PharmaSUG 2015 - Paper DS10

PhUSE De-Identification Working Group: Providing De-Identification Standards to CDISC Data Models

Jean-Marc Ferran, Qualiance & PhUSE, Copenhagen, Denmark
Jacques Lanoue, Novartis, East Hanover, New Jersey

ABSTRACT

In this era of Data Transparency and sharing data with researchers, companies are defining their processes and de-identification guidance in order to comply with data privacy regulations. In particular, it is possible for researchers to request access to data across sponsors and both the difference of data models and de-identification techniques may make the analyses cumbersome and error-prone.

While the CDISC data models are now adopted in the industry, PhUSE launched in July 2014 a dedicated Working Group to define de-identification standards for CDISC data models starting with SDTM. Participants from Pharmaceuticals, CROs, Software Vendors, CDISC specialists, Data Privacy specialists and Academia have joined forces to define a set of rules against SDTM to provide the industry with a consistent approach to data de-identification and increase consistency across anonymized datasets.

Each domains and variables holding potentially Personally Identifying Information (PII) have been rated in terms of impact on data privacy. Based on that rating the variables are allocated standard rules of de-identification, the rationale and the impact on data utility is documented., allocated rules to apply and their rationale and impact on data utility.

This presentation will elaborate on the Working Group main findings, the current deliverables and the perspective to take this first initiative to the next stage.

INTRODUCTION

There are current efforts by regulators such as EMA to examine how to make Individual Patient Data (IPD) from clinical trials shared more widely. Sponsors have started sharing data based on request proposals from researchers and:

- Data is presented in different data models
- Each company seems to be defining their own high-level guidelines for data de-identification ([7])
- It is possible to request data from different companies within same research proposal

This new process and concept is posing a number of challenges:

- Is the applied data de-identification sufficient to protect the privacy of the patient?
- Will the researchers be able to use the data for the purpose of their research request?
- How much resources are required to both support the data de-identification process and assist researchers with their tasks?

Since its inception, PhUSE has been an advocate of sharing and exchange of knowledge and is supporting the Data Transparency initiative. A working group was set-up back in July 2014 with the goal to define data de-identification standards for CDISC data models. The first deliverable is focusing on CDISC SDTM 3.2 [1] and was finalized in April 2015 (See De-Identification Standards for CDISC SDTM 3.2 [0]).

The PhUSE de-identification standards aim at:

1. Facilitate the identification of direct and quasi identifiers (See Definitions tab of [0])
2. Provide rules to apply together with technical guidelines
3. Ensure consistency in data de-identification across sponsors

This paper elaborates on the approach, the key principles and important areas for data de-identification that are addressed in the PhUSE De-Identification Standards for CDISC SDTM 3.2.

APPROACH

This PhUSE De-Identification Standards for CDISC SDTM 3.2 has been written in the context of the data transparency initiative in the pharmaceutical industry and assumes the following main requirements are also being met as part of the data disclosure process:

1. Anonymized data is shared with researchers through a secure portal where the download and upload of data is controlled and is the responsibility of the sponsor.
2. Data sharing agreements are signed between sponsors and researchers and commit the researchers not to attempt to re-identify patients, download data or carry out analyses outside the approved research request, attempt to contact any of the participants or presumed participants, link the data with other data sets that they may have access to, and provide access to the portal to someone else.
3. The researchers have privacy practices in place at their institution.

In case of public data sharing, stricter measures must be applied in terms of data de-identification. This is discussed in paragraph "Public Data" in Appendix 2.

The PhUSE De-Identification Standards for CDISC SDTM 3.2 considers both pro-active data de-identification outside the context of an approved research request and reactive data de-identification within the context of a particular approved research request. For each variable identified as a direct or quasi identifier, both a primary and alternative rules are suggested. The primary rule was defined mostly in the context of pro-active data de-identification maximizing data utility. While the alternative rule was defined to address special cases and reactive data de-identification. It is assumed that sponsors verify the data utility is adequate in general for proactive data de-identification or for a given research request for reactive data de-identification.

It must also be noted that documents (e.g. CSR) disclosed together with the de-identified data must be redacted in a manner consistent with the data.

Low frequency of variables values poses a risk of re-identification. It is generally recommended that after the application of the rules described in this document, the data be examined a second time to identify those risks. When implementing a de-identification process sponsors should consider developing policies to address low frequencies in their studies, how to measure it, and the types of generalizations or suppressions that they would apply, as well as methods to automate that process.

The following other aspects of data de-identification and data sharing are not directly addressed in the deliverable and are the responsibility of the sponsors to define. These items are important and need to be implemented by the sponsor in conjunction with this standard:

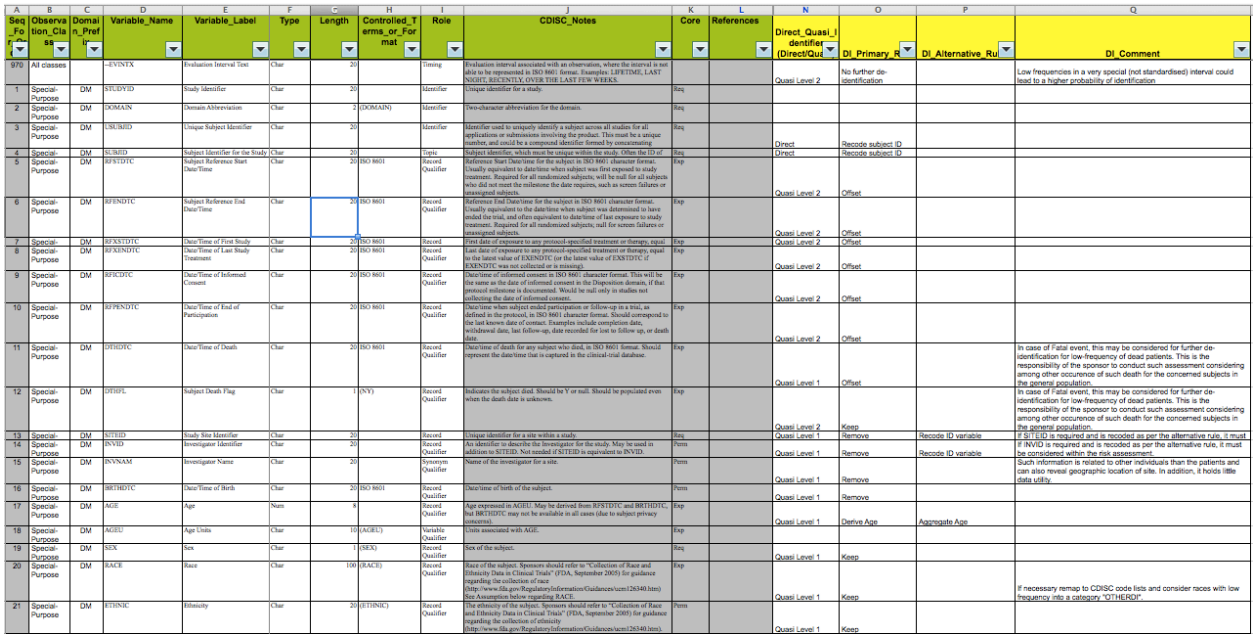
1. The processes that support data transparency implementation (Use of Independent Review Boards, deletion of the keys, provision of full database or subset fitted to the request, etc.). Note that TransCelerate [4] provides guidance in these areas.
2. The process to assess residual risk in de-identified data. One methodology is described in more detail in the IOM report [5] and some guidance is provided in section "Low frequency" in "Decisions" tab and Appendix 2 of the deliverable.
3. The actual documentation of the de-identification that was applied.
4. The definition and approach to the de-identification of sensitive data.

The definitions, decisions and assessments in the deliverable represent the consensus of the working group.

PHUSE DE-IDENTIFICATION STANDARDS FOR SDTM 3.2

The deliverable consists of an MS Excel spreadsheet (See at http://www.phuse.eu/Data_Transparency.aspx) with different tabs:

- Cover tab: Document information.
- Intro tab: Introduction including background, important considerations, out of scope, disclaimer and approach.
- Definitions tab: List of important terms with their definitions and examples when applicable.
- Decisions tab: Important areas with rational for decisions.
- Rules tab: The different rules to be applied together with technical guidance.
- SDTMIG tab: The SDTM 3.2 variables (1300+) together with their assessment for Direct/Quasi identifiers, Primary rules, Secondary rules and Comment for De-Identification. See Figure 1.
- References tab: Sources used for the elaboration of the deliverable.
- Appendices tab: Appendices for guidance on “Dates Offset” and “Low Frequency”
- Change log tab: Different versions of the deliverable together with list of changes.



	A	B	C	D	E	F	G	H	I	J	K	L	N	O	P	Q
	Seq_For	Observation_Class	Domain_Prefix	Variable_Name	Variable_Label	Type	Length	Controlled_Terms_or_Format	Role	CDISC_Notes	Core	References	Direct_Quasi_Identifier (Direct/Quasi	DI_Primary_R	DI_Alternative_Ru	DI_Comment
970	All classes		--EVINTX	Evaluation Interval Text	Char	20		Timing	Evaluation interval associated with an observation, where the interval is not able to be represented in ISO 8601 format. Examples: LIFETIME, LAST NIGHT, RECENTLY, OVER THE LAST FEW WEEKS.			Quasi Level 2	No further de-identification		Low frequencies in a very special (not standardised) interval could lead to a higher probability of identification
1	Special-Purpose	DM	STUDYID	Study Identifier	Char	20		Identifier	Unique identifier for a study.	Req					
2	Special-Purpose	DM	DOMAIN	Domain Abbreviation	Char	2	(DOMAIN)	Identifier	Two-character abbreviation for the domain.	Req					
3	Special-Purpose	DM	USUBJID	Unique Subject Identifier	Char	20		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. This must be a unique number, and could be a compound identifier formed by concatenating	Req		Direct	Recode subject ID		
4	Special-Purpose	DM	SUBJID	Subject Identifier for the Study	Char	20		Topic	Subject identifier, which must be unique within the study. Often the ID of	Req		Direct	Recode subject ID		
5	Special-Purpose	DM	RFSTDTC	Subject Reference Start Date/Time	Char	20	ISO 8601	Record Qualifier	Reference Start Date/time for the subject in ISO 8601 character format. Usually equivalent to date/time when subject was first exposed to study treatment. Required for all randomized subjects; will be null for all subjects who did not meet the milestone the date requires, such as screen failures or unassigned subjects.	Exp		Quasi Level 2	Offset		
6	Special-Purpose	DM	RFENDTC	Subject Reference End Date/Time	Char	20	ISO 8601	Record Qualifier	Reference End Date/time for the subject in ISO 8601 character format. Usually equivalent to the date/time when subject was determined to have ended the trial, and often equivalent to date/time of last exposure to study treatment. Required for all randomized subjects; null for screen failures or unassigned subjects.	Exp		Quasi Level 2	Offset		
7	Special-Purpose	DM	RFXSTDTC	Date/Time of First Study Treatment	Char	20	ISO 8601	Record Qualifier	First date of exposure to any protocol-specified treatment or therapy, equal	Exp		Quasi Level 2	Offset		
8	Special-Purpose	DM	RFXENDTC	Date/Time of Last Study Treatment	Char	20	ISO 8601	Record Qualifier	Last date of exposure to any protocol-specified treatment or therapy, equal to the latest value of EXENDTC (or the latest value of EXSTDTC if EXENDTC was not collected or is missing).	Exp		Quasi Level 2	Offset		
9	Special-Purpose	DM	RFICDTC	Date/Time of Informed Consent	Char	20	ISO 8601	Record Qualifier	Date/time of informed consent in ISO 8601 character format. This will be the same as the date of informed consent in the Disposition domain, if that protocol milestone is documented. Would be null only in studies not collecting the date of informed consent.	Exp		Quasi Level 2	Offset		
10	Special-Purpose	DM	RFPENDTC	Date/Time of End of Participation	Char	20	ISO 8601	Record Qualifier	Date/time when subject ended participation or follow-up in a trial, as defined in the protocol, in ISO 8601 character format. Should correspond to the last known date of contact. Examples include completion date, withdrawal date, last follow-up, date recorded for lost to follow up, or death date.	Exp		Quasi Level 2	Offset		
11	Special-Purpose	DM	DTHDTC	Date/Time of Death	Char	20	ISO 8601	Record Qualifier	Date/time of death for any subject who died, in ISO 8601 format. Should represent the date/time that is captured in the clinical-trial database.	Exp		Quasi Level 1	Offset		In case of Fatal event, this may be considered for further de-identification for low-frequency of dead patients. This is the responsibility of the sponsor to conduct such assessment considering among other occurence of such death for the concerned subjects in the general population.
12	Special-Purpose	DM	DTHFL	Subject Death Flag	Char	1	(NY)	Record Qualifier	Indicates the subject died. Should be Y or null. Should be populated even when the death date is unknown.	Exp		Quasi Level 2	Keep		In case of Fatal event, this may be considered for further de-identification for low-frequency of dead patients. This is the responsibility of the sponsor to conduct such assessment considering among other occurence of such death for the concerned subjects in the general population.
13	Special-Purpose	DM	SITEID	Study Site Identifier	Char	20		Record Qualifier	Unique identifier for a site within a study.	Req		Quasi Level 1	Remove	Recode ID variable	If SITEID is required and is recoded as per the alternative rule, it must
14	Special-Purpose	DM	INVID	Investigator Identifier	Char	20		Record Qualifier	An identifier to describe the Investigator for the study. May be used in addition to SITEID. Not needed if SITEID is equivalent to INVID.	Perm		Quasi Level 1	Remove	Recode ID variable	If INVID is required and is recoded as per the alternative rule, it must be considered within the risk assessment.
15	Special-Purpose	DM	INVNAM	Investigator Name	Char	20		Synonym Qualifier	Name of the investigator for a site.	Perm		Quasi Level 1	Remove		Such information is related to other individuals than the patients and can also reveal geographic location of site. In addition, it holds little data utility.
16	Special-Purpose	DM	BRTHDTC	Date/Time of Birth	Char	20	ISO 8601	Record Qualifier	Date/time of birth of the subject.	Perm		Quasi Level 1	Remove		
17	Special-Purpose	DM	AGE	Age	Num	8		Record Qualifier	Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC, but BRTHDTC may not be available in all cases (due to subject privacy concerns).	Exp		Quasi Level 1	Derive Age	Aggregate Age	
18	Special-Purpose	DM	AGEU	Age Units	Char	10	(AGEU)	Variable Qualifier	Units associated with AGE.	Exp					
19	Special-Purpose	DM	SEX	Sex	Char	1	(SEX)	Record Qualifier	Sex of the subject.	Req		Quasi Level 1	Keep		
20	Special-Purpose	DM	RACE	Race	Char	100	(RACE)	Record Qualifier	Race of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, September 2005) for guidance regarding the collection of race (http://www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm) See Assumption below regarding RACE.	Exp		Quasi Level 1	Keep		If necessary remap to CDISC code lists and consider races with low frequency into a category "OTHERDI".
21	Special-Purpose	DM	ETHNIC	Ethnicity	Char	20	(ETHNIC)	Record Qualifier	The ethnicity of the subject. Sponsors should refer to "Collection of Race and Ethnicity Data in Clinical Trials" (FDA, September 2005) for guidance regarding the collection of ethnicity (http://www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm).	Perm		Quasi Level 1	Keep		

Figure 1 - PhUSE De-Identification Standards for SDTM 3.2

The large number of variables to assess justified the choice of MS Excel as support media, also in the perspective of using the deliverable to automatize the data de-identification using software.

Every domain and variable defined in SDTM 3.2 (See SDTMIG tab of [0]) is assessed and variables that hold PII were evaluated in terms of data privacy and what rules to apply. Data privacy is defined across 3 levels, Direct Identifiers, Level 1 Quasi Identifiers and Level 2 Quasi Identifiers (See Definitions tab of [0]) of decreasing impact on data privacy and risk over patient re-identification. For these variables, rules to apply are recommended (See Rules tab of [0] for details). In some cases, it is recommended to keep them as-is as they hold critical data for analysis (e.g. Sex) or an alternative rule is suggested to address different cases (e.g. a device number may need to be recoded in a medical device study while not holding any data utility in a medicinal study where it can be removed). Note that columns A to L of SDTMIG tab of [0] come from CDISC SDTM 3.2 Implementation Guide [1]. The topics that the working group addressed and the associated decisions are included in the "Decisions" tab.

Note that the deliverable is in final draft status at the time this paper is written and may have changed. Please visit http://www.phuse.eu/Data_Transparency.aspx to download the latest version.

KEY PRINCIPLES

The PhUSE Data De-Identification Standards for SDTM 3.2 provides support in:

- Assessing the role of each variable in term of identification potential, quasi and direct identifiers across the SDTM domains
- Assigning rules for de-identification
- Understanding rational and address exceptions and special considerations

Direct & Quasi Identifiers are assessed

- **Direct identifiers**: One or more direct identifiers can be used to uniquely identify an individual. E.g. Subject ID, Social Security Number, Telephone number, Exact address, etc. It is compulsory to remove or provide a consistent substitute value for any direct identifier.
- **Quasi identifiers**: Quasi identifiers are background information that can be used in connection with other information to identify an individual with a high probability. E.g. Age at baseline, Race, Sex, Events, Specific Findings, etc.

Primary & Alternative Rules for De-Identification are assigned

- **Primary rule**: Rules maximizing data utility (at the exception of geographic location information that is key to decrease the residual risk) for pro-active data de-identification
- **Alternative rule**: Rules addressing a particular request in the case of reactive data de-identification and special cases
- **Impact on data utility** is evaluated qualitatively
- **Implementation guidance** for each rule is provided
- **Rules address different scenarios** rather than different implementation possibilities

Comments are added to guide the reader

- To explain further the **rational of a given assessment**
- To warn reader for **exceptions or special considerations**

IMPORTANT AREAS TO CONSIDER

This section addresses the different important areas to consider, the rational for the decisions that were made and the rules associated with the associated SDTM variables.

A number of variables types are at stake with regards data de-identification. HIPAA [2] provides guidance in what type of data to consider (Safe Harbor 18 identifiers). Clinical data typically holds more PII and these 18 identifiers are also extended to variables such as Race, Ethnicity, etc. In addition, Hrynaszkiewicz et al. [3] provides more guidance specific to clinical data.

As part of the development of this standard, the working group reviewed requirements from HIPAA [2], existing and available sponsors' anonymization standards ([6], [7]), ICO's Anonymisation code of practice [8] and methods proposed in the literature ([3], [5]). Furthermore, there has been continuous collaboration with the TransCelerate Working Group responsible for developing their anonymization guidance document [4].

Although data transparency is outside the scope of submissions of data to health authorities (HA), maintaining CDISC compliance of de-identified data is a nice-to-have, there are a number of scripts that have been and are being developed assuming that datasets are CDISC compliant, and these scripts could then be reused.

The spirit should be to preserve as much data as possible to make sure the data remains usable.

Dates

All time-related information is important in clinical research in particular dates, they present them self in two forms, date and relative days. Conforms with CDISC standards (both dates and relative days are present), it is preferable from a data utility perspective to keep both types of dates, describe how to offset dates, and keep study day and other relative dates as-is. Full and partial dates must be offset (guidance is provided in Appendix 1 of deliverable [0]) while relative dates such Study Day can be kept as-is in the datasets.

It was also decided that Date of Death must be offset like any other dates considering its importance in clinical research. It is also flagged in the deliverable what variables can indicate death (e.g. AEOUT) and should be considered if further de-identification is required.

In particular, date of Birth must be derived into "Age at baseline" and patients over 89 years old must be aggregated into one category. It is also possible as an alternative rule to derive into age folds (10-15, 15-20, 20-25 etc., 18-20, 20-22, 22-24, etc.) to be defined by the sponsor. Dates indicative of age >89 years, e.g., year of disease diagnosis or year a prior medication was started must be replaced by "--redacted--" (e.g. in MH or CM domains).

Recoding of unique identifiers

Subject IDs, Reference IDs, Sponsor IDs, Investigator IDs and Site IDs must be recoded.

In particular, a new random unique subject ID must be created that is not made up of any identifiable information. Site numbers must not be replicated in the recoded subject IDs. The list of original subject IDs and the recoded ones must not have any values in common. Same recoded subject ID must be used in extension study data.

Variables such as Reference ID or Sponsor ID are usually constructed using CRF page numbers or laboratory sample numbers, which are Direct Identifiers and require recoding or deletion (if only operational and are not necessary to link data across datasets). The list of original IDs and the recoded ones must not have any values in common. This applies also to Investigator ID and Site ID, among others, when applicable.

Low Frequency & Rare Events

The concept of low frequency as a means to evaluate re-identification risk is often cited and used. Low frequency is one way to measure risk. The term has the same meaning as `minimal cell size" and "maximum risk" [5]. Maximum risk is a conservative way to measure risk and is more suited to public data release. For non-public data release, which is more congruent with the manner in which data will be disclosed under many clinical trial transparency initiatives, a more suitable way to measure risk is using the concept of "average risk" [5]. Average risk is less conservative and allows the disclosure of more detailed information. Another important consideration is that the actual value of "low frequency" needs to be computed from the population.

More guidance on the topic is available in Appendix 2.

Handing of Free-text variables and Extensible code-lists

In general, free-text data must be deleted as free-text data is considered to be a Direct identifier as it can hold any data including PII. If there is no associated coded information available in the dataset that can be used instead, free-text data must be considered for review and redaction of values with PII. However such measure needs to be balanced with the criticality of such variable for future research and as alternative rule, a non-recoded free-text variable can be removed from the dataset.

For variables that are supported by extensible code lists, extra values can be added as free-text. While free-text is considered in the context of data transparency and sharing data with researchers as uncontrolled and at risk, in this particular case, there is often an extensive list of values available from the CDISC controlled terminology. As part of the data management process, sponsors would detect, query and update free-text values that would not be appropriate to keep if they include PII. Such variables are not marked as Direct or Quasi Identifiers in the SDTMIG tab, but for precaution, we assigned the rule "Review and only redact values with personal information" like for free-text variables.

Geographic Location

Country is an important Level 1 Quasi Identifier and in order to decrease residual risk by default (i.e. in the case of proactive data de-identification), country is advised to be aggregated to continent as primary rule unless critical to the analysis (e.g. Country is a fixed-effect factor in a statistical model and the results cannot be reproduced). The alternative rule is to keep country as-is. In particular the rule described in HIPAA [2] within the Safe Harbor method for geographical location is based on an empirical analysis performed by the US Census Bureau and may not be applicable globally.

Site ID and Investigator ID need to be removed because a frequency analysis would likely reveal the most highly recruiting site in a country/region (which by definition would include many of the participants). The alternative rule is to recode Site ID and Investigator ID if required for the analysis and in this case, it should be considered within the risk assessment.

Sensitive data

Although sensitive data (See Definitions tab of [0]) may not necessarily be PII and re-identifying, sensitive data may need to be deleted so that in case of data breach, further measures have already been taken. In principle if the data is not personal data anymore, such data could also be kept. This is the responsibility of the sponsors to decide what risk they are willing to take.

Variables and datasets at stake have a comment associated with such considerations.

PII of third-parties

PII of third parties (Laboratory name or address, Investigator name, etc.) must be removed from all datasets as it may provide geographical information about the patients and also could compromise the privacy of the third parties themselves. However third party roles may be kept as they can be an important factor for the analysis.

Other important quasi identifiers for data analysis

While a variable has been identified as a Quasi Identifier, it may be advisable to keep the variable as-is if it is judged critical for analysis and clinical research in general (e.g. Level 1 Quasi Identifer Sex). The rule "Keep" is to document clearly that no action should be taken although the variable should be considered in computing the residual risk and may require in some cases further de-identification (See DI_comment column of SDTMIG tab in [0]).

Also in the case of dates, while visit dates are important information with regards to data privacy, e.g. --DY/VISIT/VISITNUM/VISITDY could help re-identify all visit dates should only one be found out since planned relative dates indicates when during the study the visit is planned to occur. They are classified as Level 2 quasi identifiers and assigned the rule "No further de-identification" meaning that it is not advised to apply further de-identification (they already represent de-identified variable) but are flagged for consistency. This rationale is applied in particular to all relative dates (actual and planned).

CONCLUSION

This set of rules defined for CDISC SDTM 3.2 is written with the goal of both facilitating the assessment of direct and quasi identifiers in SDTM datasets and ensuring consistency in anonymized data shared across sponsors. The definitions of direct and quasi-identifiers represent the consensus of the working group.

However, the rules described here do not guarantee an acceptable or very small residual risk of re-identification in the data and it is the responsibility of the sponsors to define and measure what the residual risk is and define an acceptable risk threshold.

Hrynaszkiewicz et al. [3] suggests that if more than 2 quasi identifiers are left in the dataset, a risk assessment must be carried out. The recent report from the Institute of Medicine [5] suggests that this rule of 2 may not be enough in certain cases because the re-identification risk may still be high, and therefore suggests a methodology to carry out such a risk assessment in Appendix B "Concepts and Methods for De-identifying Clinical Trial Data".

SDTM being also a normalized data model, not all direct nor quasi identifiers may be captured in this deliverable and it is the responsibility of the sponsor to ensure that such assessment is conducted and reviewed according to defined internal procedures.

Following the completion of this first version of the PhUSE De-Identification Standards for SDTM 3.2 in April 2015, a number of pilots are conducted with private and public organizations in order to assess further the usability and accuracy of the rules defined. Based on the feedback, a second version may be developed and made available to the community.

The Working Group is also currently discussing the next deliverables to work on to support the data transparency initiative in the industry. CDISC ADaM or the CDISC SDTM Therapeutic Areas standards are possibilities as an extension as well as addressing the issue of data de-identification documentation or providing open-source code to de-identify studies according to the PhUSE standards.

REFERENCES

- [0]: PhUSE Data Transparency Resources, http://www.phuse.eu/Data_Transparency.aspx
- [1]: CDISC SDTM Implementation Guide (version 3.2)
- [2]: Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule - Code of Federal Regulations
- [3]: Hrynaszkiewicz I, Norton M L, et al. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. British Medical Journal 2010; 340:304–307
- [4]: "Data De-identification and Anonymization in Clinical Trials– A Model Approach", TransCelerate, 2015
- [5]: "Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk", Institute of Medecine, 2015
- [6]: A De-identification Strategy Used for Sharing One Data Provider's Oncology Trials Data through the Project Data Sphere Repository, Malin, 2013
- [7]: Sponsor's anonymisation standards published on ClinicalStudyRequest.com
- [8]: Anonymisation: managing data protection risk code of practice, ICO, 2012

ACKNOWLEDGMENTS

We would like to thank all the working group participants for their input and contribution to the deliverable.

Vinitha Arumugam & Patricia Coyle (GSK)	Jean-Marc Ferran (Qualiance & PhUSE)	Nancy Freidland (IBM)
Per-Arne Stahl (AstraZeneca)	Nick De Donder (Business & Decision Life Sciences)	Gene Lightfoot (SAS Institute)
Sherry Meeh (Johnson & Johnson)	Cathal Gallagher (d-Wise)	Jacques Lanoue & Benoit Vernay (Novartis)
Kim Musgrave (Amgen)	Nate Freimark (Theorem)	Joanna Koft (Biogen Idec)
Gary Chen (Shire)	Khaled El Emam (Privacy Analytics)	Jennifer Chin (EISAI)
Carl Herremans (Merck)	Beate Hientzsch & Sven Greiner (Accovion)	Kishore Papineni, Thijs van den Hoven & Bharat Jaswani (Astellas)
Kelly Mewes (Roche)	Kristin Kelly (Accenture)	Sarah Nolan (Liverpool University & Cochran)
Boris Grimm (Boehringer Ingelheim)	Shafi Chowdury (Shafi Consultancy)	Ravi Yandamuri (MMS Holidings)

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Jean-Marc Ferran
Enterprise: Qualiance
E-mail: JMF@qualiance.dk
Web: www.qualiance.dk
Twitter: @QualianceTwitta

Name: Jacques Lanoue
Enterprise: Novartis
Email: jacques.lanoue@novartis.com
Web: www.novartis.com
Twitter: @jacqueslanoue