**Technical Challenges**

Despite tremendous advancements made in the digitization of health data, the transfer of EHRs between systems remains a challenge to this day. The history of EHR adoption incentives helps understand why. The Medicare Electronic Health Record Incentive Program, which ran from 2011 to 2016 and paid more than $35 billion in subsidies under the Centers for Medicare and Medicaid Services (CMS), set the criteria for physicians and hospital systems to be eligible for EHR adoption bonuses. These criteria include electronic prescription management, active medication and diagnoses lists, vital sign records, and clinical summaries, as well as patient electronic access. Notably, the criteria do not mandate abiding by a particular data or transaction standard and contain no requirement for interoperability with external systems [1], [2]. With the patient security standards set in the HITECH and HIPAA laws, healthcare data companies were weary of penalties for mismanagement of personal health data and built up proprietary, private silos for storing and retrieving health data.

From the LOINC and HL7 standards came consistent low-level labeling and transfer of medical data; the missing piece is how to group data together to form a database of patient health records or visitation records. Each healthcare data system thus solved the issue of aggregating health data independently. Some systems group clinical and laboratory observations together into studies that allow rapid communication of relevant data for observation and analyses of a single diagnosis. Others index everything to the patient the data is relevant to, creating a personal, patient-centric data that allows clinicians easy access to all a patient’s health data. Typically, EHRs are transactional by nature and linked to a single visitation; healthcare providers use registries to link EHR data to a patient’s registry [3]. Under the HITECH Act’s Meaningful Use mandate, the use of EHRs should operate within the nation’s healthcare system in a meaningful manner by allowing patients more direct management of their health data; this is done through the PHR. While EHRs are associated with one or several visits, PHRs are designed to be lifelong records of a patient’s medical data [4], [5].

The focus shift towards the patient has brought to light issues in aggregating EHRs to form a cohesive and accurate patient health record. Perhaps the most essential data component to link EHRs to the patient’s PHR is the patient identifier; these data include the patient’s full name, date of birth, home address, and other contact and personal information. The record will also include a unique patient ID record number, and there are commonly three different patient IDs within a hospital system: 1) an internal ID for patient operations, 2) a network-wide ID to distinguish patients at the healthcare system’s numerous facilities, and 3) a regional or statewide ID if the health system is connected to these health information exchanges (HIEs). These patient IDs can then be used to merge EHRs into a registry or PHR [4]. A challenge with this is the registry or PHR provider must locate the patient’s master ID through the HIE and reach out to multiple providers to locate the patient’s EHRs to aggregate. In many cases, this is not attainable and PHR vendors typically use alternative methods to match patient identifiers, leading to increased mismatches and incomplete or inaccurate data. Match rates, or the rate that a patients EHR is correctly matched to an additional EHR or registry, fall from more than 90% internally to 50-60% outside of a healthcare data system [6].

The transfer of medical images presents its own unique challenges, and overcoming these barriers is essential to creating interoperable health data systems. The average healthcare provider manages more than 600 terabytes of patient information, with up to 80% of that data by storage size being unstructured medical images [7]. Moreover, medical images represent up to 90% of medical data growth, forecast to reach 2.3 zettabytes in 2022 [8]. Images are commonly stored as either the raw or proprietary file formats output from the medical imaging device, or as rendered images of a standardized format (typically DICOM). Currently, challenges exist for healthcare systems that maintain both raw and rendered DICOM images in linking files together from the same study [9]. Additionally, it can be difficult to maintain the data viability of proprietary data formats, which can be necessary to elucidate insights for a radiologist examining the image [10]. With medical data increasing exponentially in size, health data systems are moving medical data to offsite cloud computing storage, the images must either be encrypted or anonymized to comply with HIPAA. This presents a challenge in storing raw images, as they must be de-encrypted on-site to render the image, removing much of the benefit of cloud computing power. Anonymized images are also a possibility, particularly for sharing images or data with a broader audience (for example, in creating a public medical imaging dataset for machine learning). However, linking these images back to their patient presents a challenge for clinical use, and anonymization techniques have proven at times ineffective at truly removing personal identifiable information [9].

Despite advancements in the storage and standardization of medical images, transferring these objects outside of a health data system remains a challenge. The image sharing network proposed by the RSNA has attracted few participants, and as such the physical transfer of medical images via CDs or DVDs remains viable to this day [11]. The challenges are largely associated with ensuring accurate patient identification between systems and correctly linking raw and DICOM-formatted medical images together with their patient or study in both the sending and receiving platforms.

**Patient Privacy Considerations**

Notably, the technical challenges associated with medical data sharing between systems are not infeasible to solve. The issue in promoting interoperability comes from the growth of these data systems themselves and the financial incentive to upkeep the status quo. Governments, including the US government, poured tens of billions of dollars to incentivize EHR adoption with minimal interoperability requirements. Now, with large EHR data silos already in place, there are no requirements, mandates, or subsidies to convert these platforms to interoperable standards, including HL7’s FHIR standard. Additionally, consider that these health data systems only make money on the patients and hospitals that continue to use their platform; they are incentivized to minimize interoperability and keep customers on their service. Hospital systems are satisfied with this; so long as the data service is adequate at providing internal health data management, it can ensure that existing patients remain with the hospital system. This lack of interoperability harms the patients, as they do not have easy access to the best care available if that care is outside of their existing care provider’s system. Thus, the parties with the onus to promote interoperability in their health data systems are simultaneously incentivized to maintain the “walled garden” standard we have grown accustomed to, while the patients with minimal say in the matter are most negatively impacted.

Patient considerations don’t stop there, though. When a patient is admitted to a hospital or care provider for the first time, they are asked to sign a release form for their personal health information. HIPAA mandates the patient be made aware of their privacy rights: what information will be shared, for what purpose, and with whom; once signed, the healthcare provider does not need to ask for consent or authorization again [12]. The healthcare provider may also ask patients to allow electronic health information exchange by granting access permission to EHR vendors and service providers [13]. Additionally, hospitals do not need patient consent to transfer or even sell de-identified personal health information, resulting in the founding of Truveta, a company selling near real-time clinical data on over 50 million patients in the US [14].

With health data moving to primarily digital storage and retrieval, giving patients granular access controls of their data is more possible than ever. With patient’s personal health data being sensitive and innately personal, this must be considered to maintain patient’s trust and consideration in the healthcare system. Patients with rapid electronic control over their data may be more likely to share clinically applicable data with relevant research studies without requiring the data to be de-identified, giving the researchers more variables to consider. However, there are key cybersecurity concerns with giving each patient access control over their own data. Data breaches have stolen more than 100 million patient’s records through existing secure storage mechanisms; granting patients access controls could open them to additional phishing scams already prevalent in financial accounts [15]. It is worth noting that the current state of the art, based on these figures, is far from perfect and the personal data considerations of giving patients control of their data may outweigh the feasible increase in phishing attacks. This warrants at least investigating the benefits and drawbacks of patient control. Such a solution must nonetheless try to minimize the potential for these phishing scams.

* Generalized Adversarial Networks (GANs) – *does this fit?*
  + Use clinical data to create artificial data sets
  + Ultimately still sourced from patient data
  + Patients not compensated or recognized for providing essential, private, sensitive data