



CENTER FOR GLOBAL CLINICAL RESEARCH DATA

Anonymization and Data De-Identification Guide

Background

This document applies *only* to clinical trial data shared on the Vivli platform. Other organizations or industries may take different approaches. The Future of Privacy Forum's [A Visual Guide to Practical Data De-Identification](#) offers an overview of data identifiability more broadly.

From Source Data to Anonymized Data – What are the Steps?

1

Not Coded

SOURCE DATA

Source data is the information the investigator directly records in the patient record, direct observations, lab results, or data which the patient records directly into questionnaires

Jonathan Bob Jones

23April 1966 DOB

58 years old

Chicago, IL Investigative site

Site 45678

04Aug2017 First dose of study drug

04 Aug 2017

14Aug2017 2nd study visit

14 Aug 2017

Decreasing risk of patient re-identification

From Source Data to Anonymized Data – What are the Steps?

Bolded information is
in clinical databases

1

Not Coded Source Data

Jonathan Bob Jones
23 April 1966 DOB
Chicago, IL Investigative site
04Aug2017 First dose of study drug
14Aug2017 2nd study visit

Each investigative site maintains a listing of the patient's name and the assigned study subject number.

2

Pseudonymized or De-identified (aka key coded)

Subject 123
58 years old
Site 45678
Day 0
Day 10

Pseudonymized is popular in Europe
De-identified is popular in N. America

Many privacy laws allow the data to be disclosed or shared for secondary purposes with consent once data is de-identified / pseudonymized

De-identification standard in the U.S is Safe Harbor (as part of HIPAA)

We can track backwards that Subject 123 is Jonathan Bob Jones as the key is preserved between name and Subject ID

Decreasing risk of patient re-identification

From Source Data to Anonymized Data – What are the Steps?

2

Pseudonymized or De-identified (aka key coded)

Subject **123**
58 years old
Site **45678**
Day 0
Day 10



Sever link and Key Scrambled
and destroyed
now we cannot link
Subject 890 to 123 or Jonathan
Bob Jones

3

Anonymized

Multiple methods
can be applied



Subject 890
Aged 50-60
US
Study Day 0
Study Day 10

Applied a risk
reduction algorithm
to mitigate re-id
Put age in 10 yr age
bins for example,
day offsets, redact
AE narratives etc

Vivli requires Anonymized data to
be contributed to the platform

Decreasing risk of patient re-identification

From Source Data to Anonymized Data – What are the Steps?

Source data is the information the investigator directly records in the patient record.

The methodology applied is known and maintained as a record. It can be tracked backwards that subject 123 was Jonathan Bob Jones

The methodology is applied and no record of it is maintained. Subject 890 can NOT be tracked back as 123. Vivli requires anonymized data to be contributed

1

Source Data

Jonathan Bob Jones
23 April 1966 DOB
Chicago, IL Investigative site
04Aug2017 First dose of study drug
14Aug2017 2nd study visit

2

Pseudonymized or De-identified

(aka key coded)
Subject 123
58 years old
Site 45678
Day 0
Day 10

3

Anonymized

Subject 890
Aged 50-65
US
Study Day 0
Study Day 10



Decreasing risk of patient re-identification

Providing Anonymized Data— How Vivli can help



PIs and their teams will need to decide if they have the capability and capacity to anonymize their data before uploading to Vivli.



To assist in this process, Vivli partners with d-wise and Privacy Analytics, industry leaders in health data-sharing methods and technologies, data security and de-identification. Our partners offer anonymization services to individuals and organizations who wish to share clinical research data on the Vivli platform. For more information, contact support@vivli.org.