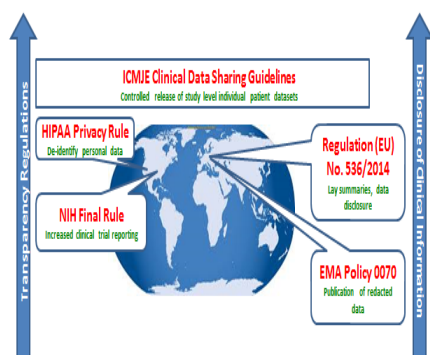


Optimizing Synergies between Data and Document Anonymization

Peggy Zorn, Rashmi Dodia and Ravi Yandamuri



Background



Problem Statement

There are several major differences that affect anonymization processes between clinical trial individual patient datasets and clinical trial study documents:

Different Resources Involved in Anonymization

Regulatory/Clinical vs. Programming and Statistics

Different Purposes

Public Awareness vs. Advancing Health Research

Different Release Strategies

General Public/Industry vs. Controlled Release to Researchers

Different Regulations/Standards

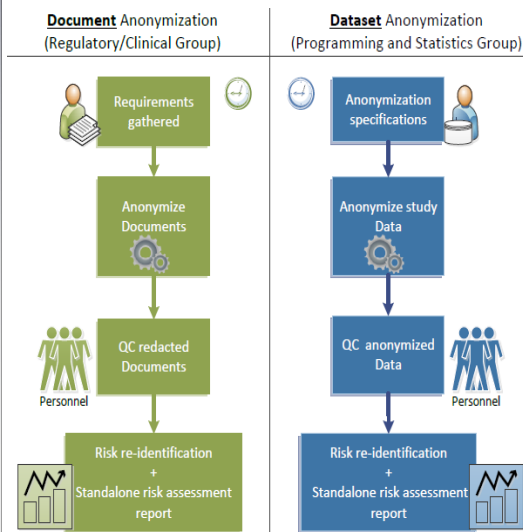
General Agency Guidance vs. PhUSE Specifications

Different Input Formats – Automation Limitations

Unstructured Text/PDFs vs. Well-structured Datasets

These differences can lead to inconsistencies in anonymization approaches and outputs for a single study. Inconsistencies in anonymization approaches could lead to a greater risk of protected personal data being disclosed.

Standalone Anonymization Processes



Recommendations

Increased Interaction

Increase Consistency and Efficiency

Better Defined Study Eligibility Criteria

Reduce Risk of Re-identification

Integrated Set of Specifications

Reduce Overall Efforts

Use of Dataset Inputs for Key Variable Identification in Documents

Increase Accuracy and Alignment Between Outputs

Integrated Risk Assessment

Reduce Risk and Costs

Need for Comprehensive Standards

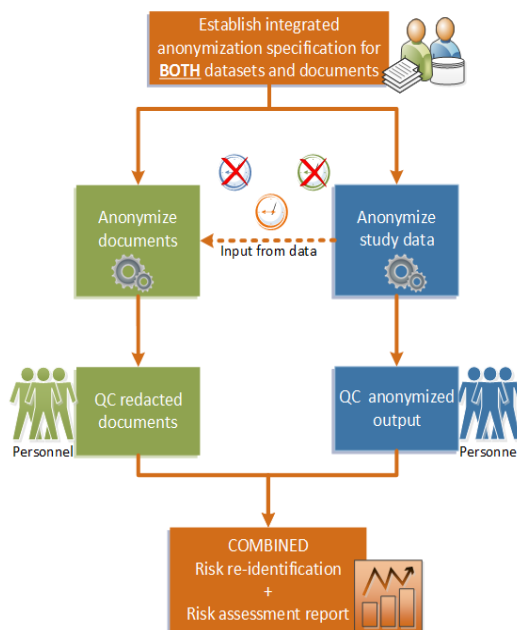
Increase Dialog Within the Industry

Need for Advances in Technology

Increase Automation Based on Standards

Proactive Approach Produces the BEST RESULTS!

Integrated Anonymization Processes Documents and Datasets



Conclusions

Integration of anonymization processes for both datasets and documents could result in:

- **Precise/consistent outputs** – as similar standards would be utilized for both
- **Increased efficiency** – as the overall time spent in establishing integrated specifications and combined risk-assessment would be reduced
- **Lesser risk of re-identification** – as a consistent approach would be used for both anonymizations
- **Reduced costs** – Combined efforts would lead to the overall reduction of resources used for anonymization