ETH zürich

Roche

patient data security

Cyber threats in healthcare - are we at risk?

Virtual Tower Talks
27 + 28 May 2020
12.15 - 13.00

Thank you again for your participation and engagement in the recent Roche Tower Talks virtual mini-series focused on patient data and its value, as well as cyber systems in healthcare.

During the event, we received a number of questions to which our panelists did not have time to respond. Please find a small selection of responses here, for your information:

- **Are the Clinical Data Interchange Standards Consortium (CDISC) standards used to enable the interoperability?**
  CDISC is one possible data model that can be used in the framework of the Swiss Personalized health Network (SPHN). Concerning standards, of course, CDISC supports and requires certain ontologies, such as Logistical Observation Identifiers Names and Codes (LOINC). This is also how the Interoperability Strategy of SPHN is conceptualized.

- **Who is the owner of the data? Can it be bought and sold?**
  Swiss legislation does not allow the commercial sale of human data.

- **Does SIB (Swiss Institute of Bioinformatics) also link the data on the same patient across different databases? If yes, what is the identifier? The AHV-number (AHV is a Swiss state-run insurance plan)?**
  It is not allowed to use the AHV-number for research purposes, unless there is a legal basis for it (as e.g., in the Cancer Registration Law). Therefore, we can currently not link the data on the same patient across systems.

- **Have you thought about the meta-data standards and who should enforce this?**
  Yes, we have several Swiss Personalized health Network (SPHN) working groups dealing with the harmonization of meta-data and the deployment of meta-data standards.

- **Do you believe that Switzerland, with its political system and population size is better suited to such data collection endeavors than other countries are?**
  No, as Katrin Crameri stated at the end of the discussion that Switzerland with a bit more than 8 million inhabitants is clearly not suited for a large genomic project like Genomics England or All of Us in the U.S. However, we could occupy a niche by going deep instead of going broad: a multi-omics (genomics, transcriptomics, proteomic, metabolomics, etc.) approach combined with high quality phenotypic and imaging data could be a possibility. The federal system should not stop us here.

- **How is the general knowledge level in Swiss hospitals when it comes to generating reliable, comparable and thereby interoperable data sets?**
  We cannot judge the general knowledge level, but the SPHN contributes to building capacity in this regard in the University Hospitals. We hope to be able to derive smart ways to apply the gained knowledge and experience in smaller hospitals in Phase 2 of SPHN (2021-2024).

- **Wenn die Daten anonymisiert sind, wieso keine big data "bibliothek" errichten, um individuelle research zu ermöglichen ... nicht nur z.B. für hackathons oder Thesis-Arbeiten?**
  (English: If the data is anonymized, why not build a big data "library" to enable individual research ... not just for hackathons or thesis work?)
  The data is not anonymized, but pseudonymized (i.e., this allows hospitals to communicate with the patient) in cases, for example, of reportable findings. Personalized Medicine relies on longitudinal information, anonymization would cut this short, but we are also working on the generation of large anonymized data sets to be shared with the research community.

- **Why is Switzerland lagging behind other countries by such a large extent when we look at the huge number of data silos in the healthcare system? Is this due to our Federal System?**
  The Federal System is probably one reason, but cultural dimensions also play an important role in this regard.

- **What is the main challenge in the handling of data? Structured vs. not structured, amount of data, and relevance of the data, etc.**
  Interoperability is the greatest challenge; and the quality (accuracy; comparability; completeness; consistency, etc.) of care data is often not sufficient for research purposes.

- **Can you give an example of who or which organizations are interested in stealing patient data, and to what ends?**
  Patient data is made up of many data fields, all of which could create value depending on who is ingesting the data and what other data sets are available to them. It is very difficult to pinpoint the motive an entity might have when acquiring some of this data, because this requires application, to a degree, of a perspective of malice and the impact is not often immediate, but rather manifests over time. Motive can range from opportunistic / for fun, direct or indirect criminal financial gain to national security focused outcomes.

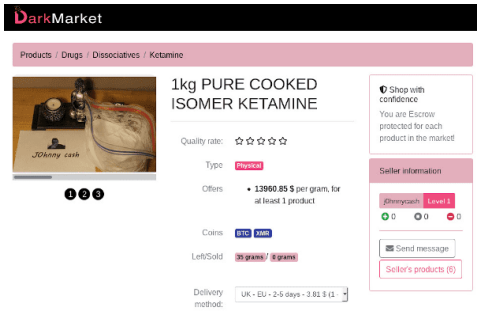
  Nevertheless, the best way to answer this question is through some examples:

  1. A topical example currently is the theft of data for the purposes of extorting an organization. This is sometimes reported as ransomware but effectively the outcome is the same. The data is stolen for financial gain by depriving access to the data or by threating public disclosure. This could have significant impact on patient safety. https://fortune.com/2020/04/01/hackers-ransomware-hospitals-labs-coronavirus/

  2. Patient Data (e.g., Electronic Medical Records (EMR), which include personally identifiable information, protected health information, are valuable to facilitate identity theft and impersonation. The attributes in this type of data, depending on how rich the data is, can be used wholly to create or impersonate an identity (e.g., first name, last name, DOB and other unique attributes). This then fuels the secondary fraud market. https://healthitsecurity.com/news/maze-ransomware-hackers-post-patient-data-stolen-from-2-providers

3. Patient data can be used to enrich data sets obtained from other breaches to build complete pictures of individuals to facilitate other criminal activities. For example, patient data includes attributes that we usually rely on for other services e.g. DOB, email address, telephone number to reset an online bank account. A good example of one service that does breach aggregation is Have I Been Pwned. This is a platform set up by researchers to educate people about breaches.

4. Patient data includes conditions and treatments an individual will be receiving. These data points facilitate the creation of counterfeit prescriptions (e.g., for controlled drugs that are then either resold / reused illegally). Example markets: https://darknetlive.com/markets/ and https://darknetlive.com/markets/darkmarket/



5. Data sets facilitate the execution of activities like spear phishing/ business email compromise as an actor with malicious intent can draft more convincing enticement for a victim. For example, the general population is very aware of the phishing email that that promises money and will more than likely avoid clicking a malicious link. However, an email drafted to refer to a specific medical condition from which a victim may be suffering will be more than likely succeed due to the relatability of the content.

   It is also worth noting that the attribution behind some of the breaches related to patient data is sometimes linked to nation state actors. This adds another dimension to understanding. Why? Because the motivation may scale beyond the financial, gain and extends to matters of national security.

- **Data security is most likely the key issue as to why users/patients are reluctant to share data. You mentioned data leaks in hospitals etc., but even at a much bigger, professional security level with much higher budgets, leaks have happened. So, why should patients trust institutions that handle data?**
  As people continue to become increasingly technology, data security and privacy aware, coupled with wider media coverage of data breaches (and their impact for individuals), it is important that a significant level of digital trust needs to be established across the ecosystem that could consume this data to drive improved patient outcomes.

  In addition, the benefits of the data driven approaches need to be crystallized to data subjects/patients along with articulation of who could benefit from their data. In addition, data subjects/patients having control of their data and transparency into the data security practices their data will be benefit from across the data ecosystem should create the trust needed to allow for a coherent risk-benefit analysis by data subjects/patients that could encourage them to share data.

  Suffice to say irrespective of budgets and advanced security levels, organizations have evolved to deal with security as strategic risk and as a resilience topic. This is a nod to the fact that the threat landscape is evolving so quickly that it is very difficult to be 100% secure and operating under the assumption that it's a case of "if" not "when" a breach will happen creates an approach that allows for timely response and recovery to/from a

breach. A segment of this approach is ensuring timely communication with data subjects/patients is carried out in the event of a breach and should aid more weight to digital trust.

Additionally, building digital trust with patients also requires acknowledgement that data security is a shared responsibility with each party in the data ingestion ecosystem having a role to play. The successful execution of each party's role and responsibilities requires transparency and collaboration to ensure that each party is enabled to perform their role.

- **Are people nowadays more hesitant to give their data?**
  The rate of patients that agree to the use of their data for research purposes is about 80-85%.

- **Will only Swiss Hospitals contribute to the data pool or will other hospitals contribute as well? Where will the database be kept and who decides who gets access to this data? Will the data collection go on indefinitely?**
  There will not be a central database (data in hospitals is generally stored for 10 years). Data is only compiled from its original sources in the context of concrete research projects.