STORAGE, ACCESS, LEGAL REQUIREMENTS, OH MY!

GATHERING, TRANSMITTING, AND ARCHIVING PASSIVE DATA RESPONSIBLY AND LEGALLY

Personal Background

- Passive Data: Data gathered without involvement of subject. Web histories, locations, sensors, etc.
- Majority of experience collecting passive data in American mental health research contexts for correlations and predictions.
- Systems created: Pennyworth, Mobilyze, Purple Robot, IntelliCare, Passive Data Kit.
- Current work includes projects outside mental health & research contexts.

Relevant Regulations

- HIPAA (USA): Regulates Personal Health Information (PHI) collection, management, and disclosure.
  - Practitioners are responsible for secure storage, transmission, and disclosure.
  - Breach notification requirements, monetary penalties for failure.
  - Patients' rights to view & modify PHI.

Relevant Regulations

- GDPR (EU+): Data privacy regulation protecting Personally Identifiable Information (PII).
  - Practitioners are responsible for secure storage, transmission, and disclosure.
  - More general than HIPAA.
  - Explicit subject notification, justification, and consent requirements.

“The Nightmare Letter: A Subject Access Request under GDPR”

- Short URL: https://bit.ly/2pvqFaO
- Written from the perspective of a subject exercising their GDPR rights to the fullest extent possible.
- (Partially) applicable in HIPAA contexts.
- Great checklist for validating system data privacy implementations.

“Nightmare Letter” Greatest Hits

- 1a: Enumerate types of data gathered about me.
- 1b: Where (physically) is my data stored?
- 1c: Give me a copy of my data.
- 2: List specific uses of my data.
- 3. List third parties who have access to my data.
- 3c. How specifically are you protecting my data?
- 4. How long will you store my data?

Passive Data Lifecycle

1. Identify.
2. Gather.
3. Transmit.
4. Store.
5. Analyze.
6. Take action.
7. Discard.

Lifecycle may be continuous.

(Or not!)

“Identify” Challenges

- Discovering and selecting subjects.
- Gathering informed consent.
- Communicating rights and responsibilities.

“Gather” Challenges

- Obtaining additional consent to data source.
  - System permissions.
  - Third-party interactions (wearables, online communities, etc.).
- Limiting collection to relevant data.
- Preventing leakage while gathering.
  - Don’t leak data to system logs!
- Storing data securely before transmission.
  - Data protection against hostile third parties in possession of device.

“Transmit” Challenges

- Third-party networks and channels.
- Passive eavesdropping.
- Active “men in the middle”.
- Knowing the strengths and limits of various channels.
  - No PHI via SMS!
- Technical limitations of selected platforms.
  - “Can client-side JavaScript encryption be secure?”

“Store” Challenges

- Detecting and protecting against hostile third parties.
- Balancing security with usability and performance.
- Procuring sufficient resources.
- Managing and tracking legitimate third party access.
- Determining appropriate level of third party access.
- Processes and methods for dealing with intrusions.
- Don’t forget subjects’ rights to their own data!

“Analyze” Challenges

- Appropriate third party access.
- To anonymize or not?
- Verifying whether anonymization is sufficient.
- Controlling access to products of analysis.

“Take Action” Challenges

- Ensuring that actions taken are consistent with spirit of original data collection.
- Are actions taken ethical?
- Are subjects aware of actions taken?
- Can subjects control actions taken with their data?

“Discard” Challenges

- Standard technical issues with secure disposal of electronic information.
- Identifying and discarding any derivative transformations.
- Validating that third parties have discarded data.
- Pervasive encryption as an answer?
  - “Lock the door and throw away the key.”

Shameless Self-Promotion

Audacious Software is currently building a “Nightmare Letter”-compliant framework for secure passive data collection: Passive Data Kit.

- Strong encryption for storage on mobile devices.
- Secure transmission via public-key cryptography.
- Fine-grained event logging on the server for creating access audit trails.
- Free & open source (Apache License).

Goal: GDPR & HIPAA compliance by default.