NIH NATIONAL CANCER INSTITUTE

Conventional Approaches to De-Identification

Setting the Stage

UAMS
College of Medicine

Fred Prior, PhD
Professor and Chair
Department of Biomedical Informatics
Professor of Radiology
University of Arkansas for Medical Sciences

Image De-identification for Open Access Data Sharing

- Open access or shared research data must comply with regulations that govern patient privacy.
  - Health Insurance Portability and Accountability Act (HIPAA) in the US
  - General Data Protection Regulation (GDPR) in the EU
- These regulations require the removal of protected health information (PHI) and other personally identifiable information (PII) from datasets before they can be made publicly available.
- Covered entities (US) or Data Controllers (EU) are legally responsible for compliance, even if the data publisher is exempt.

De-Identification, Anonymization, Pseudonymization

- **De-identificatio**n of medical record data refers to the removal or replacement of personal identifiers so that it would be **difficult** to re-establish a link between the individual and his or her data. (Kushida, et al. https://doi.org/10.1097/mlr.0b013e3182585355 (2012).)
  - the removal of **specified individual identifiers** as well as **absence of actual knowledge** by the covered entity that the remaining information could be used alone or in combination with other information to identify the individual. (HIPAA, 45 CFR Part 160 and Part 164.)
- **Anonymization** refers to the irreversible removal of the link between the individual and their medical record data to the degree that it would be virtually impossible to reestablish the link
  - To achieve anonymization under GDPR, **re-identification of a data subject must be impossible**.
  - **Anonymized data is excluded from GDPR regulation** altogether because anonymized data is no longer "personal data."
- **Pseudonymization** replaces personal identifiers with nonidentifying references or keys so that anyone working with the data is unable to identify the data subject without the key.
  - This type of data may enjoy fewer processing restrictions under GDPR.

https://www.jdsupra.com/legalnews/the-edata-guide-to-gdpr-anonymization-95239/

12052:2017

- The Digital Imaging and Communications in Medicine (DICOM®) Standard (**ISO 12052:2017)** is the international standard for the exchange of digital medical images and related information.
- DICOM® includes a standard for medical image anonymization: **PS3.15 2016a - Security and System Management Profiles**
- This standard defines profiles that detail what data elements need to be modified and in what manner to achieve specified levels of anonymization and pseudonymization.
- Such profiles are not generally available for other image data formats

What Does it Take to De-Identify Image Data?

- Legal agreements must be in place between the covered entity / data controller and the data processor (data publisher)
- Tools for identifying, removing or remapping PHI and PII
  - Most de-identification tools in the US focus on compliance with the HIPAA Safe Harbor method rather than the Expert Determination method
- Secure data transport protocols
- Procedures for ensuring nothing is missed