Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240351

Smart FOX - Enabling Citizen-Based Donation of EHR-Standardised Data for Clinical Research in Austria

Klaus DONSA [a,1], Karl KREINER [a], Dieter HAYN [a], Angelika RZEPKA [a], Sofia OVEJERO [b], Michaela TOPOLNIK [c], Andreas ZIEGL [d], Bernhard PFEIFER [e,f], Sabrina NEURURER [e,f], Saskia KALTENBRUNNER [g], Elisabeth KLAGER [h], Kurt ZATLOUKAL [i], Bernhard ZATLOUKAL [j], Thomas SCHABETSBERGER [k], Manuel L. GARCIA [l], Nikola TANJGA [m] and Günter SCHREIER [a]
[a] *AIT Austrian Institute of Technology GmbH, Austria*
[b] *Dedalus HealthCare Ges.m.b.H., Netherlands*
[c] EIT Health Austria GmbH, Austria
[d] *telbiomed Medizintechnik und IT Service GmbH, Austria*
[e] *UMIT TIROL GmbH, Austria*
[f] *Tirol Kliniken GmbH, Austria*
[g] *University Vienna, Austria*
[h] *Ludwig Boltzmann Institute Digital Health and Patient Safety, Austria*
[i] *Medical University of Graz, Austria*
[j] *fragmentiX Storage Solutions GmbH, Austria*
[k] *Siemens Healthcare Diagnostics GmbH, Austria*
[l] *Probando GmbH, Austria*
[m] *ELGA GmbH, Austria*
ORCiD ID: Klaus Donsa https://orcid.org/0000-0003-4671-6218

Abstract. Access to healthcare data for secondary use in clinical research is often restricted due to privacy concerns or business interests, hindering comprehensive analysis across patient pathways. The Smart FOX project seeks to address this challenge by developing concepts, methods, and tools to facilitate citizen/patient-driven donations of health data for clinical research. Leveraging the groundwork, laid by the national Electronic Health Record implementation in Austria (called ELGA), Smart FOX aims to harness structured datasets from ELGA for research purposes through an opt-in approach. With funding secured from the Austrian Research Promotion Agency, the project embarks on innovative solutions encompassing governance frameworks, community engagement, and technical infrastructure. The Smart FOX consortium, comprising key stakeholders across various healthcare-associated domains, will evaluate these efforts through demonstrators focusing on clinical registries, patient-generated data, and recruitment services. The project targets to accompany the development of future data donation infrastructure while ultimately advancing clinical research efficiency and bolstering Austria's preparedness for the European Health Data Space. This paper presents the first systematic evaluation of the technical concept and proposal for the federated system architecture of the Austrian Health Data Donation Space, which is the socio-technical goal of Smart FOX.

---

[1] Corresponding Author: Klaus Donsa, AIT Austrian Institute of Technology; E-mail: klaus.donsa@ait.ac.at.

Keywords. Data Donation, Electronic Health Records, Privacy Preservation, Real-World Data, Clinical Research, Citizen Science

1. Introduction

Data from healthcare often remains unavailable for secondary use in clinical research due to privacy or business protection reasons. This often results in difficulties analyzing data along the continuous patient pathway, especially involving various healthcare providers. One part of the complexity, especially in Austria, lies in the federal healthcare system with different financing modalities, responsibilities, and lobbies.

As a result, from an institutional perspective, the attempt to analyse data along a continuous patient journey involving different healthcare institutions produces several blind spots and missing links.

The common denominator in accessing real world data (RWD) on continuous patient journeys is the patient him/herself. Many patients affected with severe illness have a detailed medical history, but no possibility to help others in form of data donations for secondary use. Various studies have shown a high degree of willingness by patients regarding data donations if certain conditions are met [1-3].

Against this backdrop, the Smart FOX project (Smart and Federated Open data eXchange of citizen-based data donations for clinical research) targets to develop concepts, methods, and tools to enable citizens/patients in Austria to donate their health data for clinical research. The Smart FOX consortium comprises 19 partners from research, governmental organizations, national and international industry, healthcare, and patient advocacy. Within the implementation of the national Electronic Health Record (EHR) in Austria – ELGA - Elektronische Gesundheitsakte – a decade has been invested in data harmonization. Smart FOX aims to leverage this potential and utilize donations of highly structured ELGA datasets for clinical research. Since current legislation does not allow for an opt-out-based use of ELGA data for research, this potential will be tapped for the first time with an opt-in-based approach.

This paper presents the first systematic evaluation of the technical concept and proposal for the federated system architecture of the Austrian Health Data Donation Space (AHDDS), which is the socio-technical goal of Smart FOX.

2. Methods

The project requires the development of innovations on several levels, particularly:

a) governance framework for ethical, legal, social, and economic issues,
b) community and capacity building to ensure sustainability and population-wide acceptance,
c) technical solutions for the donation of health data by citizens/patients and their utilization by clinical researchers in real world demonstrators.

The AHDDS works towards combining an open independent architecture with components and services from the industry. On the one hand, to prevent vendor lock-in and maximize trust and sustainability, on the other hand with the intention to establish a living data space ecosystem, facilitating participation of third-party services and intermediaries. The AHDDS ecosystem envisages to support the following modes of data donation to increase efficiency in clinical research using real world data:

- Directed data donation takes place within a newly or an already established research context. A research context is defined as medical research, a clinical registry or biobank or a clinical trial, where the regulatory approval for the usage of the data is clearly defined within e.g. an ethics approval.
- Undirected data donation takes place in a not yet specified research context. The data donation is based on ethics approval as well as data protection impact assessment according to GDPR and will be provided by the patient/citizen knowing that the data will be used in a future research context. The patient/citizen will be asked to provide informed consent that the data will be used under predefined conditions.

3. Results

The proposed software architecture (Figure 1) follows the original federated approach by ELGA but is extended with components and services for enabling data space ecosystems. Implementation of the proposed software architecture will be performed within the project up to a Technological Readiness Level of 4 (Demonstrator).

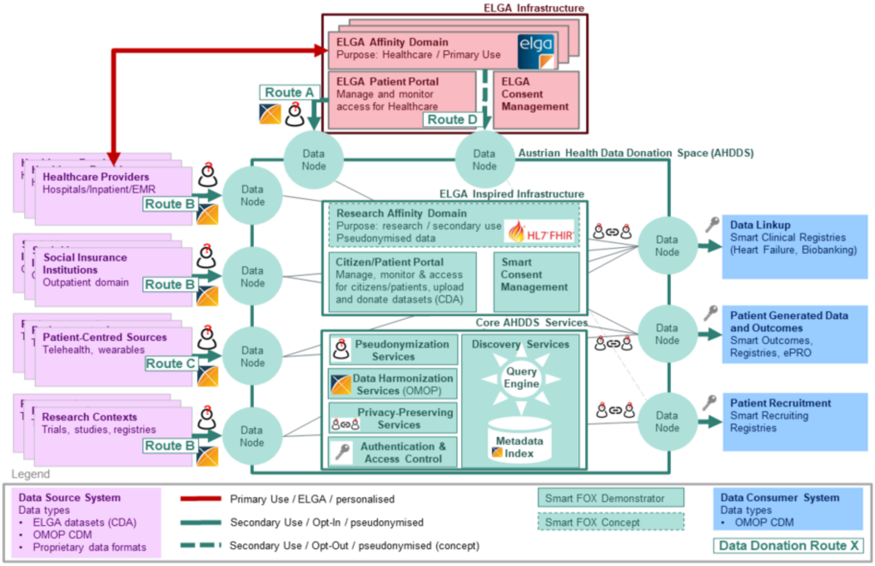

Figure 1. Overall architecture and components of the Austrian Health Data Donation Space (AHDDS).

ELGA inspired infrastructure: Following the concept of ELGA, a federated infrastructure for access and usage of citizen/patient-based data donations is envisaged. One implementation, which is investigated in the project, is the setup of a dedicated, fully pseudonymized research affinity domain, which builds up on the IHE Cross-Enterprise Document Sharing (XDS) based infrastructure. Within the AHDDS, Clinical Document Architecture (HL7 CDA) and Fast Healthcare Interoperability Resources (HL7 FHIR) will act as standardized data formats. Through a citizen/patient portal, citizens/patients have access to their donated data. The portal also acts as one access point to the smart consent management and the data donation service, and it enables monitoring of access to their data. This will also require the development and testing of a consent model for patients based on the XACML (eXtensible Access Control Markup Language) infrastructure used in large-scale health information exchange networks [4].

Distributed data nodes and data donation routes: The AHDDS is linked to primary data from ELGA (top, red, Figure 1) and from various data providers (left, purple, Figure 1). The architecture is based on distributed data nodes, holding pseudonymized and harmonized data [5]. The nodes are located at the edge of the AHDDS, within the source data's jurisdiction. Data are transferred to the distributed data nodes along four possible data donation routes:

- Route A: Citizens/patients access and download their datasets from within the ELGA Portal and upload them to a dedicated donation platform (driven by citizens/patients).
- Route B: Citizens/patients give additional consent to use their EHR data during an established research context (e.g., biobanking, clinical registries and studies).
- Route C: Citizens/patients enrol into a dedicated "patient generated data and outcomes" service to regularly collect and donate data on outcomes via questionnaires, vital signs (industry driven).
- Route D: Citizens/patients manage their data donation in a portal dedicated to secondary use of health data, e.g., an additional ELGA affinity domain linked to the existing shared EHR infrastructure (concepts developed).

Existing ELGA reports (e.g., hospital discharge letters, medication, lab results …) will be used to extract data for further use. A dedicated data processing pipeline for automated conversion from HL7 Clinical Document Architecture (CDA) to HL7 FHIR and Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) will be designed and demonstrated, featuring policy enforcement rules including updating, versioning and accessing in a granular rights/roles system.

Core AHDDS services: Once a patient donates data, the pseudonymization services are called and the patient's record undergoes a de-identification protocol before being transferred to the storage component, where data harmonization and standardization is performed. All source data will be mapped to the OMOP CMD.

Linkage of data between the distributed data nodes is done by privacy-preserving services after ethical approval for a research context. While different privacy-preserving and cybersecurity technologies will be kept in mind for defining the AHDDS architecture, two specific approaches will be explored in the demonstrators. The European Patient Identity Management Services (EUPID) will be applied for privacy preserving record linkage [6]. To support even very sensitive data, like genomes from Next Generation Sequencing, highly secure storage solutions will be evaluated [7].

Access to linked data is managed by the authentication & access control services. For authentication of citizens/patients we will explore the identity management services of ID Austria. Different methods for electronic consenting, authentication schemes, and authorization schemes in secondary use will be explored [7]. The discovery services allow researchers to find data that are of potential interest for dedicated research questions. A query engine is envisioned that searches the metadata provided by the data nodes (e.g. Data Catalogue vocabulary Application Profile – DCAT AP), which are stored in the metadata index.

Data consumer systems: Any access to data within the AHDDS is done through data nodes, based on services, user interfaces and APIs. The data nodes ensure that the respective consent is granted, that the user is authenticated to access the requested data, that the required level of privacy is preserved, that the data is standardized, harmonized, stored and transferred. During Smart FOX, selected ways to access the AHDDS will be explored based on three demonstrators:

1. Data Linkup will explore the linkage of ELGA-standardized data with two clinical registries: a) heart failure registry and b) biobanking registry.
2. Patient Generated Data and Outcomes will explore the donation of patient generated data generated from patients with chronic diseases, e.g., from telehealth services, and their linkage with in-clinic and out-patient data.
3. Patient Recruitment will explore smart recruiting based on metadata and donated data. This involves (further) developments of the industry services (mappings to meta data repository, standardization, consent management, pseudonymization) and the collaborative work for connecting the interfaces.

4. Conclusions

The requirements for the next generation of data donation infrastructure will be derived from the results and experiences gained from these demonstrators. Ultimately, Smart FOX aims to pave the way for large-scale and trusted citizen/patient-driven donation of health data to enhance the efficiency of clinical research and contribute to Austria's readiness for the upcoming European Health Data Space.

Acknowledgements

Smart FOX is funded by the Austrian Research Promotion Agency (FFG).

References

[1] Seltzer E, Goldshear J, Guntuku SC, Grande D, Asch AD, Klinger EV and Raina RM, "Patients' willingness to share digital health and non-health data for research: A cross-sectional study," BMC Med. Inform. Decis. Mak., vol. 19, no. 1, pp. 1–8, 2019, doi: 10.1186/s12911-019-0886-9.

[2] Silber H,. Gerdon F, BachR, Kern C, Keusch F, and Kreuter F, "A preregistered vignette experiment on determinants of health data sharing behavior," Polit. Life Sci., vol. 41, no. 2, pp. 161–181, 2022, doi: 10.1017/pls.2022.15.

[3] Pilgrim K and Bohnet-Joschko S, "Donating Health Data to Research: Influential Characteristics of Individuals Engaging in Self-Tracking," Int. J. Environ. Res. Public Health, vol. 19, no. 15, 2022, doi: 10.3390/ijerph19159454.

[4] Mangesius P, Saboor S and Schabetsberger T. Patient Summaries in Context of Large Scale EHR Networks with Fine Granular Access Control Restrictions. Stud Health Technol Inform. 2017;237:91-96. PMID: 28479549. https://doi.org/10.3233/978-1-61499-761-0-91

[5] Baumgartner M, Kreiner K, Lauschensky A, Jammerbund B, Donsa K, Hayn D, Wiesmüller F, Demelius L, Modre-Osprian R, Neururer S, Slamanig G, Prantl S, Brunelli L, Pfeifer B, Pölzl G and Schreier G (2024) Health data space nodes for privacy-preserving linkage of medical data to support collaborative secondary analyses. Front. Med. 11:1301660. doi: 10.3389/fmed.2024.1301660

[6] Nitzlnader M and Schreier G, "Patient identity management for secondary use of biomedical research data in a distributed computing environment," Stud. Health Technol. Inform., vol. 198, pp. 211–218, 2014, doi: 10.3233/978-1-61499-397-1-211.

[7] Zatloukal B, Müller H, Strasser, W, Zatloukal K. (2022). Biobanks for Enabling Research and Development by Trusted Patient Data Environment. In: Sargsyan K, Huppertz B, Gramatiuk S (eds) Biobanks in Low- and Middle-Income Countries: Relevance, Setup and Management. Springer, Cham. https://doi.org/10.1007/978-3-030-87637-1_27

[8] Hampf C, Bialke M, Geidel L, Vass A, Bahls T, Blasini R, Blumentritt A, Boeker M, Bruns C, Jandrig B, Fünfgeld, M, Heinrich P, and Hoffmann W"A survey on the current status and future perspective of informed consent management in the MIRACUM consortium of the German Medical Informatics Initiative," Transl. Med. Commun., vol. 6, no. 1, pp. 1–12, 2021, doi: 10.1186/s41231-021-00086-1.