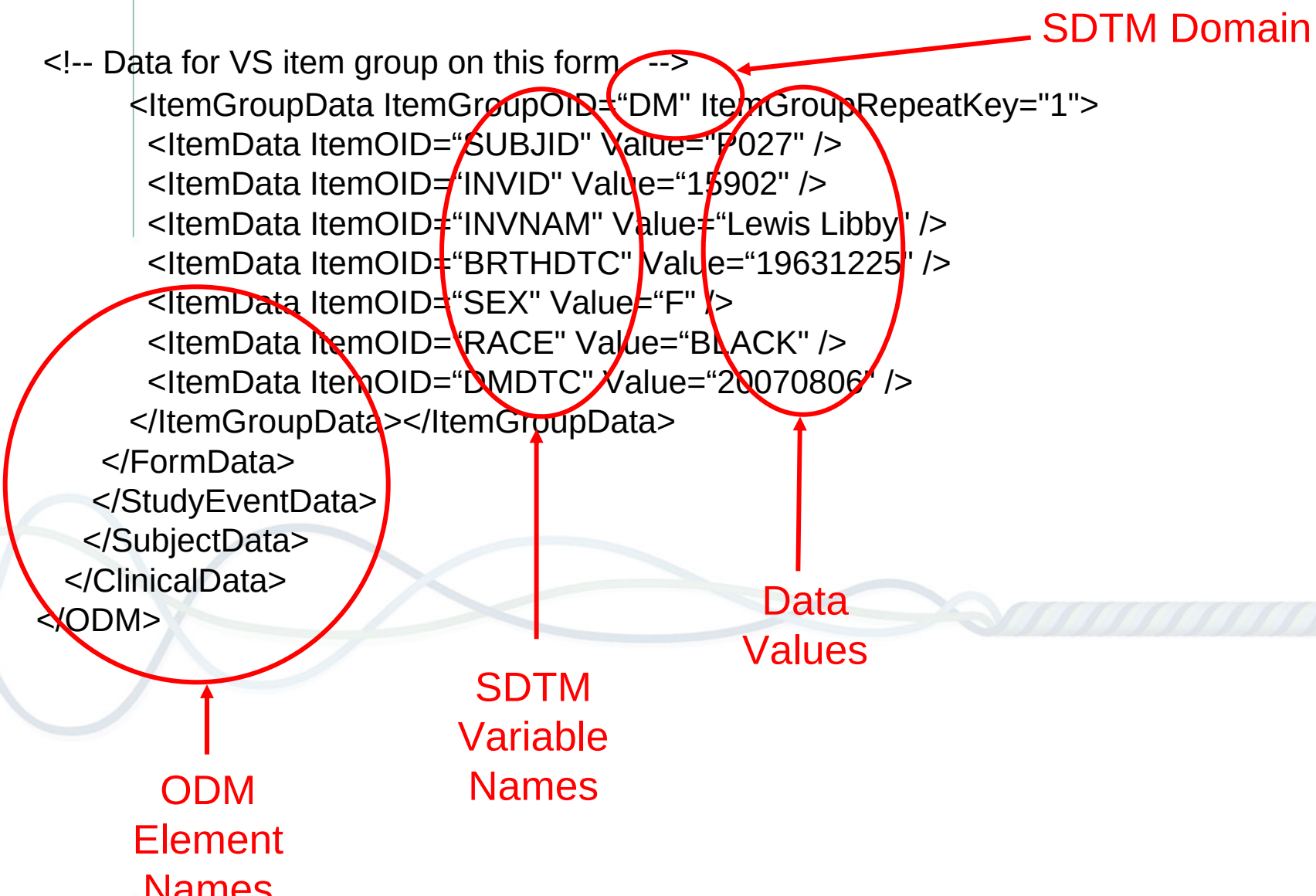
Wayne Kubick
CDISC Chief Technology Officer
Nov. 5, 2012

CDISC Operational Data Model

- Vendor neutral XML Schema for exchange and archive of Clinical Trials metadata and data: snapshots, updates, archives
- In global production use since 2000 – currently at v1.3.1
- Supports Part 11 compliance and FDA Guidance on Computerized Systems
- Hierarchical metadata structure: Study, protocol, events, forms, item groups, items
- Basis for Define-xml metadata description document used in submissions
- CDASH-ODM form metadata available
- SDM-xml represents BRIDG protocol/study design model (structure, workflow, timing)
- CT-xml delivers NCI-EVS controlled terminology
- Includes vendor extension capability
- Human and machine readable



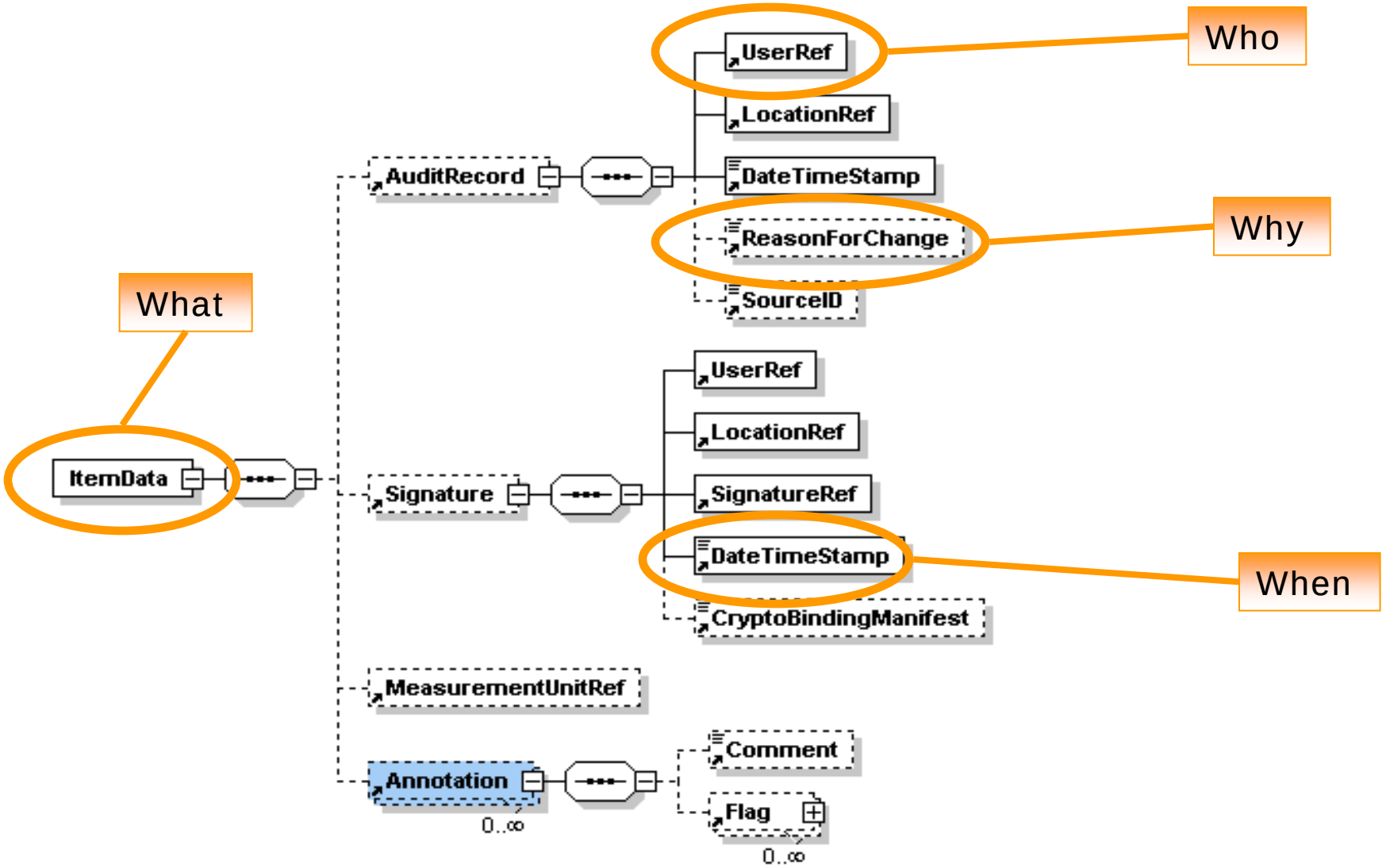
Example: SDTM Data in ODM



Benefits of CDISC ODM

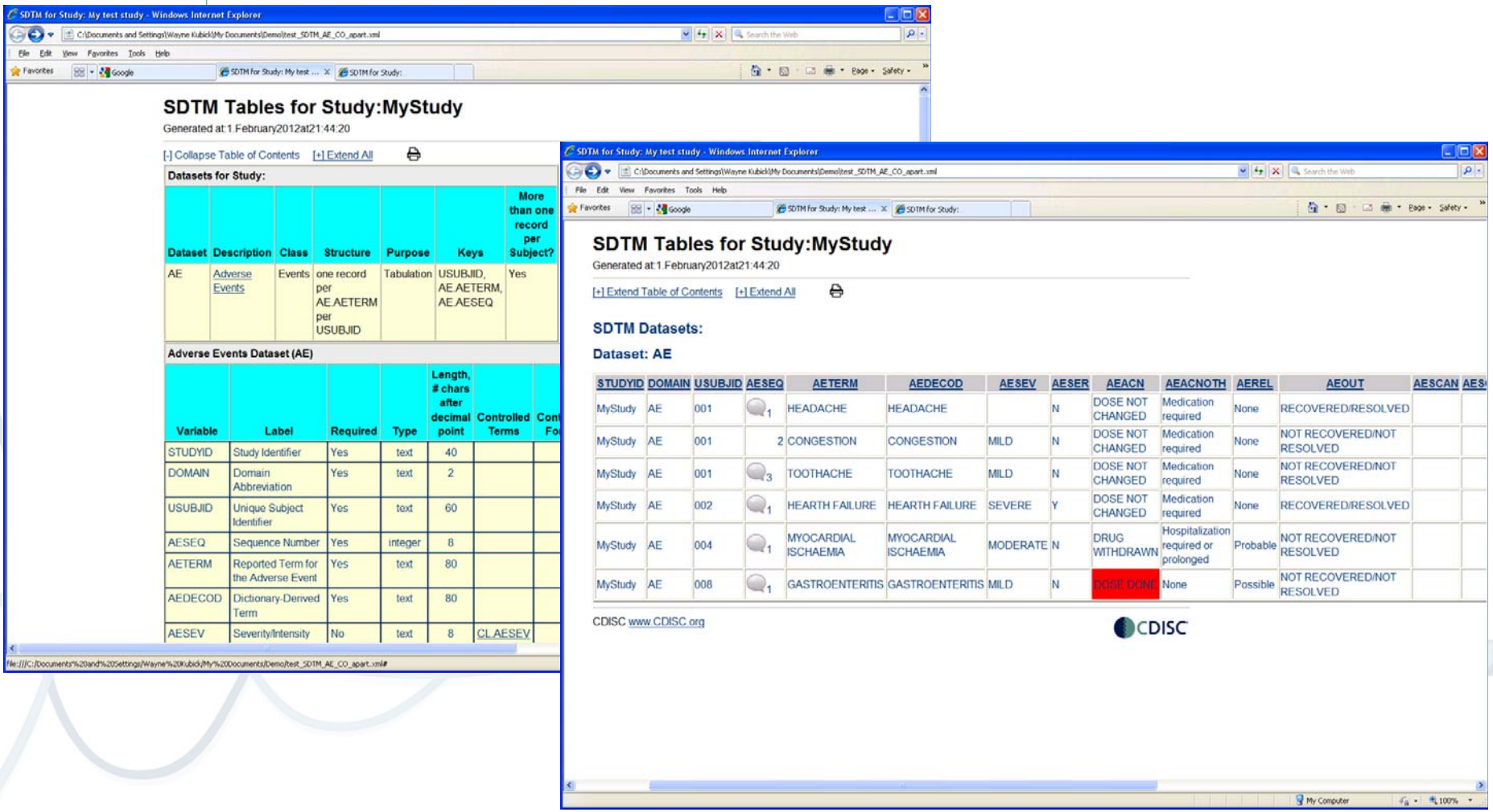
- Supports metadata-driven data transport with traceability from protocol through analysis
- Capitalizes on existing knowledge, business processes, tooling and investment, so it can be rapidly deployed more rapidly for submissions
- Complies with 21 CFR Part 11 and regulatory guidance
- Can faithfully reproduce research domain datasets, analysis files and eCRFs with audit trail
- Compatible with current define.xml metadata submission standards already accepted by FDA and CDISC controlled terminology in EVS
- Can be used with with HL7 C-CDA from EHRs to support research and OSI audit under an HHS sponsored interoperability spec
- Can represent more complex relationships between data events recorded per the research protocol.
- Can be easily extended to address new requirements as needed.
-

ODM & Audit Trail



Source: Dave Iberson-Hurst

Working with ODM Files



ODM can be viewed in CRF or table format with style-sheets
Flat files or SAS datasets can be easily extracted from ODM XML

March 13, 2007 Federal Register

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration

[Docket No. 2007N-0064]

Electronic Case Report Form Submission; Notice of Pilot Project

AGENCY: Food and Drug Administration, HHS.

ACTION: Notice.

SUMMARY: The Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) in the Food and Drug Administration (FDA) are seeking sponsors interested in participating in a pilot project to test the submission of case report form (CRF) data provided electronically in extensible markup language (XML) based on the Operational Data Model (ODM) developed by the Clinical Data Interchange Standards Consortium (CDISC). This pilot will test the ability of a new data format to support all

[[Page 11371]]

Strength *through collaboration.*

Solutions for Study Data Exchange Standards

Fred Wood, Ph.D.
Octagon Research Solutions/Accenture

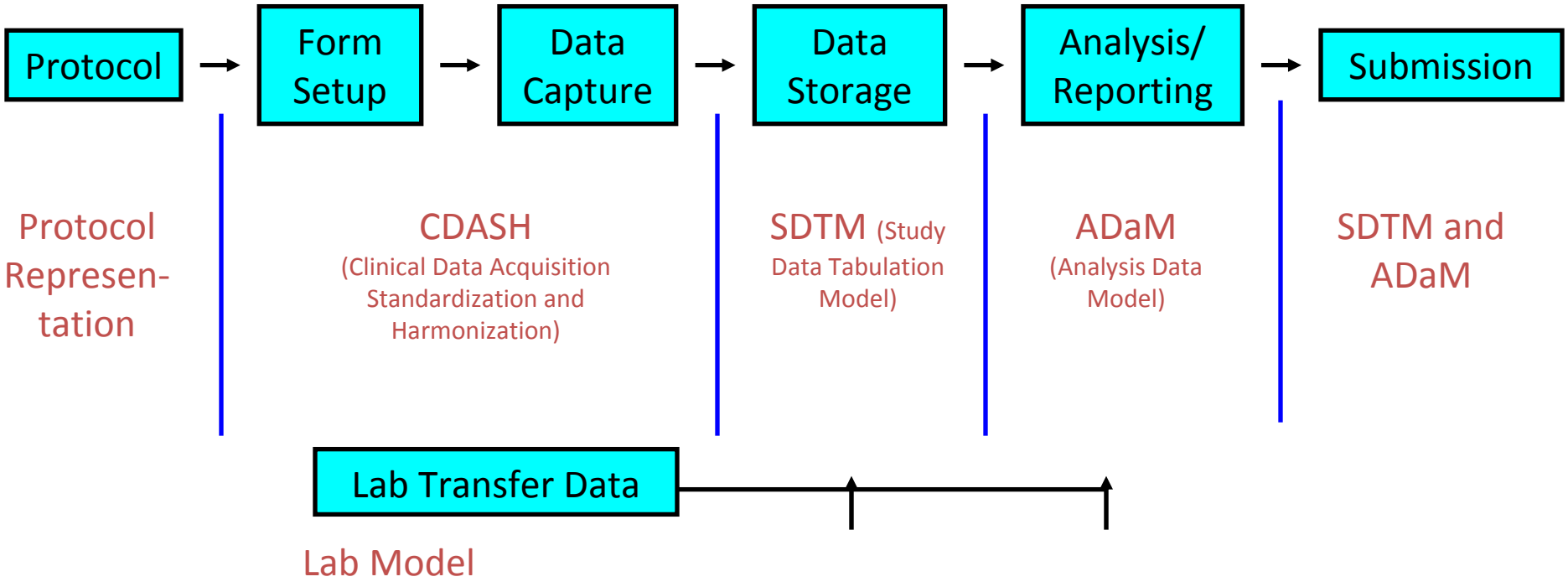
Question 1

- Please address each of the following areas:
- What opportunities/solutions exist to meet each challenge?

(a) Study design/set-up
(b) capture
(c) integration
(d) analysis
(e) reporting
(f) regulatory submission.

Response: Opportunities and Solutions Provided by CDISC Models

- The ODM has been demonstrated to be an efficient mechanism for reusing metadata across the data lifecycle.



Question 2

- How could FDA's regulatory requirements make the study data management process more efficient?

Response:

- Many deficiencies in SDTM-based submissions are related to poor data-collection and data-management practices.
- Deficiencies are easier to see with the submission of standardized data.
- Standards for submission (content or format) will not address problems that occur earlier in the clinical data flow.

Question 5

- Would Health Level Seven [HL7] v3 (e.g., messages, structured documents and Clinical Data Architecture) be a viable study data exchange standard? Please explain advantages and disadvantages. What would be the impact (e.g., financial, technical, or in terms of implementation or change in business processes).

Response: HL7 Standards (1)

- Both the clinical and nonclinical communities have much more experience and knowledge of CDISC standards than those of HL7.
  - The SDTM has been a Study Data Specification since 2004.
  - Since 2004, CDISC and its partners have trained thousands of people within the pharmaceutical industry on various CDISC standards.
  - CDISC standards have been implemented widely across the industry.

Response: HL7 Standards (2)

- Potential adverse consequences to the pharmaceutical industry associated with the implementation of HL7 standards:
  - A significant negative economic impact compared to the use of the CDISC ODM.
  - A slowing in the adoption of the above standards for companies who are “still waiting for the standards to stabilize.”
  - A delay in submission timelines as companies spend time developing and/or implementing technologies to change to an unfamiliar exchange format.

Response: HL7 Standards (3)

- The introduction of a whole new set of compliance and validation issues whose resolution make take tremendous amounts of time, resources, and money on the part of the industry as well as that of the Agency .

Question 8

- What would be a reasonable phased implementation period for each recommended exchange standard? And should supporting multiple, concurrent study data exchange standards be evaluated (please explain advantages and disadvantages of this approach)?
- What can FDA do to help industry to be more prepared for, or reduce burden of, a migration to a new study data exchange standard?

Response: Phasing

- This is a challenging question. My observations:
  - The support of multiple versions of an exchange standard needs to be evaluated.
    - This is almost mandatory, since neither industry or the FDA will be prepared for an instantaneous switch.
  - Phasing depends upon:
    - FDA's ability to support multiple versions of the same standard, and the infrastructure to support a new one.
      - FDA technologies for validation have generally lagged behind standards development.
    - Industry's ability to adopt standards with which they are familiar compared to standards with which they are not.

Response: Reducing Burden

- Any new exchange standard should:
  - Be relatively familiar
  - Be relatively easy to implement
  - Be consistent with systems and processes used for internal data flow
  - Have minimal economic impact
  - Not delay the data submission process

Solutions for Study Data Exchange Standards

Diane Wold, Ph.D.

GlaxoSmithKline

Short Term Solutions to SAS V5 Transport File Limitations

- ODM would be orders of magnitude easier than HL7 V3 for industry to implement
- Both are xml formats, so files would be much larger than currently. How will size issues be handled?
- ODM has audit trail capability. Is this needed/desirable?

Long Term Solutions to Relationships between Data Items

- This problem goes far beyond data exchange format
- Capture of relationships has been difficult and expensive in the past
- HL7 V3 has had limited adoption
- CDISC SHARE aims to provide rich metadata based on the BRIDG model and a robust datatype standard

Developing Solutions

- Requires partnership of FDA with industry
- Requires a full understanding of the problems to be solved
  - My comments have addressed limitations of SAS V5 and capturing of relationships
  - Are these the main problems?
  - Are there other problems?