

Privacy Considerations of Using RWD from Statistical Programming Perspective

Joe Xi, Lorraine Fang

Mar 2026

Disclaimer

This presentation shares only the author's experience and opinions, and not necessarily those of Bristol Myers Squibb (BMS). BMS does not guarantee the accuracy or reliability of the information provided in the slides.

Outline

- **Introduction**
- **Policy and Strategy**
- **Data Storage and Access Control**
- **De-identification**
- **Conclusion**

Introduction

Regulatory Landscape

- Real-World Data (RWD) brings **Privacy Challenges**.
- **FDA/EMA** and **CDISC/PHUSE** set high standards for protecting patient identities and handling data securely.

Diverse Data Sources

- Data vendors with specific privacy requirements.

Objective

- Statistical programmers approach to be compliant and applicable.

Policy and Core Strategy

Industry regulation

- Policy covers data use, access/share, storage, and anonymization to meet legal regulations and vendor requirements.
- HIPPA and GDPR.
- BMS enforces a comprehensive policy on RWD privacy, focusing on internal controls.

Workflow

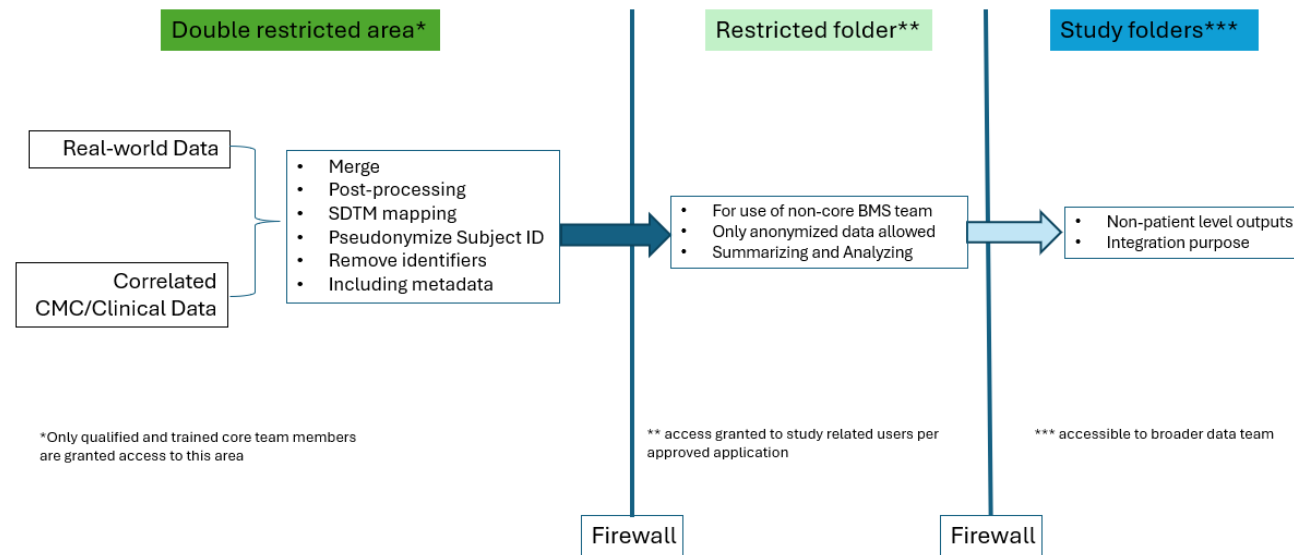
- Daily practice for statistics and programming teams: Access restriction and de-identification

Data Storage & Access

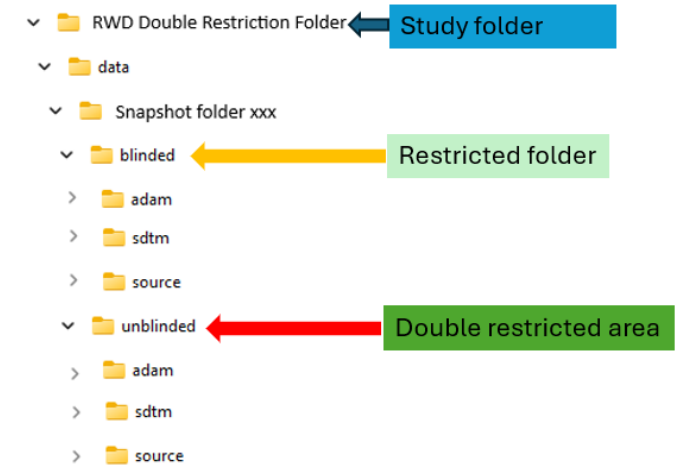
- Data stored on **dedicated, validated servers** or regular servers with specially designed folder access.
- Restricted to **authorized users**.
- Physical servers in regions (US, EU, etc.) complying with vendor requirements.

Data Storage & Access –cont.

Double Restriction



Folder Structure



De-identification

Regulatory Guidelines

- **HIPAA** requires removal or masking of 18 specific identifiers (e.g., names, detailed geography smaller than a state, and specific dates except years).
- **GDPR** requires anonymization and pseudonymization of patient information that can directly or indirectly identify a person.
- Company level requirements: patient identity protected through anonymization/de-identification practices; Vendor requirements also fulfilled

Original Patient Data



Name: John Doe
ID: xxx-014
Site: Madison
Country: USA
Address: 1 BMS Rd., Madison, NJ 07111
DOB: 1990-01-01
First TX Date: 2025-01-01
Visit 1 Date: 2025-02-01

Masked Patient Data



Name: -----
ID: 0001
Site: Site 1
Country: USA
Address: New Jersey
DOB: -12784 (days)
First TX Date: 0
Visit 1 Date: 31 (days)

De-identification – Subject ID

Purpose

- Implement additional layer of **pseudonymization** before data sharing to any non-core team.
- Algorithm/crosswalk table stored in restricted areas, and accessible to authorized personnel only, preventing end-users from reversing the process.

Methodology

- **Alphanumeric ID:** Map using a one-to-one cryptographic crosswalk table.
- **Numeric ID:** Transform via a secure mathematical algorithm.

Alphanumeric ID

Orig. ID		Pseu. ID
ABC012	➡	001
DEF453	➡	002
...		...
UVW103	➡	004
XXZ554	➡	005

Numeric ID

Orig. ID		Pseu. ID
001		14
003		100
...		...
199		8528
200		8571

x43-29 ➡

De-identification - Dates

Purpose & Concerns

- Maintain feasibility for clinical analysis, while simultaneously ensure patient privacy.
- Mitigate **re-identification** risk.

Approach

- Convert to **relative day** format referencing either first study treatment or another significant milestone.
- In SDTM/ADaM datasets sharing, suppress of absolute dates variables.

De-identification – Other Personal Information

General Concerns

- Per **HIPAA**, geographic information be pseudonymized or masked up to the state level in the US.
- Sensitive information NOT shared with any third party.

Methods

- **Critical information:** e.g. site information for site analysis pseudonymized using aliases (e.g., 'site 1', 'site 2') to protect personal geographic information from being released.
- **Masking:** irrelevant geographic information masked to US state or equivalent level.
- **Removal:** Examples: contact information, account information, technical identifiers, biometrics, images, and job titles.
- Study team reviews and agrees on information to be de-identified/masked for analysis purpose or removed

Conclusion

- *Two Pillars to manage RWD Privacy:*
 - ❑ Pillar 1: Strict Access and Storage
 - ❑ Pillar 2: Rigorous De-identification
- *Outcomes:*
 - ❑ keep up with global regulations and industry standards.
 - ❑ Data remain useful for research w/o releasing patient privacy.
 - ❑ makes our evidence ethical, secure, and ready for regulatory audit.

Reference

- (1) Jaffe D.H., Malin B.A., Hendricks-Sturup R.M. (2025). “A real-world data challenge: guidance for aligning data privacy compliance and fit-for-purpose usability”. Health Affairs Scholar, 2025 Nov 12;3(11).
- (2) Ferran J.M., Lanoue J. (2015). “PhUSE De-Identification Working Group: Providing De-Identification Standards to CDISC Data Models”. PharmaSUG 2015 -Paper DS10.
- (3) FDA Guidance (2021) “Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry”
- (4) GDPR <https://gdpr-info.eu/> (accessed 12/02/2025).
- (5) HIPAA: <https://www.hhs.gov/hipaa/index.html> (accessed 12/02/2025).

Acknowledgement

Great thanks to Joy Wang and Samar Noor

Thank you