Ensuring Privacy and Data Security in Medical Research

Jon Crowcroft
University of Cambridge &
Turing Institute,
UK

Ethics & Law

- Do no harm
- GDPR (Europe) & HIPAA (USA)

Trust & Eco-System

- Approaches to security&privacy for health data (for research) depend
    - on how healthcare is provided - state v. private v. mix
- on who holds and maintains data - central, distributed, decentralised
- Impacts where, when, how processing can be done
    - and what security&privacy technologies can be employed
- Assume, at a minimum, data is encrypted at rest and in transit
    - and that there are strict access control and audit logs

Privacy Enhancing Technologies - PETs

- Secure enclaves (Trusted Execution Environments)
    - training classifiers, acquiring models, causal inferencing
    - care as trained classier can still leak training data (set membership)
- Differential Privacy
    - pseudanonymous & fuzzed data
    - quantify privacy/precision tradeoff in access to data
    - see also synthetic data
- Federated Machine Learning
- Secure Multiparty Computations (Zero Knowledge systems)
- Homomorphic Encryption.

PET readiness

- Secure enclaves (Trusted Execution Environments)
    - not perfect, but help
- Differential Privacy
    - very useful framework
- Federated Machine Learning
    - also very useable technology today
- Secure Multiparty Computations
    - Complex to design & explain - depends on use case
- Homomorphic Encryption.
    - quite expensive computational but becoming more usable

PET tech additional complexity

- key management
- role management
- verification
- assurance/certification

Conclusions

- Royal Society Report on PETS

https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/

- UK Opensafely project

http://www.thedatalab.org/blog/189/opensafely-the-origin-story/