Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200163

Incorporation of Multiple Sources into IT - and Data Protection Concepts: Lessons Learned from the FARKOR Project

Doris LINDOERFER[a,1], Ulrich MANSMANN[a] and Isabel REINHARDT[a,b]
[a] *Institute for medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Munich, Germany*
[b]*Medical center of the university of Munich*

Abstract. The IT- and data protection concept of the FAmiliäres Risiko für das KOloRektale Karzinom (FARKOR) project will be presented. FARKOR is a risk adapted screening-project in Bavaria, Germany focusing on young adults with familial colorectal cancer (CRC). For each participant, data from different sources have to be integrated: Treatment records centrally administered by the resident doctors association (KVB), data from health insurance companies (HIC), and patient reported lifestyle data. Patient privacy rights must be observed. Record Linkage is performed by a central independent trust center. Data are decrypted, integrated and analyzed in a secure part of the scientific evaluation center with no connection to the internet (SECSP). The presented concept guarantees participants privacy through different identifiers, separation of responsibilities, data pseudonymization, public-private key encryption of medical data and encrypted data transfer.

Keywords. Data integration, HIC data, pseudonymization, data encryption

1. Introduction

The *FAmiliäres Risiko für das KOloRektale Karzinom* (FARKOR) project assesses the efficacy and safety of a risk adapted screening-program focusing on familial colorectal cancer (CRC) in young adults (aged 25-50 years of age) [1]. It is implemented in Bavaria, Germany [2]. The project integrates individual data from different sources: (1) individual diagnostic and treatment data from medical practitioners, (2) individual patient-reported lifestyle data, and (3) individual outcome data from health insurance companies (HIC).

FARKOR has two parts: (1) Efficient selection of persons with familial colorectal cancer risk within a population of 3 million people, (2) offering safe and efficient screening approaches for the selected high risk population. FARKOR is a prospective, population-based cohort study. Its efficiency is assessed by (1) the number of advanced adenoma or carcinoma detected in the selected high-risk population, (2) a health economic evaluation that estimates avoided morbidity and mortality.

A large number of practices enrolled into the project. All 35 health insurance companies working within Bavaria participate. They provide individual outcome data

[1] Corresponding Author: Doris Lindoerfer, lindoerf@ibe.med.uni-muenchen.

regarding CRC mortality and morbidity. This allows to have control outcome of persons not enrolled in FARKOR.

To implement and to run this project successfully, a reliable IT infrastructure and robust data protection concept has to be established that is also compliant with the new European General Data Protection Regulation (EU-GDPR) [3].

2. Methods

2.1. Several institutions and responsibilities

The FARKOR project integrates data from multiple sources and exchanges the data between different institutions (see Figure 1).

2.2. Specific challenges and used methods

Specific challenges regarding the data protection concept are: 1.) Integrating data from different sources; 2.) Protecting the privacy of participating individuals 3.) Protect business issues between the participating HICs (which patient is insured by which HIC); 4.) Providing the legal reasons for collecting the data (informed consent for subjects enrolled in FARKOR, permission to use individual data of persons not enrolled in FARKOR – the controls); 5.) Performing person rights: Informing about collected data or deleting the records on request; 6.) Providing a 24/7 availability of the IT-infrastructure.

The FARKOR IT- and data protection concept is built on several strategies [4-10]: Informational separation of powers (Informationelle Gewaltenteilung [11]) by using different identifiers for data of different origin [4]; Pseudonymization of identifiers [5] and encryption of personal health information [6]; a trust center [7] and concepts of asymmetric encryption and pseudonymization [8-10].

Integration of HIC data over multiple databases uses standardized terminologies like the International Classification of Diseases, 10th Revision (ICD-10), *Operationen- und Prozedurenschlüssel* (OPS) and *Einheitlicher Bewertungs Maßstab* (EBM).

3. Results

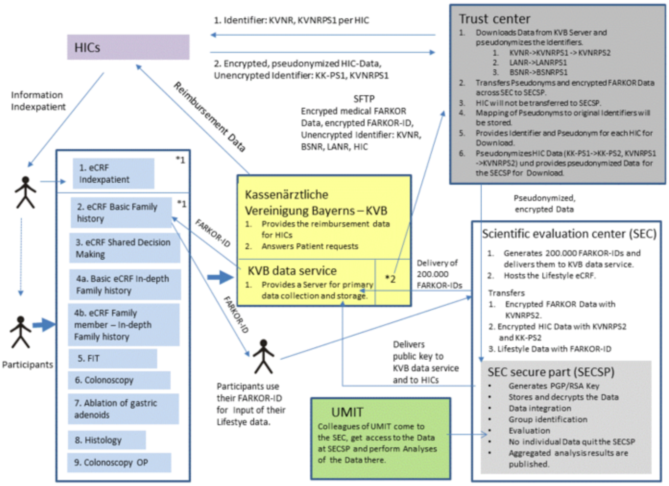
3.1. IT- and data protection concept with data flow

The program started in October 2018. By end of December 2019 about 9000 persons were enrolled. The data capturing, data encryption, data flow (see Figure 1), pseudonymization and data decryption work appropriately.

A specific data service by the resident doctor's association *(1) KVB data service* administrates the individual project data. It is separated from the institution's routine data and general activities. The *(2) trust center* pseudonymizes the data identifiers.

The *KVB data service* provides an internet portal to document the *medical data*: demographic, diagnostic, and treatment data of individuals enrolled in FARKOR. After informed consent, participants are enrolled by their primary care physician who does a basic familial CRC risk assessment. In case of a positive assessment, the participant will be referred to specialists for further staging and potential screening measures. The *KVB data service* encrypts the data and makes it available to the trust center for download.

Record linkage between the *medical data* and the individual health insurance data uses the social security numbers (KVNR). The trust center downloads the medical data from the *KVB data service*, removes the specific HIC IDs, pseudonymizes the individual IDs, and makes the data available for download to the secure evaluation environment (SECSP). The SECSP is a secure part in the secure evaluation center (SEC) with no connection to the internet.


HICs
1. Identifier: KVNR, KVNRPS1 per HIC
2. Encrypted, pseudonymized HIC-Data,
Unencrypted Identifier: KK-PS1, KVNRPS1
Trust center
1. Downloads Data from KVB Server and pseudonymizes the Identifiers.
1. KVNR->KVNRPS1 -> KVNRPS2
2. LANR->LANRPS1
3. BSNR->BSNRPS1
2. Transfers Pseudonyms and encrypted FARKOR Data across SEC to SECSP.
3. HIC will not be transferred to SECSP.
4. Mapping of Pseudonyms to original Identifiers will be stored.
5. Provides Identifier and Pseudonym for each HIC for Download.
6. PseudonymizesHIC Data (KK-PS1->KK-PS2, KVNRPS1 ->KVNRPS2) und provides pseudonymized Data for the SECSP for Download.
Information
Indexpatient
Reimbursement Data
SFTP
Encryped medical FARKOR
Data, encrypted FARKOR-ID,
Unencrypted Identifier: KVNR,
BSNR, LANR, HIC
1. eCRF
Indexpatient
*1
2. eCRF Basic Family history
*1
3. eCRF Shared Decision Making
4a. Basic eCRF In-depth Family history
4b. eCRF Family member – In-depth Family history
5. FIT
6. Colonoscopy
7. Ablation of gastric adenoids
8. Histology
9. Colonoscopy OP
Participants
FARKOR-ID
FARKOR-ID
Kassenärztliche
Vereinigung Bayerns – KVB
1. Provides the reimbursement data for HICs
2. Answers Patient requests
KVB data service
1. Provides a Server for primary data collection and storage.
*2
Delivery of
200.000
FARKOR-IDs
Pseudonymized,
encrypted Data
Scientific evaluation center (SEC)
1. Generates 200.000 FARKOR-IDs and delivers them to KVB data service.
2. Hosts the Lifestyle eCRF.
Transfers
1. Encrypted FARKOR Data with KVNRPS2.
2. Encrypted HIC Data with KVNRPS2 and KK-PS2
3. Lifestyle Data with FARKOR-ID
Delivers
public key to
KVB data service
and to HICs
Participants use
their FARKOR-ID
for Input of their
Lifestye data.
UMIT
Colleagues of UMIT come to the SEC, get access to the Data at SECSP and perform Analyses of the Data there.
SEC secure part (SECSP)
• Generates PGP/RSA Key
• Stores and decrypts the Data
• Data integration
• Group identification
• Evaluation
• No individual Data quit the SECSP
• Aggregated analysis results are published.

*1 Data acquisition through different Physicians per Participant.

*2 Defined area for storing the FARKOR data which only the trust center has access to.

Figure 1. Data flow in the FARKOR project.

The trust center provides at scheduled points in time, the social security numbers (KVNR) and a pseudonym of the KVNR of the FARKOR participants to the HICs for download. The HICs match individual data with related ICD-10, OPS and EBM codes according to the specified identifiers. The HICs pseudonymize the data accordingly, remove the KVNR and deliver the pseudonymized data back to the trust center. The trust center pseudonymizes the HIC data, performs the record linkage and allows to download the data to the SECSP. The HICs data is now doubly pseudonymized.

The concept of informational separation of powers requests, that all medical data is encrypted by the KVB data service and has to be decrypted in the SECSP. Encryption uses public-private-key RSA encryption. Patient-related data are encrypted with the public key. The identifying variables are not encrypted for pseudonymization purposes. It was requested by the LMU's IRB that the trust center cannot read the medial data.

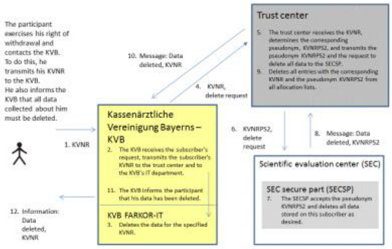
The SECSP downloads the data and decrypts the individual medical data with the private key. Then the SEC performs data storage, data integration and analysis. All data transfers are transport-encrypted.

Furthermore, lifestyle data are important for the analysis. The SEC created individual FARKOR-IDs, which are delivered to the *KVB data service*. These IDs are provided to the participants when they enroll and allows login to the lifestyle questionnaire website. Participants use their internet browser to answer the lifestyle questionnaire. The corresponding server is hosted in the SEC. These data will be transferred to the SECSP and integrated to the project data with the FARKOR-ID.

The health economic evaluation is carried out by partners at the Private University for Health Sciences, Medical Informatics and Technology (UMIT). They perform their analysis within the SEC. No data leaves the SECSP.

3.2. Warranty of the participants rights

Participants may request at any time a copy of their individual data stored. They can also withdraw their consent of participation in FARKOR at any time. Their requests are handled in a standardized manner. The process for withdrawal is shown in Figure 2. To date, four participants have withdrawn their consent. The corresponding data was deleted from the analysis data. So far, no participant has requested access to their data records.


Trust center
Kassenärztliche Vereinigung Bayerns – KVB
KVB FARKOR-IT
Scientific evaluation center (SEC)
SEC secure part (SECSP)
1. KVNR
4. KVNR, delete request
6. KVNRPS2, delete request
8. Message: Data deleted, KVNRPS2
10. Message: Data deleted, KVNR
12. Information: Data deleted, KVNR

The participant exercises his right of withdrawal and contacts the KVB. To do this, he transmits his KVNR to the KVB. He also informs the KVB that all data collected about him must be deleted.

2. The KVB receives the subscriber's request, transmits the subscriber's KVNR to the trust center and to the KVB's IT department.
3. Deletes the data for the specified KVNR.
5. The trust center receives the KVNR, determines the corresponding pseudonym, KVNRPS2, and transmits the pseudonym KVNRPS2 and the request to delete all data to the SECSP.
7. The SECSP accepts the pseudonym KVNRPS2 and deletes all data stored on this subscriber as desired.
9. Deletes all entries with the corresponding KVNR and the pseudonym KVNRPS2 from all allocation lists.
11. The KVB informs the participant that his data has been deleted.

Figure 2. Data flow when participants withdraw their consent.

4. Discussion

The FARKOR IT- and data protection concept enables the integration of multi-source data. To our knowledge, this paper describes for the first time an EU-GDPR conform simultaneous integration of individual HIC as well as project data during the course of the study. The integration of HIC data allows to have control groups of persons not enrolled in FARKOR, as well as data for subjects enrolled that supplements data documented during medical examinations. This makes it necessary to collect and store social security numbers (KVNR). This record linkage process is complex. The KVNR is stored and pseudonymized in the trust center and communicated to the HICs. The KVNR is not transmitted to the SECSP. The medical data will be encrypted in the KVB data service and decrypted in the SECSP. The medical data and the FARKOR-ID are only available in encrypted form to the trust center.

5. Conclusion

To integrate the multi-source data, a strong IT infrastructure as well as an appropriate data protection concept is needed to guarantee participating persons privacy. This is achieved by using different identifiers for data of different origin, pseudonymization of identifiers in a trust center, encrypted data transfer and encryption of patient-related data.

Ethics and approval

The Institutional Review Board (IRB) of the Medical Faculty of the Ludwig-Maximilians Universität München, Munich, Germany has given ethics approval for this study (18-545). The FARKOR data protection concept has been approved by the Bavarian State Data Protection Commissioner. Furthermore, it was approved by the supervisory authorities of the HICs, the Bavarian State Ministry of Health and Nursing and the Federal Insurance Office, this was necessary because also data about persons not participating in FARKOR are provided by the HICs. The FARKOR IT- and data protection concept was also presented and discussed in the (TMF) e.V. [11], working group data protection (AG DS).

Acknowledgements

The FARKOR project is funded by the Innovation Committee of the Federal Joint Committee (G-BA) https://innovationsfonds.g-ba.de/): 01NVF17026. DL received funding from FARKOR. IR received funding from DIFUTURE. DIFUTURE is funded by the Bundesministerium für Bildung und Forschung (BMBF): 01ZZ1804B.

References

[1] KVB, Projekt, FARKOR: Vorsorge bei familiärem Risiko für das kolorektale Karzinom, Available at: https://www.kvb.de/abrechnung/verguetungsvertraege/farkor/ (Accessed: Jan 7, 2020).

[2] S. Hoffmann, A. Crispin, D. Lindoerfer, G. Sroczynski, U. Siebert, U. Mansmann, FARKOR: Evaluating the effects of a risk-adapted screening program for familial colorectal cancer in individuals between 25 and 50 years of age in a prospective population-based intervention study, *BMC Gastroenterology*. [submitted].

[3] General Data Protection Regulation GDPR, Available at: https://gdpr-info.eu/ (Accessed: Jan 7, 2020).

[4] D. Lindoerfer, U. Mansmann, IT Infrastructure of an Oncological Trial Where Xenografts Inform Individual Second Line Treatment Decision*, Stud Health Technol Inform*, **235** (2017), 226-230.

[5] T.M. Deserno, D. Haak, V. Brandenburg, V. Deserno, C. Classen, P. Specht, Integrated Image Data and Medical Record Management for Rare Disease Registries. A General Framework and its Instantiation to the German Calciphylaxis Registry, *J Digit Imaging*, **27:6** (2014), 702-713.

[6] S. Sherman, O. Shats, E. Fleissner, G. Bascom, K. Yiee, M. Copur, K. Crow, J. Rooney, Z. Mateen, M.A. Ketcham, J. Feng, A. Sherman, M. Gleason, L. Kinarsky, E. Silva-Lopez, J. Edney, E. Reed, A. Berger, K. Cowan, Multicenter breast cancer collaborative registry, *Cancer Inform* **10** (2011), 217–226.

[7] H. Aamot, C.D. Kohl, D. Richter, P. Knaup-Gregory, Pseudonymization of patient identifiers for translational research. *BMC Med Inform Decis Mak* **13**, 75 (2013).

[8] T. Michelsen, C. Lins, S. Gudenkauf, A. Hein, C. Lüpkes, Privacy by Design for Integrated Case and Care Management: Receiver-Oriented Encryption in STROKE OWL, *Stud Health Technol Inform,* **258** (2019), 110-114.

[9] N. Adam, T. White, B. Shafiq, J. Vaidya, X. He, Privacy preserving integration of health care data, *AMIA Annu Symp Proc,* **11** (2007), 1-5.

[10] M. Bialke, P. Penndorf, T. Wegner, T. Bahls, C. Havemann, J. Piegsa, W. Hoffmann, A workflow-driven approach to integrate generic software modules in a Trusted Third Party*, J Transl Med.* **13,** 176 (2015).

[11] TMF e.V., Available at: http://www.tmf-ev.de/ (Accessed: Jan 7, 2020).