Development of a patient-centric eHealth data exchange using distributed ledger technology

Nico Heiniger[1], Pascal Moriggl[1], Petra Aspiron[1] and Quy Vo-Reinhard[2]

[1] University of Applied Sciences Northwestern Switzerland, Peter Merian – Strasse 86, 4002 Basel, Switzerland
[2] dHealth Foundation, Dammstrasse 16, 6300 Zug, Switzerland

Abstract

In Medical Informatics, the growing value of data is driven by advancements in Artificial Intelligence and Machine Learning, aiding Drug Discovery and Patient Outcome prediction. Traditional data collection methods, such as costly and time-consuming clinical trials, are supplemented by the substantial growth of real-world health data obtained from wearables with lifestyle tracking. This study explored the potential of leveraging distributed ledger technology to facilitate data exchange between patients and researchers. The introduction section identified research gaps and stakeholder requirements in this field. Supported through literature analysis, objectives for a prototypical novel application were defined. A prototype using the dHealth network and IPFS was developed and tested with HL7-format blood sample data. The results demonstrate the technical feasibility of patient-researcher data exchange via distributed ledger technology without apparent security or performance limitations.

Keywords

Medical Informatics, Data Exchange, Distributed Ledger Technology, Real-world Health Data, Blockchain

1. Introduction

Usually, behavioral data is collected by tracking users of applications, tools, and websites. Even though the user has to opt-in and consent to share their data, this is usually the only option to access the selected service or system. This approach cannot be applied in medical research due to data sensitivity and not at least data privacy/protection laws (e.g., GDPR). Patients have no benefit from sharing their data, and even with the motivation of sharing for the greater good, there is no technical solution to do so [1].

New medical informatics trends such as precision medicine or clinical decision support systems heavily depend on health data's amount, quality, access, and availability. At the same time, the volume of health data generated and collected through various stakeholders and systems is increasing rapidly. Hospital information systems collect digitized data on patients' diseases and treatments, while wearables generate longitudinal data such as the patient's heart rate, sleeping rhythm, or eating habits [2].

1.1. Research gap

The identified research gap shows a lack of secure and fast connection options between data creators, data owners (patients), and potential beneficiaries (researchers), that are peer-to-peer based. Despite the growing importance of real-world data (RWD), and unsolved challenges and risks [3] there is still a shortage of solutions that allow patients to voluntarily share their data without being dependent on

Companion Proceedings of the 16th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling and the 13th Enterprise Design and Engineering Working Conference, November 28 – December 1, 2023, Vienna, Austria

nico.heiniger@protonmail.com (N. Heiniger); pascal.moriggl@fhnw.ch (P. Moriggl); petra.asprion@fhnw.ch (P. Asprion); quy.voreinhard@gmail.com (Q. Vo-Reinhard)
0009-0001-2620-6061 (N. Heiniger); 0000-0003-1774-3591 (P. Moriggl); 0000-0001-6582-5087 (P. Asprion); 0009-0001-0970-3453 (Q. Vo-Reinhard)

a single, potentially commercial platform provider. Therefore, it is crucial to explore possible solutions where patients can share their data directly with specific research organizations without relying on storing their raw data on a platform outside of their control. This research focuses mainly on investigating how patients can consent to share data with researchers using for example the dHealth network, while respecting the needs of the involved stakeholders. The dHealth network was selected as ecosystem for the prototype due to its open-source nature, tailored design for healthcare-related transactions, and its extensive toolbox, which provided a flexible foundation for developing the required features and functionalities.

Stakeholders have different requirements for health data based on their needs and position along the care journey. Besides the general data security attributes such as integrity, confidentiality, and availability there are specific requirements on patient and research side of the sharing process [4]. Patient-oriented sharing solutions must prioritize in addition to traditional data security attributes data privacy and transparency, ensuring rigorous protection of personal health information and adherence to data protection laws [4]. These systems should be in addition highly usable, catering to a diverse user base, and emphasizing ease of use and perceived usefulness to promote widespread adoption [5].

Researchers on the receiving side of data sharing rely on high-quality data for their studies. With the integration of patient data, ensuring data integrity and verification becomes paramount. One approach is distinguishing patient-generated data, acknowledging its potential variability, and considering third-party validation, such as involving general practitioners in the data validation process [6]. Data quality also hinges on maintaining data integrity to prevent unauthorized modifications, which is crucial when accommodating patient-initiated data changes [4]. Moreover, researchers emphasize system availability to access data promptly, ensuring data preservation and appropriate removal. Efficient performance, characterized by fast data access, transmission, and retrieval, further enhances the usability of data sharing for research purposes [4].

1.2. Method

The research followed the Design Science Research (DSR) principles outlined by Peffers et al. [7]. A prototype solution was created to address specific challenges in healthcare research, such as data integrity or other defined security attributes. The prototype is designed to simulate the functionality of allowing patients to choose the designated research purpose for their data. This aspect showcases the prototype's potential to support different research objectives and enhance patient engagement in the research process. The DSR approach emphasizes iterative design and evaluation cycles. By following DSR principles, this research aims to advance healthcare research by providing an applicable solution that addresses the requirements and complexities of data management, achieving specific security attributes for data, and research purposes in the healthcare domain. In summary, a technical proof-of-concept prototype is programmed using blood sample data. The prototype is validated based on a blend of authors' objectives and derived requirements from existing frameworks.

2. Prototype design

Two key concerns arise from the perspective of data privacy. Firstly, patients should be able to determine which individuals or entities can access their data. Additionally, they need to have confidence in the system's ability to uphold these preferences. Leveraging blockchain technology as a facilitator for these transactions ensures a very high level of transparency. Given that all transactions are visible to the public, patients can always ascertain who their data has been shared with and verify that their privacy settings have been respected.

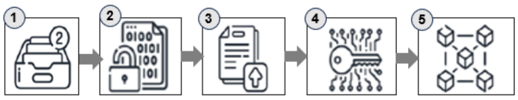

Figure 1: Patient application process

Two generic processes for the prototype were outlined – the patient side and the research side. The patient-side process illustrated in Figure 1 consists of the following steps:

1. Patient aims to send a file to another party.
2. The file needs to be secured (encrypted).
3. The file needs to be stored somewhere.
4. The keys for file decryption must be sent with.
5. All is sent to the recipient.

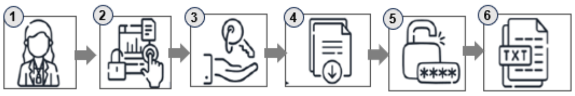

Figure 2: Research application process

The research side process is mirroring the recipient steps (Figure 2):

1. A researcher receives a message.
2. The researcher proves the right recipient.
3. The researcher retrieves the location and keys needed to access the file.
4. The file is downloaded from somewhere.
5. The file gets decrypted.
6. The file is stored locally in plain text.

The prototype does not store information in pdf format but adheres to the HL7 standard, an international standard used to store machine-readable health data. Consequently, the data processed in Figure 1 and Figure 2 becomes machine-readable and easily transferable between systems or users. This characteristic enables the solution's scalability, accommodating additional stakeholders, such as insurance companies, entering the process. These stakeholders would require an interface compatible with one widely adopted standard.

2.1. Requirements

From a patient or data sender perspective, the main requirement is to ensure data privacy. Since the patient initiates the data sharing, complying with the needs and requirements of patients is most important to ensure sufficient participation. On the other end of the sharing process are researchers, who need the data shared by patients. However, the data will be useless if the shared data sets do not underline standardization and regulation.

Therefore, data quality is the second main requirement to ensure the usability of the shared data. If the first two requirements are fulfilled, but the solution's usability is not guaranteed due to a lack of performance, neither side will use it. Hence, the third requirement focuses on performance. Based on the literature, these requirements can be translated into the objectives in Table 1.

Table 1
Prototype objectives

Objectives	Literature
(1) The patient can see where their data has been sent.	[8]–[10]
(2) The patient can control whom they share their data with.	[9], [11], [12]
(3) The data needs to be in a machine-readable format.	[13], [14]
(3.1) Meta-data is available: • When and where data was created • Clinical setting	Boeske et al., 2004; Kalra et al., 2017; Loane & Wootton, 2002; Van Doornik, 2013
(4) Performance:	
Access to data should be fast. The number of transactions per second (TPS) should be above 60.	Ash & Bates, 2005; Humphreys, 2000; Moriggl et al., 2021
The system should respond to any user input with acceptable performance. The average loading time should be below 4 seconds.	Engelbrecht et al., 2005; Galletta et al., 2004; Hier et al., 2005

2.2. Architecture

This study builds a proof-of-concept emphasizing security, efficiency, and decentralization. Blockchain technology forms the foundation of this architecture, renowned for its secure and decentralized data storage capabilities. Utilizing cryptography to link data blocks, blockchain technology is exceptionally suitable for handling sensitive health data, providing a tamper-proof environment and ensuring data integrity [23]–[26]. The blockchain utilized is the Decentralized Health (dHealth) Network. This blockchain network, designed explicitly for healthcare data, employs a Proof-of-Stake Plus (PoS+) consensus mechanism. It facilitates secure, efficient health data transactions and incentivizes participation through its digital currency, DHP, making it a valuable tool for healthcare data management [27].

The InterPlanetary File System (IPFS) plays a crucial role in the stack. IPFS revolutionizes data storage and retrieval by shifting from traditional location-based addressing to content-based addressing. This method enhances data integrity and efficiency, employing unique hashes for data, thus avoiding redundancy and ensuring easy access to information [28].

Health Level Seven (HL7) is incorporated for structured data exchange in healthcare. HL7 standardizes the messaging format, streamlining the communication of patient details, test results, and other crucial health data, facilitating effective and standardized data exchange within the healthcare domain [29], [30].

The development environment includes TypeScript for the backend and React.js for the frontend. The prototype leverages the existing dHealth SDK[2] for account management, transaction creation, and signing, and Node.js serves as the run-time environment. To ensure data privacy, the prototype encrypts HL7 files using an AES-256 encryption algorithm and uploads them to the decentralized IPFS network. This approach ensures data redundancy, availability, and resilience against single points of failure.

[2] https://github.com/dhealthproject/

Figure 3 shows the sender's perspective. The prototype prompts the end-user to select a research organization as the data recipient (1). The prototype allows a selection between academic and commercial research. The sender authenticates themselves with the private key (2). The file gets encrypted with the chosen encryption key (3). The encrypted file is then uploaded to IPFS (4). IPFS returns a Content Identifier (CID), which acts as an address to the file. The CID is added to the transaction as an encrypted message along with the encryption key.

The sender's private key is used for authentication and signing the transaction before it is broadcasted to the dHealth network using the SDK. As soon as the transaction is confirmed, it is publicly visible in the dHealth explorer[3] and can be found through the transaction hash, the sender's address, or the recipient's address.

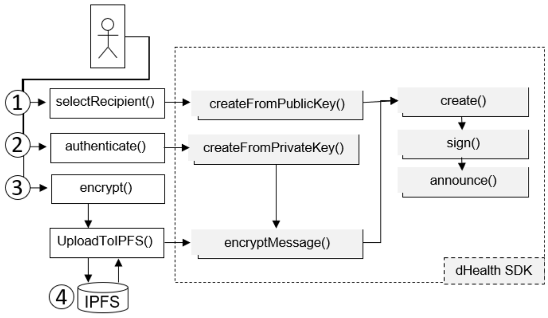

Figure 3: Sender application process

On the recipient's side (Figure 4), the prototype enables data retrieval by entering the transaction hash and providing the recipient's private key (1). Therefore, the recipient needs to notice an incoming transaction by monitoring their account in the explorer or enabling notifications in the dHealth wallet, where the transaction hash can be identified. The sender's public and recipient's private keys are used to decrypt the attached message, containing the IPFS Content Identifier (CID) and file encryption key. The retrieval and decryption of the encrypted message are done using the SDK (3). Due to the message encryption being part of the dHealth functionality, only the recipient's account can decrypt it, allowing access to the file location and its decryption key. The decrypted HL7 file is stored locally on the recipient's machine.

Overall, the architecture leverages the dHealth network, IPFS, the React.js frontend, and the provided SDK functions to ensure secure data transmission, privacy, and interoperability in the healthcare domain.

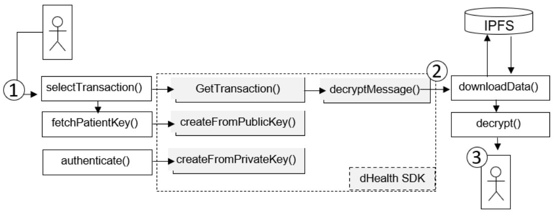

Figure 4: Recipient application process

3. Evaluation

The objectives in Table 1 are derived from stakeholder requirements, encompassing data privacy and patient transparency, data quality for research, and performance measures. Data privacy and quality

[3] http://explorer.dhealth.cloud/

objectives are assessed separately from performance objectives because privacy and quality are assessed qualitatively while performance is assessed quantitatively.

3.1. Data privacy and quality requirements

The evaluation of the first four objectives of Table 1 has been conducted systematically in a binary manner, employing a checkbox-like assessment approach. The objectives are considered achieved when they apply to the prototype, thus highlighting the successful fulfillment of the corresponding criteria. This methodical evaluation ensures a comprehensive appraisal of the prototype's adherence to the established objectives.

1. **The patient can see where their data has been sent**: The transparency provided by the applied blockchain technology allows patients to trace and verify the journey of their data. They can view the transaction history and track which research organizations or recipients have accessed their data.
2. **The patient can control whom they share their data with**: The prototype includes a user interface that enables patients to select the desired research organization as the recipient of their data. This control empowers patients to decide whom to share their health data with, ensuring privacy and consent.
3. **The data is in a machine-readable format**: HL7 format ensures that the health data is structured and machine-readable, meeting the requirement of data accessibility and interoperability.

3.1. **Meta-data is available**: The HL7 format includes meta-data such as information on when and where the data was created and the clinical setting, providing additional context and details about the health data.

4. **Performance**: The validation of the performance was very complex within the scope of the study and is therefore described in the following section.

3.2. Performance requirements

The prototype underwent rigorous examination regarding its performance within a controlled local environment, executed through command-line interfaces. A dedicated script systematically evaluated the prototype's performance across various file sizes. This method allowed for meticulous measurement of processing times, enabling a comprehensive analysis of system behavior and responsiveness under varying workloads.

The performance evaluation involved measuring the loading time and execution time of the prototype. Loading time was measured as the duration from the initial launch of the application to the point where it became responsive to user input. This metric determined how quickly users could access and interact with the system. Execution time refers to the interval from the last user input to the completion of all processing steps within the application. This metric helped assess the efficiency of the prototype in handling user requests. A range of test scenarios were executed to ensure a comprehensive evaluation. One hundred test runs were conducted, each with different file sizes. These scenarios included:

- A regular blood test HL7 file with a size of 3KB.
- An HL7 file of larger dimensions, measuring 1024KB (1MB).
- An even larger HL7 file of 30720KB (30MB).

This allowed a thorough assessment of the prototype's performance across a spectrum of data sizes, providing insights into its scalability and efficiency under varying conditions. The main components were programmed using TypeScript and Node.js in Visual Studio Code. The file format was HL7, using AES-256 encryption algorithm. The dHealth network served as blockchain for the data exchange, while the IPFS network provided distributed data storage. For the regular-sized HL7 file, the average loading time for the patient application over 100 runs was 2.21 seconds, while the research application had an average loading time of only one millisecond. The execution time for both applications averaged around 0.25-0.27 seconds. As shown in Table 2, the larger file sizes mainly impacted the execution time for all stakeholders.

Table 2
Prototype performance requirements

	Patient		Researcher	
	Load time	Execution time	Load time	Execution time
HL7 file 3KB	2.207s	0.250s	0.001s	0.271s
HL7 file 1024KB	2.138s	1.046	0.001s	2.448s
HL7 file 30720KB	5.999s	1.483s	0.001s	7.024s

The maximum transaction fee of 0.05 DHP was identified as a good balance between approval time and cost. The dHealth network has a maximum transaction throughput of 200 transactions per second, with each transaction having a fixed size of 82 bytes. The application prototype itself did not impose any limitations regarding potential bottlenecks as it ran locally and independently. The main potential bottleneck could be the IPFS network, but no specific upload limits regarding speed or size were found. The limitations would depend on the hardware specifications of the IPFS node. The prototype currently utilizes an API (Application Programming Interface) limited to 60 transactions per second.

3.3. Overall results

The prototype successfully addresses the objectives of patient visibility in data usage and control over transaction recipients, having machine-readable data and available meta-data. In addition, the evaluation demonstrates that the prototype, in terms of performance, effectively achieves the requirement of fast data access with a transaction throughput above 60 TPS and acceptable performance with an average loading time below 4 seconds.

4. Limitations

One of the prototype's main limitations is the data deletion challenge. As blockchain and IPFS are immutable technologies, once data is stored on these platforms, it becomes difficult to delete or modify it. This limitation arises from the inherent design principles of these technologies, which prioritize data immutability and integrity. Deleting sensitive patient data in healthcare is crucial to comply with data privacy regulations and individual preferences. However, due to the nature of blockchain and IPFS, achieving complete data deletion is challenging. It requires additional mechanisms and considerations beyond the scope of this prototype. Addressing the limitation of data deletion would necessitate exploring alternative approaches, such as implementing privacy-enhancing techniques like zero-knowledge proofs or utilizing off-chain storage solutions with more flexible data management capabilities. These approaches could provide more comprehensive control over data retention and deletion while maintaining the desired level of security and privacy.

5. Conclusion

This study explores the feasibility of a blockchain-based platform for sharing health data between patients and researchers. The prototype application demonstrates that combining blockchain technology and distributed file storage, specifically the IPFS network, enables secure and efficient data transfer. Patients can select recipients, encrypt files, and upload them to IPFS, sharing the encrypted IPFS Content Identifier (CID) and encryption key with the recipient. Researchers can retrieve and decrypt the files, enhancing data value for research projects. Patient autonomy and compliance with data protection/privacy regulations are emphasized. In order to ensure continuous quality, (external) audits are recommended. Overall, the study extends existing research by including research organizations in the data exchange process, affirming the suitability of distributed ledger and storage systems for healthcare applications.

The technical feasibility of a blockchain-based platform for health data sharing is confirmed through the prototype and its evaluation. Future research should explore blockchain options, encryption methods, and economic considerations while ensuring compliance with evolving privacy regulations.

References

[1] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using blockchain for medical data access and permission management," *Proc. - 2016 2nd Int. Conf. Open Big Data, OBD 2016*, pp. 25–30, 2016, doi: 10.1109/OBD.2016.11.

[2] M. J. Kaur, V. P. Mishra, and P. Maheshwari, *The Convergence of Digital Twin, IoT, and Machine Learning: Transforming Data into Action*. Springer International Publishing, 2020.

[3] F. Grimberg, P. M. Asprion, B. Schneider, E. Miho, L. Babrak, and A. Habbabeh, "The Real-World Data Challenges Radar: A Review on the Challenges and Risks regarding the Use of Real-World Data," *Digit. Biomarkers*, vol. 5, no. 2, pp. 148–157, 2021, doi: 10.1159/000516178.

[4] A. Hoerbst and E. Ammenwerth, "Electronic health records: A systematic review on quality requirements," *Methods Inf. Med.*, vol. 49, no. 4, pp. 320–336, 2010, doi: 10.3414/ME10-01-0038.

[5] E. C. Ling, I. Tussyadiah, A. Tuomi, J. Stienmetz, and A. Ioannou, "Factors influencing users' adoption and use of conversational agents: A systematic review," *Psychol. Mark.*, vol. 38, no. 7, pp. 1031–1051, 2021, doi: 10.1002/mar.21491.

[6] W. Van Doornik, "Meaningful use of patient-generated data in EHRs.," *J. AHIMA*, vol. 84, no. 10, pp. 1–5, 2013.

[7] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.

[8] C. C. Aggarwal, *Data classification : algorithms and applications*. 2015.

[9] M. J. Ball, C. Smith, and R. S. Bakalar, "Personal health records: empowering consumers.," *J. Healthc. Inf. Manag.*, vol. 21, no. 1, pp. 76–86, 2007.

[10] C. Safran *et al.*, "Toward a National Framework for the Secondary Use of Health," *Jounal Am. Med. Informatics Assoc.*, vol. 14, no. 1, pp. 1–9, 2007, doi: 10.1197/jamia.M2273.Introduction.

[11] J. G. Beun, "Electronic healthcare record; a way to empower the patient," *Int. J. Med. Inform.*, vol. 69, no. 2–3, pp. 191–196, 2003, doi: 10.1016/S1386-5056(03)00060-1.

[12] W. Pratt, K. Unruh, A. Civan, and M. Skeels, "Personal health information management," *Commun. ACM*, vol. 49, no. 1, pp. 51–55, 2006, [Online]. Available: https://doi.org/10.1145/1107458.1107490.

[13] R. Dettwiler, N. Fhnw, S. Märke, and V. Roth, "Elektronisches Patientendossier:Aktuelle Probleme, Potenziale undwie die Institutionen damit umgehen.," no. September, 2022.

[14] B. L. Humphreys, "Electronic Health Record Meets Digital Library: A New Environment for Achieving an Old Goal," *J. Am. Med. Informatics Assoc.*, vol. 7, no. 5, pp. 444–452, Sep. 2000, doi: 10.1136/jamia.2000.0070444.

[15] M. Boeske, H. Franz, Goetz, and Haibach, "Managementpaper 'Elektronische Patientenakte.'" Aktionsforum Telematik im Gesundheitswesen, p. 54, 2004, [Online]. Available: https://docplayer.org/4231065-Managementpapier-elektronische-patientenakte.html.

[16] M. Loane and R. Wootton, "A review of guidelines and standards for telemedicine," *J. Telemed.*

Telecare, vol. 8, no. 2, pp. 63–71, 2002, doi: 10.1258/1357633021937479.

[17] D. Kalra *et al.*, “The European Institute for Innovation through Health Data,” *Learn. Heal. Syst.*, vol. 1, no. 1, pp. 1–8, 2017, doi: 10.1002/lrh2.10008.

[18] J. S. Ash and D. W. Bates, “Factors and forces affecting EHR system adoption: Report of a 2004 ACMI discussion,” *J. Am. Med. Informatics Assoc.*, vol. 12, no. 1, pp. 8–12, 2005, doi: 10.1197/jamia.M1684.

[19] P. Moriggl, P. M. Asprion, and B. Schneider, “Blockchain Technologies Towards Data Privacy—Hyperledger Sawtooth as Unit of Analysis,” in *Studies in Systems, Decision and Control*, Cham: Springer, 2021, pp. 299–313.

[20] R. Engelbrecht, A. Geissbuhler, C. Lovis, and G. Mihalas, “Connecting medical Informatics and Bio-Informatics,” *Stud. Heal. Technol. Informatics, Vol. 116*, vol. 116, pp. 1–1027, 2005.

[21] D. B. Hier, A. Rothschild, A. LeMaistre, and J. Keeler, “Differing faculty and housestaff acceptance of an electronic health record,” *Int. J. Med. Inform.*, vol. 74, no. 7–8, pp. 657–662, 2005, doi: 10.1016/j.ijmedinf.2005.03.006.

[22] D. Galletta, R. Henry, S. McCoy, and P. Polak, “Web Site Delays: How Tolerant are Users?,” *J. Assoc. Inf. Syst.*, vol. 5, no. 1, pp. 1–28, Jan. 2004, doi: 10.17705/1jais.00044.

[23] M. Swan, *Blockchain for a New Economy.* Cambridge: O’Reilly, 2015.

[24] S. Haber and W. S. Stornetta, “How to Time-Stamp a Digital Document BT - Advances in Cryptology-CRYPTO’ 90,” pp. 437–455, 1991, [Online]. Available: https://link.springer.com/content/pdf/10.1007/3-540-38424-3_32.pdf.

[25] A. Sunyaev, *Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies.* 2020.

[26] L. Lantz and D. Cawrey, *Mastering Blockchain: Unlocking the Power of Cryptocurrencies, Smart Contracts, and Decentralized Applications*, vol. 1. 2020.

[27] D. Foundation, “dHealth Network,” 2021. [Online]. Available: https://uploads-ssl.webflow.com/62434be6096bbb00e80dbf0d/6253e75695e36ce5aa800dd8_Whitepaper-dHealth-Network.pdf.

[28] J. Benet, “IPFS - Content Addressed, Versioned, P2P File System,” no. Draft 3, 2014, [Online]. Available: http://arxiv.org/abs/1407.3561.

[29] J. Klauber and M. Geraedts, *Krankenhaus-Report 2010*, vol. 39, no. 01. 2010.

[30] T. Benson and G. Grieve, *Principles of Health Interoperability*, vol. 53, no. 9. Cham: Springer International Publishing, 2016.