

De-identification

Safe Harbor Method

- Remove 18 HIPAA identifiers
- Dates shifted or generalized
- Ages >89 aggregated
- Geographic areas >20,000

Expert Determination

- Statistical/scientific analysis
- Very small re-identification risk
- Certified expert assessment
- More data retained

Automated Tools

- NER for PHI detection
- Date shifting algorithms
- Scrubber software (Philter, BoB)
- Validation required

Re-identification Risk & Synthetic Data

- K-anonymity, l-diversity
- Differential privacy
- Synthetic data generation (GANs)
- Trade-off: utility vs privacy