Approaches and Challenges in Implementing CDISC Standards
DC DCISC User Group Meeting
2007-09-13

Musa Nsereko, BDS MPH
Sr. Manager, Clinical Data Programming
Cephalon, Inc.
Member CDISC SDS Team
Co-leader CDISC SDS/ADaM Pilot
Presentation Outline

• Introduction
• Overview of CDISC Implementation Strategies
• Overview of Cephalon CDISC Implementation
  – One sponsors worked example
• Overview of CDISC SDTM/ADaM Pilot
  – An Industry worked example
• Conclusions
• Questions/Comments
## Introduction

<table>
<thead>
<tr>
<th>CDISC Standard</th>
<th>Implementation Version Release Date</th>
<th>Cephalon Implementation Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDTM, SEND</td>
<td>2004*</td>
<td>2005</td>
</tr>
<tr>
<td>ODM</td>
<td>2001*</td>
<td>-</td>
</tr>
<tr>
<td>LAB</td>
<td>2002</td>
<td>-</td>
</tr>
<tr>
<td>ADaM</td>
<td>2004</td>
<td>Pending</td>
</tr>
<tr>
<td>Protocol</td>
<td>In progress</td>
<td>-</td>
</tr>
<tr>
<td>Representation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminology</td>
<td>In progress</td>
<td>Partial</td>
</tr>
<tr>
<td>Codelist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Define.xml</td>
<td>2005*</td>
<td>2006</td>
</tr>
</tbody>
</table>

* Specification in FDA Guidance
CDISC Implementation Strategies

- **Linear Method**: DBMS Extract → SDTM → Analysis Files
- **Retrospective Method**: DBMS Extract → Analysis files → SDTM
- **Parallel Method**: DBMS Extract → SDTM → Analysis Files
- **Hybrid Method**: DBMS Extract → SDTM +/- → SDTM Final = Analysis Files

From Strategies for Implementing CDISC by Susan Kenny, PhD
Cephalon CDISC Implementation

Mission:
- Build platform to harness potential of Cephalon’s Clinical information

Implementation Team:
- Champions/Sponsors
  - Vice President Biometrics Operations & Associate Director Clinical Programming
- Implementation Oversight Team
  - Sr. Manager Clinical Programming/Repository (Project Leader)
  - Sr. Manager Clinical Database Programming & Sr. Biostatistics Trial Manager
- Implementation Team
  - Clinical Data Repository Programming Team

Implementation Timeline:
- Internal CDISC SDTM (v3.1) Pilot Project – Q3 2004
- Target Model Database Design and Mapping Specifications – Q2 2005
- CDISC Process documentation – Q3 2005
- First CDISC SDTM (v3.1.1) Submission – Q4 2005
- Define.XML design completion – Q2 2006
- CDISC CRF Design Completion – Q1 2007
- First CDISC CRF Study – Q3 2007
Cephalon CDISC Implementation

- **Implementation Considerations:**
  - Part of Clinical Data Repository Project
  - Minimize Impact on existing processes
  - Implement stable and approved standards first
  - Implement standards with most corresponding internal expertise
  - Implement internal processes e.g Reporting tools/Define.XML/Utility macros

- **Implementation Risks:**
  - Early Adopter risks
  - Adequate training for project teams
  - Standards Gaps
    - considerations of earlier guidances (e.g. 1999) not addressed by CDISC
  - FDA readiness for new standards
  - CROs and Vendors readiness for new standard
  - Cost/Quality/Timeliness
Cephalon CDISC Implementation

Implementation Type
Hybrid Retrospective SDTM + Method
Cephalon CDISC Implementation

Implementation Plan
Phased approach

1) Cephalon CDISC Submission Standard
   - Supplemental Data files
   - Transport Data files
   - Define XML
   - TLG
   - Stats
   - Patient Profiles

2) CDISC CDMS/RAW Standard
   - Vendor Data e.g. Lab, ECG, Sleep &PK

3) CDISC Annotated CRF

Protocol
Cephalon CDISC Implementation

- Implementation Decisions:
  - How to address cross-boundary issues
    - SDTM vs ADaM
  - Cost of mapping Internal standard to CDISC
  - Role of Vendors and CROs
  - Achievable incremental approach
  - Process driven implementation versus technology driven solution
  - Determine a process of selecting studies for implementation
  - Change management
    - New versions of published standards
    - Corrections to implementations
    - Selectively using unpublished stable components
  - Communication with all stakeholders
    - Medical writing/Regulatory Operations/Regulatory Affairs/Clinical
    - Regulatory Agencies (Pre-NDA, Data Sample)
Cephalon CDISC Implementation

- Implementation Metrics:
  - Major reduction in new hire training time
  - Moderate reduction in Integrated database design time
  - Moderate reduction in production of TLFs for safety
  - Moderate increase in production of TLFs for efficacy
  - Efficiency gains:
    - Communication with programmers and vendors simplified
    - Improved checking of CDMS data using standard metadata
  - Think it is currently cost neutral (no hard numbers yet)
    - Improvements in time and efficiency offset by
    - addition validation steps due to hybrid retrospective method and
    - development/modification costs for tools/processes
    - Future savings anticipated
  - Invaluable communication with FDA prior to submission
    - Minimize assumptions on submission data
    - Clarify FDA expectations regarding data
Cephalon CDISC Implementation

• **Legacy Studies:**
  – All with Cephalon CRF Design
  – Most with Cephalon submission database
  – One study with CDISC submission database
  – Mapped to SDTM using standard target metadata
  – Full validation of TLF output
  – Validation versus Integrated analyses where possible
  – Re-annotation of CRFs to CDISC database
  – Generation of new data definition files (Define.XML)
  – Validation using WebSDM software
  – Used external validator for TLF
    • Acted like reviewer
Cephalon Data Integration Process

Clinical Data Repository Metadata

- Compound level is a virtual level within the metadata
- Contents of CDISC metadata repository:
  - Published SDTM 3.1.1 domains metadata (e.g. Adverse events)
  - Custom Domains metadata (e.g. Sleep Latency data)
  - Templates for ETL of source data into SDTM 3.1.1 structures
  - Templates for loading transformed data into by compound directory within the CDR
  - User written transformations (e.g. subject sequence (SEQ) generator)
Cephalon Data Integration Process

Clinical Data Repository Metadata

- Compound level is a virtual level within the metadata
- Contents of Study metadata repository:
  - Library definitions to the source, transformed and CDR data
  - ETL jobs that transform source data into SDTM 3.1.1 structures
  - ETL jobs that load transformed jobs into by compound directory within the CDR
Cephalon Data Integration Process

ETL Process
• Two parallel processes used to load CDR
  – ETL Tool
  – Traditional SAS programming
• ETL Tool
  – Only for legacy studies
  – Created reusable ETL job templates at CDISC level
  – Modified jobs per study
    • Add Inputs
    • Add user-written transformations
• Traditional SAS Programming
  – In-house programming group macros developed for CDISC conversions.
  – Used for all ongoing/current studies
  – Used for some legacy conversions
• All Clinical Data Repository load jobs done via ETL Tool
Cephalon Data Integration Process

ETL Process
- ETL Job Template:
Cephalon Data Integration Process

ETL Process

- ETL Data load Job:
Cephalon Data Integration Process

ETL Process

- ETL Data load Job:
Cephalon Data Integration Process

Traditional SAS Process

1a. CDMS/RAW Data

1b. Standards Database

2. SAS Mapping Program

3. Informs Process

4. Map Document

5. Register Metadata in CDR

- Versus Mapping database
- Apply Standard Domain/Variable Attributes
- Drives Process
- Informs Process
- Drives Process
Proposed CDR Use (s)

- Clinical Study Reports
- Integrated Analyses
- FDA Questions
- Marketing/Publications
- Safety Updates
- Data mining
- Study Planning
- Phase IV

CDISC SDTM CLINICAL DATA REPOSITORY

Study A
Closed

Study B
Closed

Study C
Open

Integrated Database

CDISC

METADATA
Proposed CDR Use (s)

Study A
Closed

Study B
Closed

Study C
Closed

Study D
Open

CDISC SDTM
CLINICAL DATA
REPOSITORY

Integrated
Database

Clinical Study Reports

Integrated Analyses

FDA Questions

Marketing/Publications

Safety Updates

Data mining

Study Planning

Phase IV

CDISC SDTM
CLINICAL DATA
REPOSITORY

CDISC SDTM
CLINICAL DATA
REPOSITORY

METADATA
Proposed CDR Use(s)

Information Delivery Portal
Proposed CDR Use (s)

Information Delivery Portal
• Overview of CDISC SDTM/ADaM Pilot
Goal of the Pilot Submission

• Produce a worked example implementation of the available CDISC standards.

• Illustrate how the various CDISC components can be used to result in a submission of electronic data that are in a format that is acceptable to... and meets the needs of both medical and statistical reviewers.
Don't let what you can't do interfere with what you can do.

- Use the tools currently available (with very minor modifications if any) to produce the pilot submission.

- SDTM IG Version 3.1.1
- SDTM Version 1.1
- ADaM Version 2.0
  - (for public comment in March, 2006)
- CRT-DDS version 3.1.1
- ODM version 1.3
  - (public comment closed May 2, 2006)
- Custom stylesheet
  - developed by team members
- Datasets as XPT not XML
FOCUS: the package not the process

- Choices/decisions guided by
  - timeline
  - realities of a team of volunteers from multiple companies
  - goal was the submission package and the FDA review
  - quick, efficient, effective - not necessarily the most preferred option
  - are not making recommendations re process!
    • But project report will discuss lessons learned
Pilot Submission Deliverables

• Submission package
  – Includes SDTM datasets, ADaM datasets, all relevant metadata, analysis results, abbreviated study report
  – Review package tied together using metadata in DEFINE.XML

• Summary report of the pilot submission project
  – issues encountered, strengths and weaknesses
  – incorporate what we learned from the FDA feedback

• Both to be made available to the public on the CDISC website
Criteria for success of the Pilot Project

- FDA statistical and medical reviewers will evaluate the submitted datasets (SDTM and AdaM), metadata and documentation
  - Usable with their tools?
  - Reproducibility of analyses, derivations?
  - Navigable?
  - Contents – what and where are OK?
- Assessment by FDA reviewers as “reviewable” and “meets expectations”
The Pilot Project Team

CDISC teams represented: ADaM, SDS, ODM, some with no affiliation

- Cathy Barrows (GSK)
- Musa Nsereko (Cephalon)
- FDA Co-Leaders:
  - Lonnie Smith (previous)
  - Chris Holland
  - Mina Hohlen
- Greg Anglin (Lilly)
- T Friebel (SAS)
- John Gorden (Quintiles)
- Tom Guinter (Octagon)
- Joel Hoffman (Insightful)
- Susan Kenny (Inspire Pharm.)
- Sandy Lei (J&J)
- Richard Lewis (Octagon)
- Arline Nakanishi (Amgen)
- Gregory Steffens (Lilly)
- Gary Walker (Quintiles)
- Aileen Yam (Sanofi-Aventis)
- Yuguang Zhao (Sanofi-Aventis)
FDA

- Unprecedented level of involvement
- Co-Leadership of the project
- 18-20 FDA employees involved
- includes medical and statistical reviewers
- ≈ 12 consistently in contact with team
- Interactions:
  - regular team teleconferences
  - Feb. face-to-face meeting to define the project (expectations/requirements)
  - Pre-submission encounter
  - Feedback from review
Project Timelines

18 Nov 05
1st team teleconference

25 Jan 06
Planning meeting w/ Board reps.

28 Feb 06
Team kick-off f2f meeting

17 Feb 06
Redacted study docs delivered

19 Apr 06
Received data

10 Apr 06
Presubmission Encounter w/FDA

30 June 06
Submission to FDA

26 Sep 06
Results at Interchange

End of Aug 06
FDA Feedback
Kick off Meeting with FDA representatives

• 12 FDA employees (including medical and statistical reviewers) met with us in February
• Key messages:
  – Consistency, accuracy, completeness are extremely important - follow the specifications!
  – Define file crucial, but needs to be accurate
  – Clear mapping between the protocol-specified analysis plan, the data, and the analyses performed
  – Programs necessary if define file inadequate
  – SDTM and Analysis datasets should be available for both medical and statistical reviewers
Presubmission “Encounter”

- Described study to FDA team
  - because no time for briefing package to be pre-submitted
- FDA made specific requests, including:
  - Hy’s Law analysis dataset (Liver hepatotoxicity)
  - Population flags in all analysis datasets
  - All levels of the MedDRA coding in SDTM datasets
  - Prefer to avoid by-patient listings
Where Did We Get the Data?

- Real clinical trial data, provided by Eli Lilly
  - Legacy data from failed trial (and shelved compound)
- Randomized, double-blind, placebo-controlled, parallel-group study
- Transdermal study drug (low dose and high dose and placebo)
- Approximately 300 patients with mild to moderate Alzheimer’s disease, multiple centers
- 24 weeks of active treatment
How Was the Work Organized?

• Our process was to create ADaM from SDTM
  – Map legacy data to SDTM first
  – Create ADaM datasets from SDTM
  – Used ADaM derivation logic to populate derived items from ADaM into SDTM
Creating SDTM

- SDTM sub-team developed a mapping document to illustrate how legacy data was mapped into SDTM domains
- Octagon team members used SAS ETL Studio to do the mapping and QC of mapping
- Other team members created annotated CRF
Creating ADaM

- Began with the end in mind!
  - Metadata spreadsheets were used to capture all information before programming was started
    - Variable metadata
    - Allowed values
    - Variable descriptions / exceptions

- Suite of proprietary macros were used that were developed by Lily (Greg Steffens)
MetaData Driven Approach for ADaM

- Spreadsheets were used to create 0 observation SAS datasets
  - These were the ‘gold standard’
- Last step of ADaM creation programs was to do a PROC APPEND with 0 obs datasets
  - All variable metadata was set
  - Value lists were checked
  - Missing variables were highlighted
  - Extra variables were dropped
- Define file used spreadsheets as input
ADaM Data included in Pilot Submission Package

• Primary efficacy variables -
  – ADAS-Cog (11-item)
  – CIBIC+
• Secondary variables –
  – NPI-X
• Safety –
  – Adverse events
  – Vital signs
  – Laboratory evaluations

• Not all of the legacy study data elements
• Representative set of analyses
CDISC Pilot Submission Package Content

- submission format: eCTD/eNDA hybrid
- PDF TOCs and eCTD folder structure
- kept the submission “simple” and the focus on the components, not on the eCTD backbone
Reviewer’s Guide

• A separate document to highlight organization of submission and ‘nice to know’ items
  – What was included in submission
  – Which ADaM datasets were used to produce which results
  – What derived data was in SDTM
  – Naming conventions used in ADaM
  – Navigation tips
  – Helpful tips for using annotated CRF with pop up comment boxes
Overall Impression/General Comments

• Submission was generally well done
• Standards have great promise!
  – Reviewers will need experience with standardized data
  – Tools will need development to assist with reviewer needs
• Some “kinks” need to be worked out
  – The Define.XML file
  – Analysis dataset structures
ADaM Comments

- Very important component since SDTM datasets are not analysis ready! ALL reviewers used ADaM datasets
  - Core variables such as treatment group, center, age, gender, etc. are not within each SDTM file.
- Overall, the files were very useful
  - Many analyses were “one PROC away”
- Biggest issue was with the structure of the efficacy data sets (ADQSADAS, ADQSCIBC, and ADQSNPIX)
- Reviewers really liked the Analysis Results Metadata
FDA Conclusions

- Great job, overall
- Standards have great promise
  - Efficiencies will come with:
    - Reviewer training
    - Reviewer experience
    - Adaptation and development of review tools
- ADaM files are critical when submitting SDTM data
FDA Conclusions

- **Items to work on:**
  - Define.XML
    - Standardized stylesheet?
      - Something to improve printing and navigating
    - Software kinks (beyond our control)
    - Solutions need to be explored
  - ADaM data
    - Transparency is the key
      - Allows reviewers to understand (and trust) what was done
      - Allows reviewers to examine the sensitivity of what was done to alternative methodologies
ADaM Conclusions

- Some already incorporated in the published ADaM v 2 document
- Additional points to be considered in building analysis datasets:
  - Variables to include in analysis datasets
    - DOMAIN (in all)
    - Dictionary name and version
    - Consistent set of treatment variables
    - Dosing start and stop dates in AD files
    - Flag for on-treatment versus off-treatment values
  - Logical ordering for the variables (not alphabetical)
  - In list of analysis datasets, denote which datasets contain the primary and secondary efficacy variables
ADaM Conclusions

- Primary efficacy involved creating windowed visits
  - FDA had difficulty identifying which visits were used for the windowed visit when multiple visits were available for a given window
  - Traceability from SDTM visit to ADaM visit is important
- Vertical structure in ADaM is ok
SDTM Conclusions

• Trial Design datasets need better descriptions
• Dictionary names and versions in relevant domains AE and CM?
• Don’t need lots of derived data in SDTM if it is available in ADaM
  – Redundancy isn’t necessarily a good thing
SDTM and ADaM

- Three different variables to indicate day of study:
  - VISITDY (SDTM)
    - Planned study day of visit
  - --DY (SDTM)
    - Actual study day of Visit/Collection/Exam
    - RFSTDTC in DM domain is Study Day 1
    - Day before RFSTDTC is Day -1
  - ANLDY (ADaM)
    - Analysis study day of Visit/Collection/Exam
    - RFSTDTC is Analysis Day 1
    - Day before RFSTDTC is Day 0
DEFINE Conclusions

- Pilot team members had to write an extension to ODM in order to accommodate the ADaM analysis results metadata
  - Shows the power of ODM extensibility
- Pilot team created a Reviewers’ Guide but many reviewers didn’t know it was there
  - Put link to Reviewers Guide at front of Define file
Next Steps:
Wrap up tasks for this iteration

• Some revisions to current package
  – Implement some of the FDA feedback
  – Fix a few things that are errors or oversights
  – Incorporate some things we wish we had done

• Project report undergoing final review
  – Final team review

• Plan to publish both package and report by CDISC interchange 2007
Conclusions:

• Both case studies the CDISC standards facilitated communication about the data
• Both case studies showed that some degree of re-tooling was needed to implement the standards
• Consider both cost and efficiency when implementing CDISC standards
• Potential Industry benefit of this standard is enormous
• Standards Implementations requires resolve and ongoing commitment
• Train and be patient!
Questions/Comments?

- Musa Nsereko
- Senior Manager, Clinical Programming
- Cephalon, Inc.
- mnsereko@cephalon.com
- 610-883-5675