

An openEHR Virtual Patient Template for Pancreatic Cancer

Samer ALKARKOUKLY^{a,1} and Abdul-Mateen RAJPUT^a

^aCECAD, Faculty of Medicine and University Hospital Cologne, University of Cologne, Joseph-Stelzmann-Strasse 26, 50931 Cologne, Germany

Abstract. openEHR is an open-source technology for e-health, aims to build data models for interoperable Electronic Health Records (EHRs) and to enhance semantic interoperability. openEHR architecture consists of different building blocks, among them is the “template” which consists of different archetypes and aims to collect the data for a specific use-case. In this paper, we created a generic data model for a virtual pancreatic cancer patient, using the openEHR approach and tools, to be used for testing and virtual environments. The data elements for this template were derived from the “Oncology minimal data set” of HiGHmed project. In addition, we generated virtual data profiles for 10 patients using the template. The objective of this exercise is to provide a data model and virtual data profiles for testing and experimenting scenarios within the openEHR environment. Both of the template and the 10 virtual patient profiles are available publicly.

Keywords. openEHR, data model, data architecture, openEHR Templates, semantic interoperability, HiGHmed, virtual patient, test data, oncology.

1. Introduction

openEHR is a technology for e-health, consisting of open specifications, clinical models, and software that can be used to create standards, and build information and interoperability solutions for healthcare (1). It provides an open-source healthcare information modelling environment that enables modelling of interoperable Electronic Health Records (EHRs). The standard is maintained by a community of healthcare professionals and software developers. Since its inception, it has been used widely across the globe (2–4).

openEHR maintains a vendor-independent standard based on two-level modelling: Reference Model and Archetype Model. In addition, Terminology binding is considered as a third major model of the openEHR (3). Archetypes are combined and nested in templates to represent specific use cases. Typically, templates express entire clinical documents containing different information modelled as several archetypes such as discharge letters, result reports, or medical history (5).

HiGHmed (6) is one of the medical informatics projects funded by the German Federal Ministry of Education and Research (BMBF) and it perceives semantic interoperability as a prerequisite for enabling a meaningful exchange of the data in

¹ Samer Alkarkoukly, CECAD, Faculty of Medicine and University Hospital Cologne, University of Cologne, Joseph-Stelzmann-Strasse 26, 50931 Cologne, Germany; E-mail: mohammad.abboud-alkarkoukly@uk-koeln.de.

federated cross-institutional settings by adopting the openEHR approach for semantic modelling (7).

To test the capabilities of an openEHR system, a complete profile of an individual patient was required. The available templates on Clinical Knowledge Manager (CKM) (8) can be used to generate partial clinical data and it has been shown that large datasets can be generated by following the semi-automated approach (9) but a complete representation of a virtual patient in a particular disease is still missing. The complete profile would provide the insights of a clinical course of the disease and all the available data elements associated with a particular disease in an openEHR format. The modelling of the virtual representation in biomedical sciences is not new. There have been many initiatives e.g., Virtual Brain (10), Virtual Cell (11), etc. to either simulate the patient phenotypes or generate data for further processing.

The objective of our current work in this paper is to build a virtual patient data model in openEHR format, that could perform as a generic data architecture for testing purposes and also for virtual environments of electronic medical records. To the author's knowledge, the approach described here is new to this field.

2. Methods

An openEHR template for a virtual patient of pancreatic cancer is developed by following the guidelines and using the available archetypes associated with use-case oncology in HiGHmed CKM (8).

The template was created to represent the minimal dataset of oncology use-case within HiGHmed (8).

The online platform "openEHR Archetype Designer" (12) was used to design the template. All the used archetypes, clusters, sections are available publicly and mentioned later in this paper. Figure 1 illustrates the architecture of the template in Archetype Designer.

The template was used as a base to generate a full set of test data for 10 virtual patients by automating the REST API calls using locally installed Better platform Think!EHR (13) endpoint and following the same methodology as described in our previous paper (9). The patient's data were generated automatically and were not exported from any medical information system.

Figure 1. Overview of the architecture of the template

3. Results

The template was created using the German language, basically to follow the language of the source dataset. It covers the following data modules: patient personal information, diagnosis (ICD, ICD-O, TNM), histology grading and analysis, clinical history, surgical operations, radiotherapy, systemic therapy, local interventional therapy, medications, risk factors, disease progress, tumour board conferences, and clinical studies.

The virtual patient template, and also the used archetypes, are now available publicly on GitHub: “https://github.com/abdulmateenraj/VirtualPatient_openEHR” and could be used as a testing structure in any openEHR system.

A test data set for 10 virtual pancreatic cancer patients was created based on the template, and also available on the same URL mentioned above in .json format. Those files could be used also for testing purposes not only in openEHR systems but also in any other electronic health record environment.

```

1 {
2   "ctx/language": "de",
3   "ctx/territory": "US",
4   "ctx/composer_name": "Silvia Blake",
5   "ctx/id_namespace": "HOSPITAL-NS",
6   "ctx/id_scheme": "HOSPITAL-NS",
7   "ctx/participation_name": "Dr. Marcus Johnson",
8   "ctx/participation_function": "requester",
9   "ctx/participation_mode": "face-to-face communication",
10  "ctx/participation_id": "199",
11  "ctx/participation_name1": "Lara Markham",
12  "ctx/participation_function1": "performer",
13  "ctx/participation_id1": "198",
14  "ctx/health_care_facility/name": "Hospital",
15  "ctx/health_care_facility/id": "9091",
16  "csv_export_test/context/personenidentifizierer/pid:0": "bd3b63fa-d598-4c41-a2b7-a0f0e6082544",
17  "csv_export_test/context/personenidentifizierer/pid:0issuer": "Issuer",
18  "csv_export_test/context/personenidentifizierer/pid:0assigner": "Assigner",
19  "csv_export_test/context/personenidentifizierer/pid:0type": "Prescription",
20  "csv_export_test/context/personenidentifizierer/geografischer_geltungsbereich:0code": "at0011",
21  "csv_export_test/context/fallidentifikation/fallnr": "FALLNR 31",
22  "csv_export_test/context/name/name_strukturiert/patienten_vornamen": "Patienten_Vornamen 7",
23  "csv_export_test/context/name/name_strukturiert/patienten_nachname": "Patienten_Nachname 92",
24  "csv_export_test/context/geschlecht/patienten_geschlecht": "Sonstiges/Intersexuell",
25  "csv_export_test/context/daten_zur_geburt/patienten_geburtsdatum": "2020-07-16",
26  "csv_export_test/context/adresse/adresse:0/artcode": "at0011",
27  "csv_export_test/context/adresse/adresse:0/strukturierte_adresse/ort:0": "ORT 7",
28  "csv_export_test/context/adresse/adresse:0/strukturierte_adresse/land:0": "Land 96",
29  "csv_export_test/context/adresse/adresse:0/patienten_plz": "Patienten_PLZ 53",
30  "csv_export_test/alter/patientenalter": "P2DT12H43M",
31  "csv_export_test/diagnose/diagnose/primaertumor_icd_code": "Primaertumor_ICD_Code 97",
32  "csv_export_test/diagnose/diagnose/primaertumor_diagnosetext": "Primaertumor_Diagnosetext 60",
33  "csv_export_test/diagnose/diagnose/diagnosedatum": "2020-07-16T12:52:17.86982+02:00",
34  "csv_export_test/diagnose/diagnose/codesystem_icd/primaertumor_icd_version": "Primaertumor_ICD_Version 75",
35  "csv_export_test/diagnose/diagnose/tumorklassifikation_icd-o/primaertumor_topographie_icd_o": "Primaertumor_To",
36  "csv_export_test/diagnose/diagnose/tumorklassifikation_icd-o/primaertumor_topographie_icd_o_freitext": "Primaertumor_Topographie_ICD_O_Freitext 1",

```

Figure 2. Example of a json representation of one test patient

4. Discussion and Conclusion

In this paper, we address a challenge that EHR systems face, especially during the establishment phase, which is the need to test the infrastructure with the patient data. The dataset of 10 virtual patients covers the various aspects of the healthcare system (starting from patient demographics and clinical history, through diagnostics and diagnosis, ending in treatment and clinical studies). It includes data coming from different clinical systems, and on the other hand covers the different data types (Boolean, integer, string, date, datetime) for a pancreatic cancer use-case.

Additionally, the virtual patient template represents a full set of data parameters and can perform as a generic data architecture for testing and virtual environments.

Through our work, we made sure to maintain a high level of interoperability with the other EMR systems at both semantic (by providing a nationally used dataset for oncology) and syntactic levels (by following the international architecture and rules of openEHR).

The limitation of this virtual patient dataset is that it is not a real clinical data thus it is not suitable for some clinically oriented scenarios such as Natural Language Processing, however, it has the similar characteristics of real patient data and can be used in the other scenarios. Although the template is based on a pancreatic cancer profile, it could be used for similar cancers. Additionally, it represents the first step for future work on patient profiles from the other diseases by following the same methodology and implementing modifications to the dataset elements.

Acknowledgement

This work was partially supported by the HiGHmed Consortium funded by the German Federal Ministry of Education and Research (BMBF, funding code [01ZZ1802U]).

References

- [1] openEHR International. What is openEHR?; 2021 [cited 2021 Jan 29]. Available from: URL: https://www.openehr.org/about/what_is_openehr.
- [2] Maia TA, Muiylder CF de, Reis ZSN. Archetype Development Process: A Case Study of Support Interoperability among Electronic Health Record in the State of Minas Gerais, Brazil. *J Med Syst* 2019; 43(3):57.
- [3] Min L, Tian Q, Lu X, Duan H. Modeling EHR with the openEHR approach: an exploratory study in China. *BMC Med Inform Decis Mak* 2018; 18(1):75.
- [4] Mascia C, Frexia F, Uva P, Zanetti G, Pireddu L, Giacomelli G et al. The openEHR Genomics Project. *Stud Health Technol Inform* 2020; 270:443–7.
- [5] Wulff A, Mast M, Hassler M, Montag S, Marschollek M, Jack T. Designing an openEHR-Based Pipeline for Extracting and Standardizing Unstructured Clinical Data Using Natural Language Processing. *Methods Inf Med* 2020; 59(S 02):e64-e78.
- [6] Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. *Methods Inf Med* 2018; 57(S 01):e66-e81.
- [7] Wulff A, Haarbrandt B, Marschollek M. Clinical Knowledge Governance Framework for Nationwide Data Infrastructure Projects. *Stud Health Technol Inform* 2018; 248:196–203.
- [8] HiGHmed Modelling Group. Project: Use Case Onkologie: HiGHmed; 2020 [cited 2021 Jan 29]. Available from: URL: <https://ckm.highmed.org/ckm/projects/867.116.5>.