臨床試験データへのオープンアクセス

中嶋　優一
ノバルティスファーマ株式会社
データサイエンス・サイエンティフィックオペレーション部

A New Era: Open Access to Clinical Trial Data

Yuichi Nakajima
Data Sciences and Scientific Operations Dept.,
Novartis Pharma K.K.

要旨：

臨床試験データの二次利用は医療の発展に貢献する一方で、被験者データのプライバシー保護の問題が重要となる。本発表では弊社におけるSASを用いたデータの匿名化プロセスについて紹介する。

キーワード：Anonymization, deidentification

Today's topics

1- Fundamentals

- Background and Multisponsor site
- Scope of Documents and Data shared

2- Challenges

- Informed Consent
- Data Anonymization Standards
- Cross-divisional approach
- New Business process and documents involved

3- Implementation

- Overview of Data Sharing Process
- De-identification (vs) Anonymization
- Overview of Anonymization process
- Modes of Anonymization
- Result of Anonymization process

4- Operating Model

- The tripartite strategy: Novartis / MMS Holdings Inc. / The SAS Institute

Today's topics

1- Fundamentals

- Background and Multisponsor site
- Scope of Documents and Data shared

2- Challenges

- Informed Consent
- Data Anonymization Standards
- Cross-divisional approach
- New Business process and documents involved

3- Implementation

- Overview of Data Sharing Process
- De-identification (vs) Anonymization
- Overview of Anonymization process
- Modes of Anonymization
- Result of Anonymization process

4- Operating Model

- The tripartite strategy: Novartis / MMS Holdings Inc. / The SAS Institute

Data Transparency is coming

BMJ OPEN DATA CAMPAIGN

Transparency measures forced on pharma
Chemistry World - Sep 30, 2014
The same is not true for clinical **data**, where **transparency** will have legal force. This is despite the EFPIA and its US counterpart PhRMA ...

Google data transparency

Web Images News Videos

About 96,700,000 results (0.34 seconds)

Legislative reform of China's healthcare
Lexology (registration) - Aug 31, 2014
While developments in high-profile **anti-bribery** investigations into both domestic ...
Indeed, all **pharmaceutical** and medical device companies ...

EU Ombudsman: Clinical trials face new **transparency** challenges
EurActiv - Sep 30, 2014
The Ombudsman warning came just month after the EU adopted a new clinical trials regulation, which obliges **pharmaceutical** companies to share scientific **data** on new medications submitted for approval to the European Medicines Agency (EMA). O'Reilly ...

New EMA policy on access to clinical trial **data** set to be finalised this week
Out-Law.com - Sep 30, 2014
Expert in life sciences Helen Cline of Pinsent Masons said: "There is a whole spectrum of benefits that could arise from making the clinical trial process and clinical trial **data** more **transparent** from improving the efficiency of drug discovery to the public health benefits." Earlier this year, the Council of Ministers and the European Parliament voted in favour of a new Clinical Trials Regulation which will require **pharmaceutical** companies and other medical researchers to post results of all their European clinical trials on a ...

Data Transparency is coming

厚生労働省 個人情

「日本再興戦略」などにおける「医療情報データベースの活用」に関する記述

日本再興戦略（平成25年6月14日）…62p

第Ⅱ　3つのアクションプラン
二．戦略市場創造プラン
テーマ1：国民の「健康寿命」の延伸
(2)　個別の社会像と実現に向けた取組み
①　効果的な予防サービスや健康管理の充実により、健やかに生活し、老いることができる社会
Ⅱ）　解決の方向性と戦略分野（市場・産業）及び当面の主要施策
○医療・介護情報の電子化の促進
・医薬品の副作用データベースシステムについて、データ収集の拠点となる病院の拡充や地域連携の推進を図ることにより、利活用できる十分な情報を確保し、医薬品の有効性・安全性評価や健康寿命の延伸につなげる。

世界最先端IT国家創造宣言（平成25年6月14日）…12p

Ⅲ．目指すべき社会・姿を実現するための取組
2．健康で安心して快適に生活できる、世界一安全で災害に強い社会
（1）　適切な地域医療・介護等の提供、健康増進等を通じた健康長寿社会の実現
②　現役世代からの健康増進等、医療・健康情報等の各種データの活用推進
・・・・・・・・・・・・・・・医療情報データベースを活用した医薬品等の安全対策に関する取組を推進できるようにするなど、2016年度までに、地域や企業における国民の健康増進・健康管理に有効な方策を確立し、それを踏まえて、全国展開を図る。

10

Novartis joins *ClinicalStudyDataRequest.com*

ClinicalStudy DataRequest.com

Registered Users, Please Login

HOME | STUDY SPONSORS | STEP BY STEP | MY REQUESTS | LOGIN OR CREATE AN ACCOUNT | METRICS | HELP

About

This site

Access to clinical trial data provides opportunities to conduct further research that can help advance medical science or improve patient care. This helps ensure the data provided by research participants are used to maximum effect in the creation of knowledge and understanding.

Researchers can use this site to request access to anonymised patient level data and/or supporting documents from clinical studies to conduct further research.

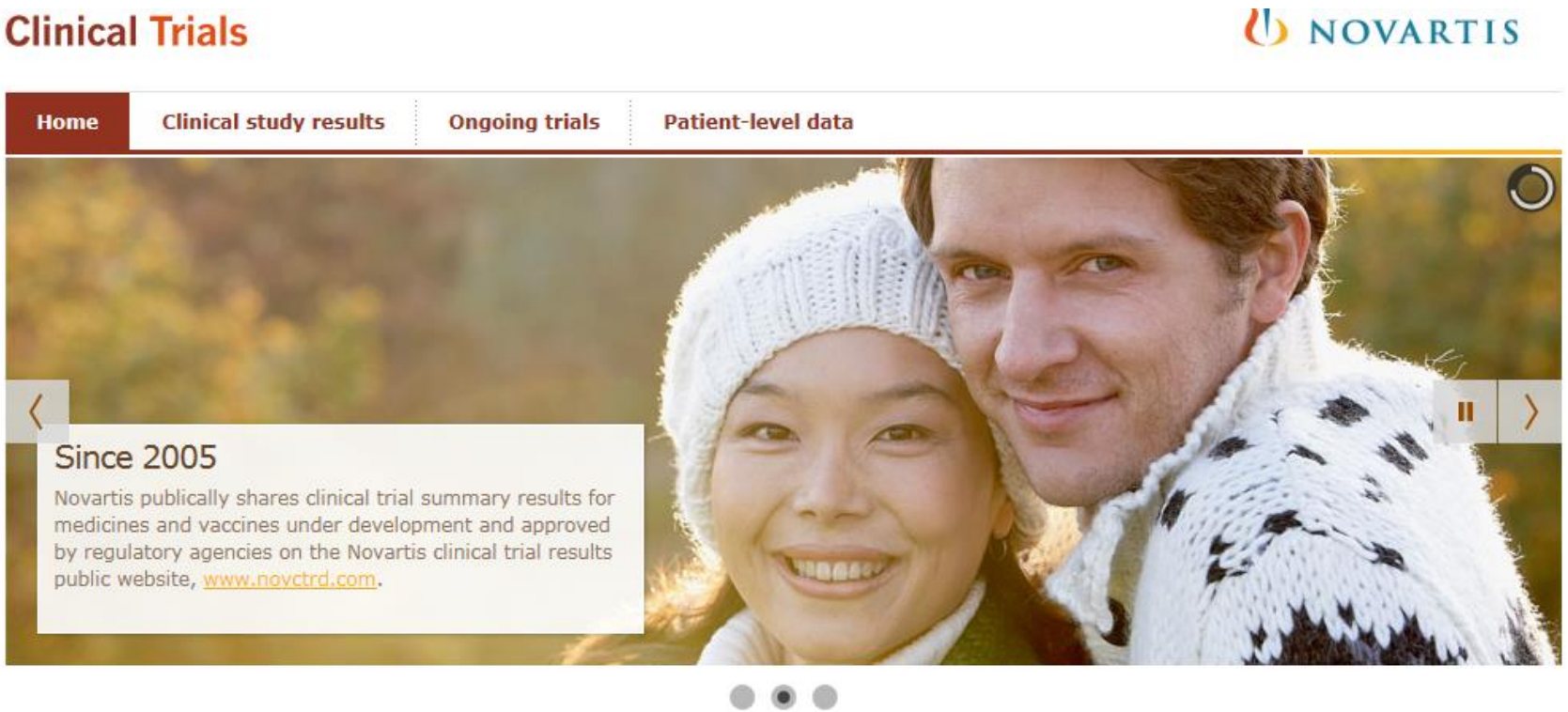
Next steps

Study sponsors who have committed to use this site are **Astellas, Bayer, Boehringer Ingelheim, Eisai, GSK, Lilly, Novartis, Roche, Sanofi, Takeda, UCB** and **ViiV Healthcare**.

Other clinical trial sponsors and funders are invited to join with the aim of transitioning to a fully independent system which allows access to data from clinical trials conducted by multiple companies and organisations. It is hoped that such a system will be put in place as soon as possible.

If you are a study sponsor interested in listing studies on this site, contact information is provided here.

Novartis Internet Site



Scope of Documents and Data shared

- After a careful review of the request by an independent review panel, the following data and accompanying trial documentation will be shared with qualified external researchers when available.

Document	Divisional Responsibility
Annotated CRF	Data Management
Original Protocol and any amendments	Clinical Office
Dataset Specifications	Stats & Programming
Original Reporting and Analysis Plan	Stats & Programming
Anonymized raw study datasets	**Stats & Programming**
Anonymized analysis-ready datasets	**Stats & Programming**
CSR excluding appendices since these are in patient level data	Medical Writing

Today's topics

1- Fundamentals

- Background and Multisponsor site
- Scope of Documents and Data shared

2- Challenges

- Informed Consent
- Data Anonymization Standards
- Cross-divisional approach
- New Business process and documents involved

3- Implementation

- Overview of Data Sharing Process
- De-identification (vs) Anonymization
- Overview of Anonymization process
- Modes of Anonymization
- Result of Anonymization process

4- Operating Model

- The tripartite strategy: Novartis / MMS Holdings Inc. / The SAS Institute

Challenges in Data Sharing

Informed Consent

Points of consideration

- Informed Consents have changed over time and may be restricting the use of the data only for the study in question – additionally individual Ethics Committees can propose alterations to ICF and these are not tracked.
- Does anonymization remove the issue of ICF? Currently a legal discussion. Moving forward consent to anonymize the data for use beyond the scope of the trial is added as an option to patients in the ICF. 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.

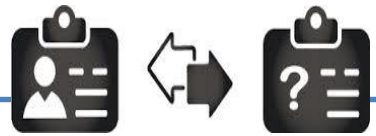
Challenges around Anonymization

Points of consideration for maintaining within study/patient data relations

Points of consideration

NVS Data
nymization Standa

- It is easy to de-identify data but to create an anonymized database that maintains the within study and within patient data relations needed for analyses is more difficult.
- To de-identify the data, we could simply drop all identifying data variables but that would result in a database that is not useful for analyses. There would be no dates, subject IDs, ages, etc.
- So, data needs to be anonymized while retaining the within study and within patient data relations. Dates need to be changed and the intervals between any two dates needs to be the same as in the original (/identified) data.
- Novartis approach is to anonymize (as opposed to de-identify) and destroy translation tables



Define a new Business process from scratch

Setup SOP/WP, including Standardization

Requires documentation that covers all the divisions

- This involves:
  - SOP and related Working Practices
  - Training and User guidance documents
  - Standards documents
  - Macro validation supportive documentation and Risk assessment (against hacking the lock box and future risk of re-identification with increased access to data through other channels, e.g. Social networks (Article 29 of Directive 95/46/EC Data Protection Working Party WP216 Anonymization Techniques released April 2014))

A single model for all Novartis Divisions*

Define a global standard but acknowledging Division specifics

Type and level of anonymization

- Define **Global Standards** for all Divisions
- Establish **one process flow** to anonymize study data from all divisions, involving different programming environment with multiple operating systems
- Heterogeneity of data format within and across Novartis divisions makes it difficult to establish a single model for data anonymization. **Data is shared in its native format**. No conversion to match CDISC or SDTM Standards

NOVARTIS ONCOLOGY

NOVARTIS GenMeds

Alcon

Today's topics

1- Fundamentals

- Background and Multisponsor site
- Scope of Documents and Data shared

2- Challenges

- Informed Consent
- Data Anonymization Standards
- Cross-divisional approach
- New Business process and documents involved

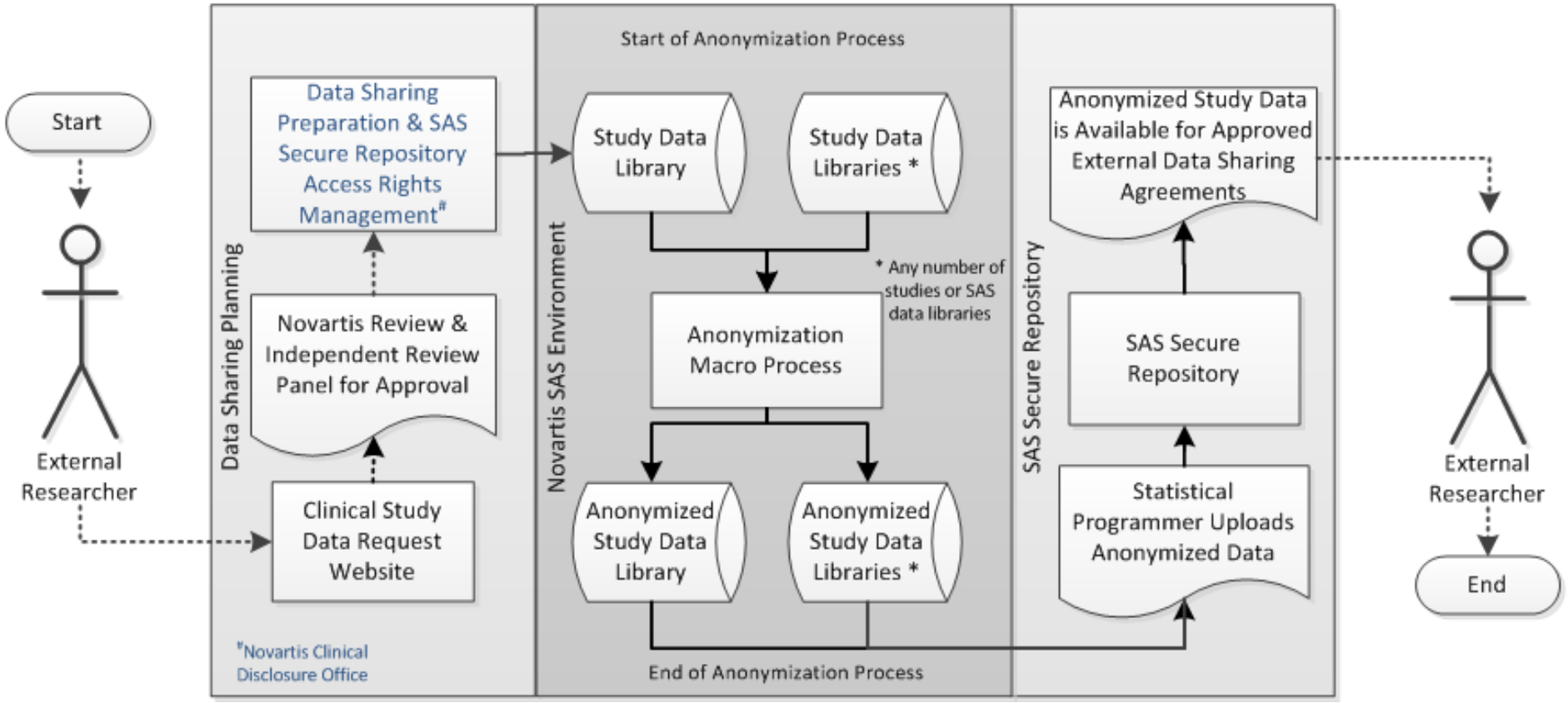
3- Implementation

- Overview of Data Sharing Process
- De-identification (vs) Anonymization
- Overview of Anonymization process
- Modes of Anonymization
- Result of Anonymization process

4- Operating Model

- The tripartite strategy: Novartis / MMS Holdings Inc. / The SAS Institute

Overview of Data Sharing Process



De-identification (vs) Anonymization

Original Dataset

VAR1	VAR2
SUBJ012	01
SUBJ012	02
SUBJ045	03
SUBJ045	04

VAR1	VAR1
SUBJ012	XXXXnn1
SUBJ012	XXXXnn1
SUBJ045	XXXXnn2
SUBJ045	XXXXnn2

Keys table

De-identified Dataset

VAR1	VAR2
XXXXnn1	
XXXXnn1	
XXXXnn2	
XXXXnn2	

Novartis Approach

Original Dataset

VAR1	VAR2
SUBJ012	01
SUBJ012	02
SUBJ045	03
SUBJ045	04

VAR1	VAR1
SUBJ012	XXXXnn1
SUBJ012	XXXXnn1
SUBJ045	XXXXnn2
SUBJ045	XXXXnn2

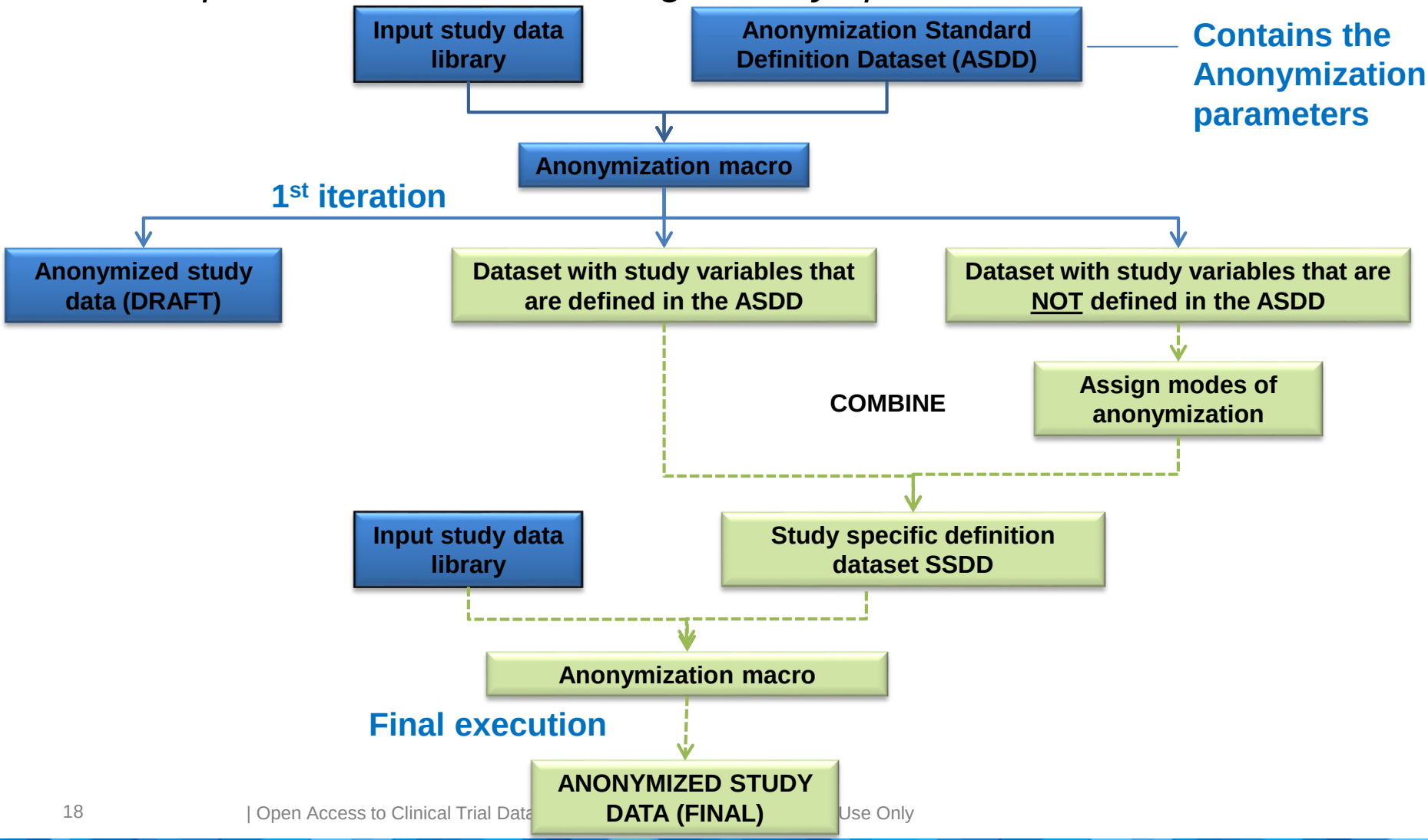
Keys table

Anonymized Dataset

VAR1	VAR2
XXXXnn1	
XXXXnn1	
XXXXnn2	
XXXXnn2	

Overview of Anonymization process

Proposed solution for creating a Study specific definition dataset



An outlook of the Definition Dataset

- List of all variables from all datasets to be anonymized

DATASET	VARIABLE	LABEL	MODE of anonymization	Identifier_ type	date_type	full_date	TYPE	FORMAT
AEFF2EVT	ACSDT	Date of 1st Hosp. for ACS	DATE	date	sasdate		1	DATE
ACMDATC	CMDEND1O	Concomitant med. end date (Oracle date)	DATE	datetime	sasdatetime		1	DATETIME
AAEV	AGECL65	Age group (<65,>=65)	DROP				1	AGE1F_
ACMD	AGECL65	Age group (<65,>=65)	DROP				1	AGE1F_
ACMDATC	AGECL65	Age group (<65,>=65)	DROP				1	AGE1F_
ACMP	ACTTRTC	Actual treatment code	NONE				2	$
ACOM	ACTTRTC	Actual treatment code	NONE				2	$
ADAR	ACTTRTC	Actual treatment code	NONE				2	$
AADJ	CTR1N	Center number	TRANSLATE				1	
AAEV	CTR1N	Center Number	TRANSLATE				1	
ABIO	SID1A	Subject Identifier	TRANSLATE				2	$
ABKG	SID1A	Subject Identifier	TRANSLATE				2	$
ACMPDTH	SID1A	Subject Identifier	TRANSLATE				2	$

- This is the place where the modes of anonymization are entered
- The Definition Dataset is standardized at the Division level and is maintained through a change management system (version history and approval process)

Modes of Anonymization

Examples

- 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	XXXXnn
2	XXXXnn
3	XXXXnn

Modes of Anonymization (cont.)

Examples

- 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

- 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 (straight copy of the variable)

** By default in the macro, if no mode is defined for a variable, the variable is dropped*

Result of Anonymization process

- **Example**

Study data example on top and anonymized data on bottom after ***modes of anonymization*** were applied.

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	TRANSLATE	MISSING	TRANSLATE	DROP	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		Ay12		57	16FEB2093	17MAR2093		Headache
Xnn10	nnnXn10		Ay12		57	28FEB2093	25MAY2093		Bronchitis
Xnn10	nnnXn10		Eb65		90 or Older	13MAY2093	30SEP2093		Nasopharyngitis
Xnn10	nnnXn10		Eb65		90 or Older	16MAY2093	19MAY2093		Influenza
Xnn10	nnnXn10		Eb65		90 or Older	19APR2093	03JUL2093		Pain
Xnn11	nnnXn11		Nz97		64	02DEC2092	08DEC2093		Pain
Xnn11	nnnXn11		Nz97		64	12JUL2093			Bronchitis
Xnn11	nnnXn11		Nz97		64	19APR2093	03MAY2093		Pain

Results: Real data Vs. anonymized data

Table 14.3-x.x (Page 1 of 1)
Overall exposure to study drug
Safety set

Statistic		Trt 1 N=338	Trt 2 N=339	Trt 3 N=328
Exposure (in days)	n	338	339	328
	Mean(SD)	322.5 (102.54)	337.2 (87.58)	340.1 (83.51)
	Median	365.0	365.0	365.0
	Min - Max	5.0 - 407.0	1.0 - 400.0	2.0 - 400.0
Categorized exposure				
<= 4 weeks	n(%)	16 (4.7)	12 (3.5)	12 (3.7)
> 4 - 12 weeks	n(%)	11 (3.3)	8 (2.4)	4 (1.2)
> 12 - 28 weeks	n(%)	19 (5.6)	12 (3.5)	8 (2.4)
> 28 - 40 weeks	n(%)	8 (2.4)	3 (0.9)	5 (1.5)
> 40 - 53 weeks	n(%)	264 (78.1)	286 (84.4)	281 (85.7)
> 53 weeks	n(%)	20 (5.9)	18 (5.3)	18 (5.5)

		trt 1	trt 2	trt 3
N		338.0	339.0	328.0
Exposure (in days)	n	338	339	328
	Mean	322.5	337.2	340.1
	Standard Deviation	102.54	87.58	83.51
	Median	365.00	365.00	365.00
	Min	5.0	1.0	2.0
	Max	407.0	400.0	400.0
Categorized exposure				
<= 4 weeks	n	16.0	12.0	12.0
	%	4.7	3.5	3.7
> 4 - 12 weeks	n	11.0	8.0	4.0
	%	3.3	2.4	1.2
> 12 - 28 weeks	n	19.0	12.0	8.0
	%	5.6	3.5	2.4
> 28 - 40 weeks	n	8.0	3.0	5.0
	%	2.4	0.9	1.5

Today's topics

1- Fundamentals

- Background and Multisponsor site
- Scope of Documents and Data shared

2- Challenges

- Informed Consent
- Data Anonymization Standards
- Cross-divisional approach
- New Business process and documents involved

3- Implementation

- Overview of Data Sharing Process
- De-identification (vs) Anonymization
- Overview of Anonymization process
- Modes of Anonymization
- Result of Anonymization process

4- Operating Model

- The tripartite strategy: Novartis / MMS Holdings Inc. / The SAS Institute

The Tripartite Strategy

Novartis, MMS Holding, SAS

Define Strategy

- Establish overall approach
- Get alignment from all Divisions
- Define technical solution
- Create company Anonymization guidelines
- Overall Project management

Implementation

- Macro validation (include URS, IQ/OQ/PQ test scripts, WP etc).
- Develop Training
- Run pilot studies
- Perform UAT's on 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.

Technical solution

- Configure "SAS Lock box"
- Train and assist MMS and Novartis personnel
- Available for trouble shooting any future issues.

Conclusion

- In today's world, Data Transparency is both a risk and competitive advantage
- The question is no longer to comply or not but how will you prepare for it