Session 3: Data Governance and Privacy Methodology

Moving from confusion to clarity: In this fireside chat, our experts will discuss how to navigate the complex data protection governance process, and how existing resources like TransCelerate’s methodology can help companies efficiently enable data privacy while improving secondary use.

Liz Roberts
Senior Director
Head, Data Policy and Privacy
UCB

Luk Arbuckle
Chief Methodologist
Privacy Analytics

Jeppe G. Manuel
Principal R&D Data Privacy Specialist
Novo Nordisk

www.crdsalliance.org
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.

This is the trigger to your process. You are here because you have a new reason or place to share data, or there's a material change (process, regulatory, etc.) that affects something you are doing today.
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.

**Data Environment**

Define business cases and determine the environment(s) that will enable your data sharing use case. For example, where will you share your data? Are there cross-border considerations (such as patients from different countries)?

Start

Success
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.

This is the trigger to your process. You are here because you have a new reason or place to share data, or there’s a material change (process, regulatory, etc.) that affects something you are doing today.

1. **Data Environment**
   - Define business cases and determine the environment(s) that will enable your data sharing use case. For example, where will you share your data? Are there cross-border considerations (such as patients from different countries)?

2. **Regulatory Environment**
   - Which data protection and privacy regulations apply (e.g., GDPR, HIPAA)? If more than one regulatory framework applies, which one will govern?

Success
Data Protection & Privacy Decision Flow

Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.

1. Data Environment
   - Define business cases and determine the environment(s) that will enable your data sharing use case. For example, where will you share your data? Are there cross-border considerations (such as patients from different countries)?

2. Regulatory Environment
   - Which data protection and privacy regulations apply (e.g., GDPR, HIPAA)? If more than one regulatory framework applies, which one will govern?

3. Lawful Processing
   - Do you have the grounds for processing patient data within the applicable regulatory framework? For example, do you have appropriate patient consent for secondary research use or other legal basis under the regulation?

Start

This is the trigger to your process. You are here because you have a new reason or place to share data, or there’s a material change (process, regulatory, etc.) that affects something you are doing today.

Success
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.

1. **Data Environment**
   - Define business cases and determine the environment(s) that will enable your data sharing use case. For example, where will you share your data? Are there cross-border considerations (such as patients from different countries)?

2. **Regulatory Environment**
   - Which data protection and privacy regulations apply (e.g., GDPR, HIPAA)? If more than one regulatory framework applies, which one will govern?

3. **Lawful Processing**
   - Do you have the grounds for processing patient data within the applicable regulatory framework? For example, do you have appropriate patient consent for secondary research use or other legal basis under the regulation?

4. **Data Classification**
   - How will the protected data be classified to meet your use case aims and fulfill any regulatory requirements? For example, at the end of the process, do you consider the patient data to be anonymized, de-identified, or pseudonymized?

**Start**

This is the trigger to your process. You are here because you have a new reason or place to share data, or there’s a material change (process, regulatory, etc.) that affects something you are doing today.

**Success**
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.

1. **Data Environment**
   - Define business cases and determine the environment(s) that will enable your data sharing use case. For example, where will you share your data? Are there cross-border considerations (such as patients from different countries)?

2. **Regulatory Environment**
   - Which data protection and privacy regulations apply (e.g., GDPR, HIPAA)? If more than one regulatory framework applies, which one will govern?

3. **Lawful Processing**
   - Do you have the grounds for processing patient data within the applicable regulatory framework? For example, do you have appropriate patient consent for secondary research use or other legal basis under the regulation?

4. **Data Classification**
   - How will the protected data be classified to meet your use case aims and fulfill any regulatory requirements? For example, at the end of the process, do you consider the patient data to be anonymized, de-identified, or pseudonymized?

5. **Privacy Methodology**
   - Which privacy-protecting methodology will you employ to meet your business goal and satisfy your regulatory environment? Examples include rules-based anonymization (qualitative risk assessment) and risk-based anonymization (quantitative risk assessment).

This is the trigger to your process. You are here because you have a new reason or place to share data, or there's a material change (process, regulatory, etc.) that affects something you are doing today.
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.

### 1. Data Environment
Determine the environment(s) that will enable your data sharing use case. For example, where will you share your data? Are there cross-border considerations (such as patients from different countries)?

### 2. Regulatory Environment
Which data protection and privacy regulations apply (e.g., GDPR, HIPAA)? If more than one regulatory framework applies, which one will govern?

### 3. Lawful Processing
Do you have the grounds for processing patient data within the applicable regulatory framework? For example, do you have appropriate patient consent for secondary research use or other legal basis under the regulation?

### 4. Data Classification
How will the protected data be classified to meet your use case aims and fulfill any regulatory requirements? For example, at the end of the process, do you consider the patient data to be anonymized, de-identified, or pseudonymized?

### 5. Privacy Methodology
Which privacy-protecting methodology will you employ to meet your business goal and satisfy your regulatory environment? Examples include rules-based anonymization (qualitative risk assessment) and risk-based anonymization (quantitative risk assessment).

### 6. Operational Considerations
How will you execute? What internal and external resources will you need? Are the right processes in place to manage risk and support your initiative?

The result is a data protection approach that supports your use cases and responsibly protects patient privacy.
Proposed Privacy Methodology to Improve Cross-Industry Data Sharing

29-February 2024

CRDSA Summit

Jeppe G. Manuel
PRESENTER

Jeppe Manuel
Principal R&D Data privacy Specialist
Privacy Methodology Initiative Co-Lead
Novo Nordisk

AGENDA

• Intro to TransCelerate
• Privacy Methodology Initiative Solution Overview
• Public Comment Period Highlights
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®

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

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

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
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.
Key Considerations for Users of the Privacy Methodology

Users of the Privacy Methodology remain responsible for their own compliance with all applicable laws and regulations.

TransCelerate did not develop this methodology as a replacement for current anonymization approaches used by sponsors, vendors, and other relevant stakeholders. Application of the Privacy Methodology alone may not be sufficient to anonymize clinical trial data.

Users should consider how the Privacy Methodology works in connection with their current approaches to increase data utility.
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

Describe in detail that there may be a significant variance in selected data on events that potentially could uniquely identify a study participant, either through a unique combination of variables or anomalous behavior or a unique variable value. It is recommended that unique identifiers be included in a study's data to help increase scientific utility while still safeguarding the participant's data.

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

- **Example**
  - Consideration of the variable type "Before" and "After" following the recommended approach.

- **Considerations**
  - Providing additional insights and factors of importance.

#### 6.1.2 Considerations

**Table 1: Example of recoding study unique identifier**

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>ID</td>
</tr>
<tr>
<td>Smith</td>
<td>12345</td>
</tr>
<tr>
<td>Jones</td>
<td>67890</td>
</tr>
</tbody>
</table>

**Note:**

Complementary consideration to contextualize the variable type.
Methodology Example: Section 4.6 Data with Low Frequencies

Example of Redacting Low Frequency Sex and Race

<table>
<thead>
<tr>
<th>USUBJID</th>
<th>DOMAIN</th>
<th>SEX</th>
<th>SEX</th>
<th>BEFORE</th>
<th>AFT</th>
<th>RACE</th>
<th>RACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234-5678-USA003-10001</td>
<td>DM</td>
<td>F</td>
<td>-- Redacted --</td>
<td>WHITE</td>
<td>W</td>
<td>WHITE</td>
<td></td>
</tr>
<tr>
<td>1234-5678-POL002-10003</td>
<td>DM</td>
<td>M</td>
<td>-- Redacted --</td>
<td>WHITE</td>
<td>W</td>
<td>WHITE</td>
<td></td>
</tr>
<tr>
<td>1234-5678-GER002-10004</td>
<td>DM</td>
<td>F</td>
<td>-- Redacted --</td>
<td>UNKNOWN</td>
<td>--</td>
<td>Redacted --</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>USUBJID</th>
<th>DOMAIN</th>
<th>AEDECOD</th>
<th>AEDECOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234-5678-POL002-10003</td>
<td>AE</td>
<td>Oligospermia</td>
<td>-- Redacted --</td>
</tr>
<tr>
<td>1234-5678-POL002-10003</td>
<td>AE</td>
<td>Headache</td>
<td>Headache</td>
</tr>
<tr>
<td>1234-5678-POL002-10003</td>
<td>AE</td>
<td>Diarrhea</td>
<td>Diarrhea</td>
</tr>
</tbody>
</table>
**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.
What are the key takeaways from the Public Review?
Diverse Participation to the Public Review

Privacy Methodology

• 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

% of Responders per Stakeholder Group

- Academia: 7%
- Data sharing platform: 7%
- Biopharma/Pharma Companies: 14%
- Trade Association/Industry Consortia: 57%
- Vendor: 7%
- Regulators: 7%
Additional References

TransCelerate Data Privacy Website


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)
Q&A
Thank you!

For additional resources and information, please visit:
https://crdsalliance.org/resources