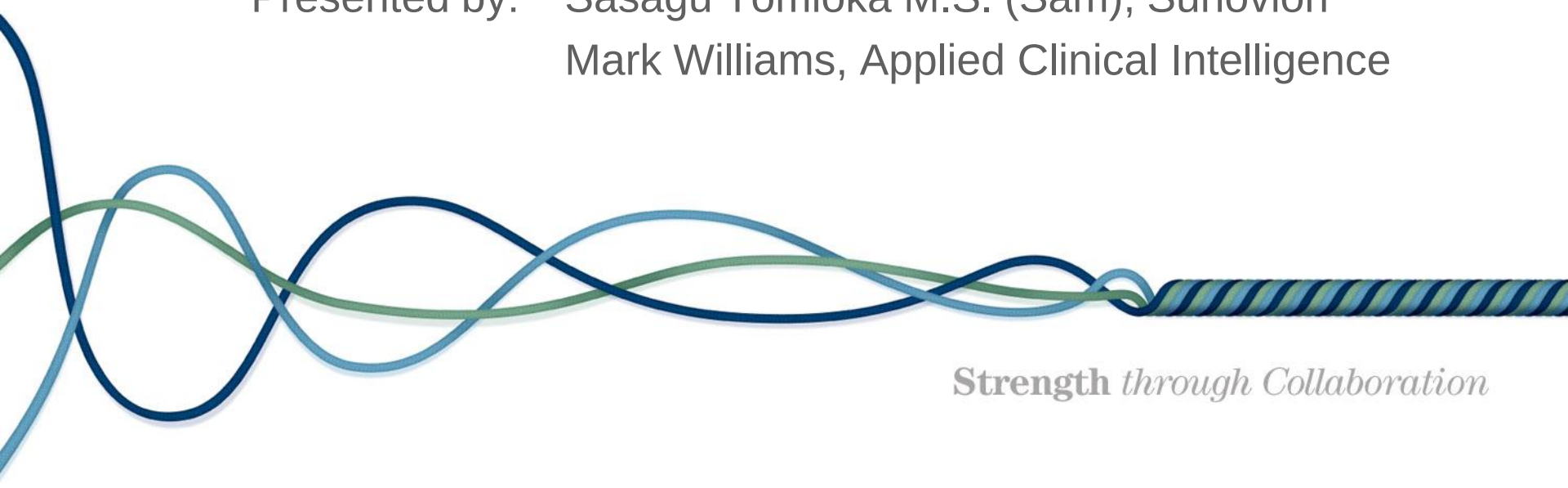
Reflections and Lessons Learned from a Recent NDA Submission Using CDISC Standards

Presented by: Sasagu Tomioka M.S. (Sam), Sunovion
Mark Williams, Applied Clinical Intelligence



Overview

- Case study
- Challenges (Organization, Project, and Study Level)
- Implementation Strategy
- QC Process
- Conclusion

Case Study

- Submission in 4Q 2009
- Studies
  - 19 Phase 1 (Subjects)
  - 7 Phase 1 (Patients)
  - 7 Phase 2/3 (DBL Efficacy, Safety)
  - 10 Phase 2/3 (OL)
- Activities
  - Translation of Japanese datasets
  - Conversion of all studies to SDTM 1.1/SDTM IG 3.1.1
  - Generation of ISE/ISS from SDTM datasets
  - Generation of CRT-DDS for SDTM and ISE/ISS.

SUNOVION

Challenges: Organization Level

- The team in the US was relatively new and small when the NDA project began.
- No prior NDA submission experience as a company. New processes had to be established.
- There had been no central repository for clinical data prior to the initiation of the project.

SUNOVION

Challenges: Project Level

- The timeline was relatively tight as the timing of submission was cut by 1 year from the original plan.
- Multiple vendors involved in the submission preparation activities.
- All of legacy studies conducted in US and Europe spanning 14 years were outsourced studies with essentially no standards and limited data management and biostatistics oversight.
- Approximately 20% of studies were in Japanese.
- Annotations of raw datasets were missing for many studies.
- For all of the legacy studies, we had to obtain source data from each vendor that we worked with before. Some of the studies were more than 10 years old.

SUNOVION

Challenges: Study Level

- Some data had not been entered for some legacy studies and had to be entered after mapping had started.
- The unpredictable nature of data from legacy studies caused the mapping process to be extremely difficult.
- There were several cases where the unique subject was identified by more than one Subject ID within a study (cross over studies) and across the studies (core & extension studies).

SUNOVION

Implementation Strategy

- Build an internal programming environment and the repository in a controlled environment.
- Build a secure information sharing and document collaboration web portal.
- Re-obtain legacy study source datasets from each vendor and store in the controlled environment.
  - Upload all source datasets, documentation, and deliverables to the secure web portal.
- Select the programming vendor from CDISC certified vendors.
- Prepare SDTM mapping specifications and annotations using one vendor to maintain consistency across all studies.
- Review the mapping specifications and annotations by the same individuals at Sunovion for all studies.

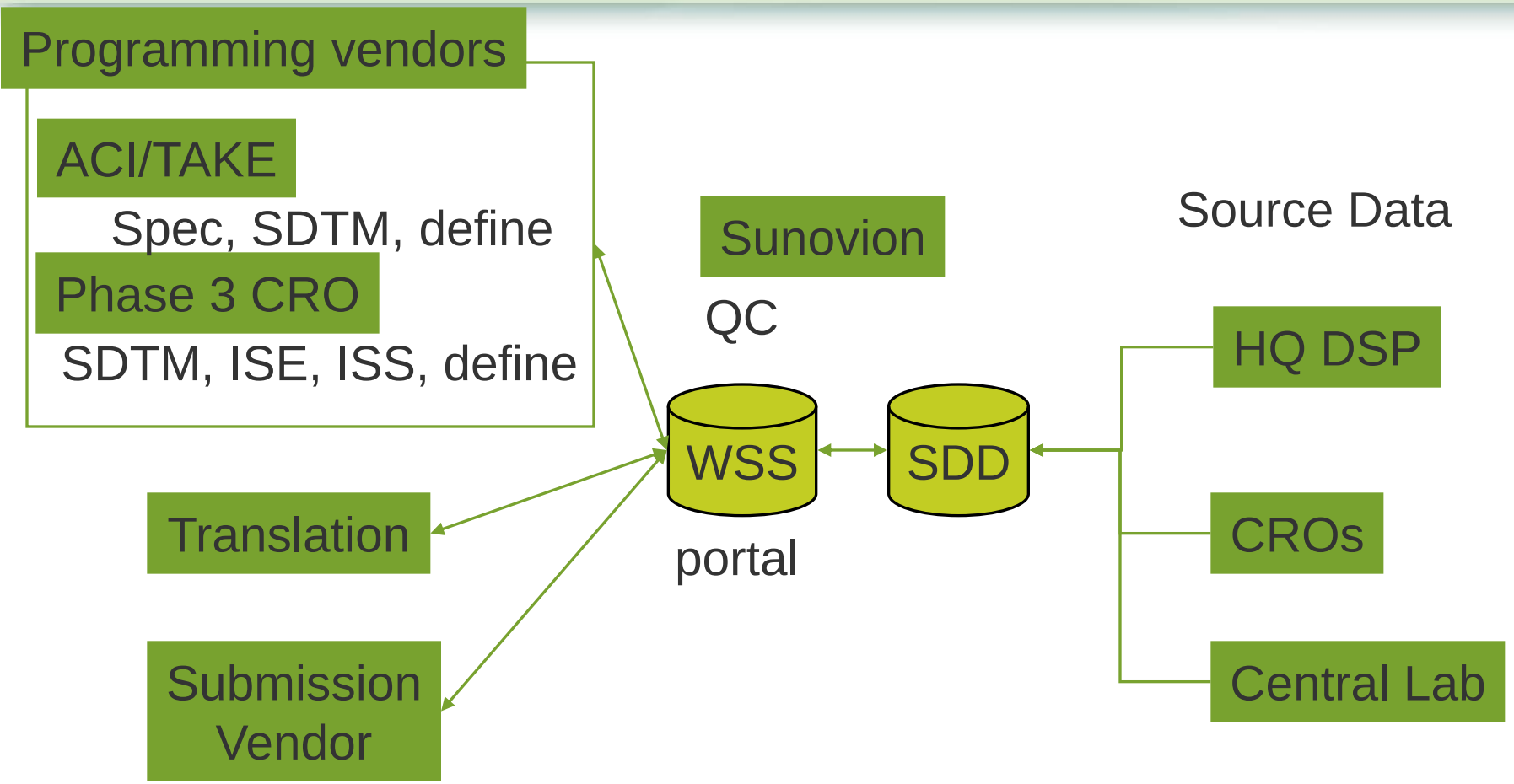
SUNOVION

Implementation Strategy

- Prepare SDTM and Define.xml for all legacy studies and Japanese studies using one vendor.
- Prepare SDTM and Define.xml for Phase 3 trials as well as ISE and ISS using the CRO contracted for Phase 3 trials.
- Perform translation using one vendor after unique strings are extracted from SAS datasets. Use synonym table.
- Perform validation and QC of all deliverables using the vendors.
- Perform in-house parallel programming of deliverables from the vendors.

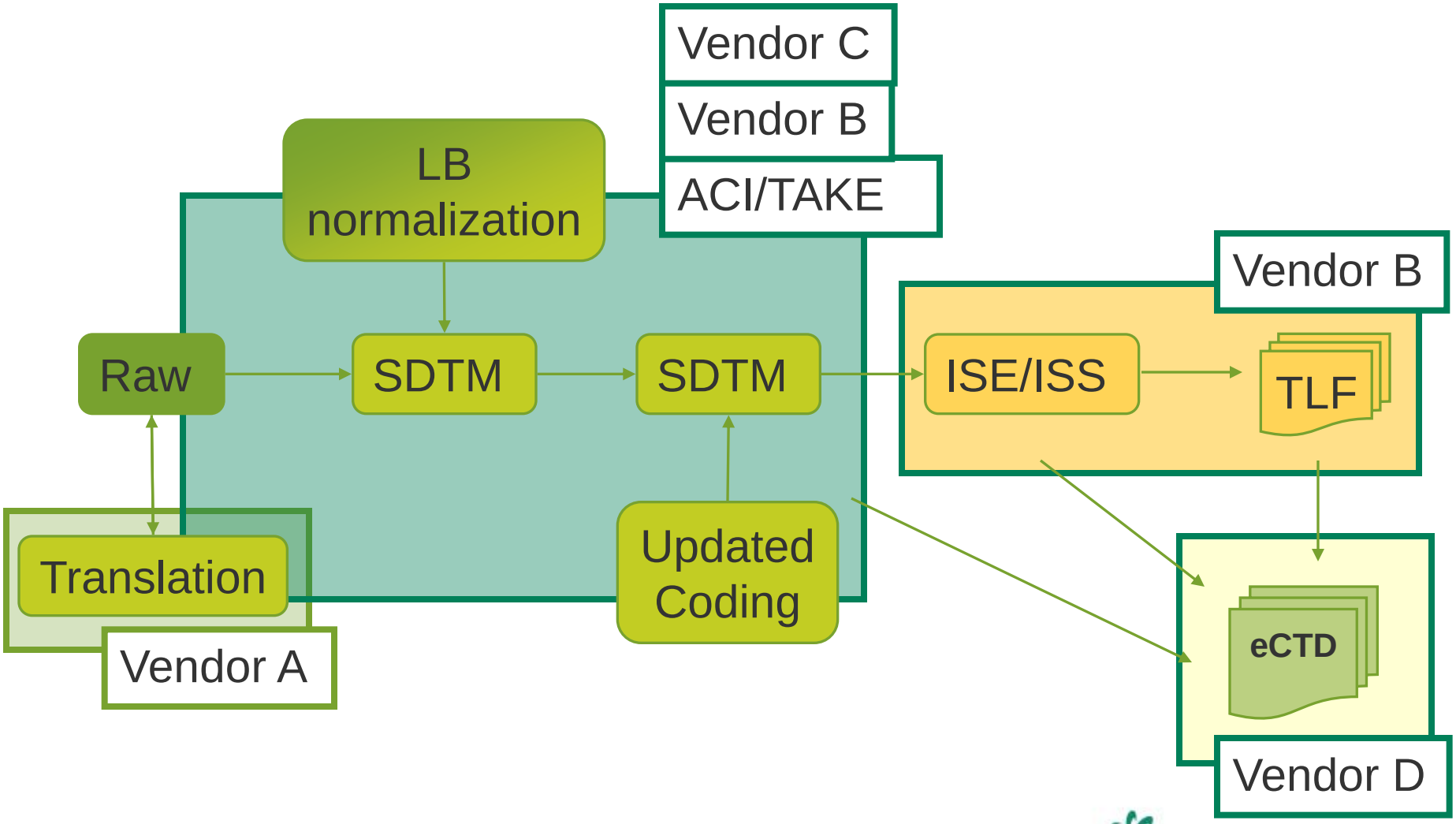
SUNOVION

Multiple Vendors



One PM managed/tracked timeline daily

Data Flow


Vendor C
Vendor B
ACI/TAKE
LB normalization
Raw
SDTM
SDTM
Translation
Vendor A
Updated Coding
Vendor B
ISE/ISS
TLF
eCTD
Vendor D

Type of datasets prepared

- **Study Level SDTM Datasets**
  - Compliant with SDTM and original Clinical study reports
    - --TEST, --TESTCD were standardized across all studies.
    - CDISC Controlled Terminology were used.
    - --DECOD were standardized at the study level.
      - AEDECOD, MHDECOD, CMDECOD, DSDECOD etc
    - --STRESC, --STRESN, --STRESU were standardized across all studies.
      - LB, EG etc.

- **Integratable SDTM datasets**
  - Compliant with SDTM
    - --DECOD, --BODYSYS are standardized across all studies.
    - Terms for ISS and ISE are mapped.
      - DM.RACE, AE.AEREL etc
    - Ensured that variable attributes were standardized across all studies

SUNOVION

Type of Non-SDTM datasets

- Study level analysis datasets
  - Efficacy, demographic, exposure, and randomization datasets were provided.
- ISS, ISE analysis datasets
  - Each dataset contains integratable SDTM data and derived data
  - Each dataset may not contain all studies.
- Datasets used for POP PK analysis

SDTM Datasets QC

Vendor's QC and QA

- QC
  - Performed parallel programming for all datasets
  - Performed SDTM conformance checks for all SDTM datasets
  - Compared SDTM against the CSR Tables for the legacy studies.
    - Recreated key safety and efficacy tables in original CSR from SDTM
      - Essential traceability check
- QA
  - Provide documentation of QC processes.
  - Review Program and Macro comments, Controlled Checklists, SAS Logs, Worksheets, Signatures
    - Who-What-When and HOW

SUNOVION

Sunovion's QC and QA

- **In-house Resource**:
  - 4 SDTM programmers, 1 ISS, ISE programmer, 4-5 statisticians

- **In-house QC Approach**:
  - **Annotation, mapping specs (3 days/study)**
    - Team review
  - **SDTM (5 days/study)**
    - Verified if all the mapped variables are present in the datasets.
    - Performed parallel programming for key datasets, mapped at least 3 key variables to SDTM and compared with the datasets from the vendors.
    - For findings domains such as EG, LB, QS and VS, compared the numeric results of the population using PROC UNIVARIATE, PROC MEANS, and PROC FREQ from the source and SDTM datasets.

SUNOVION

Sunovion's QC and QA

- **In-house QC Approach (continue)**:
  - **SDTM Conformance Check (1 days/study)**
    - Reviewed the reports

	A	B	C	D	E	F
1	ID=IR5108					
2	ID	Rule	Subject ID	Visit	Findings	Response
3	IR5108	Identifies records that violate the condition [Start Relative to Reference Period is in codelist STENRF or is null]	D1050289-0001-00001	MISSING	The value for [Start Relative to Reference Period]=PRIOR is not found in Codelist [C66728] for the medication LORAZEPAM.	
		Identifies records that violate the				

    - Documented any errors which cannot be resolved. Some errors could not be fixed because some errors were originated from the source and even from the CRF design.

- **In-house QA Approach:**
  - QA group compared data from SDTM and the CRFs.

SUNOVION

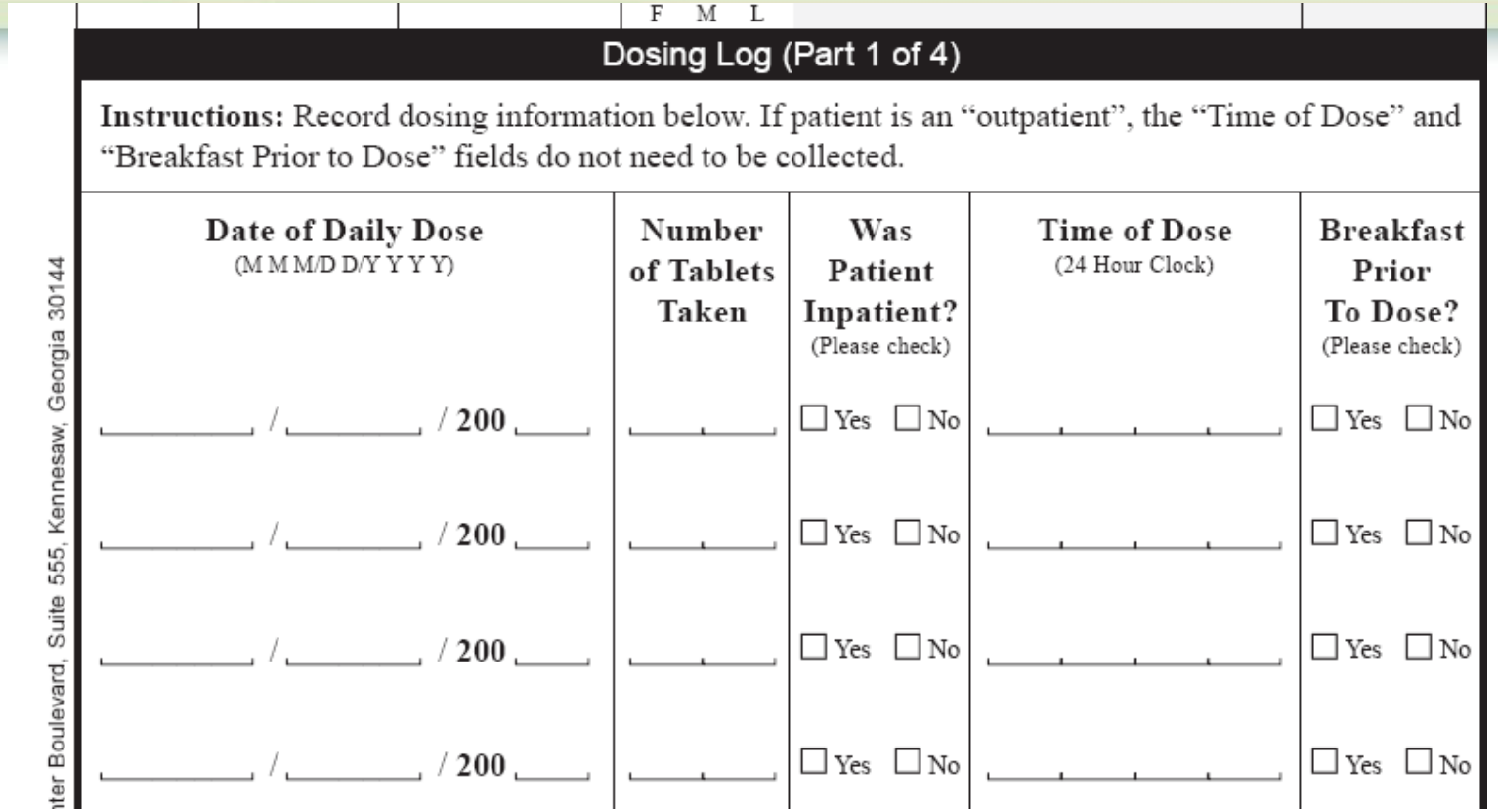
Examples of issues we identified

Issue identified during the preparation of specifications

Background

- The study had two forms to capture the exposure data.
  1. Doselog --- Captures exposure log from placebo run-in to the last dose
  2. Dose --- Captures the first and last dose

Doselog dataset



F M L

Dosing Log (Part 1 of 4)

Instructions: Record dosing information below. If patient is an "outpatient", the "Time of Dose" and "Breakfast Prior to Dose" fields do not need to be collected.

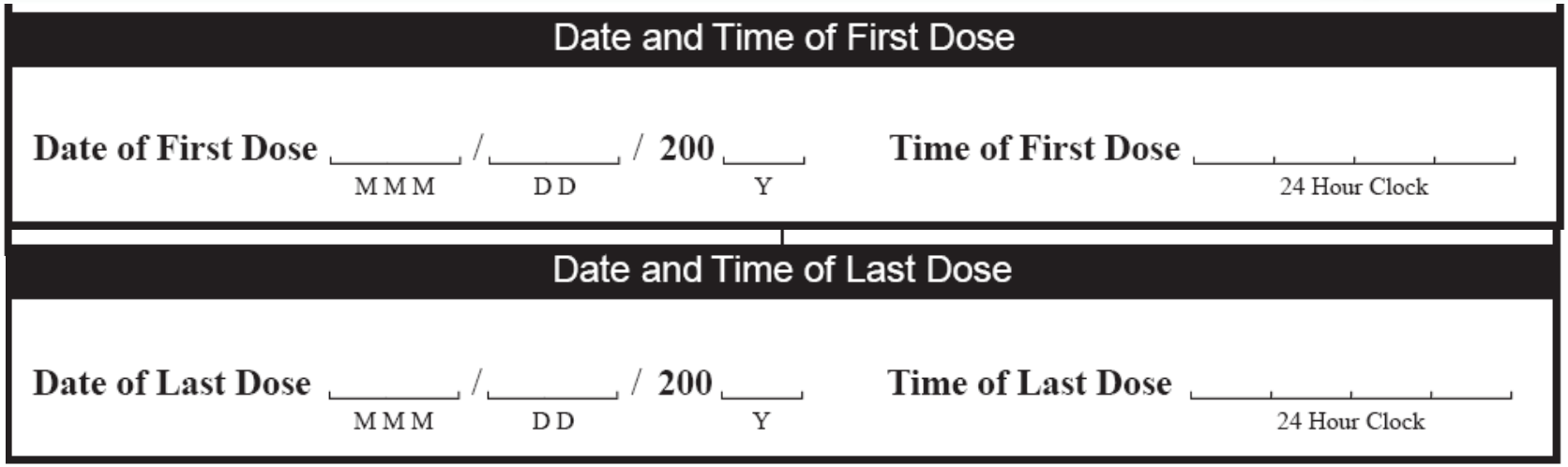
Date of Daily Dose (M M M/D D/Y Y Y Y)	Number of Tablets Taken	Was Patient Inpatient? (Please check)	Time of Dose (24 Hour Clock)	Breakfast Prior To Dose? (Please check)
____ / ____ / 200 ____	____	☐ Yes ☐ No	____:____	☐ Yes ☐ No
____ / ____ / 200 ____	____	☐ Yes ☐ No	____:____	☐ Yes ☐ No
____ / ____ / 200 ____	____	☐ Yes ☐ No	____:____	☐ Yes ☐ No
____ / ____ / 200 ____	____	☐ Yes ☐ No	____:____	☐ Yes ☐ No

ıter Boulevard, Suite 555, Kennesaw, Georgia 30144

1. There were discrepancies between DOSE and DOSELOG.
2. DOSELOG does not have EXTRT (treatment name) or EPOCH information, thus there is no clarity between the treatment phase and placebo-run phase.

SUNOVION

Dose dataset



Date and Time of First Dose

Date of First Dose _____ / _____ / 200____
MMM DD Y

Time of First Dose _____
24 Hour Clock

Date and Time of Last Dose

Date of Last Dose _____ / _____ / 200____
MMM DD Y

Time of Last Dose _____
24 Hour Clock

- CSR used **DOSE** dataset to derive the exposure duration, thus we mapped these two dates to RFSTDTC and RFENDTC.
- There were about 12 subjects whose last dose was not recorded in **DOSE**, thus the exposure duration for these subjects became 1 day in the CSR, but **DOSELOG** dataset indicate that the duration is longer than 1 day.

Initially, we generated SDTM to match the CSR, but in this case, the method used for the CSR was obviously incorrect, thus we updated the SDTM

SUNOVION

Example of WebSDM errors 1

High ← JANUS severity rating

- R4096 Identifies records that violate the condition [Subject Reference Start Date/Time is not null], limited to records where [Arm Code doesn't equal 'SCRNFAIL']

Medium

- IR4506 Identifies Demographics subjects where no record for the subject is found in the EX domain

Situations:

- Subjects whose study arm was assigned but never took the study medication.

This error is not fixable during the study and at the time of SDTM creation.

SUNOVION

Example of WebSDM errors 2

Medium

- IR4008 Identifies records where Serious Event='Y' but none of it Involves Cancer, Congenital Anomaly or Birth Defect, Persist or Signif Disability/Incapacity, Results in Death, Requires or Prolongs Hospitalization, Is Life Threatening, Other Medically Important Serious Event, or Occurred with Overdose equals 'Y'.

This error is due to the CRF design.
The question was raised by the FDA regarding this issue at the sample SDTM datasets submission.
It can NOT be fixed at the time of SDTM creation.

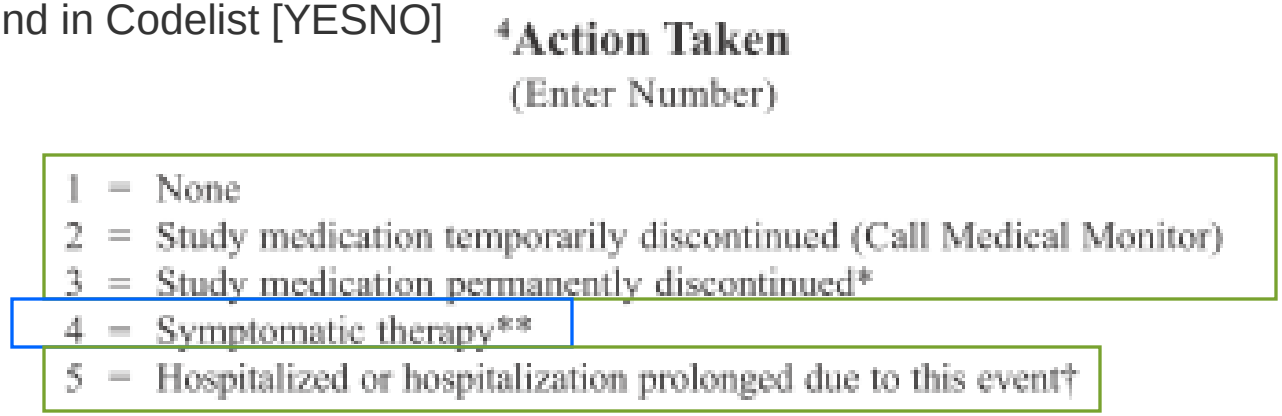
SUNOVION

Example of WebSDM errors 3

Medium

- R4043 Identifies records where value for [Concomitant or Additional Trtmnt Given] is not found in Codelist [YESNO]

Situation:



1-3 is mapped to AEACN “Action taken to the study medication”

4 is mapped to AECONTRT “Concomitant or Additional Trtmnt Given”

5 is mapped to AESHOSP

- R4043 expects AECONTRT to have either “Y” or “N”. When 1-3 or 5 is selected for the AE, we set AECONTRT=“” because we are not supposed to derive “N” .

This error is due to the CRF design.
It can NOT be fixed at the time of SDTM creation.

SUNOVION

Non-SDTM datasets QC

Datasets requirements

- Documents we referenced SDTM ver. 1.1, ver. 3.1.1 and Study Data Specifications (SDS) ver.1.4 and ver. 1.5.

Datasets	SDTM	Non-SDTM
Size	unlimited	< 400MB
File format	SAS transport file (ver 5)	SAS transport file (ver 5)
Length of variable	8	8
Length of variable label	40*	40
Length of dataset label	40	40

* Few variables have over 40 characters label in SDTM 3.1.2. which appear to be error.

Variables longer than 8 characters and the label longer than 40 characters were shortened for xpt files.
Example:

Variable	Label
AIMS10	Subject's awareness of abnormal movements
AIMS10	Subject's awareness of abnorm. Movements

SUNOVION

Datasets requirements per SDS Ver 1.5 (1)

- Dataset names and labels must be unique.
- Dataset label must be the identical to the name shown in the Description filed of Define.XML.
- Key variables (subject identifier and visit) should appear first in the dataset.
- Each subject must be identified by a unique ID.

Datasets requirements per SDS Ver 1.5 (2)

- Dataset with multiple records per subject should have
  - A variable for relative day of measurement or event e.g. AE.AEDY, VS.VSDY, EX.EXSTDY
  - Visit information in at least 2 variables (char and numeric) e.g. VISIT, VISITNUM
- For unscheduled visits or measurements, numeric visit should be distinct from other visit numbers but retain the chronological order. E.g. VISITNUM follows this rule.

SUNOVION

Datasets requirements per SDS Ver 1.5 (3)

- Add core variables in each dataset after the key variables. Example of core variables, STUDYID, SITEID, COUNTRY, ARM (treatment assignment), SEX, AGE, RACE, analysis population flags
- Standardize variable attributes (name, code, format, label, length etc.) across datasets.
- If textual data is coded, provide codes and decode values.
- Dates in numeric

QC of Define files

Define.PDF / Define.XML – TOC or Dataset

	1999 Guidance	CRT-DDS (Define.XML) Version 1.0
	For Non-SDTM datasets	For SDTM datasets
Metadata Field		
Dataset - Dataset name	Required	Required
Description - Dataset label	Required	Required
Structure –detail of individual record ex. “one record per subject per event”	NA	Optional
Purpose – ex. Tabulation or analysis	NA	Optional
Keys – variables to uniquely identify each record	NA	Optional
Location – path in eCTD	Required	Required

NA = Not Applicable

Example – TOC of the Define.XML

Datasets for Study D1001036						
Dataset	**Description**	**Class**	**Structure**	**Purpose**	**Keys**	**Location**
TA	Trial Arms	Trial Design	One record per planned element per arm	Tabulation	STUDYID, ARMCD, TAETORD	ta.xpt
TE	Trial Elements	Trial Design	One record per element	Tabulation	STUDYID, ETCD	te.xpt
TI	Trial Inclusion/Exclusion Criteria	Trial Design	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	ti.xpt
TS	Trial Summary	Trial Design	One record per trial summary parameter	Tabulation	STUDYID, TSPARMCD, TSSEQ	ts.xpt
TV	Trial Visits	Trial Design	One record per planned visit per arm	Tabulation	STUDYID, ARMCD, VISITNUM	tv.xpt
SE	Subject Elements	Trial Design	One record per actual element per subject	Tabulation	STUDYID, USUBJID, SESTDTC	se.xpt
SV	Subject Visits	Trial Design	One record per subject per actual visit	Tabulation	STUDYID, USUBJID, VISITNUM	sv.xpt
DM	Demographics	Special Purpose	One record per subject	Tabulation	STUDYID, USUBJID	dm.xpt
CO	Comments	Special	One record per comment per subject	Tabulation	STUDYID, USUBJID, RDOMAIN,	co.xpt

Programmatically compared against the SDTM IG

SUNOVION

Define.PDF / define.xml
- Value Level Metadata

	1999 Guidance	CRT-DDS Version 1.0
	For Non-SDTM datasets	For SDTM datasets
Metadata Field		
Variable Name	Required	Required
Variable Label	Required	Required
Data Type	Required	Required
Controlled Terms or Format	NA	Optional
Origin	NA	Optional (hyper links to the aCRF page)
Role	NA	Optional
Comment	Required	Required

NA = Not Applicable

SUNOVION

Example - Value Level Metadata of the Define.XML

Demographics Dataset (DM)						dm.xpt
Variable	**Label**	**Type**	**Controlled Terminology**	**Origin**	**Role**	**Comment**
STUDYID	Study Identifier	text		CRF Page i	Identifier	
DOMAIN	Domain Abbreviation	text		Derived	Identifier	Hardcode ("DM")
USUBJID	Unique Subject Identifier	text		Derived	Identifier	Concatenate 'D1001001' and SUBJID with "-"
SUBJID	Subject Identifier for the Study	text		CRF Page i	Topic	
RFSTDTC	Subject Reference Start Date/Time	text	ISO 8601	Derived	Timing	Date of first dose of study drug
RFENDTC	Subject Reference End Date/Time	text	ISO 8601	Derived	Timing	Date of last dose of study drug
SITEID	Study Site Identifier	text		CRF Page i	Record	

Programmatically compared against the SDTM datasets and IG

SUNOVION

Define.PDF / define.xml – Controlled Terminology

	1999 Guidance	CRT-DDS Version 1.0
	For Non-SDTM datasets	For SDTM datasets
Metadata Field		
Code Value	Required	Required
Code Text	Required	Required

Example – Controlled Terminology

SEX, Reference Name (SEX)	
Code Value	**Code Text**
F	FEMALE
M	MALE

Programmatically compared against the SDTM datasets and format.cat

SUNOVION

Define.xml – XML fields

XML fields hold more information than what you will see in the front end.

- example for the dataset

<ItemGroupDef OID="DM" Name="DM"
Repeating="No" IsReferenceData="No"
Purpose="Tabulation" def:Label="Demographics"
def:Structure="One record per subject"
def:DomainKeys="STUDYID, USUBJID"
def:Class="Special Purpose"
def:ArchiveLocationID="location.DM">

No for domains with 1 record per subject

No for subject level data

Class of domains

Define.xml – XML fields

Variable matadata are provided in two ODM fileds, ItemRef and ItemDef.

- Example for ItemRef

<ItemRef ItemOID="DM.STUDYID" OrderNumber="1" Mandatory="Yes" Role="Identifier" RoleCodeListOID="RoleCodeList" />

- Example for ItemDef

<ItemDef OID="DM.STUDYID" Name="STUDYID" DataType="text" Length="8" Origin="CRF Page i" Comment="" def:Label="Study Identifier" />

- Additional information such as significant digits, display format, computational method OID can be added to ItemDef.

SUNOVION

Conclusion

- Implement CDISC based standards across the entire data lifecycle would save cost and time when preparing submission and building ISE and ISS.
- Identifying the CDISC certified programming vendor early in the stage and including them as part of overall strategy saved time (by a year in our case).

- Creating analysis datasets, (ADaM) from SDTM would keep the transparency and consistency.
- CDISC conformance is important but the quality of content is more important and must be checked during each clinical trial.

SUNOVION