

Why Data and Document Anonymization Should Carpool

Ravi Yandamuri, Rashmi Dodia – MMS

Abstract / Introduction

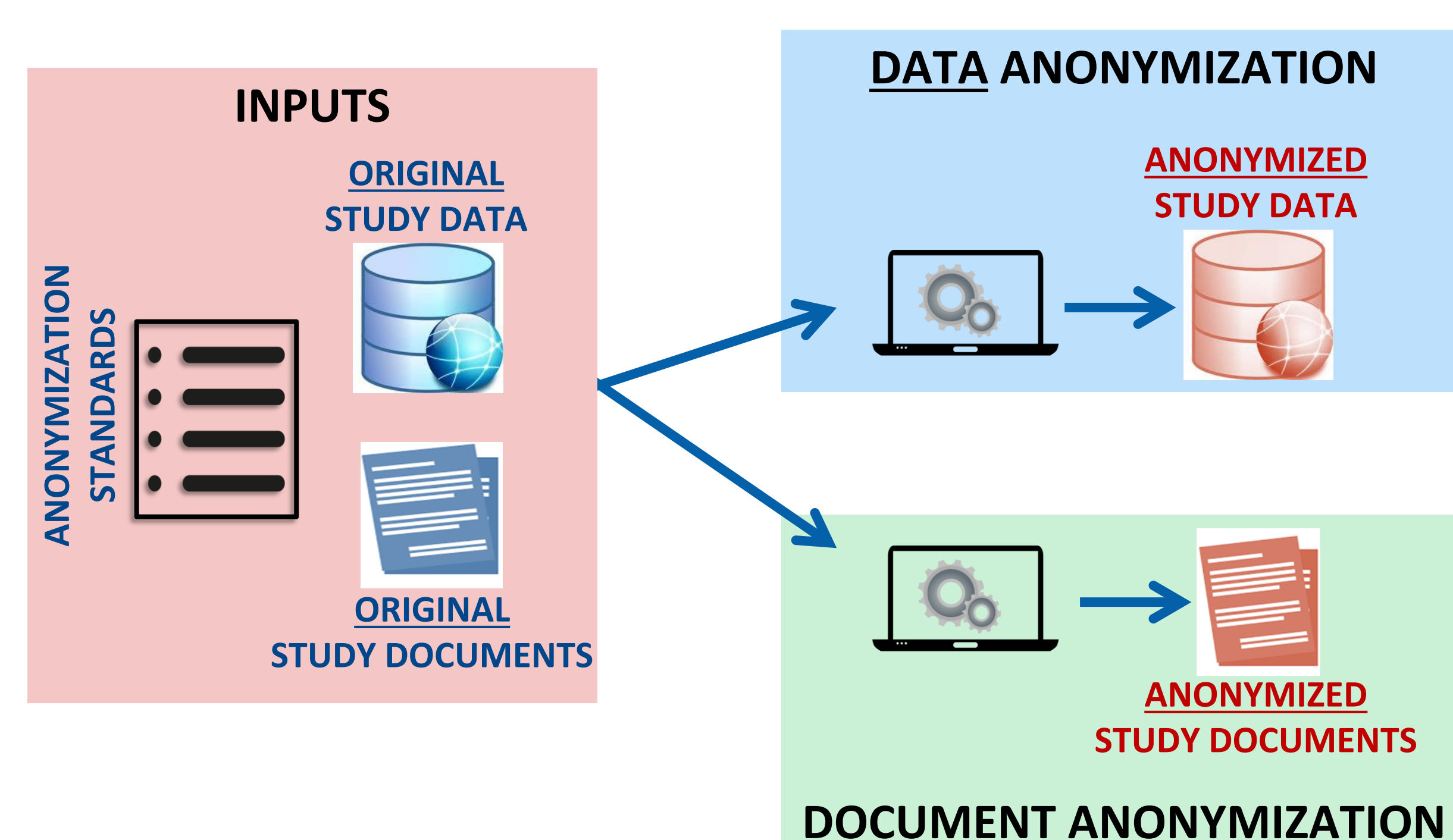
Two folks, **Data anonymization** and **Document anonymization**, have long since travelled separately to their common destination of transparency and data sharing. Each one had their own standards, techniques, and audience. But, as the regulations evolve and the traffic intensifies, it is now highly recommended that these two should carpool to anonymization.

In this poster, we propose and compare three models to carpooling with a goal to improve efficiency and quality of deliverables, along with added cost benefits to the sponsor. The best suited can be implemented as per the sponsor's need. Differences in audience, sharing times and sharing platforms are considered. Compliance to EMA Policy 0070, PhUSE data de-identification standards and guidance are also considered throughout the proposed models.

Problem

Although anonymized data and documents contain similar information, historically their disclosures are handled separately. Due to requirements and guidance coming from different directions, many are blinded of the efficiencies that could be gained by taking a concerted approach for performing these two tasks.

Model # 1: Parallel Work Flow

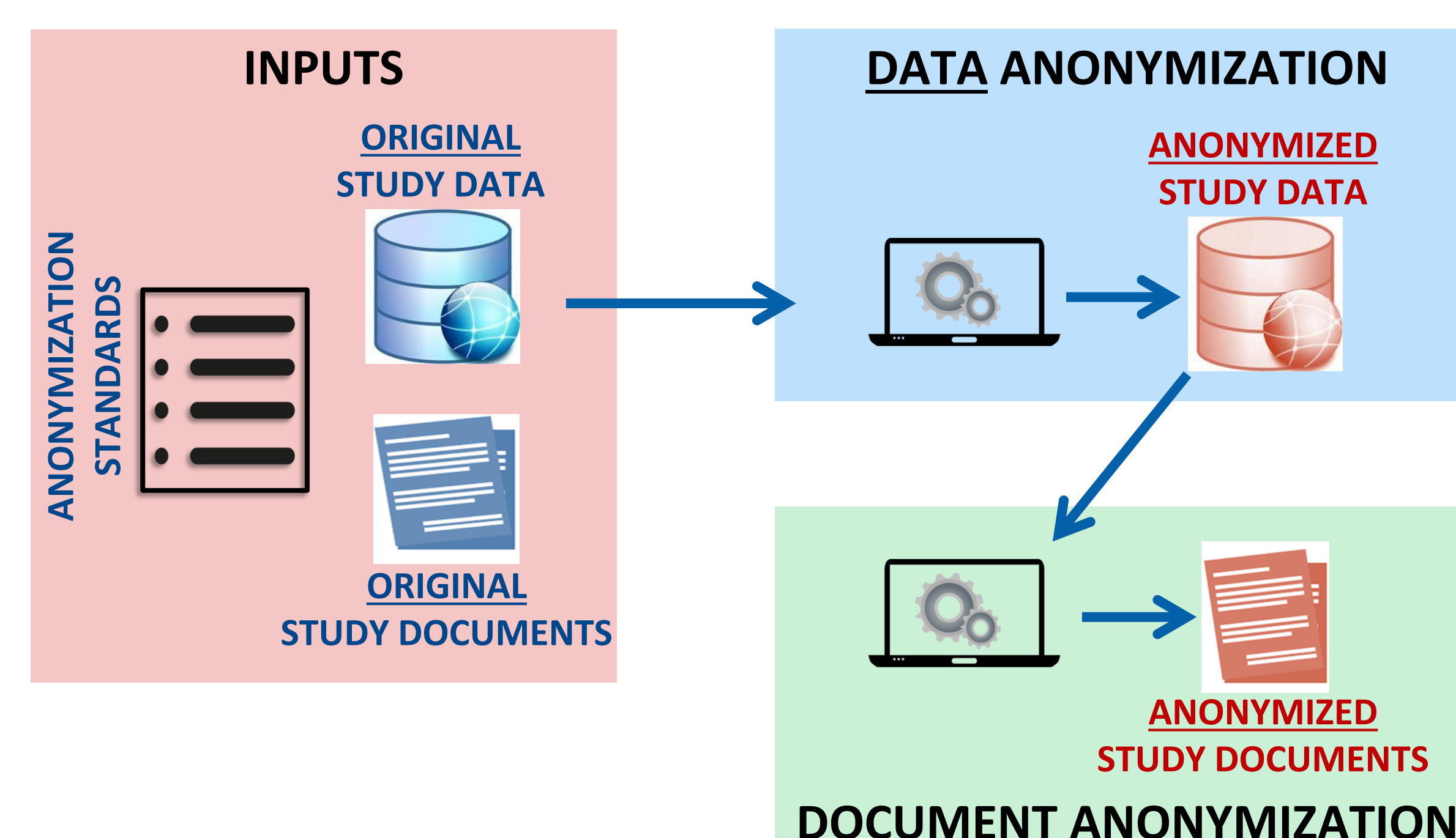


- ✓ Simple and straight forward
- ✓ Deliverables are not dependent on one another
- ✓ Low risk of re-identification



- × Inconsistency in anonymized values. Same values in data are anonymized differently in documents and vice-versa
- × Requires access to data even when only document anonymization is required

Model # 2: Continuous Work Flow

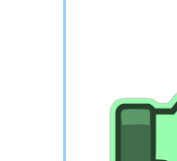
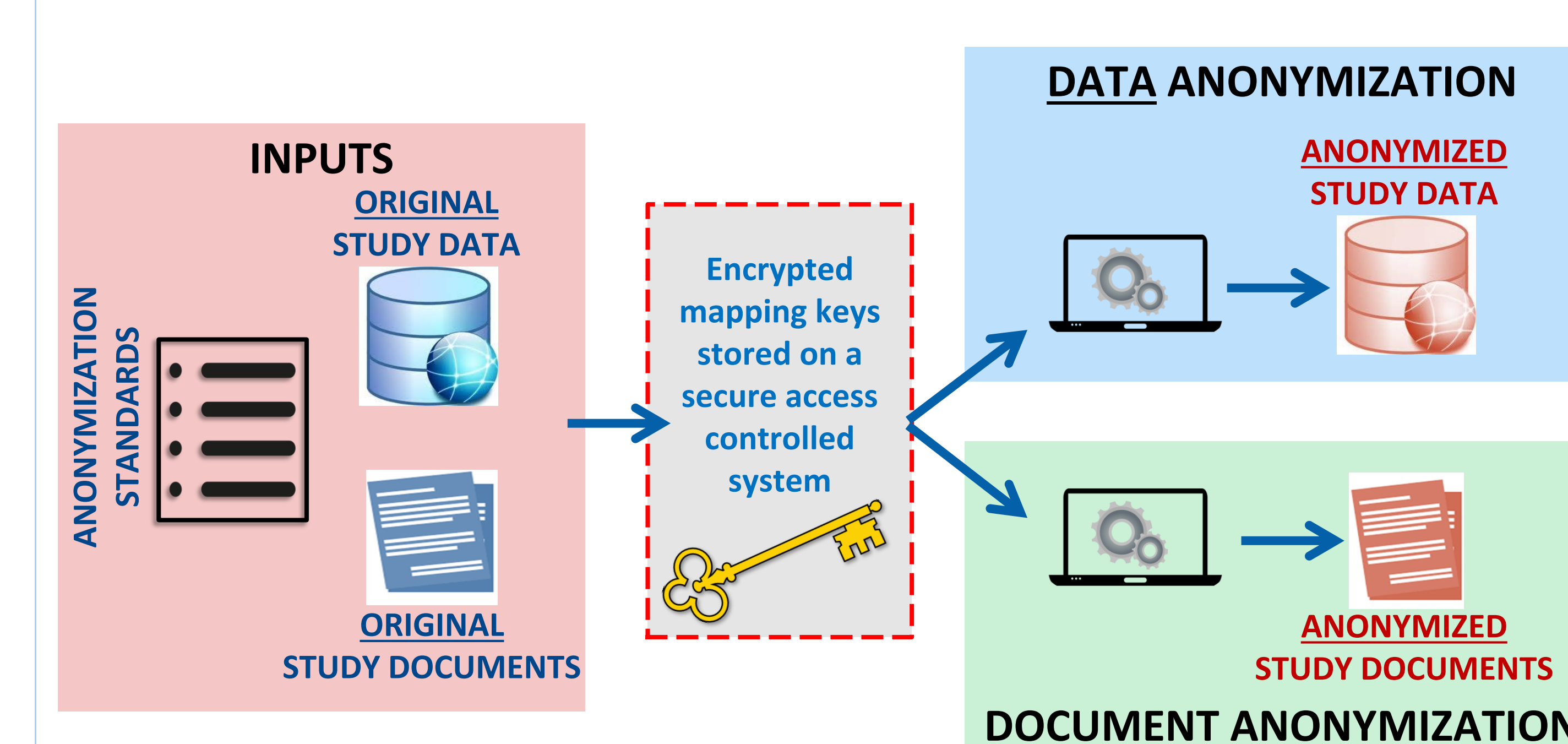


- ✓ Improved data utility
- ✓ True and consistent anonymization = Lower risk of re-identification compared to model # 1

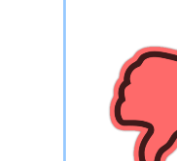


- × Requires more advanced toolset and set-up
- × Redundant effort as document anonymization requires data to be anonymized first.
- × Sharing for different audience requires re-do.

Model # 3: Parallel with Benefits of Continuous



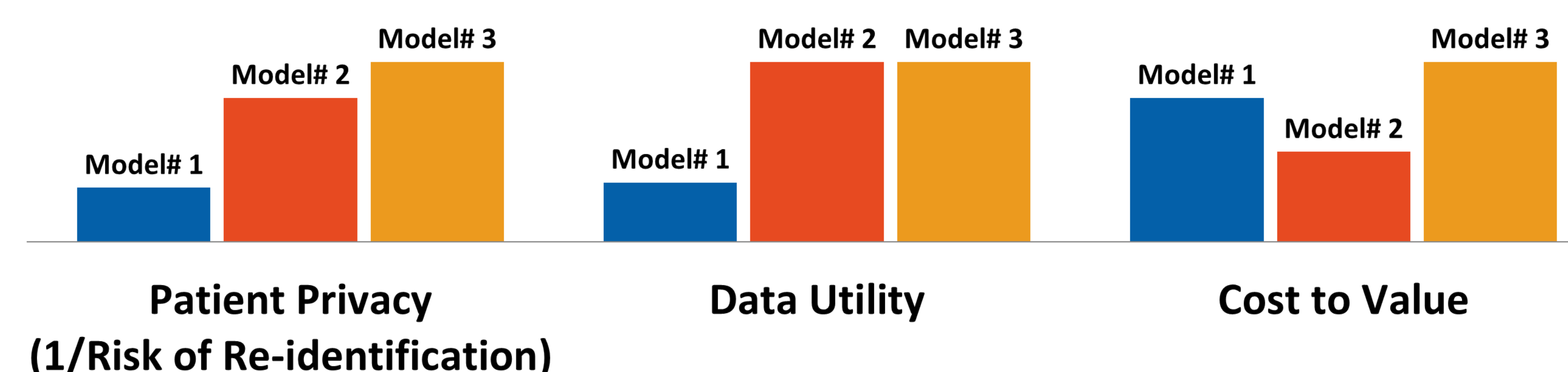
- ✓ All Pros of Model # 1 and Model # 2
- ✓ Consistent anonymization across deliverables, even when delivered at different times = Lowest risk of re-identification of all models
- ✓ Ability to handle ongoing and/or core and extension trials



- × Stored mapping keys challenges the traditional definition of anonymization - irreversible engineering.

Comparison of Estimated Cost, Data Utility and Risk of Re-identification Levels

- A survey was conducted with experts in areas of data and document anonymization who provided inputs on key aspects of data sharing
- Estimations are based on a scale of 0 - 1
- Ideal scenario = High Patient Privacy + High Data Utility + High Cost to Value



Conclusion

All three models provide a combined approach for data and document anonymization. Model # 3 provides the most benefits for patient privacy, data utility and cost aspects of data sharing. Depending on the sponsor's requirement, with proper justification, any of the above three models can be adopted effectively.