Introduction

The previous two decades of the information age saw the transition from paper to digital health records. The Clinton administration foresaw the change and passed the Health Insurance Portability and Accountability Act (HIPAA) in 1996, establishing modern security and privacy guidelines for personal health information [1]. More recently, the Affordable Care Act (ACA), the American Recovery and Reinvestment Act, and the Health Information Technology for Economic and Clinical Health (HITECH) Act passed under the Obama administration in 2009-10 sought to accelerate the transition to electronic health records (EHRs) [2]–[4]. The acts paid out more than $35 billion in subsidies and incentives from 2011 to 2016 to increase EHR adoption across the country [5]. The hope was the digitization of health data would reduce the cost of healthcare and modernize the US' health data systems.

However, the acts were shortsighted in prioritizing EHR adoption and security over a unified, standardized, interoperable system [6]. Hospital systems, keen to cash in on incentives and weary of legal consequences for security flaws, built up proprietary silos to handle EHR data storage and retrieval [7]. The lack of interoperability guidelines meant these data silos largely adopted their own methods for segmenting and aggregating health data. While organizations such as Health Level Seven International curated medical data transaction standards, the complexity of medical data and particularly medical imaging data meant that the storage solutions adopted by hospital systems did not necessarily interoperate with other proprietary systems when transferring data [8]. As a result, in 2014 while more than 80% of office-based physicians had adopted EHRs, 15% of patients had to personally bring a test result to their physician and 5% had to repeat a test due to the unavailability of prior results [9], [10].

To mitigate these issues, new proposals have arisen to prioritize health data interoperability. The non-profit Health Level Seven International group proposed the Fast Healthcare Interoperability Resources (FHIR) standard to exchange resources based on an application programming interface (API) data format standard [11], [12]. The use of API's is particularly important as hospital data has grown tremendously and become outsourced from local servers to cloud data centers [13]. Recent advancements in blockchain technology could enable a peer-to-peer network of medical data transfer that incentivizes interoperability and data security. Here, we will explore the difficulties associated with medical data transfer, with a particular focus on medical images, opportunities for FHIR API standards and blockchain networks to improve interoperability, and challenges in implementing these technologies while complying with government-mandated privacy standards.

Background

Converting medical data to its digital format is no trivial task. Patient health data represents some of the most complex and sensitive information, and transitioning decades of paper-based records to digital formats is a monumental task. The 1996 HIPAA law mandated that patient medical information is only accessible to the patient and authorized representatives as denoted by the patient themselves [1]. As such, governments around the world sought to incentivize this transition by providing billions of dollars of subsidies with the promise of faster, cheaper, and better health care in the digital age. Ten years after 2009, when the HITECH and American

Recovery and Reinvestment Acts were passed with EHR adoption incentives, EHR adoption doubled [4], [9]. As of writing this paper, 88% of office-based physicians have adopted EHRs. What the governments didn't provide was a standardized system for indexing and segmenting the medical data; this was left to private industry. Eventually, the US government created the United States Core Data for Interoperability (USCDI) in 2016 via the 21st Century Cures Act that mandated the standards medical data and APIs must abide by. The USCDI falls under the US Office of the National Coordinator for Health Information Technology (ONC) and largely draws from commercially created standards and documentation [14]–[16].

The Logical Observation Identifies Names and Codes (LOINC) universal medical laboratory data standard was created in 1994 by the non-profit Regenstrief Institute to aid in increasing demand for electronic health databases [17]. It is publicly available with no cost and is cited by the US government in the USCDI documentation. It details the data types and medical terminology to be used in electronic healthcare data. The terminology is split between laboratory and clinical data. For medical images, which comprise a majority of health data by storage size and growth in the 21st century, the Digital Imaging and Communications in Medicine (DICOM) format is most widely used [18]. It details a metadata header with a 2D array of pixels that allows the images to be viewed and important information such as the imaging device and any necessary variables to appear in the header. Notably, DICOM does not store the raw data and the image cannot be re-processed once in DICOM format. For some referrals or consultations where re-rendering the raw image file is necessary, the exchange of DICOM images alone can be a severe downside. Additionally, to comply with HIPAA privacy requirements, the images must either be encrypted or anonymized before sending to cloud server storage, an increasing occurrence as medical data storage and particularly medical images increase in size [13], [19].

The most notable and widely used standard for medical data transfers in clinical settings is Health Level Seven International's version 2 system. The standard defines a messaging system for exchanging information in a clinical workflow, ranging from patient administration to pharmacy & billing systems. It uses a non-XML syntax based around line segments and character delimiters. The HL7 version 2 standards are desired to be interoperable among all record keeping systems of a hospital workflow and is implemented in every major hospital system in the US [20], [21, p. 7]. Another medical data transfer initiative is Integrating the Health Enterprise (IHE), a non-profit organization created in 1998 to sponsor projects aiming to improve health information sharing. The organization has aided in the US Department of Veterans Affairs health system development and in developing a cross-enterprise document sharing (XDS) model using the LOINC and HL7 standards. They also developed a standard for retrieving medical documents across domains, called the ITI-43 transaction standard [10], [22].

The interoperability of EHRs between vendors and healthcare systems has become a major focus of the ONC in the US. However, as recently as 2019 the transfer of a patient's EHR from one healthcare vendor to another is commonly done via fax or print copies sent via mail. Even for transferring a subset of results for a consultation or referral, the electronic exchange of these documents is only possible roughly 50% of the time, depending on if the different healthcare systems use a vendor with electronic health exchange capabilities between the two systems [8]. To overcome these challenges, the Radiological Society of North America (RSNA) developed the image share network, a clearinghouse-based system where participating healthcare systems

send their medical images for sharing to the clearinghouse operator, who stores the images indexed by a cryptographic hash for 30 days. Personal Health Record (PHR) vendors are able to download the patient's information after they authorize it by divulging the token needed to reproduce the hash. While this system eradicates the physical exchange of medical images, it introduces two new, centralized organizations with access to sensitive personal health information. Additionally, early results show that patients are not likely to view or authorize their medical images within the 30-day window, and the image share network is dominated by a small number of radiological centers, and a small number of PHR vendors (who can also act as a clearinghouse operator) control authorized retrieval of medical images [10], [23].

To help mitigate these issues, the Health Level Seven International group proposed the FHIR standard, first drafted in 2011 and officially published in 2017. The standard revolves around APIs to incentive the transfer of data between and within systems; it is built around resources, such as clinical observations, that can be aggregated into FHIR profiles. However, despite endorsements of the standard by the US government, FHIR is still dwarfed in adoption by HL7's version 2 [11].

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 [6], [24]. 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. The complicated nature of health data and the choices present for EHR vendors to make means clinicians handling EHRs often deal with data fragmentation and missing data [25], [26]. 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 [27]. 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 [16], [28].

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 [16]. 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 [29].

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 [30]. Moreover, medical images represent up to 90% of medical data growth, forecast to reach 2.3 zettabytes in 2022 [31]. 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 [19]. 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 [18]. 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 [19].

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 [10]. The technical 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 (as of July 2021, the CMS does require availability of patient records in the FHIR standard, but only for Medicare and Medicaid patients, compromising around 135 million individuals or roughly 40% of the American population [32]–[34]). 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 [35]. The healthcare provider may also ask patients to allow electronic health information exchange by granting access permission to EHR vendors and service providers [36]. 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 aggregated from 14 healthcare systems [37].

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 [38]. 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

Blockchain-based Solutions

- Blockchain technology applications for EHR data transfer
    - Native cryptographic security of on-chain data
    - Network-wide consensus on data transfers
        - Reduces likelihood of massive data breaches by allowing participants to flag suspicious-looking transfers and block them
    - Technologies enable temporary access control and revocation
        - Proxy re-encryption
    - Peer-to-peer network across health data systems can enable standardized interoperable transfer standards
        - data transfer can be standardized and required in order to participate in network

- Problems blockchain cannot solve
    - Storage or computation of health data
        - Blockchain data is notoriously computationally expensive; infeasible to store health data on chain (only transactions would be stored)
        - Security of health data is thus not guaranteed
    - Does not prevent phishing attacks
        - Leaking private key enables attacker to control access of a patient(s) data

- Existing blockchain initiatives
    - HealthChain
        - Patient access control with PRE
        - Hybrid block schema
    - MIFS
        - Health data systems access control
        - Designed for increased availability of data for ML research
    - MedRec
        - Ethereum smart contract-based data control
        - Forms basis for many EHR blockchain solutions

Discussion

Conclusion

References

[1] P. F. Edemekong, P. Annamaraju, and M. J. Haydel, "Health Insurance Portability and Accountability Act," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2022. Accessed: Nov. 06, 2022. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK500019/

[2] S. F. Fontenot, "The Affordable Care Act and electronic health care records: can technology help reduce the cost of health care?," *Physician Exec.*, vol. 40, no. 1, pp. 68–72, Feb. 2014.

[3] S. Freymann Fontenot, "The Affordable Care Act and electronic health care records. Does today's technology support the vision of a paperless health care system?," *Physician Exec.*, vol. 39, no. 6, pp. 72–74, 76, Dec. 2013.

[4] M. Butler, "Top HITECH-HIPPA compliance obstacles emerge," *J. AHIMA*, vol. 85, no. 4, pp. 20–24; quiz 25, Apr. 2014.

[5] "CMS EHR Incentive Program - November 2016 Report," p. 5, 2012.

[6] A. Wright, J. Feblowitz, L. Samal, A. B. McCoy, and D. F. Sittig, "The Medicare Electronic Health Record Incentive Program: Provider Performance on Core and Menu Measures," *Health Serv. Res.*, vol. 49, no. 1 Pt 2, pp. 325–346, Feb. 2014, doi: 10.1111/1475-6773.12134.

[7] H. Liu, X. Xiao, X. Zhang, K. Li, and S. Peng, "MIFS: A Peer-to-Peer Medical Images Storage and Sharing System Based on Consortium Blockchain," in *Bioinformatics Research and Applications*, Cham, 2021, pp. 336–347. doi: 10.1007/978-3-030-91415-8_29.

[8] T. H. Payne *et al.*, "Status of health information exchange: a comparison of six countries," *J. Glob. Health*, vol. 9, no. 2, p. 020427, doi: 10.7189/jogh.09.020427.

[9] "Office-based Physician Electronic Health Record Adoption | HealthIT.gov." https://www.healthit.gov/data/quickstats/office-based-physician-electronic-health-record-adoption (accessed Nov. 07, 2022).

[10] V. Patel, "A framework for secure and decentralized sharing of medical imaging data via blockchain consensus," *Health Informatics J.*, vol. 25, no. 4, pp. 1398–1411, Dec. 2019, doi: 10.1177/1460458218769699.

[11] M. Ayaz, M. F. Pasha, M. Y. Alzahrani, R. Budiarto, and D. Stiawan, "The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities," *JMIR Med. Inform.*, vol. 9, no. 7, Art. no. 7, Jul. 2021, doi: 10.2196/21929.

[12] E. R. Pfaff *et al.*, "Fast Healthcare Interoperability Resources (FHIR) as a Meta Model to Integrate Common Data Models: Development of a Tool and Quantitative Validation Study," *JMIR Med. Inform.*, vol. 7, no. 4, Art. no. 4, Oct. 2019, doi: 10.2196/15199.

[13] Y. Duan, Y. Li, L. Lu, and Y. Ding, "A faster outsourced medical image retrieval scheme with privacy preservation," *J. Syst. Archit.*, vol. 122, p. 102356, Jan. 2022, doi: 10.1016/j.sysarc.2021.102356.

[14] S. K. Magid, K. Cohen, and L. S. Katzovitz, "21st Century Cures Act, an Information Technology-Led Organizational Initiative," *HSS J.*, vol. 18, no. 1, pp. 42–47, Feb. 2022, doi: 10.1177/15563316211041613.

[15] "United States Core Data for Interoperability (USCDI)." https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi (accessed Nov. 07, 2022).

[16] V. Ehrenstein, H. Kharrazi, H. Lehmann, and C. O. Taylor, *Obtaining Data From Electronic Health Records*. Agency for Healthcare Research and Quality (US), 2019. Accessed: Nov. 14, 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK551878/
[17] C. J. McDonald *et al.*, "LOINC, a Universal Standard for Identifying Laboratory Observations: A 5-Year Update," *Clin. Chem.*, vol. 49, no. 4, pp. 624–633, Apr. 2003, doi: 10.1373/49.4.624.
[18] M. Larobina and L. Murino, "Medical Image File Formats," *J. Digit. Imaging*, vol. 27, no. 2, pp. 200–206, Apr. 2014, doi: 10.1007/s10278-013-9657-9.
[19] S. G. Langer, "Challenges for Data Storage in Medical Imaging Research," *J. Digit. Imaging*, vol. 24, no. 2, p. 203, Apr. 2011, doi: 10.1007/s10278-010-9311-8.
[20] "Standards Organizations for the NHII," *ASPE*. https://aspe.hhs.gov/standards-organizations-nhii (accessed Nov. 07, 2022).
[21] "Introduction to HL7 Standards | HL7 International." http://www.hl7.org/implement/standards/index.cfm?ref=nav (accessed Nov. 07, 2022).
[22] D. A. Clunie, D. K. Dennison, D. Cram, K. R. Persons, M. D. Bronkalla, and H. "Rik" Primo, "Technical Challenges of Enterprise Imaging: HIMSS-SIIM Collaborative White Paper," *J. Digit. Imaging*, vol. 29, no. 5, pp. 583–614, Oct. 2016, doi: 10.1007/s10278-016-9899-4.
[23] S. G. Langer *et al.*, "The RSNA Image Sharing Network," *J. Digit. Imaging*, vol. 28, no. 1, pp. 53–61, Feb. 2015, doi: 10.1007/s10278-014-9714-z.
[24] "Medicare & Medicaid EHR Incentive Program Registration & Attestation System." Centers for Medicare and Medicaid Services, 2013a. [Online]. Available: https://ehrincentives.cms.gov/hitech/login.action
[25] M. Quinn *et al.*, "Electronic Health Records, Communication, and Data Sharing: Cchallenges and Opportunities for improving the diagnostic process," *Diagn. Berl. Ger.*, vol. 6, no. 3, pp. 241–248, Aug. 2019, doi: 10.1515/dx-2018-0036.
[26] B. J. Wells, K. M. Chagin, A. S. Nowacki, and M. W. Kattan, "Strategies for Handling Missing Data in Electronic Health Record Derived Data," *eGEMs*, vol. 1, no. 3, 2013, doi: 10.13063/2327-9214.1035.
[27] C. A. Caligtan and P. C. Dykes, "Electronic Health Records and Personal Health Records," *Semin. Oncol. Nurs.*, vol. 27, no. 3, pp. 218–228, Aug. 2011, doi: 10.1016/j.soncn.2011.04.007.
[28] M. Lester, S. Boateng, J. Studeny, and A. Coustasse, "Personal Health Records: Beneficial or Burdensome for Patients and Healthcare Providers?," *Perspect. Health Inf. Manag.*, vol. 13, no. Spring, p. 1h, Apr. 2016.
[29] G. Morris *et al.*, "Patient Identification and Matching Final Report," p. 93, 2014.
[30] H. Jin, Y. Luo, P. Li, and J. Mathew, "A Review of Secure and Privacy-Preserving Medical Data Sharing," *IEEE Access*, vol. 7, pp. 61656–61669, 2019, doi: 10.1109/ACCESS.2019.2916503.
[31] "Cloud migration for medical imaging data using Azure Health Data Services and IMS." https://azure.microsoft.com/en-us/blog/cloud-migration-for-medical-imaging-data-using-azure-health-data-services-and-ims/ (accessed Nov. 21, 2022).
[32] "CMS Interoperability and Patient Access final rule | Guidance Portal." https://www.hhs.gov/guidance/document/cms-interoperability-and-patient-access-final-rule-0#CMS-Interoperability-and-Patient-Access-Final-Rule (accessed Nov. 28, 2022).

[33] “CMS Releases Latest Enrollment Figures for Medicare, Medicaid, and Children’s Health Insurance Program (CHIP) | CMS.” https://www.cms.gov/newsroom/news-alert/cms-releases-latest-enrollment-figures-medicare-medicaid-and-childrens-health-insurance-program-chip#_ftn1 (accessed Nov. 28, 2022).
[34] “NA_EST2021_POP: Monthly Population ... - Census Bureau Table.” https://data.census.gov/table?tid=PEPNATMONTHLY2021.NA_EST2021_POP&hidePreview=false (accessed Nov. 28, 2022).
[35] R. A. Tariq and P. B. Hackert, “Patient Confidentiality,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2022. Accessed: Nov. 26, 2022. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK519540/
[36] “Health Information Privacy Law and Policy | HealthIT.gov.” https://www.healthit.gov/topic/health-information-privacy-law-and-policy (accessed Nov. 27, 2022).
[37] “Our Approach to Data Quality.” Truveta, Spring 2022. Accessed: Nov. 27, 2022. [Online]. Available: https://www.truveta.com/wp-content/uploads/2022/06/whitepaper-data-quality.pdf
[38] K. P. Seastedt *et al.*, “Global healthcare fairness: We should be sharing more, not less, data,” *PLOS Digit. Health*, vol. 1, no. 10, p. e0000102, Oct. 2022, doi: 10.1371/journal.pdig.0000102.