

Authors Introduction



Shiva Gogu

Sr. Statistical Programmer



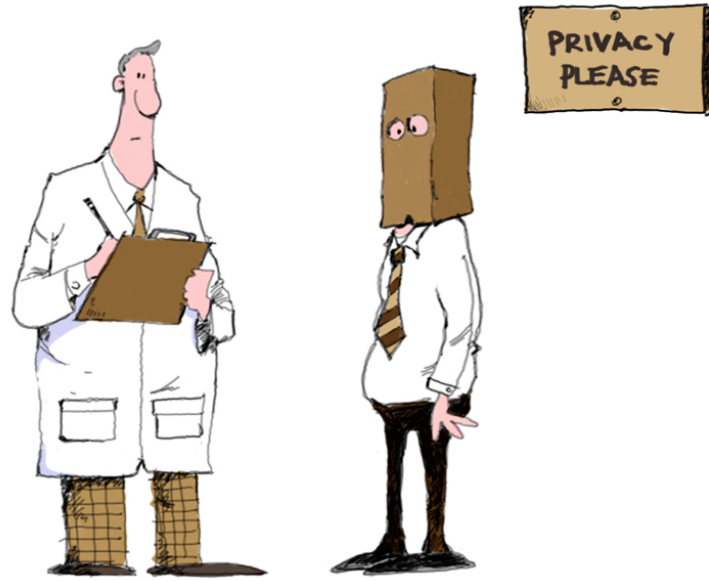
Patient Privacy – The Topmost Priority!

PhUSE US Connect 2020
11th Mar, 2020
Paper # CT08

Why Patient Privacy is important?



- To protect individual control over their personal health information
- To protect from Identity theft
- To protect from data breach
- To protect Patient information by keeping it safe and confidential



Unlike CSR (clinical study report) result disclosure, data disclosure has the high risk for patient privacy

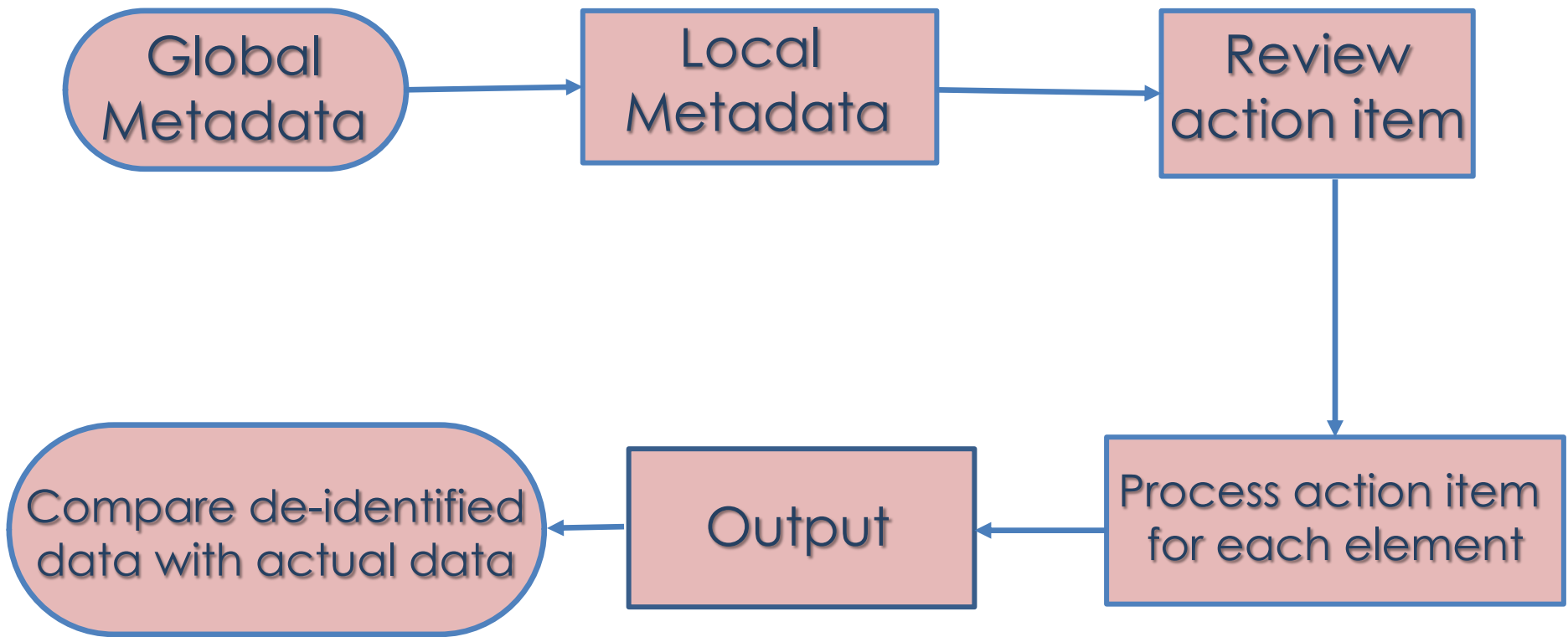
Health Insurance Portability Accountability Act

HIPAA rules require healthcare organizations to adopt processes and procedures to ensure the highest degree of patient confidentiality

How do we ensure to protect patient privacy?

By implementing data de-identification process with metadata approach using SAS®

Process Flow:



Sample Metadata:

MEMNAME	NAME	TYPE	VARNUM	LENGTH	LABEL	ACTION	COMMENT
DM	STUDYID	CHAR	1	9	Study Identifier	NONE	CRF Collected
DM	DOMAIN	CHAR	2	2	Domain Abbreviation	NONE	CRF Collected
DM	USUBJID	CHAR	3	17	Unique Subject Identifier	MERGE	Randomly assigned patient number for De-Identity
DM	SITEID	CHAR	4	10	Study Site Identifier	MERGE	Randomly assigned site number for De-Identity
DM	AGE	NUM	5	8	Age	GROUP	>=90 and <90
AE	STUDYID	CHAR	1	9	Study Identifier	NONE	CRF Collected
AE	DOMAIN	CHAR	2	2	Domain Abbreviation	NONE	CRF Collected
AE	USUBJID	CHAR	3	17	Unique Subject Identifier	MERGE	Randomly assigned patient number for De-Identity
AE	AESEQ	NUM	4	8	Sequence Number	NONE	
AE	AETERM	CHAR	5	70	Reported Term for the Adverse Event	DROP	Verbatim information will be removed from deidentify dataset.
AE	AEDECOD	CHAR	6	40	Lowest Level Term	NONE	No action needed for dictionary terms
AE	AESTDTC	CHAR	7	20	Start Date/Time of Adverse Event	DATE	All dates will be dropped after deriving Relative Day (or) Offset dates are applied.

Classification of Elements:



- Direct
- Indirect
- Outlier
- Free text

Classification of Action item for each element:



- None
- Merge
- Group
- Date
- Missing
- Drop

Output?

Before:

SITEID	INVID	INVNAME	USUBJID	AGE	AESTDTC	AEENDTC	AETERM	AEDECOD
2125	1254	James	ABCD_2343	65	18Dec2019	19Dec2019	Stomach ache	Abdominal pain upper
2125	1254	James	ABCD_2344	78	13Dec2019	15Dec2019	Worsening of headache	Headache
2125	1254	James	ABCD_2345	84	08Dec2019	08Dec2019	Right eye pain	Eye pain
4254	1651	Samuel	ABCD_2346	91	09Dec2019	28Dec2019	Common cold	Nasopharyngitis
4254	1651	Samuel	ABCD_2347	55	04Dec2019		Fever	Pyrexia
4254	1651	Samuel	ABCD_2348	95	15Nov2019	08Dec2019	Knee pain	Arthralgia

Site, Investigator and Subject identifiers are re-assigned with random values

After:

Age values have been categorized per HIPAA guidelines

SITEID	INVID	INVNAME	USUBJID	AGEGRP	AESTDTC	AEENDTC	AETERM	AEDECOD
1011	412		ABCD_51285	<90	17Mar2020	18Mar2020		Abdominal pain upper
1011	412		ABCD_62289	<90	12Mar2020	14Mar2020		Headache
1011	412		ABCD_73590	<90	07Mar2020	07Mar2020		Eye pain
1114	515		ABCD_84615	>=90	08Mar2020	27Mar2020		Nasopharyngitis
1114	515		ABCD_98845	<90	03Mar2020			Pyrexia
1114	515		ABCD_99586	>=90	13Feb2020	07Mar2020		Arthralgia

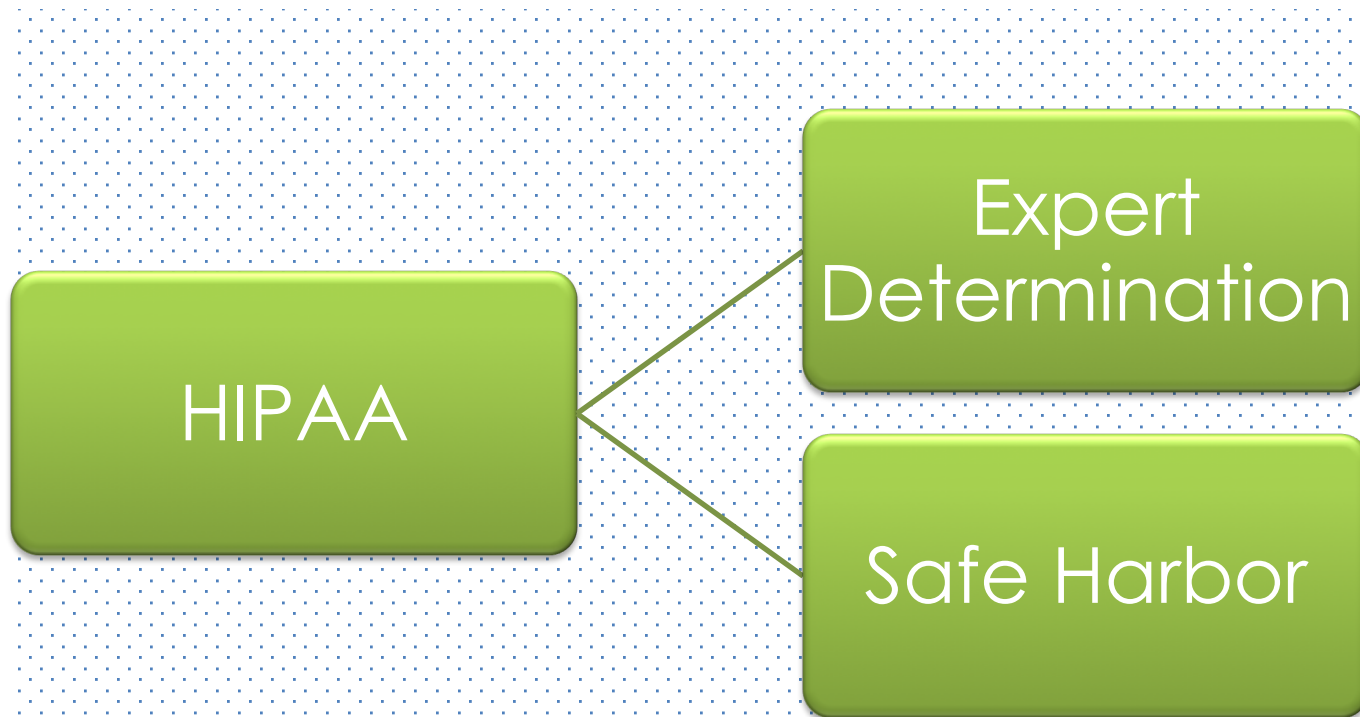
Onset of 3 months applied

AETERM set to missing

AEDECOD is retained

What does guideline say?

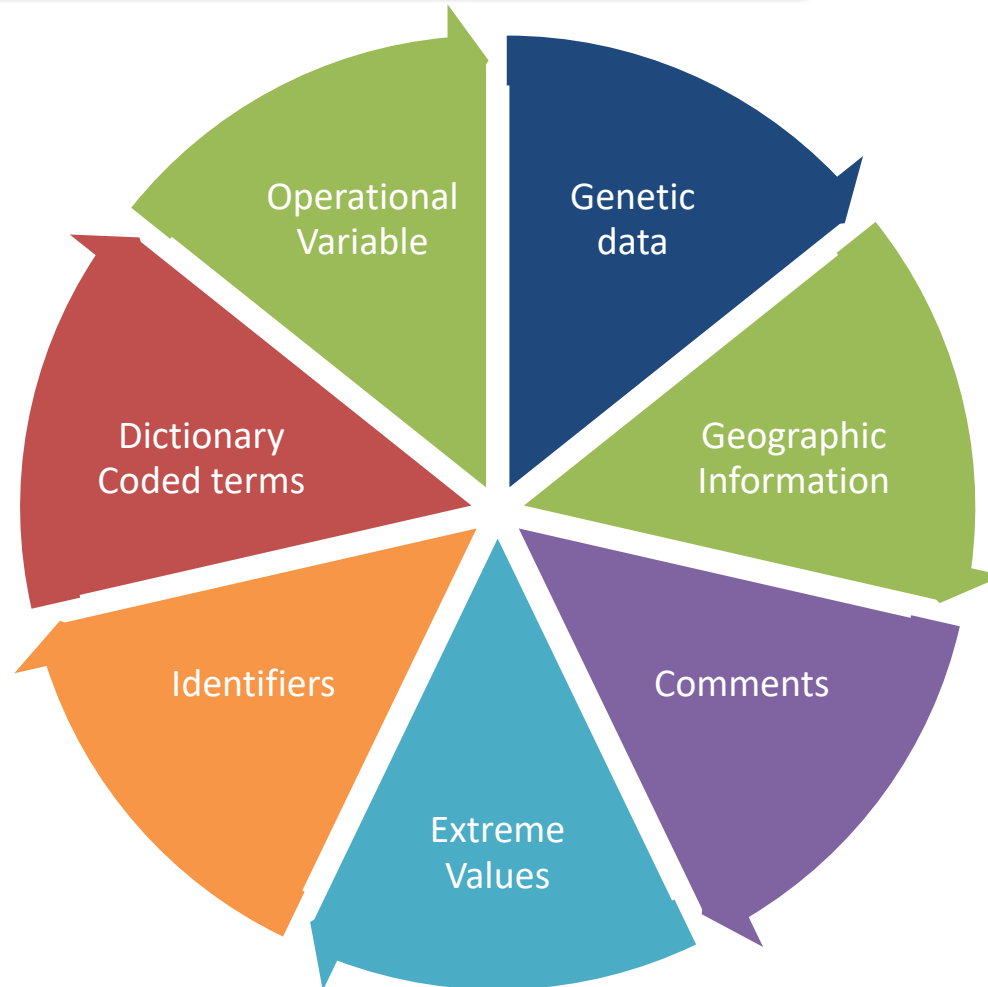
Two methods to achieve data de-identification in accordance with the HIPAA Privacy Rule:



The following 18 identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

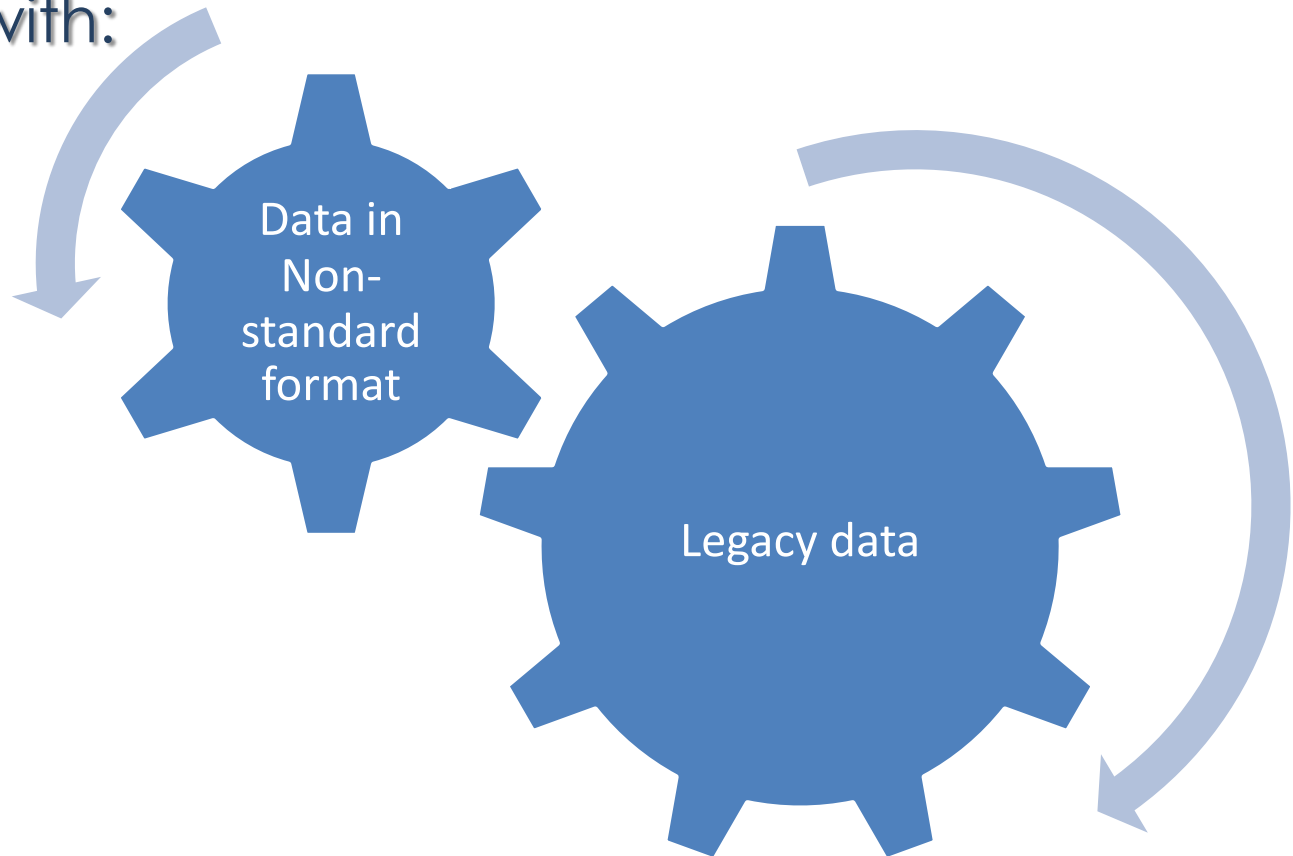
- Names
- All geographic information
- All elements of dates
- Telephone numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Fax numbers
- Device identifiers and serial numbers
- Email addresses
- Web Universal Resource Locators (URLs)
- Social security numbers
- Internet Protocol (IP) addresses
- Medical record numbers
- Biometric identifiers, including finger and voice prints
- Health plan beneficiary numbers
- Full-face photographs and any comparable images
- Account numbers
- Any other unique identifying number, characteristic, or code
- Certificate/license numbers

Important points to be considered:



Challenges:

To deal with:



Data de-identification vs Data Anonymization?



Conclusion:

- Protection of patient privacy is equally important as patient safety in clinical trials. Hence, implementing efficient steps while adhering to the applicable guidelines is important before sharing the trial data.
- Process discussed here is one of the multiple ways to de-identify the data in order to protect patient personally identifiable information.



Shiva Gogu

Vita Data Sciences

281 Winter St.

Waltham, MA

Phone: +1 (781) 833-0256

E-mail: sgogu@vitadatasciences.com