The 2024 CRDSA Summit

The Patient Data Revolution: From Promise to Realization

February 29, 2024

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

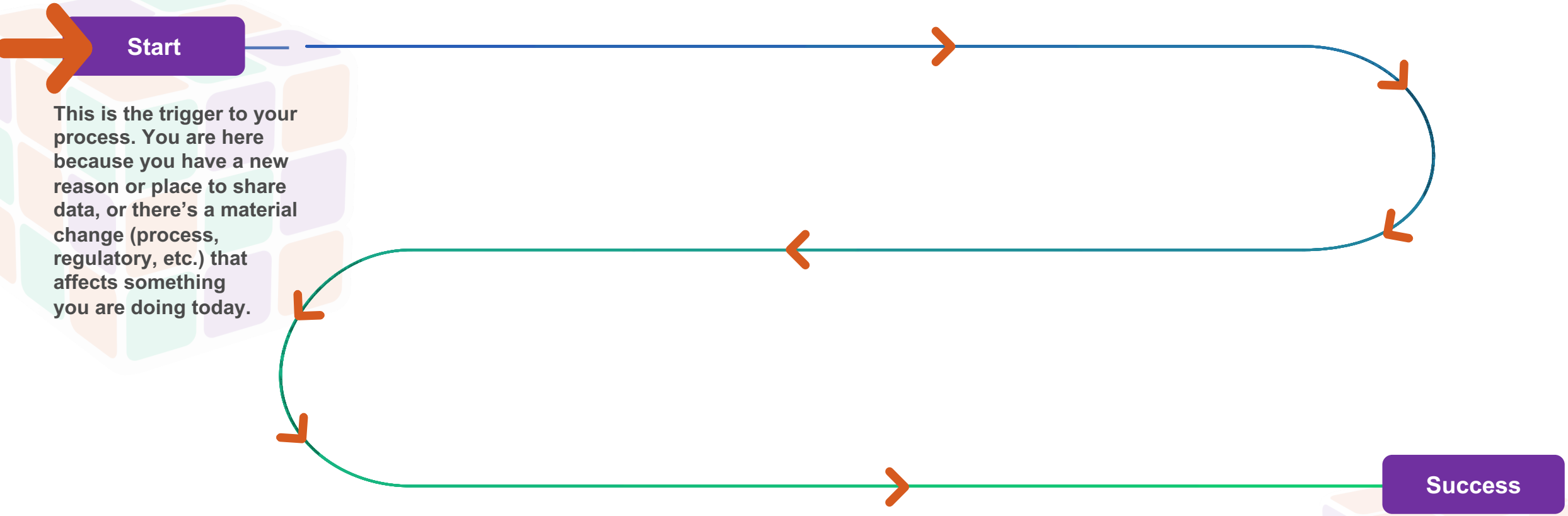
Jeppe G. Manuel
Principal R&D Data Privacy Specialist
Novo Nordisk

CRDSA
Clinical Research Data Sharing Alliance

www.crdsalliance.org

Data Protection & Privacy Decision Flow

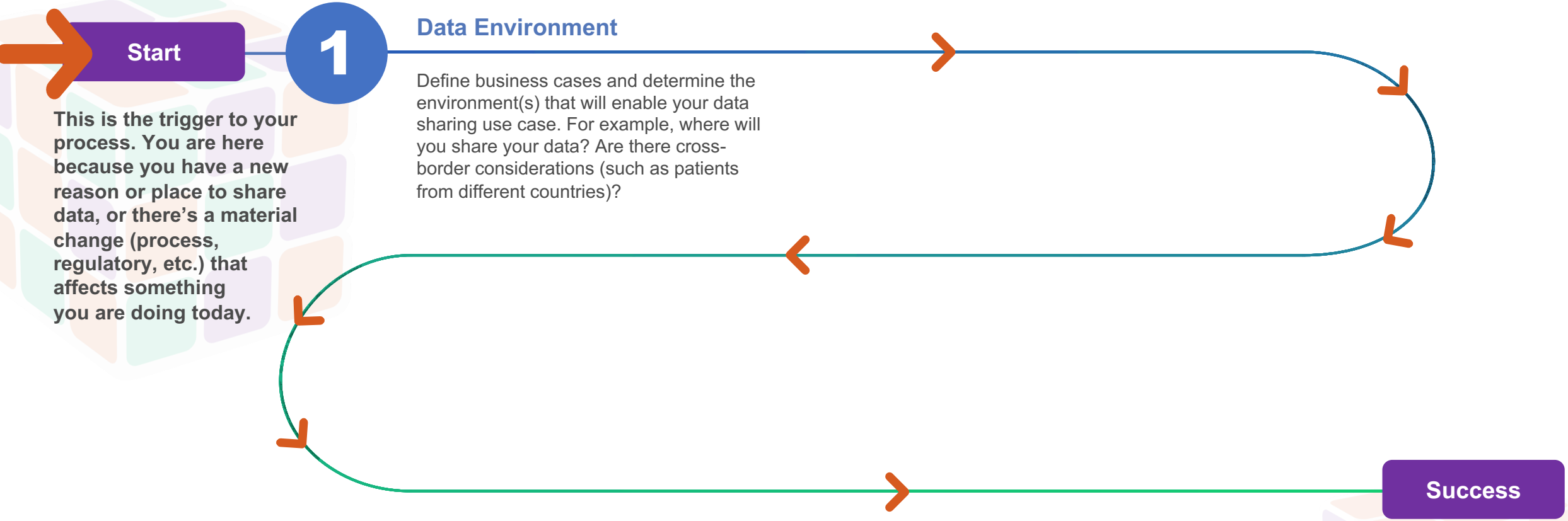
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.



CRDSA

Data Protection & Privacy Decision Flow

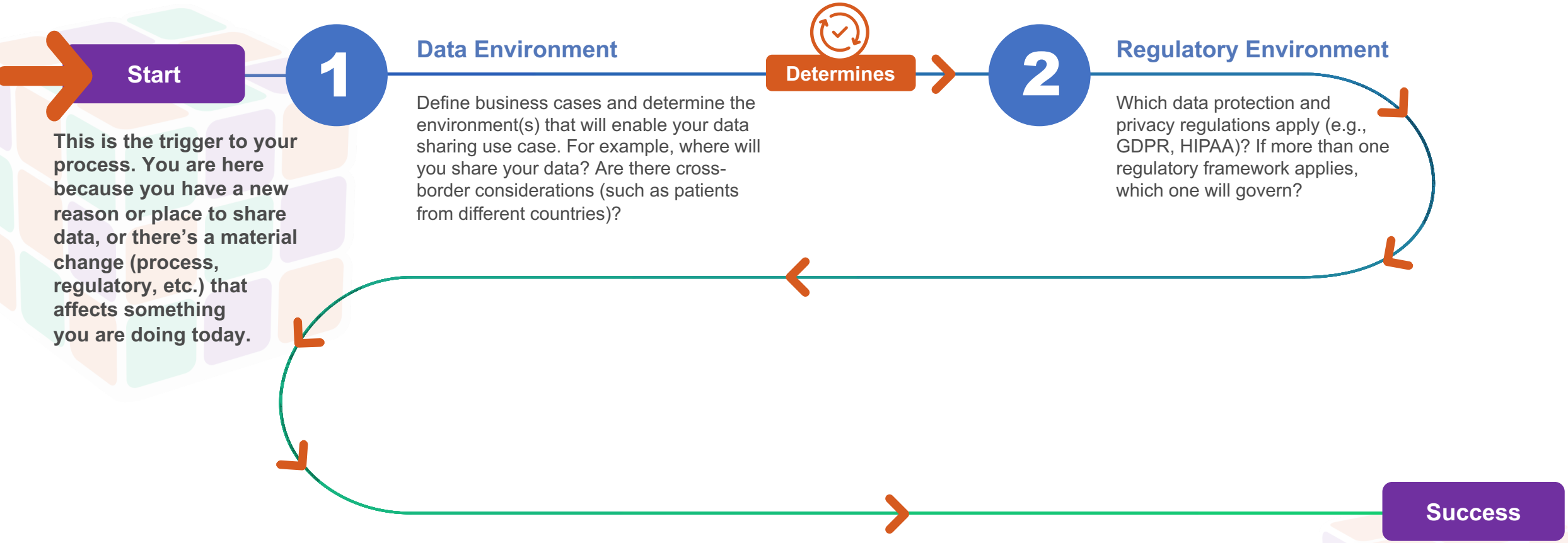
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.



CRDSA

Data Protection & Privacy Decision Flow

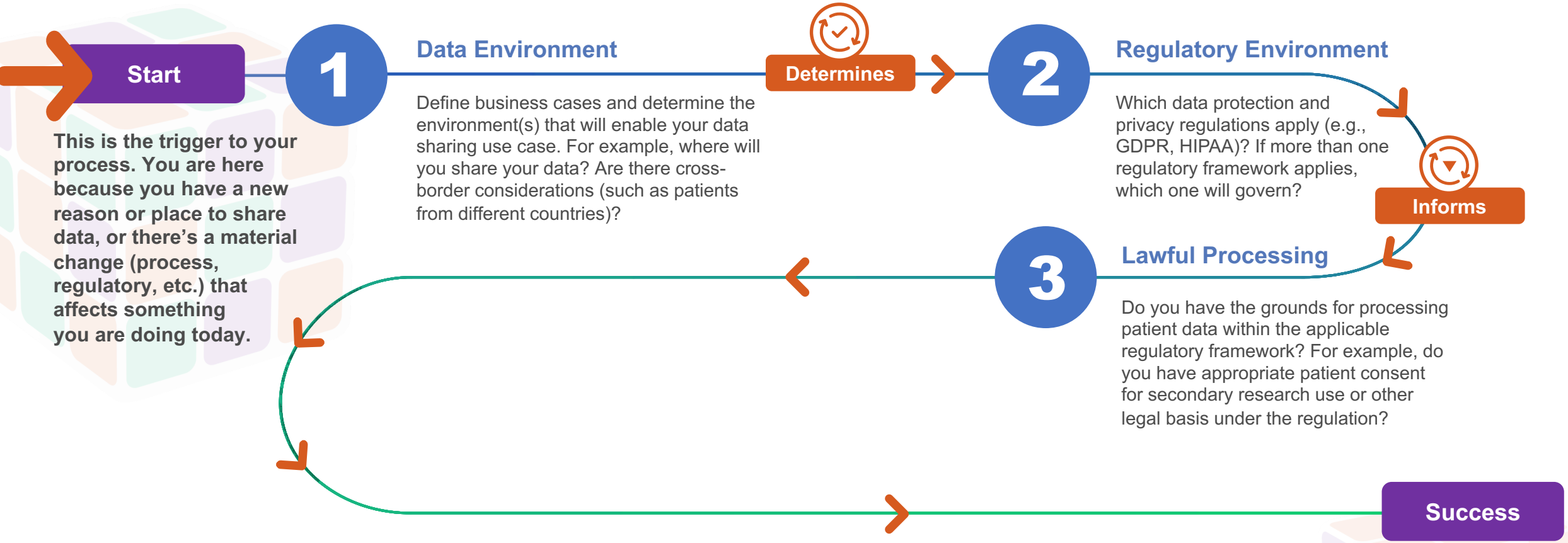
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.



CRDSA

Data Protection & Privacy Decision Flow

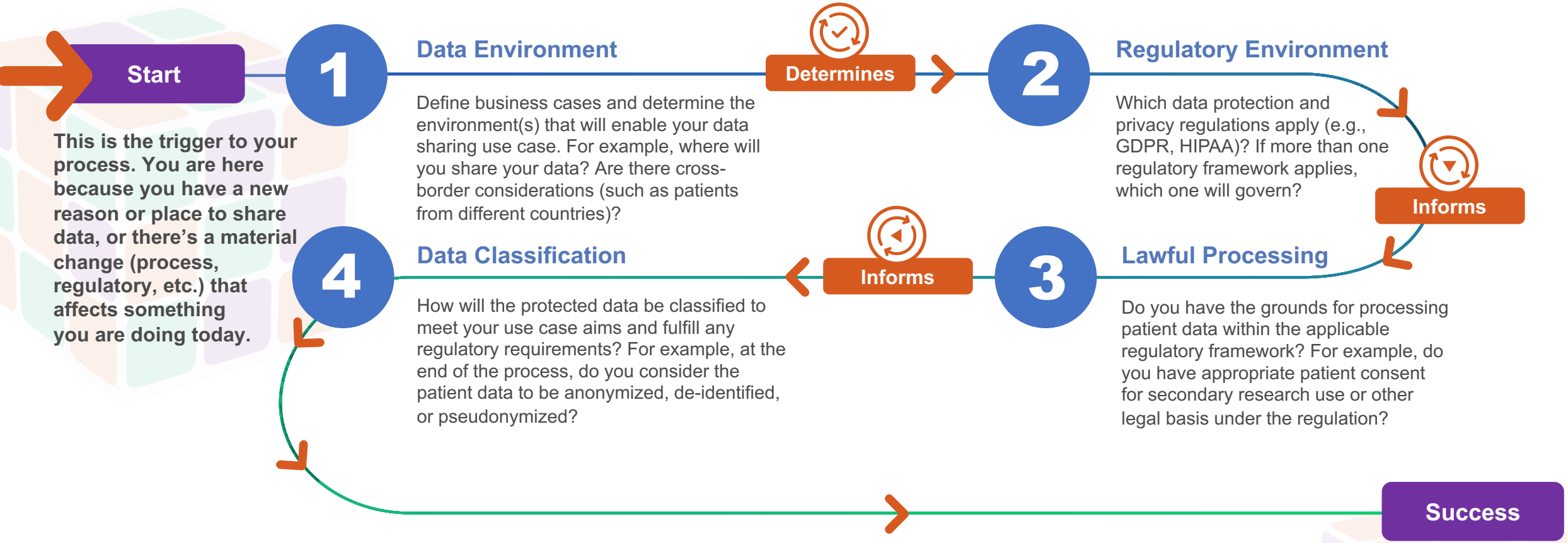
Operationalizing privacy is challenging, but there are clear steps you can take to devise an operational policy that supports your business goals. The following decision flow can bring clarity and order to the process of sharing individual patient data.



CRDSA

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.



CRDSA

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.

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.

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)?

Determines

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?

Informs

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?

Informs

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?

Informs

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).

Success

CRDSA

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.

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.

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)?

Determines

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?

Informs

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?

Informs

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?

Informs

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).

Determines

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?

Success

The result is a data protection approach that supports your use cases and responsibly protects patient privacy.

CRDSA

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

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®

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 comment

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

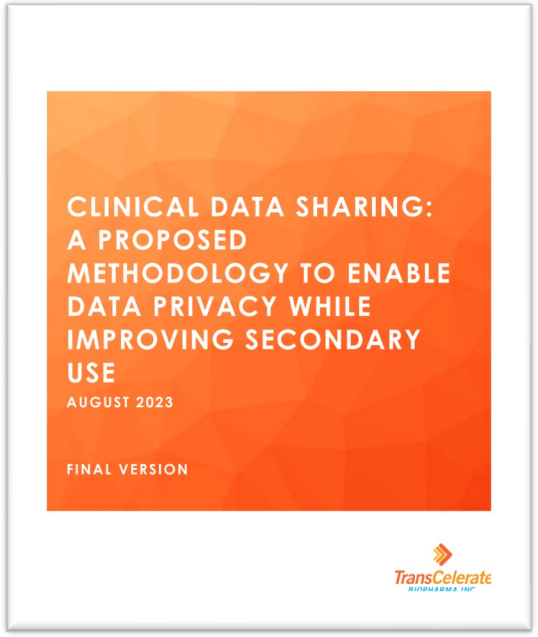
Goal of the Privacy Methodology Solution

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

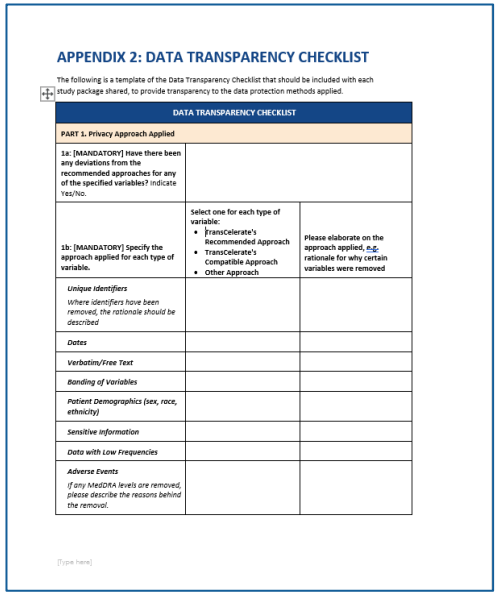
Reduce Variations that hamper cross-study analysis across the ecosystem by reducing variability of the data privacy measures applied during anonymization

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

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?

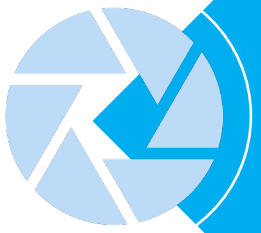
These tools are a result of a three-year collaboration involving members from more that 20 Sponsor Companies willing to share best practices and develop the privacy methodology and transparency checklist.

TransCelerate BIOPHARMA INC.

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

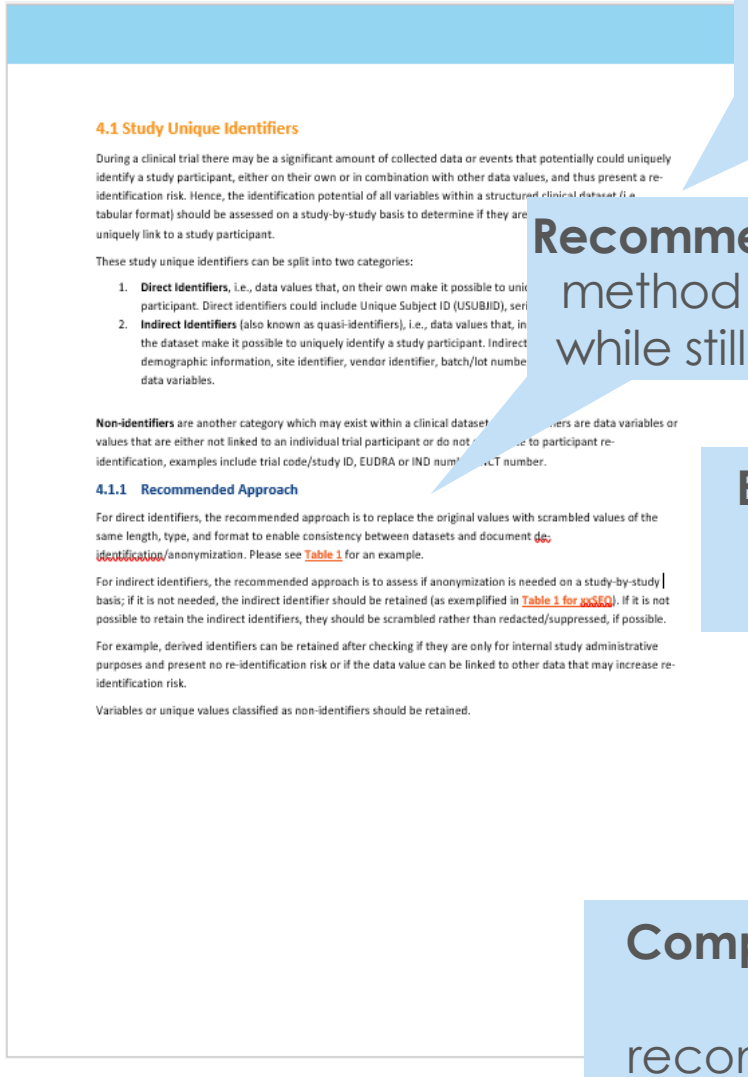
Background description to contextualize the variable type

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



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 (i.e. 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 uni participant. Direct identifiers could include Unique Subject ID (USUBJID), seri
2. **Indirect Identifiers** (also known as quasi-identifiers), i.e., data values that, in the dataset make it possible to uniquely identify a study participant. Indirect demographic information, site identifier, vendor identifier, batch/lot numbe data variables.

Non-identifiers are another category which may exist within a clinical dataset ers are data variables or values that are either not linked to an individual trial participant or do not e to participant re-identification, examples include trial code/study ID, EUDRA or IND num CT number.

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 de-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 xxSEQ). 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.

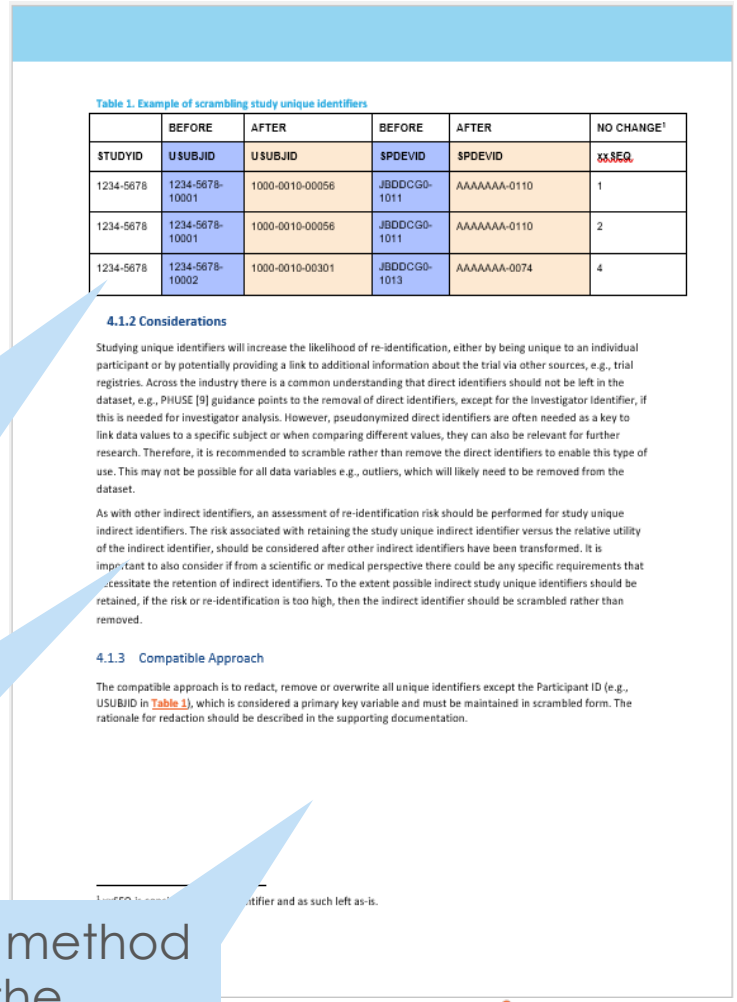


Table 1. Example of scrambling study unique identifiers

	BEFORE	AFTER	BEFORE	AFTER	NO CHANGE[1]
STUDYID	USUBJID	USUBJID	SPDEVID	SPDEVID	xx SEQ
1234-5678	1234-5678-10001	1000-0010-00056	JBDDCG0-1011	AAAAAAA-0110	1
1234-5678	1234-5678-10001	1000-0010-00056	JBDDCG0-1011	AAAAAAA-0110	2
1234-5678	1234-5678-10002	1000-0010-00301	JBDDCG0-1013	AAAAAAA-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.

tifier and as such left as-is.

TransCelerate BIOPHARMA INC.

Methodology Example: Section 4.6 Data with Low Frequencies

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 --

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

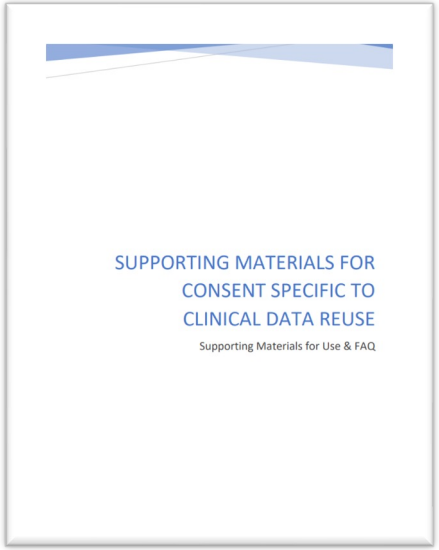
		BEFORE	AFTER
USUBJID	**DOMAIN**	**AEDECOD**	**AEDECOD**
1234-5678-POL002-10003	AE	Oligospermia	-- Redacted --
1234-5678-POL002-10003	AE	Headache	Headache
1234-5678-POL002-10003	AE	Diarrhea	Diarrhea

Additional Tools

Educational Toolkit, Poster and FAQs

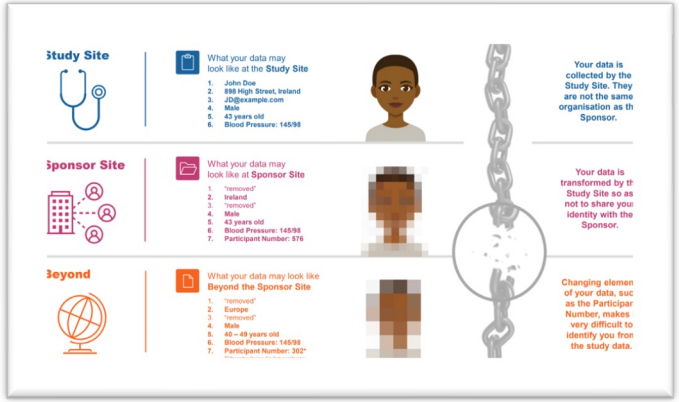
Educational Toolkit

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.



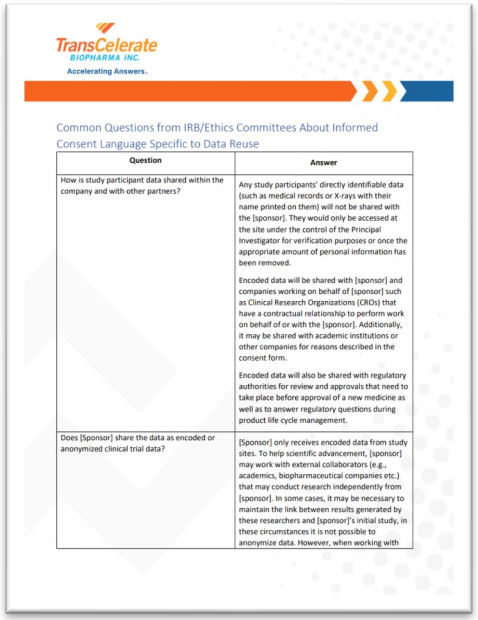
Educational Poster

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.

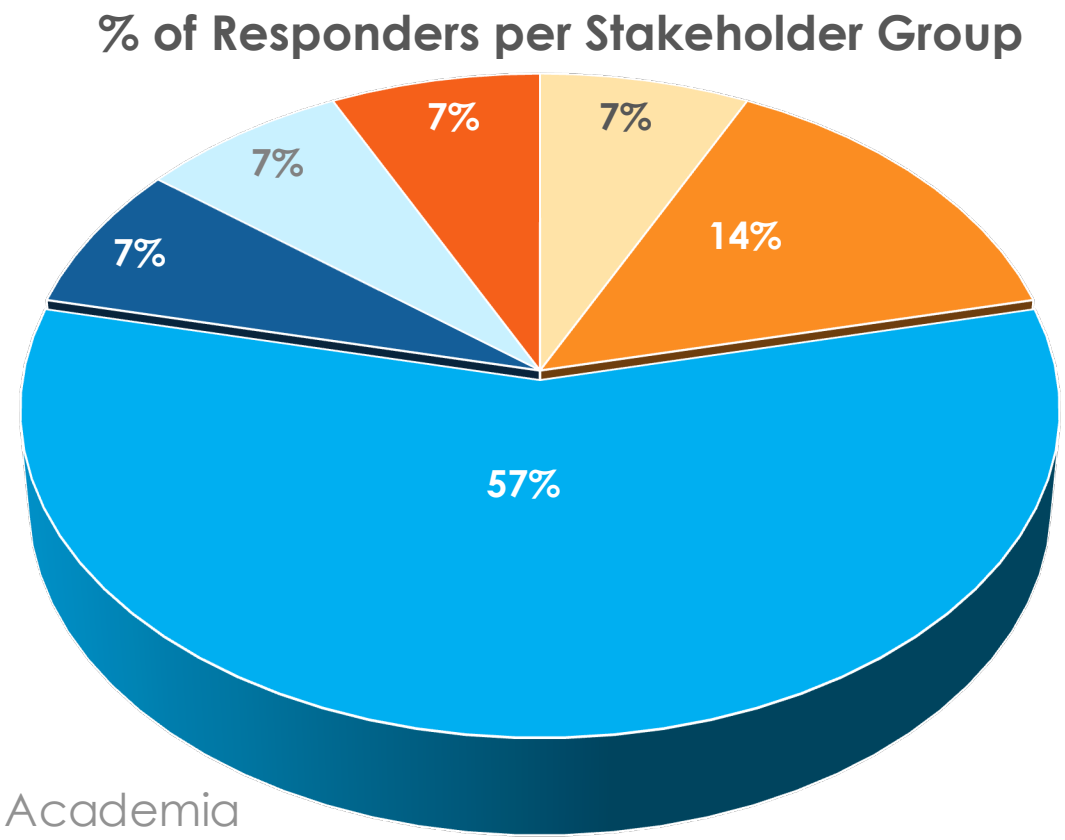


TransCelerate BIOPHARMA INC.

What are the key take aways 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

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)

Q&A

CRDSA

Clinical Research Data Sharing Alliance

Thank you!

For additional resources and information, please visit:

https://crdsalliance.org/resources