申請電子データプロセスの効率化
～メタデータ・社内ツールを中心に～
○飯島 康浩、梶原 正幹、小南 真由美、
中嶋 優一、北原 孝志
（ノバルティスファーマ株式会社）

Metadata Driven Approach for eSubmission
Yasuhiro Iijima, Yoshimasa Kajiwara, Mayumi Kominami,
Yuichi Nakajima, Takashi Kitahara
Novartis Pharma K.K.
要旨:

本発表では、申請電子データを効率的に作成するためのメタデータ及びメタデータを包括的に管理する社内のガバナンス体制、SASを用いたannotated CRFの自動作成やDefine.xml確認ツールを紹介する。

キーワード：申請電子データ, メタデータ, annotated CRF, Define.xml
Disclaimer

• The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein.
Agenda

• Introduction
• Part 1: Metadata Management
• Part 2: CRF annotation tool
• Part 3: Automated Define.xml QC Checks
• Summary
Introduction

1st e-study data submission

2nd e-study data submission

Nth e-study data submission

Oct-2016～
Start accepting e-study data

Transitional period (3.5 years)
Partial e-study data submission can be accepted

Apr-2020～
e-study data submission will be mandatory.
Data Standardization in Novartis

• Novartis Clinical Data Standard
  – The Novartis Clinical Data Standards (NCDS) contain all data elements and attributes needed for data collection, processing, analysis, reporting and submission.
Novartis’ domains Overview
Simplification and standardization

- SDTM / ADAM / Define-XML
  - Template programs
  - Comprehensive DEFINE Review

- aCRF
  - Standard CRF
  - Annotation tool

- SDRG/ADRG
  - Template and examples
  - Training material and check list

Standard governance drives consistency and quality
Part 1.

Metadata Management
What is Metadata

✓ Data about data
✓ Information describes a variable
✓ A proc contents

DATA

METADATA
Novartis’ domains Overview

Clinical Data Element View

Derivations/Imputation View

Actual contents
Novartis SDTM/ADaM process overview

Metadata Management

Global Metadata

Local Metadata

SAS

Excel

SDTM* & ADaM Templates

Create xpt file

SAS

XP

T

SDTM* / SuppQual

ADaM

* Including Novartis Clinical Data Standard
### Metadata Management

- More detail about template program generation process.

#### Standard macro library

- 

#### Contributing Element

- Derivation Name

#### Global / Local Metadata

#### Template programs
Metadata Management

CDISC Standards

Project team

Study team

Project team

Study team

Standard & Process group

Study team

Study team
# Key Benefits and challenges

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controlled and Audit trailed environment</td>
<td>Mapping metadata to the appropriate version of the SDTM IG, CDISC Controlled terminology, etc.</td>
</tr>
<tr>
<td>• <strong>Centralize</strong> all changes made to metadata</td>
<td></td>
</tr>
<tr>
<td>• <strong>Traceability</strong> of changes</td>
<td></td>
</tr>
<tr>
<td>Efficacy gained in defining and <strong>maintenance of metadata</strong> e.g.</td>
<td>While some of the Controlled terminology values are mapped to NCI values in SDTM/ADaM, it is an on-going effort to keep our Metadata in alignment with CDISC standards</td>
</tr>
<tr>
<td>Codelists, lab reference tables, etc</td>
<td></td>
</tr>
<tr>
<td>Instant access to all end users to browse matadata</td>
<td><strong>Project level metadata management</strong> is also challenge</td>
</tr>
</tbody>
</table>
Simplification and standardization

Standard governance drives consistency and quality

SDTM /ADAM / Define-XML
- Template programs
- Comprehensive DEFINE Review

aCRF
- Standard CRF
- Annotation tool

SDRG/ADRG
- Template and examples
- Training material and check list
Part 2.

Creation of SDTM Annotated CRFs
Advantage of automation tool

Automation
• Re-usable
• Keep quality and consistency

Manual
• Human errors
• Time consuming
Overview of annotation tool

**Standard Governance**

- Develop standard & annotated CRF
- Centralize information in Metadata & annotation database
- To be used within Individual studies

- Develop global / Therapeutic area standard CRF page
- SDTM annotation is manually setup per individual page.
- Domain / annotation information are loaded to metadata and annotation database respectively.
- Get standard information for SDTM annotation from global metadata
ADOBE ACROBAT FDF FILE

FDF stands for "Forms Data Format". FDF is a file format for representing form data and annotations that are contained in a PDF form.

%FDF-1.2
%ããïÓ
1 0 boy
</FDF</</Annots[2 0
R]/F(/C/SUGI/SUGI2017_example_CRF_DM.pdf)/ID[<B8EE73C02F4D874C93AAE44ACF8C2212><C2D8B7C87E8FCE47B3CE92FB8082505F>]UF(/C/PharmaSUG2014/PharmaSUG2014_example_CRF_DM.pdf)>>/Type/Catalog>>
endobj
2 0 obj
<//BS 3 0 R/C[0.0 1.0 1]/Contents(SEX )/DA(0 0 0 rg /Arial 10 )/DS(font: Arial,sans-serif 10.0pt; text-align:left; color:#000000 )/M(D:20140406095629-04'00'/NM(7c71d692-03dc-45c4-abd9-4bf182d12a4f)/Page 0/RC(<?xml version="1.0"?><body xmlns="http://www.w3.org/1999/xhtml" xmlns:xfa="http://www.xfa.org/schema/xfa-data/1.0/" xfa:APIVersion="Acrobat:10.1.5" xfa:spec="2.0.2" style="font-size:10.0pt;text-align:left;color:#000000;font-weight:normal;font-style: normal;font-family:Arial,sans-serif;font-stretch:normal">
<p>SEX </p></body>)/Rect[407.868 324.866 431.868 338.866]/Subj(TextBox)/Subtype/FreeText/T(NCDS)/Type/Annot>>
endobj
trailerr
</</Root 1 0 R>>
%%EOF
CRF Annotation Tool – Process Flow

1. **Central Annotation Database**
   - Convert to SAS dataset

2. **SAS**
   - Convert to FDF File
   - Extract study level information

3. **Excel**
   - Import Data File

4. **FDF File**
   - Export Data File

5. **Blank CRF(PDF)**
6. **Annotated CRF(PDF)**
CRF Annotation Tool
Central Annotation Database
CRF Annotation Tool – DEMO

- `master_annotation.sas7bdat`: Central Annotation Database
- `BlankCRF_Study2_Demo.pdf`: Blank CRFs to annotate for Study 2
- `crfid_pageno.xls`: CRF IDs/Page
- No specification
- `create_fdf_file.sas`: SAS program to create study specific FDF file
- `extraction_annotation.sas`: SAS program to create Secondary Annotation Database(s)
<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Kind</th>
<th>Modified</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlankCRF_Study1_Demo.pdf</td>
<td>300317</td>
<td>File El...</td>
<td>02/19/2014 21:06:56</td>
<td>vmain1</td>
</tr>
<tr>
<td>sCRF_Study1_Demo.pdf</td>
<td>150581</td>
<td>File El...</td>
<td>02/19/2014 21:31:45</td>
<td>vmain2</td>
</tr>
<tr>
<td>lcrfId_pagemath</td>
<td>31332</td>
<td>File El...</td>
<td>02/19/2014 21:37:33</td>
<td>vmain1</td>
</tr>
<tr>
<td>extract_annotation.sas</td>
<td>22360</td>
<td>File El...</td>
<td>28/05/2015 10:32:26</td>
<td>vmain2</td>
</tr>
<tr>
<td>create_full_files.sas</td>
<td>12130</td>
<td>File El...</td>
<td>27/11/2017 23:25:09</td>
<td>vmain4</td>
</tr>
<tr>
<td>Study2.fdf</td>
<td>48076</td>
<td>View...</td>
<td>06/07/2017 12:17:21</td>
<td></td>
</tr>
<tr>
<td>study3.os36416</td>
<td>2236416</td>
<td>View...</td>
<td>06/07/2017 12:21:21</td>
<td></td>
</tr>
<tr>
<td>out1#_demo1.fdf</td>
<td>42673</td>
<td>View...</td>
<td>06/07/2017 12:30:27</td>
<td></td>
</tr>
<tr>
<td>out1#_demo13.fdf</td>
<td>36778</td>
<td>View...</td>
<td>06/07/2017 12:50:30</td>
<td></td>
</tr>
</tbody>
</table>
### Vital Signs

Any clinically significant findings present prior to signing informed consent should be recorded on the Medical History page. Any new or worsening clinically significant finding noted since signing informed consent should be recorded on the Adverse Events page.

<table>
<thead>
<tr>
<th>Date of assessment</th>
<th>VSORRESU</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSJTG</td>
<td></td>
</tr>
<tr>
<td>VSJTG</td>
<td></td>
</tr>
<tr>
<td>VSJTG</td>
<td></td>
</tr>
<tr>
<td>VSJTG</td>
<td></td>
</tr>
<tr>
<td>VSJTG</td>
<td></td>
</tr>
<tr>
<td>VSJTG</td>
<td></td>
</tr>
</tbody>
</table>

#### Blood pressure

<table>
<thead>
<tr>
<th>Systolic / Diastolic</th>
<th>VSJTG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Pulse

<table>
<thead>
<tr>
<th>Beats/Min</th>
<th>VSJTG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Key Benefits and challenges

• Benefit
  – This process not only helps in reducing the cycle time by a great extent, but also helps maintain consistency of the annotations across different studies.
  – FDF file can be easily handled in SAS datasets and easy to extend its usage for other tool e.g. Define-XML.

• Challenge
  – Some manual task such as page mapping, overall QC is still necessary.
Simplification and standardization

Standard governance drives consistency and quality

- **SDTM / ADAM / Define-XML**
  - Template programs
  - Comprehensive DEFINE Review

- **aCRF**
  - Standard CRF
  - Annotation tool

- **SDRG / ADRG**
  - Template and examples
  - Training material and check list
Part 3.

SAS Application to Automate a Comprehensive Review of DEFINE and All of Its Components
Problem Statements

The DEFINE package is a large electronic document comprised of many different but interrelated components.

The define.xml acts as a road map where embedded hyperlinks allow reviewers to easily navigate between components and understand how data were collected/derived for analysis purpose.

It is a massive undertaking to review all the components to ensure accuracy, completeness, and consistency once the DEFINE package is created, especially done manually.
Automation Can Be Achieved once You

- are familiar with the DEFINE and all of its distinct but interrelated components and sections
- fully understand the scope of what a thorough review of the DEFINE entails
- construct a data structure capable of consolidating disparate metadata from each of the DEFINE components
Interrelated DEFINE Components and Structure
Scope of a Thorough Review within a Single DEFINE

Hundreds of data elements and values must be reviewed for accuracy, consistency, and traceability.

Consistency and traceability within define.xml

Define.xml consistent with Input Specifications

Define.xml, annotated CRFs, Data Reviewer’s Guide consistent with XPT files

Annotated CRFs and Data Reviewer’s Guide consistent with define.xml

Each of these data points must be re-verified each time a new draft is generated.
Interrelated DEFINEs within a Submission
Automated DEFINE Review Tool Overview

Source File Conversion

- Study Data Reviewer’s Guide
- Annotated CRF
- define.xml

Excel file
.fdf file
Input to the DEFINE

Automated DEFINE QC Checks

- SAS dataset
- SAS dataset
- SAS dataset

Issues identification & resolution spreadsheet
Automated DEFINE Review Tool Overview

Source File Conversion

Study Data Reviewer’s Guide

Annotated CRF

define.xml

Excel file

.fdf file

Input to the DEFINE

Automated DEFINE QC Checks

Excel file

SAS dataset

SAS dataset

SAS dataset

Issues identification & resolution spreadsheet
Issues Identification & Resolution Spreadsheet

- All QC issues consolidated into a single Excel file
- Separate tabs used to store related issues
- Standard columns / issue messages to facilitate resolution
Prior to Implementing the Tool

- **Line-by-line visual compare** of all components and data points
- Basic SAS code checks
- Manual documentation of findings in various formats
- Overlapping and/or blind spots by the reviewers

**Manual task**

**Inefficient review**

- Long time
- Incomplete
- Inconsistent
- More iterations of review
After Implementing the Tool

- **Spare resources** from lengthy, tedious, and repeated manual reviews
- **Focus valuable resources on tasks that require a higher level of functional expertise and knowledge**
- **Eliminate incomplete and inconsistent findings** due to factors such as level of experience, fatigue, and time constraints
- **Improve quality and efficiency of review**
What the Tool Does Not Do

- Substitute the review of derivations in Input Specifications and defile.xml, which ensures that the description is clear, complete, and accurate.
- Substitute the review of the contents in Data Reviewer’s Guide.
- Replace the conformance check.
- Consistency check within the submission.
## Project-level Input Specifications

**Well-conceived project-level input specifications** (a single file which applies to all studies contained in the submission) play a pivotal role.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Must be accurate and concise, and <strong>consolidate project-level</strong> data point attributes</td>
<td>Allow for the <strong>existence</strong> of study specific derivations and additions where necessary</td>
</tr>
<tr>
<td></td>
<td>Can be used to drive the code used to both create the tabulation and analysis datasets and produce and QC the DEFINE</td>
</tr>
</tbody>
</table>
Summary

Automation eliminates incomplete and inconsistent findings, **reduces the burden on programming and statistical resources**, and greatly improves the quality of the review of the DEFINE.

Organized metadata allow us to easily create some kind of tools like the ACRF tool or the DEFINE review tool by SAS and it **could improve our routine work drastically**.

Standardization and Simplification are important elements for eSubmission.
Reference

- Walter Hufford, Vincent Guo, Mijun Hu, ” SAS Application to Automate a Comprehensive Review of DEFINE and All of Its Components”, PharmaSUG 2017
Back up
Metadata Management

Standards
Create and Maintain Data Standards

Sites
Enter clinical data on standard eCRFs

Third party
Load clinical data based on standard Data Transfer Specification

OC/RDC
Metadata is used in CRF page setup for consistent database design

LSH
Metadata is used to map/transform, derive the data collected in CRFs to 3DTM submission accepted data formats

JReview
Reports

Spotfire

GPS
Datasets
Tables
Listings
Figures

Metadata Management System

<table>
<thead>
<tr>
<th>Data Elements definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Element Name</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>SEX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbr</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>SEX</td>
</tr>
<tr>
<td>SEX</td>
</tr>
<tr>
<td>SEX</td>
</tr>
<tr>
<td>SEX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Derivations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivation name</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>copy_element</td>
</tr>
<tr>
<td>combine_results</td>
</tr>
</tbody>
</table>
Why Do SDTM Annotations?

- SDTM annotated CRFs is one of the key CDISC format submission deliverables
- This is a blank CRF annotations that document the location of the data with the corresponding names of the datasets and the names of those variables included in the submitted SDTM datasets
- Annotations are meant to help the PMDA/FDA reviewer find the origin of data variables included in the submitted SDTM datasets
- Annotations should be text-based and searchable using standard PDF viewers
CRF Annotation Tool
SAS Programs

• Create_fdf_file.sas (SAS dataset to FDF file)
  – Inputs:
    • Central Annotation Database (SAS dataset)
    • CRF IDs/Page No in a Excel spreadsheet
  – Output: Study specific FDF file that could be imported to the study blank CRFs

• Extract_annotation.sas (FDF file to SAS dataset)
  – Inputs: FDF files exported from already annotated CRF pages (FDF file)
  – Output: Central Annotation Database (SAS dataset)
Savings after Automated DEFINE Review Tool

- An estimate of about 2/3 time reduction
- **Savings are multiplied** where submissions consist of multiple studies
- Savings remain significant even after considering initial investment in the development of the tool
- The more the tool is used, the more the savings increase