Challenges in Addressing Data Sharing

Ravi, Sr. Clinical Programmer I,

MMS Holdings Inc., Canton MI, July 15th, 2014

NOVARTIS

Outline

- Overview of Data Sharing Process
  - Components
    - Request approval
    - Data Anonymization
    - Data sharing
- Overview of Data Anonymization process
  - Components
    - Input study data
    - Anonymization macro
    - Definition dataset
    - Output study data

Outline cont'd..

- Challenges in Data Sharing
  - Study request approval
  - What do we share?
  - Informed consent
- Challenges in Data Anonymization
  - Data Anonymization Standards
  - Cross-divisional approach
  - Anonymization macro
  - Definition Dataset
  - Macro Validation and Approval process
- The tripartite strategy
  - Novartis
  - MMS Holdings Inc.
  - The SAS Institute

Novartis commitment

WHY?

- In order to...
  - **Enhance Data Sharing with Researchers**
  - **Enhance Public Access to Clinical Study Information**
  - **Share lay summaries of clinical trials with participating patients**
  - **Reaffirm Commitments to Publish Clinical Trial Results**

Definition

WHAT?

- ANONYMIZATION:

  “Modifying certain elements of a study data in a way that it is impossible to identify the participating subject from the data”

Anonymization (vs) De-identification

Original Dataset

USUBJID	Phone #
SUBJ012	000-000-0000
SUBJ012	000-000-0000
SUBJ045	999-999-9999
SUBJ045	999-999-9999

USUBJID	USUBJID_anon
SUBJ012	XXXXnn1
SUBJ012	XXXXnn1
SUBJ045	XXXXnn2
SUBJ045	XXXXnn2

Keys table

Anonymized Dataset

USUBJID	Phone #
XXXXnn1	
XXXXnn1	
XXXXnn2	
XXXXnn2	

Original Dataset

USUBJID	Phone #
SUBJ012	000-000-0000
SUBJ012	000-000-0000
SUBJ045	999-999-9999
SUBJ045	999-999-9999

USUBJID	USUBJID_anon
SUBJ012	XXXXnn1
SUBJ012	XXXXnn1
SUBJ045	XXXXnn2
SUBJ045	XXXXnn2

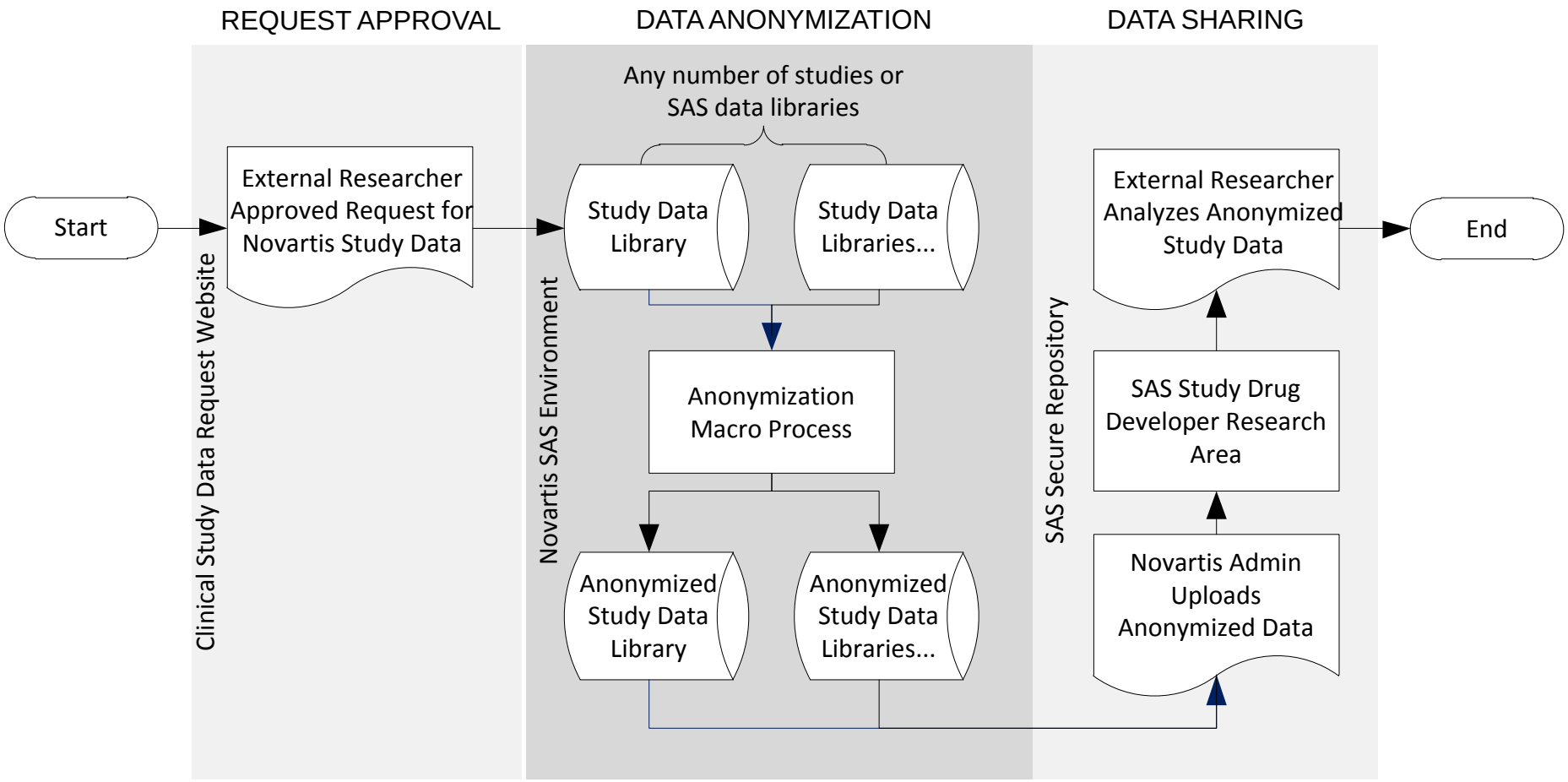
Keys table

De-identified Dataset

USUBJID	Phone #
XXXXnn1	
XXXXnn1	
XXXXnn2	
XXXXnn2	

Overview of Data Sharing Process

HOW?



NOVARTIS

Components of Data Anonymization process

- Input study data
- Definition Dataset
  - Contains modes of anonymization for all study variables.

VIEWTABLE: TMP3.st_def_dt2

	Domain	Variable_Name	mode
67	AE	project	translate
68	AE	projectid	drop
69	AE	RecordDate	drop
70	AE	RecordId	none
71	AE	RecordPosition	none
72	AE	SaveTS	drop
73	AE	SDVTier	drop
74	AE	Site	drop

- Anonymization macro
  - Applies modes of anonymization as specified in Definition dataset, to data contained in study variables.
- Output study data

Modes of Anonymization

- DROP

VIEWTABLE: Rchwork.Input

	TEST_VAR1	TEST_VAR2
1	DATA	data
2	DATA	data
3	DATA	data

VIEWTABLE: Rchwork.Output

	TEST_VAR2
1	data
2	data
3	data

- MISSING

VIEWTABLE: Rchwork.Input

	TEST_VAR1	TEST_VAR2
1	DATA	data
2	DATA	data
3	DATA	data

VIEWTABLE: Rchwork.Output

	TEST_VAR1	TEST_VAR2
1		data
2		data
3		data

- TRANSLATE

VIEWTABLE: Rchwork.Input

	TEST_VAR1
1	SUBJ01
2	SUBJ01
3	SUBJ04

VIEWTABLE: Rchwork.Output

	TEST_VAR1
1	XXXXnn1
2	XXXXnn1
3	XXXXnn2

Modes of Anonymization

- DATE

VIEWTABLE: Rchwork.Input

	TEST_DT
1	06-25-1985
2	08-16-1986
3	12-31-1991

VIEWTABLE: Rchwork.Output

	TEST_DT
1	07-26-2085
2	09-17-2086
3	01-01-2091

  - All date variables will offset by a random generated number.

- AGEINT

VIEWTABLE: Rchwork.Input

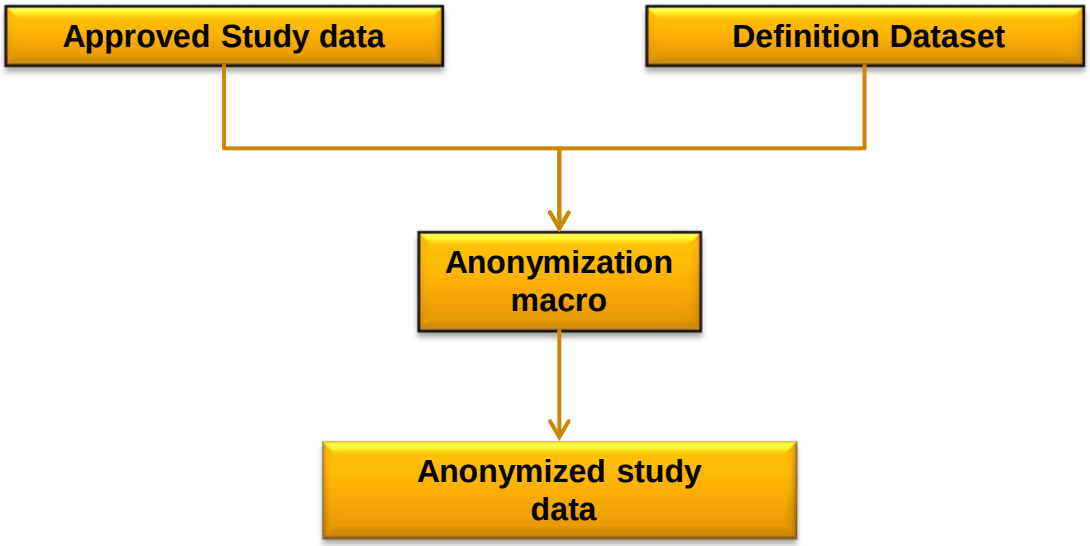
	TEST_AGE
1	60
2	96
3	55

VIEWTABLE: Rchwork.Output

	TEST_AGE
1	60
2	90 or older
3	55

- NONE

Overview of Anonymization process



Anonymization process

- **Example**

Study data example on top and anonymized data on bottom.

Study Data

Center ID	Investigator ID	Investigator name	Subject number	Date of birth	Age (yrs)	AE start date	AE end date	Verbatim term	Preferred term
T1230	279T344	Dr Smith	2002	08Aug1954	57	29DEC2010	27JAN2011	HEADACHE	Headache
T1230	279T344	Dr Smith	2002	08Aug1954	57	10JAN2011	06APR2011	BRONCHITIS	Bronchitis
T1230	279T344	Dr Smith	2004	09Aug1919	92	25MAR2011	12AUG2011	COLD	Nasopharyngitis
T1230	279T344	Dr Smith	2004	09Aug1919	92	28MAR2011	31MAR2011	FLU	Influenza
T1230	279T344	Dr Smith	2004	09Aug1919	92	01MAR2011	15MAY2011	PAIN	Pain
G5670	348G224	Dr Jones	2010	09Aug1947	64	14OCT2010	20OCT2011	ACHE NOS	Pain
G5670	348G224	Dr Jones	2010	09Aug1947	64	24MAY2011		BRONCHIAL INFECTION	Bronchitis
G5670	348G224	Dr Jones	2010	09Aug1947	64	01MAR2011	15MAR2011	CHRONIC PAIN	Pain

Translate	Trans ate	Missing	Translate	Missing	Ageint	Date	Date	Missing	None

Anonymized Data

Center ID	Investigator ID	Investigator name	Subject number		Age (yrs)	AE start date	AE end date	Verbatim term	Preferred term
Xnn10	nnnXn10		1111		57	19AUG2010	17SEP2010		Headache
Xnn10	nnnXn10		1111		57	06JUL2010	20SEP2010		Bronchitis
Xnn10	nnnXn10		1113		90 or Older	05SEP2010	23JAN2011		Nasopharyngitis
Xnn10	nnnXn10		1113		90 or Older	06SEP2010	09SEP2010		Influenza
Xnn10	nnnXn10		1113		90 or Older	29JUN2011	12SEP2011		Pain
Xnn11	nnnXn11		1101		64	16JUL2011	12SEP2011		Pain
Xnn11	nnnXn11		1101		64	04NOV2010			Bronchitis
Xnn11	nnnXn11		1101		64	01JUL2010	15JUL2010		Pain

Challenges in Data Sharing

Study request approval

- Evaluation criteria for approval of a data sharing request.
  - Description of project and its purpose
  - Role of the data requestor in the project
  - Proposed processing including publication
  - Type of personal data and if anonymization is feasible (e.g. Rare Disease indications with a small number (less than 100) of patients, Single center studies etc.
- Establish process for assessing request eligibility
  - Define a regulatory questionnaire to identify which submission or studies are eligible for data sharing.

Challenges in Data Sharing

Study request approval

- Questionnaire may include the below criteria:
  - Is the requested study complete?
  - Is the requested study in Phase II – IV?

- Independent review panel to determine the following:
  - Validity of research
  - Conflict of interest
  - Funding

Challenges in Data Sharing

What do we share?

- What do we share:
  - Original protocol and any amendments
  - Original documentation and amendments that articulate statistical methodology
  - CSR (Redacted) appendices
  - Annotated CRF
  - Dataset specifications
  - Anonymized raw study datasets – collected data from each patient in the study
  - Anonymized analysis-ready datasets – data used for analysis

Challenges in Data Sharing

Informed Consent

- Informed Consents have changed over time and may be restricting the use of the data only for the study in question.
  - Moving forward, consent to anonymize the data for use beyond the scope of the trial is added to the ICF.
- Individual Ethics Committees can propose alterations to ICF and these are not tracked.
  - Management of ICF needs to be assessed as many trials do not use the standard ICF but use the site ICF or their own version of an ICF – Need to ensure anonymization is inserted into these ICF's without exception.
- Does anonymization remove the issue of ICF? Currently a legal discussion.

Challenges in Data Anonymization

Data Anonymization Standards

- What data needs to be anonymized?
  - All variables which contain personal identifying information (PII's) should be anonymized.

- Are there rules?
  - HIPAA (only applies to US) provided some high level considerations but additional definitions and specifications are needed in EU. Eg: height/weight; dealing with small subgroups (center, gender); handling dates

- It is easy to anonymize data but to create an anonymized database that maintains the within study and within patient data relations needed for analyses is more difficult.

Challenges in Data Anonymization

Data Anonymization Standards

- When US clinical data is de-identified according to this process, it may be disclosed to a third party. De-identification requires the removal of 18 identifiers of which a few are listed below:
  - Names
  - Telephone numbers
  - Fax numbers
  - Email addresses
  - Social security numbers

Challenges in Data Anonymization

Data Anonymization Standards

- De-identification under US Privacy rules are **not acceptable** for anonymization standards in the European Union or in Switzerland.
  - Problem - Quasi identifiers (indirect identifiers)
  - 14 additional identifiers. Few listed below:
    - Initials
    - Names of relatives
    - Place of treatment and treating physician
    - Rare disease or treatment for same

Challenges in Data Anonymization

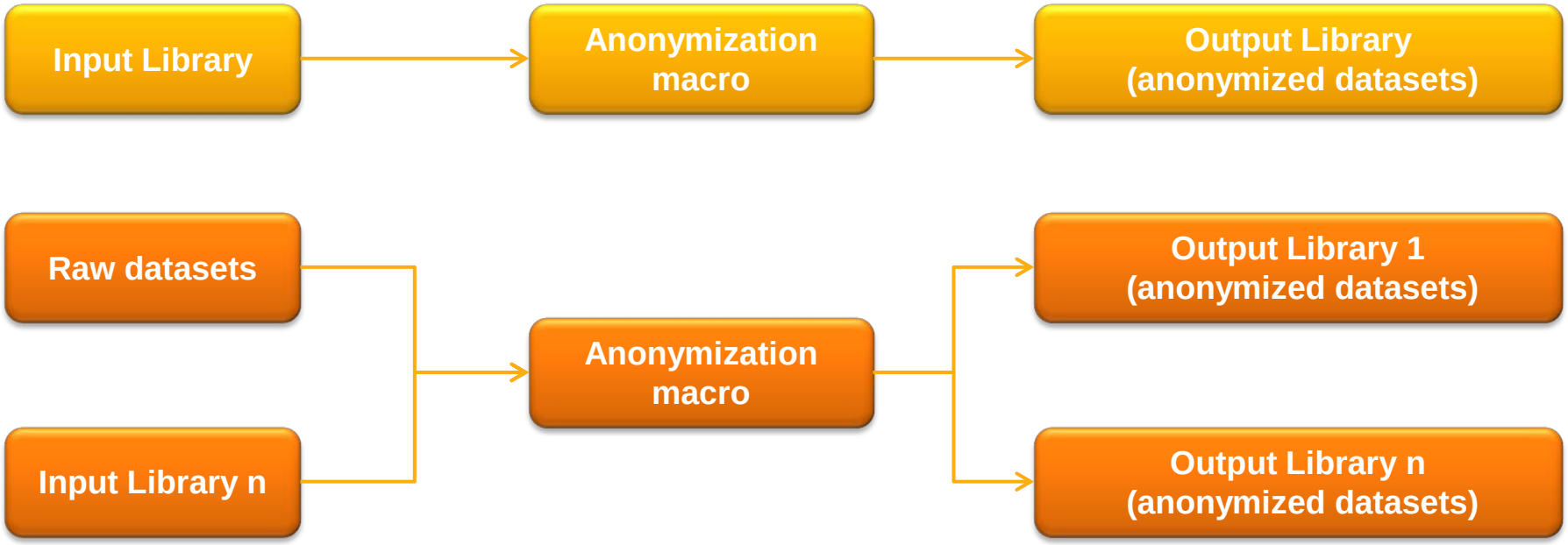
Cross-divisional approach

- One process designed to anonymize study data from all divisions. Involves:
  - Different Study environment
    - Macro is designed to work on multiple operating systems
  - Heterogeneity of data
    - Data is shared in its native format. No conversion to match CDISC or SDTM Standards
- Requires documentation that covers all the divisions.
  - Involves:
    - Work practices
    - Training and User guidance documents
    - Standards documents
    - Risk assessment (against hacking the lock box and future risk of re-identification (Article 29 of Directive 95/46/EC Data Protection Working Party WP216 Anonymization Techniques released April 2014))
    - Business impact document etc.,

Challenges in Data Anonymization

Anonymization Macro

- Requires a validated macro that can successfully anonymize any given data.
- Macro should have the ability to anonymize one or more input libraries in one run.



Challenges in Data Anonymization

Anonymization Macro

- Keep consistent translations between related datasets form different libraries.
  - Ex:
    - USUBJID of a study should be consistently anonymized across all the datasets in Raw dataset library and analysis dataset library.
    - Date offset in a core study should match the date offset value in extension study.

Challenges in Data Anonymization

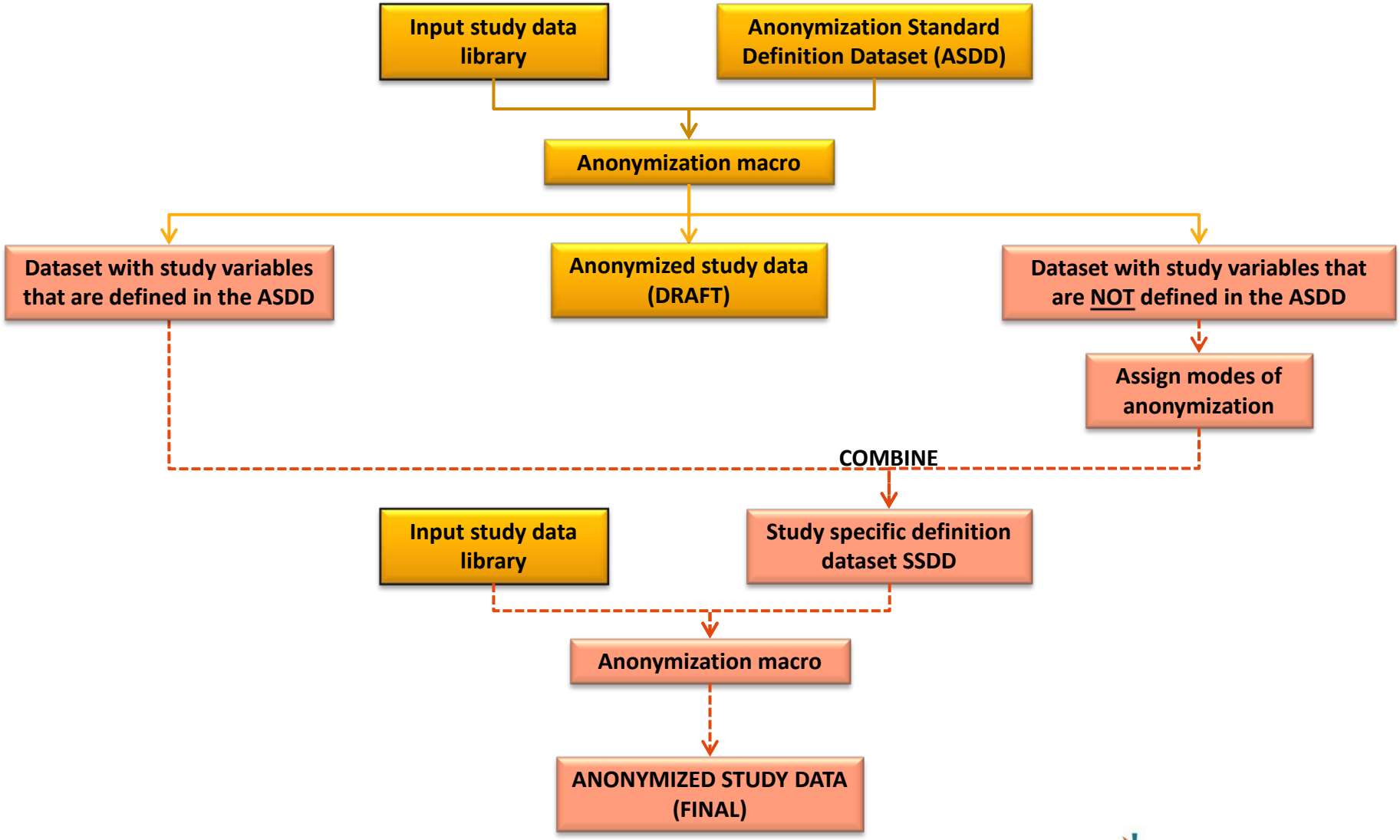
Definition Dataset

- Is it possible to build a definition dataset that applies to all studies across all the divisions?
  - Study datasets, prior to CDISC and SDTM standards, may have common variables that differ in their attributes and content between different datasets.
    - Ex: AGE can be either numeric or character
- So, a Study Specific Definition Dataset (SSDD) can be built in combination with the Standard Definition Dataset (SDD) which covers the anonymization definitions of all the study variables.
- SDD??? – A definition dataset that contains the definitions for the most recurrent variables which have a definite anonymization mode.
  - Ex: NAME – DROP

    USUBJID - TRANSLATE

Overview of Anonymization process

Proposed solution for creating a Study specific definition dataset



NOVARTIS

Challenges in Data Anonymization

Macro validation and Approval process

- Validating the process (vs) Validating the macro
  - Can we validate the macro with out validating or standardizing the input definition dataset?

- QM's perspective
  - Employing Anonymization macro to modify patient data involves the risk of revealing patient's identity to the public and hence is considered a GxP practice.
  - So, for validating a GxP system, the processes around the macro should also be defined and validated.

    Ex: A change management process for the input definition dataset needs to be performed.

The Tripartite Strategy

Novartis

- Novartis directs and controls all major activities of the Novartis Global Data Transparency initiative
- The decisions that effect other divisions are carefully reviewed and discussed with the Cross-divisional work stream in periodical Cross-divisional work stream meetings.
- The criteria for approving a data sharing request and the process of setting up a website where a researcher can view his request are set forth by Novartis.
- All the project relevant documentation is managed and controlled by the appropriate Novartis personnel.

The Tripartite Strategy

MMS Holdings Inc.

- Macro validation
  - Includes URS/FSDS/TP, IQ/OQ/PQ test scripts, test programs, Work practice documents, validation reports.
- Training
  - Creates division specific training and brings back knowledge to Novartis.
- Anonymize pilot studies
  - Co-ordinate with various divisions at Novartis to access their data and perform a pilot anonymization project on at least study in Test environment
- Perform UAT's for transferring the anonymized data to SAS Lock box.
  - Co-ordinated with the SAS Institute to attain the permissions required to successfully upload and access data from the SAS Lock-box.
  - Researcher accounts were created and tested to having controlled access and being able to perform analysis in the Secure repository – SAS Environment.

The Tripartite Strategy

The SAS Institute

- Configure Secure SAS Repository – “SAS Lock box”
- Train and assist MMS and Novartis personnel involved in uploading the anonymized data to the secure repository.
- Assist in performing User acceptance testing (UAT) of the Secure SAS repository.
- Available for trouble shooting any future issues.

<<<<<<<<<<<<<<<<<<<<<<<<<< END >>>>>>>>>>>>>>>>>>>>>>>>>>

NOVARTIS

Q & A

Thank You