Paper DH15

A Privacy Methodology Toolkit to Improve Cross-Industry Clinical Data Reuse

Boris Grimm, Boehringer Ingelheim
on behalf of the TransCelerate Privacy Methodology for Data Sharing Workstream

ABSTRACT

Secondary use of clinical trial data can help accelerate the development of medicines through reducing patient burden, generating insights, and enabling evidence-driven decision making. However, differences in the current data privacy methodologies can make the data usability extremely challenging, especially for cross-study analysis. TransCelerate Biopharma has developed a Privacy Methodology, Supporting Materials and toolkit that aim to help companies improve reusability of data and data transparency to researchers on the applied safeguards to protect patient privacy. Despite being initially developed for TransCelerate's Clinical Data Sharing Initiatives, the Privacy Methodology paper was submitted to public commenting in November 2022 - March 2023 to collect industry feedback on potential areas of improvement and determine whether the solution could be useful to other stakeholders and data sharing platforms. This paper will explore the methodology toolkit resources and a summary of stakeholder input from the public review period.

TransCelerate Biopharma is a not-for-profit entity created to foster collaboration, to accelerate and simplify the research and development of new therapies around the world, for the benefit of the patients.

INTRODUCTION / DISCLAIMER

The paper below discusses a proposed privacy methodology developed by TransCelerate for use with clinical trial data.

Nothing in the methodology or this paper should be construed to represent or warrant that persons using the methodology have complied with all applicable laws and regulations.

All individuals and organizations using this methodology bear responsibility for complying with the applicable laws and regulations for the relevant jurisdiction.

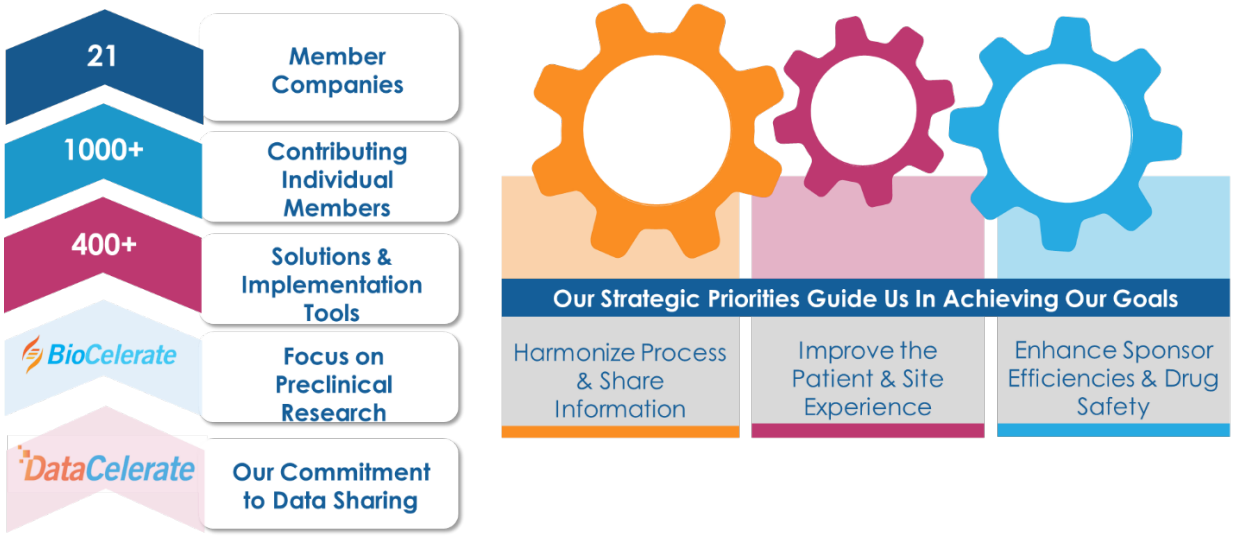
TRANSCELERATE & PRIVACY METHODOLOGY INITIATIVE SOLUTION OVERVIEW

TRANSCELERATE

TransCelerate BioPharma Inc. is a non-profit organization dedicated to improving the health of people around the world by accelerating and simplifying the research and development (R&D) of innovative new therapies.

The organization's mission is to collaborate across the global biopharmaceutical R&D community to identify, prioritize, design, and facilitate implementation of solutions intended to drive the efficient, effective, and high-quality delivery of new medicines. The vast majority of TransCelerate solutions are publicly available. Headquartered in Philadelphia,

TransCelerate has 22 member companies and 30+ initiatives focused on improving the patient and site experience, enhancing sponsor efficiencies and drug safety and, as appropriate, harmonizing process and sharing information.

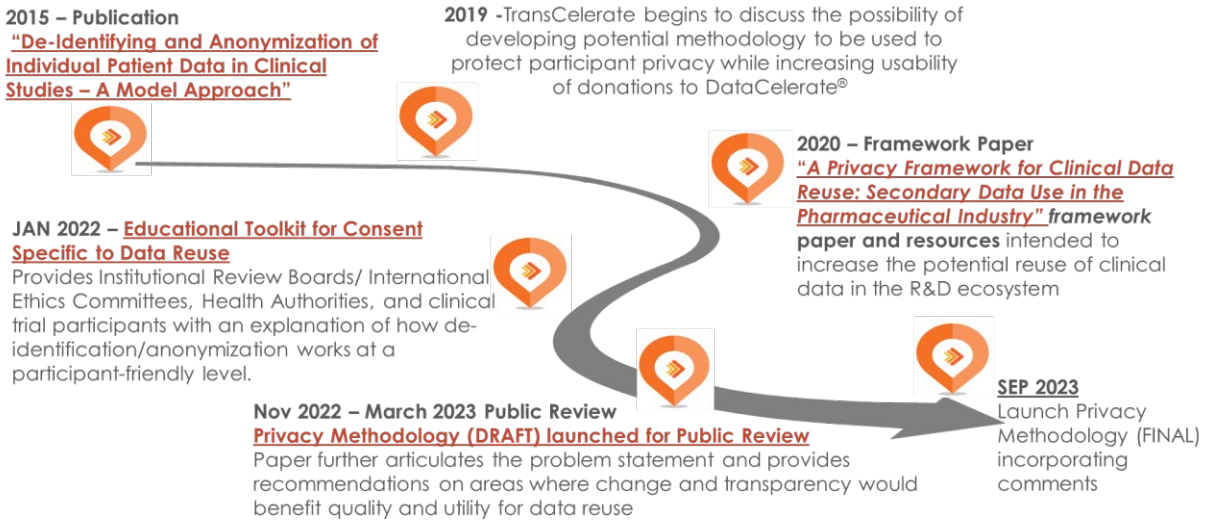


PRIVACY METHODOLOGY FOR DATA SHARING INITIATIVE

This proposed methodology is the third deliverable created by TransCelerate to address issues related to privacy regulations and compliance in connection with secondary reuse of clinical trial data.

The first deliverable was the 2020 framework "A Privacy Framework for Clinical Data Reuse: Secondary Data Use in the Pharmaceutical Industry" [1] to provide principles and best practices related to reuse of clinical trial participant data in the hopes of increasing the potential reuse of clinical data and reducing patient burden.

The second deliverable, released in 2022, was the "Educational Toolkit for Consent Specific to Data Reuse", [2] which provides Institutional Review Boards/International Ethics Committees, Health Authorities, and clinical trial participants with a customizable explanation of how anonymization works at a participant-friendly level.



PRIVACY METHODOLOGY FOR DATA SHARING SOLUTION

The proposed data privacy methodology presented here focuses on clinical trial datasets shared among the TransCelerate member companies in the DataCelerate® database. As such, the recommended approach is based on specific conditions of that project. However, the hope is that the methodology may also be useful for data sharing outside the scope of DataCelerate®.

Certain variations in the data privacy approach used across companies or studies can make the resulting datasets time-consuming and difficult to understand and use, and in some cases, can greatly reduce the potential for cross-study analyses. The challenge is further complicated when there is a lack of transparency in describing the specific

data privacy safeguards that have been applied, making it difficult for a Data User to determine whether the data are fit for an intended secondary purpose.

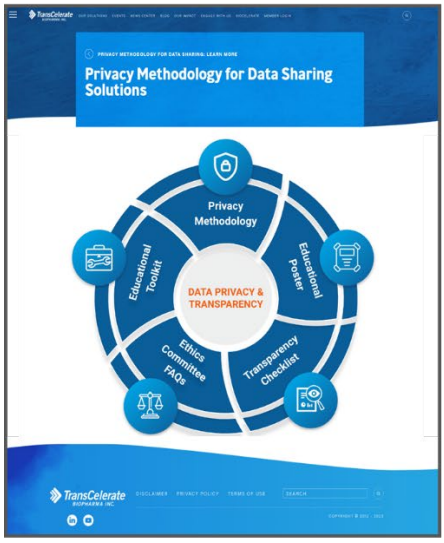
Building on existing solutions in the industry, such as those developed by PHUSE—The Global Healthcare Data Science Community [3], TransCelerate [4], and the European Union Agency for Cybersecurity (ENISA) [5], the proposed data privacy methodology outlined in this paper recommends data modification approaches for relevant data variables typically found in clinical datasets.

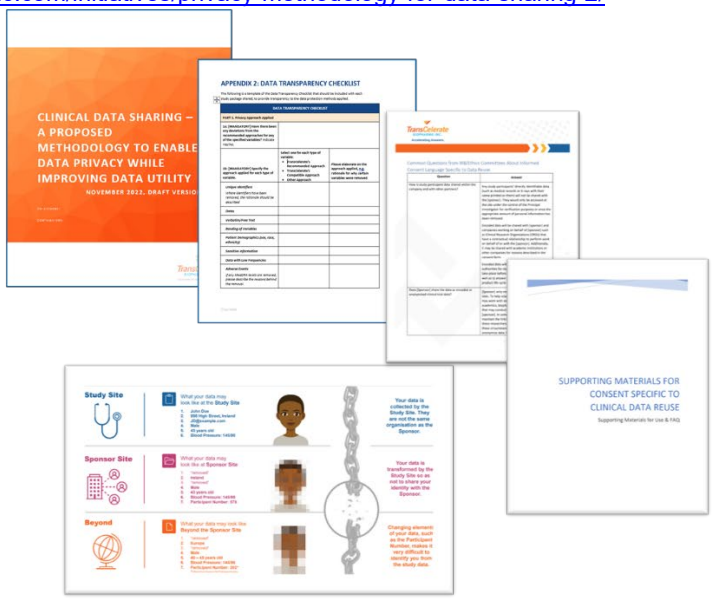
Ultimately, the proposed data privacy methodology is intended to:

- Preserve privacy and confidentiality of research participants.
- Reduce research participant burden through easier reuse of existing study data.
- Deliver faster scientific insights.
- Provide greater transparency in the privacy safeguards applied to study data.
- Increase overall data utility.

Data Providers benefit from this methodology by being able to identify critical privacy variables to protect and as well as proposed approaches for application of these safeguards. The approaches focus on key data elements for which a consistent approach better enables cross-study analyses.
Data Users benefit from this methodology by having greater transparency and confidence in the data they are receiving from Data Providers who follow this data privacy approach.

Website: https://www.transceleratebiopharmainc.com/initiatives/privacy-methodology-for-data-sharing-2/





CORE COMPONENTS OF THE PRIVACY METHODOLOGY TOOLKIT

The **Privacy Methodology** is a proposed methodology to data protection for clinical data reuse that combines the emergent thinking from the TransCelerate member companies and public commenters into practical variable type recommendations and considerations.

The **Transparency Checklist** is a template of key information, from a data user's perspective, that describes the context to the data structure and the data modifications/derivations performed. This template or similar document aids greater transparency for researchers to understand the data they are using.

These tools are a result of a three-year collaboration involving members from more than 20 sponsor companies willing to share best practices and develop the privacy methodology and transparency checklist, as well as an extended public comment period.

The Methodology provides recommended and compatible approaches to aid anonymization of 14 key variable types.

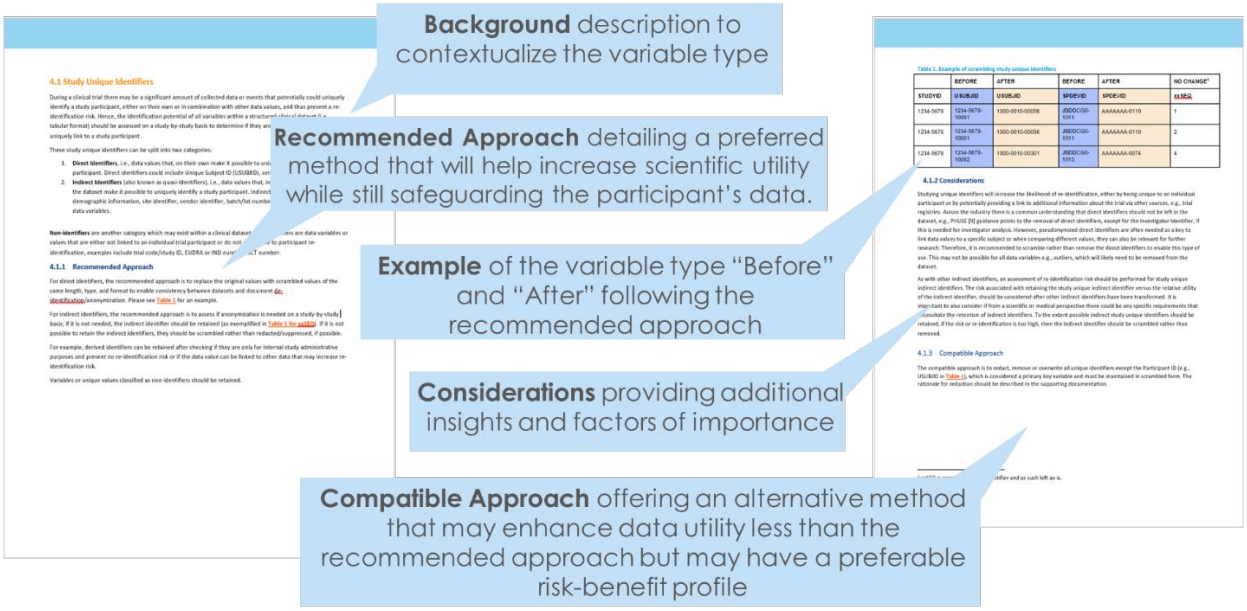
12 Common Variable Types

- Study Unique Identifiers
- Dates
- Verbatim / Free Text
- Banding of Variables
- Records of Patients Who Have Died
- Patient Demographics (sex, race, ethnicity)
- Sensitive Information
- Data with Low Frequencies
- Adverse Events & Medical History
- Medications
- Geographic Location
- Information Collected Under Copyright Licenses

2 Novel Areas

- Data Derived from Genomic Data
- Seasonality

To increase the usability of the Methodology whilst honoring data privacy and preserving data utility, each variable type has a recommended approach and a compatible approach. Compatible approaches present alternatives, where available, to the recommended approaches.
The structure of the methodology enables flexible implementation – adopters may implement one or more of the suggested approaches to drive towards greater consistency.
The use of the methodology is **voluntary** - each organization must decide on its own what, when, and how to implement it.



METHODOLOGY EXAMPLE

The example shows section 4.6 Data with Low Frequencies.

4.6 Data with Low Frequencies
Variables that contain some low frequency values may lead to an increased risk of reidentification (e.g., data that, after grouping of several variables and their expression levels, applies to a very small cell/group size). This may be further compounded by data with multiple variables of low frequencies.

4.6.1 Recommended Approach
It is recommended that variables with low frequencies should be assessed, and if there is a significant risk of re-identification, those values with low frequencies should be redacted (Table 5 and Table 6).

4.6.2 Considerations
A great deal can be learned from certain low-frequency study data so there is usually a special interest to analyze rare events and outliers in a clinical study. Careful consideration should be given to whether these study data should be retained as the risk of re-identification significantly increases when participant data falls into a very low frequency group. For a deeper dive into the topic and for additional guidance, Appendix 2 "Low Frequencies" of the PHUSE Data De-Identification Standard for SDTM 3.2 [6] is recommended.

4.6.3 Compatible Approach No alternative approach is proposed.

Table 5. Example of Redacting Low Frequency Sex and Race

		BEFORE	AFTER	BEFORE	AFTER
USUBJID	DOMAIN	SEX	SEX	RACE	RACE
1234-5678-USA003-10001	DM	F	-- Redacted --	WHITE	WHITE
1234-5678-POL002-10003	DM	M	-- Redacted --	WHITE	WHITE
1234-5678-GER002-10004	DM	F	-- Redacted --	UNKNOWN	-- Redacted --

Table 6. Example of Redacting Additional Information Revealing Trial Participants' Sex Through a Rare Event (e.g., Event of Oligospermia)

		BEFORE	AFTER
USUBJID	DOMAIN	AEDECOD	AEDECOD
1234-5678-POL002-10003	AE	Oligospermia	-- Redacted --
1234-5678-POL002-10003	AE	Headache	Headache
1234-5678-POL002-10003	AE	Diarrhea	Diarrhea

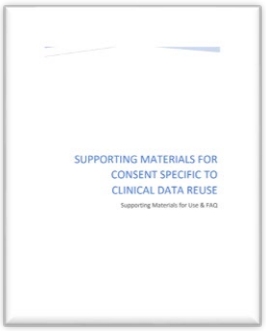
ADDITIONAL TOOLS

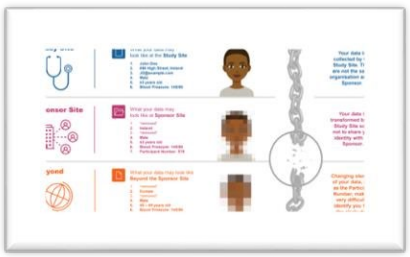
The Educational Toolkit, the Educational Poster and FAQs have been developed to help trial participants and sponsors.

Educational Toolkit
This educational toolkit can be used by study sponsors to help clinical trial participants better understand the data protection measures that will be applied to their personal data.

Educational Poster
The editable educational tool will provide a visual, consistent, succinct, easy-to-understand breakdown for the clinical trial participant outlining how patient privacy is protected by study sponsors. The tool can be edited by a sponsor as necessary to reflect the process it uses.

FAQ for Ethics Committee
The educational tool can aid study sponsors in answering common questions asked by IRBs and Ethics Committees.





PUBLIC COMMENT PERIOD HIGHLIGHTS

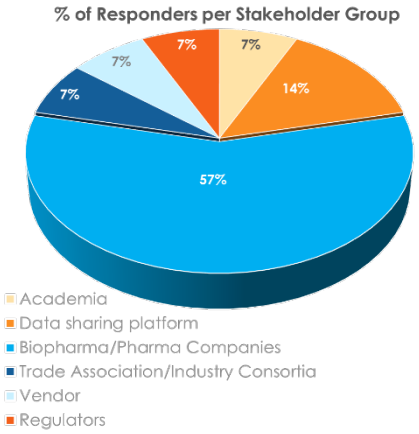
DIVERSE PARTICIPATION TO THE PUBLIC REVIEW & OUTCOME

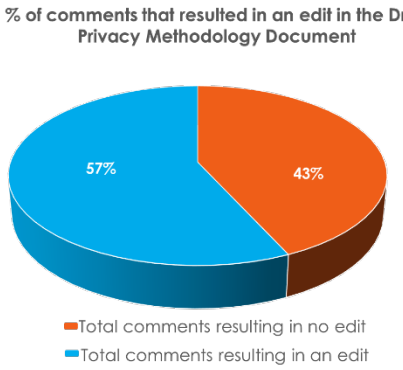
TransCelerate submitted to public review the draft Privacy Methodology for Clinical Data Sharing to solicit input from the broader R&D ecosystem to improve the draft methodology and better ensure the feasibility of the deliverable.

During a 4-month period ending in March 2023, more than 40 comments were received from 6 different stakeholder groups across the clinical research ecosystem (Academia, Vendors, BioPharma/Pharmaceutical Companies, Data Sharing Platforms, Trade Associations/Industry Consortia, Regulators).

The feedback received was positive and pointed to:

- Usefulness of having a solution like this in the industry
- Readability and comprehensibility of the variables
- Usefulness of the Data Transparency Checklist.





No major conflicting conceptual or structural feedback was part of the comments received. Most edits contributed to enhancing the robustness and the clarity of the draft Privacy Methodology document.

The full responses to the comments and the revised Privacy Methodology toolkit were made publicly available on the TransCelerate Privacy methodology website in Q3 2023.

The team incorporated these comments to improve the robustness and clarity of the privacy methodology and the revised version is now publicly available.

CONCLUSION

Reuse of clinical trial data honors a patient's contribution by using data to the fullest extent possible, while advancing the speed and quality of drug development. Unlocking the vast potential that exists in combining and repurposing clinical trial data has advanced due to technological developments, but there is still hesitancy from many organizations as they weigh up the effort involved in harmonizing the datasets and consider the requirements around protecting the privacy of participants. The growing influence of the GDPR (General Data Protection Regulation) in legislation outside of the EEA puts privacy at the forefront of the agenda when it comes to Clinical Data Sharing and Reuse.

Sponsors use varying data privacy measures (anonymization and/or pseudonymization) to safeguard patient privacy and comply with relevant laws and regulations. Those variations in approaches may result in making data difficult and time-consuming to reuse and, in some cases, may reduce data utility, especially in the context of cross-study analysis. The issue can be further compounded by a lack of transparency in describing the specific privacy safeguards that have been applied, thereby making it difficult to determine whether the data are fit for an intended secondary purpose.

The vision for the Privacy Methodology for Data Sharing is to improve transparency and reduce variability in the methodology applied by companies to protect personal data in ways that facilitate the reuse of data. The proposed approach could reduce the complexity associated with data sharing and better facilitate cross-study analysis.

REFERENCES

[1] TransCelerate Biopharma Inc. A Privacy Framework for Clinical Data Reuse: Secondary Data Use in the Pharmaceutical Industry. October 2020. Accessed October 20, 2022. https://www.transceleratebiopharmainc.com/wp-content/uploads/2020/10/TransCelerate_A-Privacy-Framework-forClinical-Data-Reuse_Interpretations-of-Guidances-and-Regulations-Clinical_October-2020.pdf

[2] TransCelerate Biopharma Inc. Privacy Methodology for Data Sharing Solutions: Educational Toolkit for Clinical Data Reuse. Accessed October 20, 2022. https://www.transceleratebiopharmainc.com/assets/privacy-methodology-for-data-sharing-solutions

[3] PHUSE Working Group. Data Transparency. Accessed October 20, 2022. https://advance.phuse.global/display/WEL/Data+Transparency

[4] TransCelerate BioPharma Inc. Welcome to TransCelerate. Accessed October 20, 2022. www.transceleratebiopharmainc.com

[5] European Union Agency for Cybersecurity. Data Pseudonymisation: Advanced Techniques and Use Cases. January 28, 2021. Accessed October 20, 2022. https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases
[6] PHUSE Working Group. De-Identification Standard for SDTM 3.2 Version 1.0. May 20, 2015. Accessed October 20, 2022.
https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Data+Transparency/Deidentification+Standard+for+SDTM+3.2+Version+1.0.xls

ACKNOWLEDGMENTS

This TransCelerate Privacy Methodology for Data Sharing Solution would not have been possible without the tireless effort of all the industry experts who have been involved in the Core and Reference team of the TransCelerate Privacy Methodology for Data Sharing Workstream.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Boris Grimm
Boehringer Ingelheim
On behalf of TransCelerate
Email: feedback@transceleratebiopharmainc.com
Web: www.transceleratebiopharmainc.com