

Proposed Privacy Methodology to Improve Cross-Industry Data Sharing (DH15)

5 - 8 November 2023

PHUSE EU Connect 2023

Boris Grimm



Disclaimer

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

Nothing in the methodology or this presentation 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.

PRESENTER



Boris Grimm

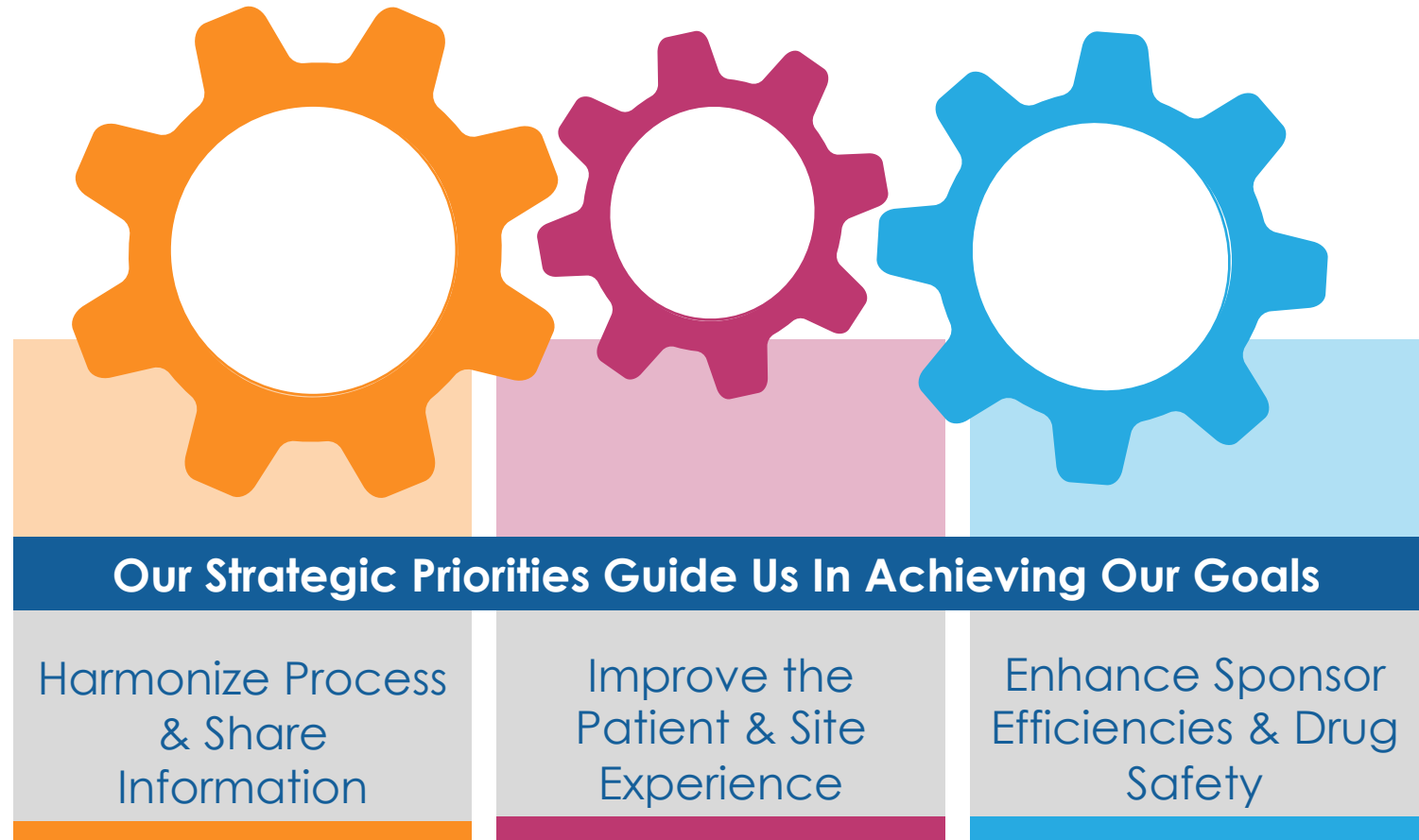
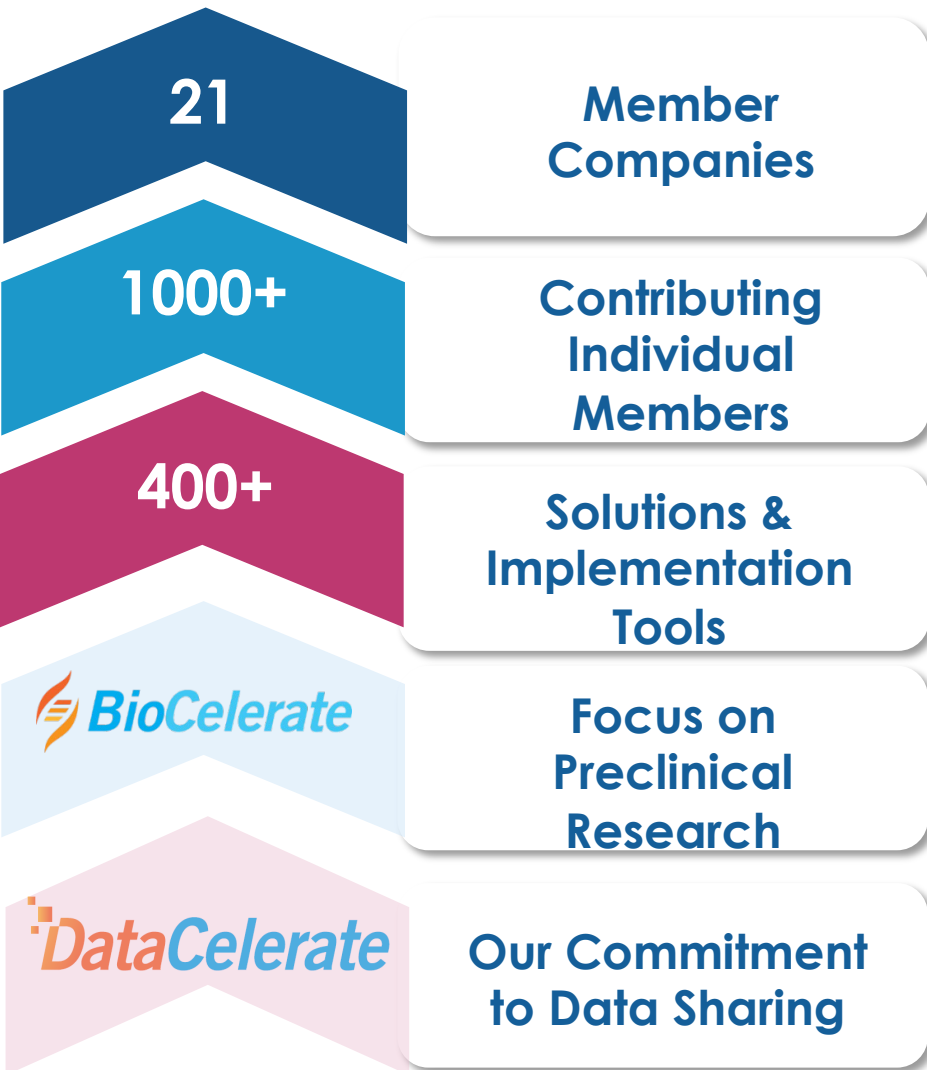
**Chapter Head and
Capability Owner of Clinical Trial
Transparency and Data Sharing**

AGENDA

- **TransCelerate & Privacy Methodology Initiative Solution Overview**
- **Public Comment Period Highlights**
- **What Comes Next?**
- **Q&A**

TransCelerate Overview

TransCelerate : A Catalyst For Industry-Wide Collaboration



Initiative Overview: Privacy Methodology for Clinical Data Sharing



TransCelerate Solutions in Data Privacy / Transparency

2015 – Publication

“De-Identifying and Anonymization of Individual Patient Data in Clinical Studies – A Model Approach”

2019 -TransCelerate begins to discuss the possibility of developing potential methodology to be used to protect participant privacy while increasing usability of donations to DataCelerate®

JAN 2022 – Educational Toolkit for Consent Specific to Data Reuse

Provides Institutional Review Boards/ International Ethics Committees, Health Authorities, and clinical trial participants with an explanation of how de-identification/anonymization works at a participant-friendly level.

Nov 2022 – March 2023 Public Review

Privacy Methodology (DRAFT) launched for Public Review

Paper further articulates the problem statement and provides recommendations on areas where change and transparency would benefit quality and utility for data reuse

2020 – Framework Paper

“A Privacy Framework for Clinical Data Reuse: Secondary Data Use in the Pharmaceutical Industry” framework paper and resources intended to increase the potential reuse of clinical data in the R&D ecosystem

SEP 2023

Launch Privacy Methodology (FINAL) incorporating comments

Solution Overview: What is the Privacy Methodology Solution?



The Privacy Methodology Solution tackles key data privacy challenges in clinical data reuse

Today's Challenge

- Regional and local data privacy laws and regulations are varied and constantly evolving
- Emerging technologies (AI, ML) present opportunities for efficiencies but also support greater customization in ways that reduce data utility.
- Lack of visibility and confidence in shared data is limiting reuse of clinical data for scientific advancement, reducing potential patient benefit

Goal of the Privacy Methodology Solution

Enable Greater Data Utility to support complex cross-study analysis, deepen insights in cross population disease profiling, and **reduce patient burden**

Reduce Variations that hamper cross-study analysis across the ecosystem by reducing variability of the data privacy measures applied during anonymization

Increase Transparency to data users and researchers of data protection approaches leading to greater confidence in the applicability of the data

Privacy Methodology Solution

Website: <https://www.transceleratebiopharmainc.com/assets/privacy-methodology-for-data-sharing-solutions/#educational-toolkit-carousel>

TransCelerate
BIOPHARMA INC.

OUR SOLUTIONS | EVENTS | NEWS CENTER | BLOG | OUR IMPACT | ENGAGE WITH US | BIOCELERATE | MEMBER LOGIN

14

PRIVACY METHODOLOGY FOR DATA SHARING: LEARN MORE

Privacy Methodology for Data Sharing Solutions

Privacy Methodology

Educational Toolkit

Educational Poster

Transparency Checklist

Ethics Committee FAQs

DATA PRIVACY & TRANSPARENCY

TransCelerate
BIOPHARMA INC.

DISCLAIMER | PRIVACY POLICY | TERMS OF USE

SEARCH

in

Copyright © 2012 - 2023

CLINICAL DATA SHARING –
A PROPOSED
METHODOLOGY TO ENABLE
DATA PRIVACY WHILE
IMPROVING DATA UTILITY

NOVEMBER 2022, DRAFT VERSION

CO-AUTHORS:
CONTRIBUTORS:

TransCelerate
BIOPHARMA INC.

APPENDIX 2: DATA TRANSPARENCY CHECKLIST

The following is a template of the Data Transparency Checklist that should be included with each study package shared, to provide transparency to the data protection methods applied.

DATA TRANSPARENCY CHECKLIST

PART 1: Privacy Approach Applied

1a. (MANDATORY) Have there been any deviations from the recommended approaches for any of the specified variables? Indicate Yes/No.

1b. (MANDATORY) Specify the approach applied for each type of variable.

Unique identifiers
Where identifiers have been removed, the reasons should be described

Verbatim/Free Text

Banding of variables

Patient demographics (sex, race, ethnicity)

Sensitive information

Data with Low Frequencies

Adverse Events
(If any MANDATA levels are removed, please describe the reasons behind the removal.)

Select one for each type of variable:

TransCelerate's Recommended Approach

TransCelerate's Compatible Approach

Other Approach

Please elaborate on the approach applied, e.g. rationale for why certain variables were removed

[Type text]

TransCelerate
BIOPHARMA INC.
Accelerating Answers.

Common Questions from IRB/Ethics Committees About Informed Consent Language Specific to Data Reuse

Question	Answer
How is study participant data shared within the company and with other partners?	Any study participants' directly identifiable data (such as medical records or X-rays with their name printed on them) will not be shared with the [sponsor]. They would only be accessed at the site under the control of the Principal Investigator for verification purposes or once the appropriate amount of personal information has been removed.
Does [sponsor] share the data as encoded or anonymized clinical trial data?	Encoded data will be shared with [sponsor] and companies working on behalf of [sponsor] such as Clinical Research Organizations (CROs) that have a contractual relationship to perform work on behalf of or with the [sponsor]. Additionally, it may be shared with academic institutions or other companies for reasons described in the consent form.
	Encoded data will be shared with [sponsor] and companies working on behalf of [sponsor] such as Clinical Research Organizations (CROs) that have a contractual relationship to perform work on behalf of or with the [sponsor]. Additionally, it may be shared with academic institutions or other companies for reasons described in the consent form.

Study Site

What your data may look like at the Study Site

1. John Doe
2. 880 High Street, Ireland
3. JD@example.com
4. Male
5. 43 years old
6. Blood Pressure: 145/98

Your data is collected by the Study Site. They are not the same organization as the Sponsor.

Sponsor Site

What your data may look like at Sponsor Site

1. "removed"
2. Ireland
3. "removed"
4. Male
5. 43 years old
6. Blood Pressure: 145/98
7. Participant Number: 576

Your data is transformed by the Study Site so as not to share your identity with the Sponsor.

Beyond

What your data may look like Beyond the Sponsor Site

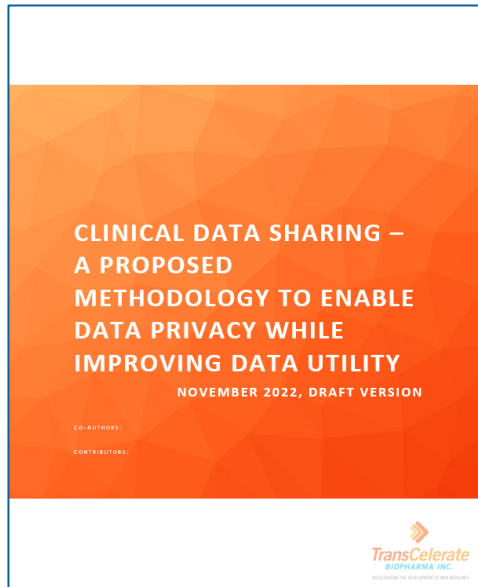
1. "removed"
2. Europe
3. "removed"
4. Male
5. 40 - 49 years old
6. Blood Pressure: 145/98
7. Participant Number: 302

Changing element of your data, such as the Participant Number, makes it very difficult to identify you from the study data.

SUPPORTING MATERIALS FOR
CONSENT SPECIFIC TO
CLINICAL DATA REUSE

Supporting Materials for Use & FAQ

Core Components of the Privacy Methodology Toolkit



**Privacy
Methodology
Paper**

APPENDIX 2: DATA TRANSPARENCY CHECKLIST

The following is a template of the Data Transparency Checklist that should be included with each study package shared, to provide transparency to the data protection methods applied.

DATA TRANSPARENCY CHECKLIST		
PART 1: Privacy Approach Applied		
1a: (MANDATORY) Have there been any deviations from the recommended approaches for any of the specified variables? indicate Yes/No.		
1b: (MANDATORY) Specify the approach applied for each type of variable.	Select one for each type of variable: - TransCelerate's Recommended Approach - TransCelerate's Compatible Approach - Other Approach	Please elaborate on the approach applied, e.g., rationale for why certain variables were removed
Unique Identifiers Where identifiers have been removed, the rationale should be described		
Dates		
Verbatim/free text		
Banding of variables		
Patient Demographics (sex, race, ethnicity)		
Sensitive information		
Data with Low Frequencies		
Adverse Events If any MedDRA levels are removed, please describe the reasons behind the removal.		

[Type here]

**Data
Transparency
Checklist**

How were these tools developed?

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.

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

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.

Use of the methodology is **voluntary** – each organization must decide on its own what, when, and how to implement it.

Structure of the methodology enables flexible implementation – adopters may implement one or more of the suggested approaches to drive towards greater consistency.

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

Each variable type includes detailed context and examples to help one apply the recommendations

Background description to contextualize the variable type

4.1 Study Unique Identifiers

During a clinical trial there may be a significant amount of collected data or events that potentially could uniquely identify a study participant, either on their own or in combination with other data values, and thus present a re-identification risk. Hence, the identification potential of all variables within a structured clinical dataset (in a tabular format) should be assessed on a study-by-study basis to determine if they are uniquely link to a study participant.

These study unique identifiers can be split into two categories:

1. **Direct Identifiers**, i.e., data values that, on their own make it possible to uniquely identify a study participant. Direct identifiers could include Unique Subject ID (USUBJID), serial number, date of birth, etc.
2. **Indirect Identifiers** (also known as quasi-identifiers), i.e., data values that, in combination with other data values, make it possible to uniquely identify a study participant. Indirect identifiers could include demographic information, site identifier, vendor identifier, batch/lot number, etc.

Non-identifiers are another category which may exist within a clinical dataset. These are data variables or values that are either not linked to an individual trial participant or do not contain information that could be used to identify a study participant. Examples include trial code/study ID, EUDRA or IND number, etc.

4.1.1 Recommended Approach

For direct identifiers, the recommended approach is to replace the original values with scrambled values of the same length, type, and format to enable consistency between datasets and document the process of identification/anonymization. Please see [Table 1](#) for an example.

For indirect identifiers, the recommended approach is to assess if anonymization is needed on a study-by-study basis; if it is not needed, the indirect identifier should be retained (as exemplified in [Table 1](#) for [USUBJID](#)). If it is not possible to retain the indirect identifiers, they should be scrambled rather than redacted/suppressed, if possible.

For example, derived identifiers can be retained after checking if they are only for internal study administrative purposes and present no re-identification risk or if the data value can be linked to other data that may increase re-identification risk.

Variables or unique values classified as non-identifiers should be retained.

Recommended Approach detailing a preferred method that will help increase scientific utility while still safeguarding the participant's data.

Example of the variable type “Before” and “After” following the recommended approach

Considerations providing additional insights and factors of importance

Compatible Approach offering an alternative method that may enhance data utility less than the recommended approach but may have a preferable risk-benefit profile

Table 1. Example of scrambling study unique identifiers

	BEFORE	AFTER	BEFORE	AFTER	NO CHANGE!
STUDYID	USUBJID	USUBJID	SPDEVID	SPDEVID	XXSEQ
1234-5678	1234-5678-10001	1000-0010-00056	JBDDCG0-1011	AAAAAAAA-0110	1
1234-5678	1234-5678-10001	1000-0010-00056	JBDDCG0-1011	AAAAAAAA-0110	2
1234-5678	1234-5678-10002	1000-0010-00301	JBDDCG0-1013	AAAAAAAA-0074	4

4.1.2 Considerations

Studying unique identifiers will increase the likelihood of re-identification, either by being unique to an individual participant or by potentially providing a link to additional information about the trial via other sources, e.g., trial registries. Across the industry there is a common understanding that direct identifiers should not be left in the dataset, e.g., [PHUSE \[9\]](#) guidance points to the removal of direct identifiers, except for the Investigator Identifier, if this is needed for investigator analysis. However, pseudonymized direct identifiers are often needed as a key to link data values to a specific subject or when comparing different values, they can also be relevant for further research. Therefore, it is recommended to scramble rather than remove the direct identifiers to enable this type of use. This may not be possible for all data variables e.g., outliers, which will likely need to be removed from the dataset.

As with other indirect identifiers, an assessment of re-identification risk should be performed for study unique indirect identifiers. The risk associated with retaining the study unique indirect identifier versus the relative utility of the indirect identifier, should be considered after other indirect identifiers have been transformed. It is important to also consider if from a scientific or medical perspective there could be any specific requirements that necessitate the retention of indirect identifiers. To the extent possible indirect study unique identifiers should be retained, if the risk or re-identification is too high, then the indirect identifier should be scrambled rather than removed.

4.1.3 Compatible Approach

The compatible approach is to redact, remove or overwrite all unique identifiers except the Participant ID (e.g., USUBJID in [Table 1](#)), which is considered a primary key variable and must be maintained in scrambled form. The rationale for redaction should be described in the supporting documentation.

Methodology Example: Section 4.6 Data with Low Frequencies

		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 5. Example of Redacting Low Frequency Sex and Race

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

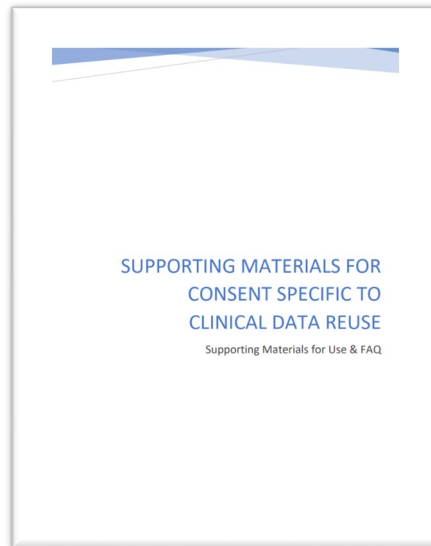
		BEFORE	AFTER
USUBJID	DOMAIN	AEDECOD	AEDECOD
1234-5678-POL002-10003	AE	Oligospermia	-- Redacted --
1234-5678-POL002-10003	AE	Headache	Headache
1234-5678-POL002-10003	AE	Diarrhea	Diarrhea

Additional Tools

Educational Toolkit, Poster and FAQs

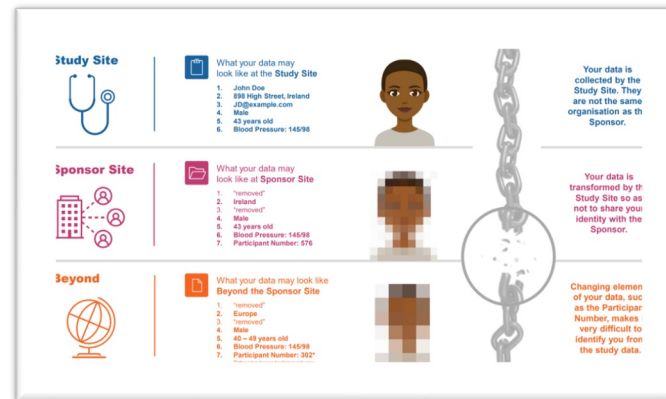
Educational Toolkit

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 and can aid study sponsors in answering common questions asked by IRBs and Ethics Committees.

The image shows a document titled "Common Questions from IRB/Ethics Committees About Informed Consent Language Specific to Data Reuse". It features the TransCelerate Biopharma Inc. logo and a table with two columns: "Question" and "Answer".

Question	Answer
How is study participant data shared within the company and with other partners?	Any study participants' directly identifiable data (such as medical records or X-rays with their name printed on them) will not be shared with the [sponsor]. They would only be accessed at the site under the control of the Principal Investigator for verification purposes or once the appropriate amount of personal information has been removed. Encoded data will be shared with [sponsor] and companies working on behalf of [sponsor] such as Clinical Research Organizations (CROs) that have a contractual relationship to perform work on behalf of or with the [sponsor]. Additionally, it may be shared with academic institutions or other companies for reasons described in the consent form. Encoded data will also be shared with regulatory authorities for review and approvals that need to take place before approval of a new medicine as well as to answer regulatory questions during product life cycle management.
Does [sponsor] share the data as encoded or anonymized clinical trial data?	[Sponsor] only receives encoded data from study sites. To help scientific advancement, [sponsor] may work with external collaborators (e.g., academics, biopharmaceutical companies etc.) that may conduct research independently from [sponsor]. In some cases, it may be necessary to maintain the link between results generated by these researchers and [sponsor]'s initial study; in these circumstances it is not possible to anonymize data. However, when working with

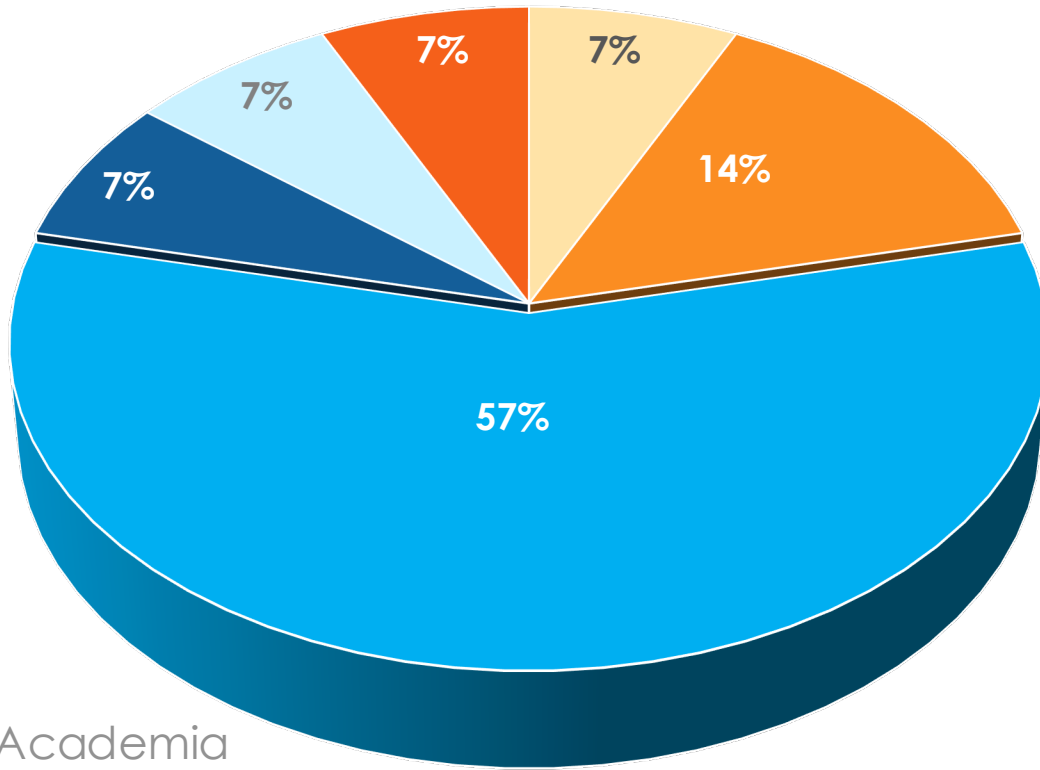
What are the key take aways from the Public Review?



Diverse Participation to the Public Review

Privacy Methodology

% of Responders per Stakeholder Group



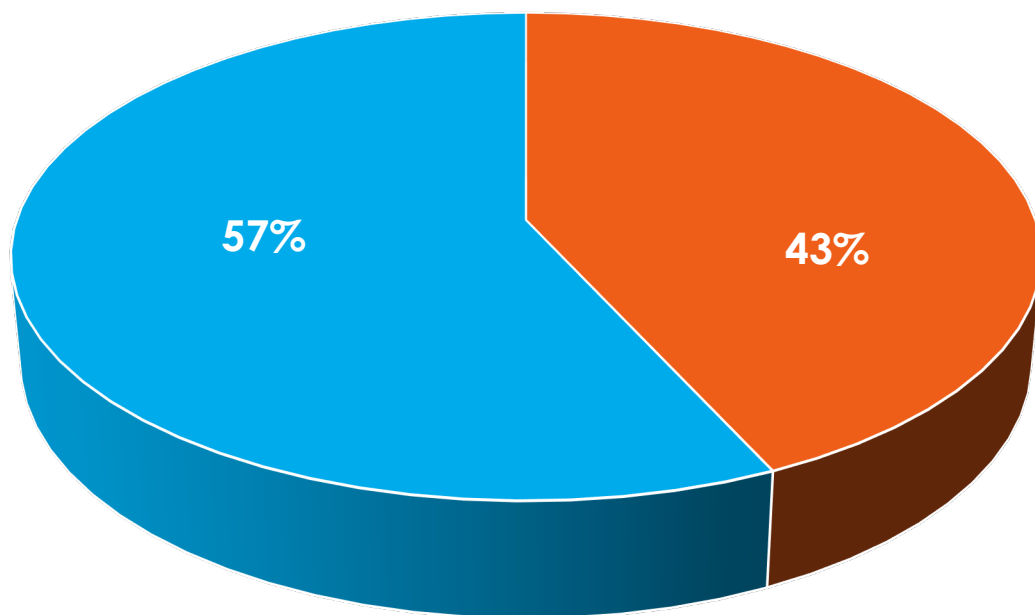
- Academia
- Data sharing platform
- Biopharma/Pharma Companies
- Trade Association/Industry Consortia
- Vendor
- Regulators

- More than 40 comments were received from 6 different stakeholder groups across the clinical research ecosystem
- 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

Comments Review Outcome

Privacy Methodology

% of comments that resulted in an edit in the Draft Privacy Methodology Document



- Total comments resulting in no edit
- Total comments resulting in an edit

- 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 3Q2023

Additional References

[TransCelerate Data Privacy Website](#)

2015 – Publication: [“De-Identifying and Anonymization of Individual Patient Data in Clinical Studies – A Model Approach”](#)

2020 - Framework Paper: [“A Privacy Framework for Clinical Data Reuse: Secondary Data Use in the Pharmaceutical Industry”](#)

Jan 2022 - [Educational Toolkit for Consent Specific to Data Reuse](#)

Nov 2022 - March 2023 Public Review: [Privacy Methodology \(DRAFT\) launched for Public Review](#)

Data Privacy Education poster: [TransCelerate Privacy Page Educational Poster_final \(transceleratebiopharmainc.com\)](#)

Thank you!

Any questions?

For further information, please contact us at:
feedback@transceleratebiopharmainc.com