Feedback from an SDTM Submission: A sponsor perspective

CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM

William J. Qubeck
February 1, 2005

CDISC℠

Setting the Global Standard for Clinical Data

Agenda

- Submission goals, characteristics, and metrics information
- Submission Challenges
- Lessons Learned
- Conclusions/Summary

SDTM Submission Goals

- To submit CDISC SDTM compliant data for a product (Aug 2004 & Jan 2005).
- 1st Submission Goals:
  - Provide the *key* safety and efficacy in SDTM
  - Develop the software to support simple mappings (e.g., many sources to one target)
  - To deliver the define.xml
  - To include the data within the original NDA with NO impact on the filing date.
- 2nd Submission Goals:
  - Expand *beyond* the key safety & efficacy data; with the inclusion of derived results.
  - Develop product level/generic software to support complex mappings (many sources to many targets).

Submission Complexity

- Submitted 5 studies:
  - 1st: 18 domains covered about 25 submission datasets (40% of the total study data).
  - 2nd: 22 domains covered about 49 submission datasets (80% of the total study data).
- Contained over 11,000 subjects worth of data.
- All 5 studies included were in parallel design:
  - 2 Blinded trials, 2 Pivotal trials, 1 Summary of Safety.

Submission Metrics

- Resources:
  - eSub programmers: 1st 5 & 2nd 3 (fulltime).
  - 1 Project programmer: 2-4 days per study.
  - 1 Reviewer: part time, both submissions.
- Time:
  - Upper management endorsement and funding.

1st
- 1 month discussing strategy, philosophy…
- 4 months of programming, documentation, QC, & publishing.

2nd
- 1 month reviewing 1st submission (+/-), redesign
- 4 months of programming, documentation, QC, & publishing.

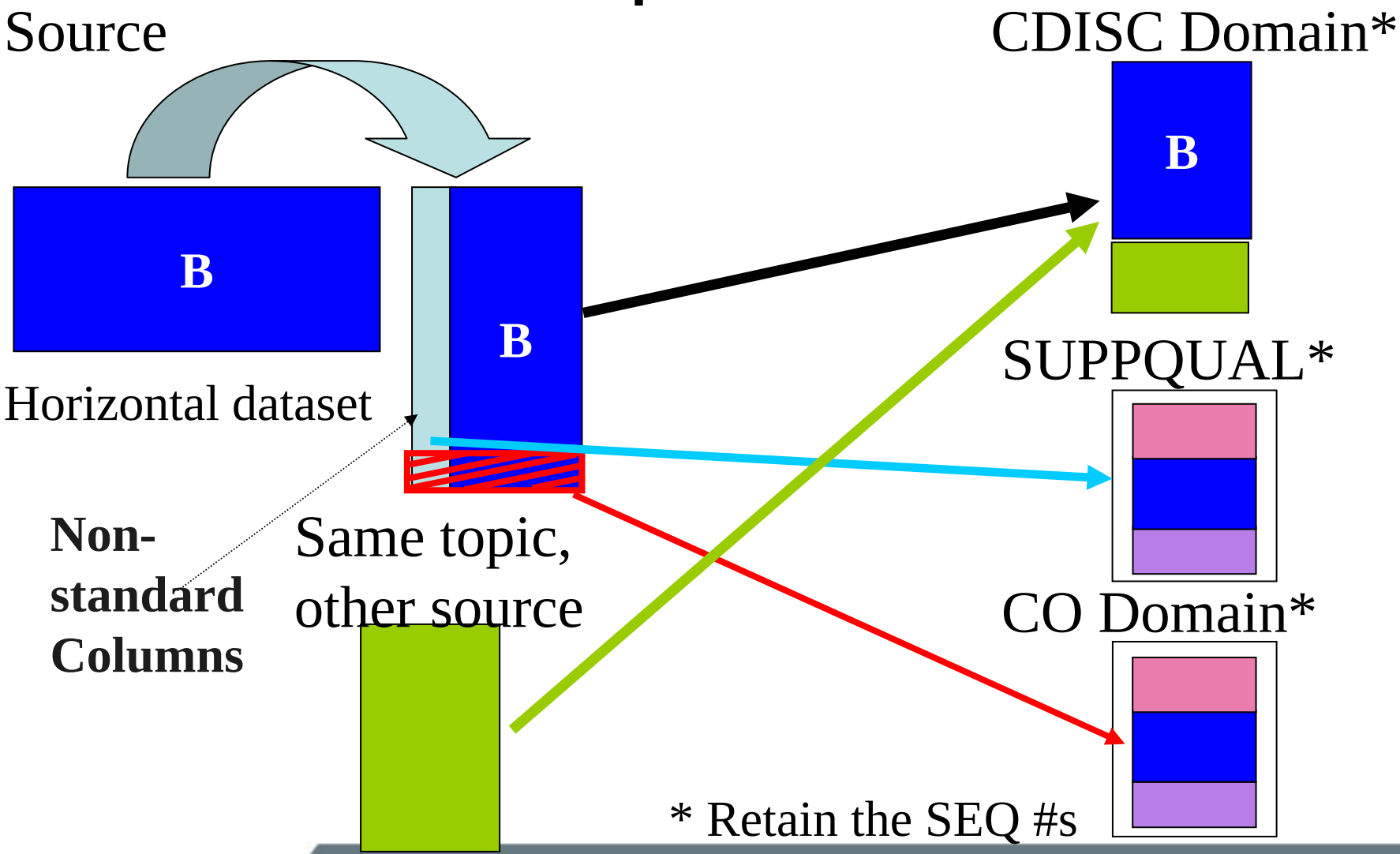
Submission Metrics (2)

- Submission Redundancies:
  - Also submitted the data per the eNDA; 2 to 4 weeks per study to deliver.
- What are the 2nd time costs?
  - Doubled the amount of data converted to SDTM with 60% of the original staff.
  - While implementing new software and quality control procedures.

Submission Challenges

- Traditional challenges with resources.
- File size issues (3gb file!).
- Documentation issues.
- Incorporation of the SDTM metadata.
- Development of define.xml.
- Engaging the project programmers/statisticians .
- Multiple file dependency issue.
  - Complex mappings increases the file dependency issue.

Data Dependencies



Lessons Learned

Lessons Learned: Special Domains

- Trial Design Datasets
  - All datasets but Trial Visits and Subject Visit were completely manually created (3 of 5)
  - Must use protocol to generate, but the information &/or location may not be consistent from protocol to protocol
  - Was quick to implement – but who should "own", review, and sign off?

Lessons Learned: Events & Interventions

- There was a 1:1 mapping for most of the Events and Intervention data (PFE to CDISC SDTM).
- All remaining E/I variables were placed in SUPPQUAL
  - SUPPQUAL became too large (exceeded our Version Control system), therefore produced 1 SUPPQUAL per dataset; updated in next version of SDTM.
  - For example, ae.xpt had ae_supp.xpt (consulted CDISC SDS leadership).
  - This made the implementation much easier because there is no longer a dependency on all datasets for SUPPQUAL.

Events & Interventions (2)

- eSub data documentation was not significantly affected (e.g., define.pdf); except for variables placed in SUPPQUAL.
- Needed CDISC SDTM metadata.
  - Used spreadsheet provided by CDISC, converted it to SAS, custom Macros that accessed the metadata.
  - Used for: labeling, validation, ordering, and additional column (define.pdf) information (e.g., Variable Roles).
- Other Technical challenges:
  - Were not able to combine Concomitant Medications (CM) with Concomitant Non-Drug Treatments because they use difference dictionaries; CDISC needs to address “how to”.

Lessons Learned: Findings

- It describes the vast majority of the data in a submission.
- eSub data documentation is affected.
- Unlike the Events & Interventions, the structure of the Findings Model is very, very flexible.
- More complicated than E/I:
  - May need to transpose data into SDTM structures.
  - Findings are stored in ‘normalized’ data structures.
  - Should provide value-level metadata (test code info).
  - It was easier provide value-level for the ‘flipped’ datasets than those previously stored in a vertical structure.

An Example: Vitals Signs (VS)

Example Dataset

USUBJID	VISIT	DIABP	SYSBP	BMI	HEIGHT
0001	1	70	110	25.3	55

CDISC SDS Version 3 stores data in vertical structures

USUBJID	VISIT	VSTESTCD	VSORRES
0001	1	DIABP	70
0001	1	SYSBP	110
0001	1	BMI	25.3

Value-level Metadata

Dataset Variables for Vital Signs - Findings						
Variable	**Label**	**Type**	**Code**	**Origin**	**Role**	**Comment**
USUBJID	Unique Subject Identifier	text		Sponsor Defined		Unique subject identifier within the submission.
VSTESTCD	Vital Signs Test Short Name	text		Sponsor Defined		Topic variable for VS.

Hypertext Linked

Appendix I: Record Value List							
Variable	**Record Value Name**	**Label**	**Data Type**	**Origin**	**Dictionary**	**Format**	**Role**
VSTESTCD	SYSBP	Systolic Blood Pressure	integer	CRF Page 5			CRT
VSTESTCD	DIABP	Diastolic Blood Pressure	integer	CRF Page 5			CRT

Findings (2)

- These datasets can become extremely large. Several source datasets may map to 1 domain target.
  - E.g., All Questionnaire data goes into the QS domain, we placed 11 different questionnaires in QS.
  - QSCAT was used to separate them.

Lessons Learned: define.xml

- Specifications will be finalized this week (Feb. 4, 2005).
- A great medium for the storage and communication of the metadata.
- It is human-readable (with a style sheet) AND machine-readable (unlike the define.pdf).
- Process changes and software development may be needed.
- Define.xml should accompany SDTM submissions, why?
  - If define.xml is not provided then generic CDISC metadata will be used in FDA applications and NOT the definitions provided within the define doc.
  - Therefore, YOUR variable and value definitions will only be available as a stand alone document.

Lessons Learned: Data Browsing

- Used WebSDM & PPV to view the SDTM data.
- Fast, easy application to generate subject listings and profiles.
- Convenient browsing and data inspection functionality.
- Submitted report templates in the resubmission (facilitate browsing).
- Submission browsers for both the FDA and sponsors?

General Implementation Related Topics

Implementation

- Had little to no difficulty flipping the datasets back & forth (e.g., de-normalized structures); variable roles (in part) determine how this can be automated.
- The devil is in the details.
  - Controlled terminology is even more important than ever before.
  - Multiple interpretations of the SDTM documentation.
  - The implementation guide will evolve over time to provide more guidance.
    - This should reduce the variability of interpretation.
    - How do you write a generic, global document that will be used by all companies, all phases of development, for all therapies and be self evident?

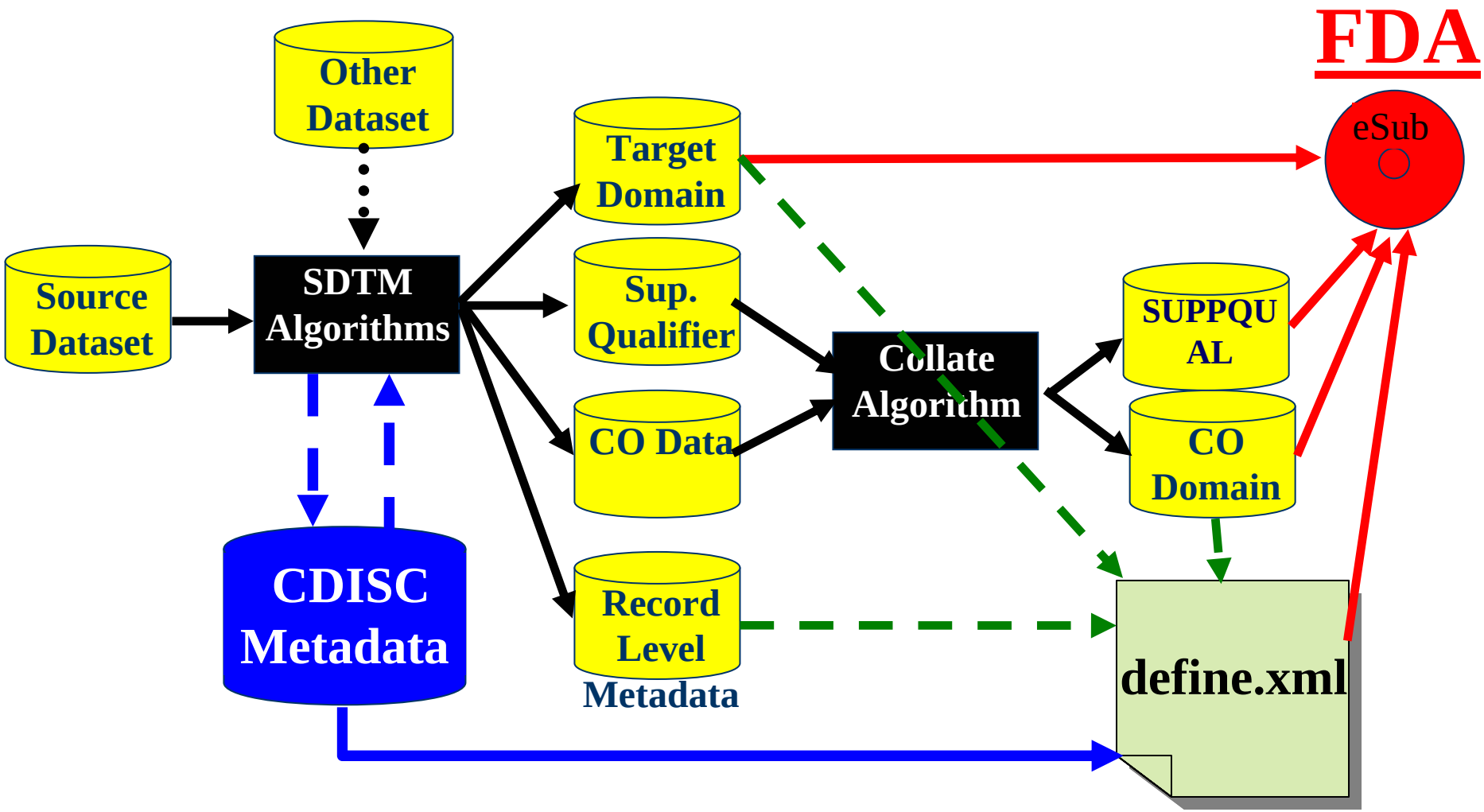
Systems Development

- Can automate the creation and validation of existing and new SDTM domains.
- The variable roles can drive the reporting or browsing of the data (can distinguish between a result and an unit).
- Some parts of the STDM metadata are domain specific:
  - "Adjust the labels of the variables only as appropriate to properly convey"; how do you do this?
- Until you develop the process and software…
  - Additional time/resources will be needed of your programmers and statisticians.
- Until SDTM submissions are 'routine'…
  - Reviewers will be learning the standard, the tools, and the data.

Systems Development (2)

- There are different benefits/costs associated with different implementation sources
  - Our short-term approach is at submission time; post-study report completion (end-game); we will convert the data to SDTM.
  - Benefit: we can do it fast and now; it does not affect any other internal process (e.g., table generation).
  - Cost: timeline considerations, conversion costs, additional QC, Rapid Responses need to be re-converted….

Systems Development (3)



Benefits of the Submissions

- We've encountered most of the obstacles.
  - Identified the internal process & software changes.
  - Baseline level of expertise and experience.
- Software reusability.
  - Will be applying both the software and learnings to other submissions.
  - The experience will be driving our global implementation.
- FDA reviewers started to look at and use the data within 2 weeks of receipt of the application.
- Providing tabulation & derived results within SDTM enabled reviewers:
  - The potential to use their tools for both types of data.
  - Reviewers can use the same data as the sponsors to make decisions (not just tabulation data).

Benefits of SDTM

- Return on Investment (ROI), depends entirely on how you use the SDTM.
- If you only use the SDTM for submissions then ROI will be limited to FDA efficiency gains (which will be balanced against development costs).
- If, however, you use the SDTM as a data exchange format (partners, vendors, etc.) then your ROI could potential be significantly greater.
  - By achieving an industry exchange standard we can reduce or eliminate non-value added activities, processes, and custom applications; thus reducing our total development costs.

Summary

- Pfizer will continue to move forward with submitting CDISC SDTM compliant data.
- Pfizer is involved in the development of industry standards (e.g., eCTD, CDISC, HL7)
  - Industry standards may necessitate process changes and result in software development costs.
  - Our goal is to achieve a positive cost/benefit ratio of implementing standards without increasing the regulatory burden.