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

doi:10.3233/SHTI240655

Integration of Oncological Data into openEHR: A Path Towards Improved Cancer Care and Research

Jan SCHLADETZKY [a,b,c,1], Niklas REIMER [a,c,d,1], Herwig NICOLAUS [b], Hauke BUSCH [c,d], Björn SCHREIWEIS [a,e] and Ann-Kristin KOCK-SCHOPPENHAUER [a,b,c]

[a] *Medical Data Integration Center, University Hospital Schleswig-Holstein, Germany*
[b] *IT Center for Clinical Research, Lübeck, Universität zu Lübeck, Lübeck, Germany*
[c] *University Cancer Center Schleswig-Holstein, University Hospital Schleswig-Holstein, Lübeck, Germany*
[d] *Medical Systems Biology Group, Lübeck Institute of Experimental Dermatology, Universität zu Lübeck, Lübeck, Germany*
[e] *Institute for Medical Informatics and Statistics, Kiel University and University Hospital Schleswig-Holstein, Germany*

Abstract. The integration of tumor-related diagnosis and therapy data is a key factor for cancer-related collaborative projects and research projects on-site. The Medical Data Integration Center (MeDIC) of the University Hospital Schleswig-Holstein, resulting from the Medical Informatics Initiative and Network University Medicine in Germany, has agreed on an openEHR-based data management based on a centralized repository with harmonized annotated data. Consequently, the oncological data should be integrated into the MeDIC to interconnect the information and thus gain added value. A uniform national data set for tumor-related reports is already defined for the cancer registries. Therefore, this work aims to transform the national oncological basis data set for tumor documentation (oBDS) so that it can be stored and utilized properly in the openEHR repository of the MeDIC. In a previous work openEHR templates representing the oncological basis data set were modeled. These templates were used to implement a processing pipeline including a metadata repository, which defines the mappings between the elements, a FHIR terminology service for annotation and validation, resulting in a tool to automatically build openEHR compositions from oBDS data. The prototype proved the feasibility of the referred mapping, integration into the MeDIC is straightforward and the architecture introduced is adaptable to future needs by design.

Keywords. health information interoperability, data integration, data management, registries, cancer epidemiology

1. Introduction

Information models are required to meet the increasing need for interoperability in the healthcare sector. openEHR is a widely used standard for modeling content from medical

[1] Both authors contributed equally. Corresponding Authors: Jan Schladetzky; E-mail: jan.schladetzky@uksh.de; Niklas Reimer; E-mail: niklas.reimer@uksh.de

records in the healthcare domain [1]. The medical data integration center (MeDIC) at the University Hospital Schleswig-Holstein [2], which originated from the Medical Informatics Initiative in Germany, uses the openEHR standard to integrate semantically enriched medical data permanently into the openEHR Clinical Data Repository (CDR). This architecture requires all data from relevant source systems to be transferred to the openEHR format based on agreed upon archetypes.

In the oncological field in Germany, it is mandatory by law to report cancer cases regularly to the cancer registries. The responsible stakeholders agreed on a standardized export for reporting, called the German oncological basis data set (oBDS) [3]. This data set, currently available in version 3.0.2, is the successor to the ADT/GEKID data set and has received an extensive structural and technical update [4]. The oBDS standardizes treatment data of cancer patients, ensuring quality assurance and comparability. The data set is highly maintained and of excellent quality, making it valuable for highly relevant scientific purposes.

Relevant projects and collaborations like the German Cancer Consortium (DKTK) rely on the ADT/GEKID data set for data analysis and quality assurance but implement a mapping to FHIR resources [5]. Carus et al. [6] mapped the oBDS elements to the OMOP common data model to enable decentralized analysis. The MeDIC needs to implement the mapping between the oBDS and openEHR templates to integrate these data into its openEHR CDR for further usage. The integration of the oBDS data set into the CDR supports the secondary use of data like populating subsequent systems like i2b2 repositories and increasing the screening capabilities of eligible patients for cancer-related studies [7,8].

2. Methods

2.1. oBDS Representation in openEHR

openEHR comprises open specifications and information models for managing, storing, and exchanging health data. Patient data are stored in an electronic health record (EHR), allowing easy transfer between systems. A multi-level modeling approach separates data representation from actual content. IT experts define the data representation through the reference model (RM), serving as the foundation for syntactic interoperability. Based on the RM, domain experts design archetypes to define possible content, serving as the basis for semantic interoperability. The archetypes can then be included in templates to constrain and utilize them for specific use cases. The use of international terminologies ensures a semantically higher level of interoperability between systems.

Nicolaus et al. [9] used the Archetype Designer to model templates for the eight report types defined by the oBDS: Diagnosis, Pathology, Surgery, Radiotherapy, Systemic Therapy, History, Death, and Tumor Conference. The mapping was performed manually, referring to predefined archetypes wherever possible. The eight templates utilize a total of 35 archetypes, reusing 25 existing archetypes. All templates are accessible via the HiGHmed Clinical Knowledge Manager [10].

2.2. Metadata Repository and Terminology Validation

A core part of the data integration pipeline is a metadata repository (MDR). It serves two primary purposes: 1) the representation and annotation of data models of source and target systems, which includes information about data types or whether the data element is mandatory. Data elements of the target system may also reference terminological resources. 2) defining relations between source and target systems. Relations can define equality or transform between two elements. As shown in Figure 1, the MDR can perform schema-level mapping using relations by referencing data attributes to the corresponding paths of the openEHR template, as well as value conversions e.g. for the ICD-10 code.

The XML schema of the oBDS dataset includes various value sets for the data elements. Those can be split into internal value sets, where the XML schema defines all possible values, and external value sets, which refer to code systems like ICD-10, where neither the code nor the description are provided within the scheme. For the latter, a FHIR-based terminology service looks up descriptions for the codes provided [11].


```
{                                                    MDR Input
  "srcProfileCode": "obds-tod",
  "srcProfileVersion": 1,
  "trgProfileCode": "openehr-obds-tod",
  "trgProfileVersion": 1,
  "values": {
          "Sterbedatum":        "2023-01-01",
          "Abschluss_ID":       "12345",
          "Tod_tumorbedingt":   "J",
          "Todesursache_ICD": {
                    "Code":     "C92.00",
                    "Version":  "10 2023 GM"
   }
  }
}
```

```
{                                                                    MDR Output
  "trgProfileCode": "openehr-obds-tod",
  "trgProfileVersion": 1,
  "values": {
     "openEHR-EHR-ADMIN_ENTRY.person_data.v0/at0024/openEHR-EHR-
        CLUSTER.death_details.v1/at0001":                 "2023-01-01",
     "openEHR-EHR-ADMIN_ENTRY.person_data.v0/at0024/openEHR-EHR-
        CLUSTER.death_details.v1/openEHR-EHR-
        CLUSTER.case_identification.v0/at0001":          "12345",
     "openEHR-EHR-ADMIN_ENTRY.person_data.v0/at0024/openEHR-EHR-
        CLUSTER.death_details.v1/at0002":                 true,
     "openEHR-EHR-ADMIN_ENTRY.person_data.v0/at0024/openEHR-EHR-
        CLUSTER.death_details.v1/at0003": {
                            "code":                       "C92.00",
                            "version":                    "2023"
  }}}
```


Figure 1. Conversion for the oBDS report type *death report* using the MDR. The value part of the MDR input comes from the XML data, while the attribute names of the MDR output correspond to the path of the attribute in the openEHR template.

3. Results

A Java-based tool was developed to map the oBDS XML data to openEHR (cf. Figure 2). For mapping paths, the content of the XML subnodes will be sent as JSON to the MDR (cf. Figure 1). Attributes that contain an object with the nodes' *code* and *version* will be sent to the terminology service for further annotation. Depending on the attribute, this can either be a *$lookup* operation to fetch the description or a *$translate* operation if mapping between two value sets is necessary. The raw MDR output has a flat structure, which must be converted into a tree-based structure for further processing. To transform the mapping result into an openEHR composition, the tool leverages components of *openEHR mapper* [12]. The required template is queried as operational template (OPT) from the openEHR repository and then used along with the mapping result to populate all fields, including labels and default values. Resulting compositions can be saved as canonical JSON to disk or sent to an openEHR repository by direct access or through an XDS.b compatible document registry [13]. The source code is available on GitHub [14].

A test environment was used for the evaluation with EHRbase v0.32.0 as the openEHR repository, due to its capability of using a FHIR Terminology Service to validate codes against the respective terminology. The pipeline successfully processed over 2000 test reports from 725 test patients containing all modules (340 Diagnosis, 0 Pathology, 370 OP, 250 Radio Therapy, 500 Systemic Therapy, 90 History, 30 Death, and 500 Tumor Conference). The test environment's implementation demonstrated the possibility to map between the two standards. By creating templates based on existing archetypes, no information was lost, and all elements were transferred.

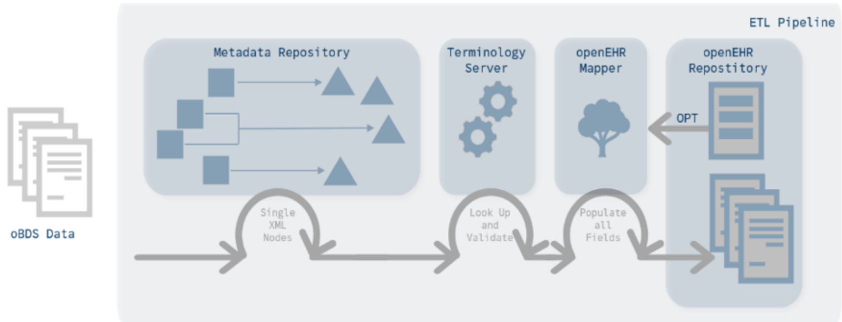

Figure 2. Tools and methods workflow of the integration pipeline: the actual mapping is centralized in the MDR, code systems are resolved via the terminology server, the individual parts are assembled into a composition in the openEHR Mapper based on the OPT, and the instances are then made available for research in the openEHR repository.

4. Discussion

This implementation shows that data can be integrated across two schemas via an MDR. The work relies on an architecture of decoupling imperative code and declarative mapping. This enables easy adoption of new revisions of both oBDS and openEHR archetypes/templates. Because the tool is not limited to input data in the oBDS format but also supports any XML input, it can easily be expanded to other formats, such as HL7 v2. Once this approach will be in productive use within the MeDIC, future developments or existing ETL routes may also be mapped using this architecture.

As the department of pathology sends oBDS pathology reports directly to the cancer registry, the oBDS pathology module was neither included in the test data nor converted to openEHR. By now, all entry types and data structures are supported, but due to the iterative development process, some openEHR reference model data types must still be implemented in the mapping tool. Implementation will evolve once a broader usage leverages additional data types. To include archetypes in templates, they must include a translation into the template's language. This requirement has hindered the use of some existing archetypes in German templates lacking a German translation during the realization of this work. The Gleason Score archetype was translated and contributed to the openEHR community to enable its inclusion in German templates in the future.

As the presented approach relies on the MDR's ability to create relations and executable conversions, it limits the choice of the MDR due to not all available implementations supporting such functionality. Using the FHIR-based terminology API enables the utilization of any available terminology service that implements the FHIR API allowing the free choice of implementations.

Declarations

Author contributions: JS, *NR*, *AKS*: conception of the work; data acquisition and interpretation; *HN*, *JS*: openEHR design; *JS*, *NR*: implementation of pipeline; *JS*, *NR*, *AKS*: writing the manuscript. All authors approved the manuscript in the submitted version and take responsibility for the scientific integrity of the work.

Acknowledgment: This research was funded by German Federal Ministry of Education and Research, grant numbers 01KX2121, 01ZZ2322N.

References

[1] Beale T. openEHR [Internet]. openEHR Foundation; [cited 2024 Mar 11]. Available from: https://www.openehr.org/

[2] Kock-Schoppenhauer A-K, Schreiweis B, Ulrich H, Reimer N, Wiedekopf J, Kinast B. Medical Data Engineering – Theory and Practice. In: Bellatreche L, Chernishev G, Corral A, Ouchani S, Vain J, editors. Advances in Model and Data Engineering in the Digitalization Era. Vol. 1481. Cham: Springer International Publishing; 2021. p. 269–84. doi:10.1007/978-3-030-87657-9_21.

[3] Basisdatensatz [Internet]. [cited 2024 Mar 13]. Available from: https://basisdatensatz.de/

[4] Katalinic A, Halber M, Meyer M, Pflüger M, Eberle A, Nennecke A. Population-Based Clinical Cancer Registration in Germany. Cancers 2023;15:3934. doi:10.3390/cancers15153934.

[5] Deppenwiese N, Delpy P, Lambarki M, Lablans M. ADT2FHIR - A Tool for Converting ADT/GEKID Oncology Data to HL7 FHIR Resources. Stud Health Technol Inform 2021;283:104–10. doi:10.3233/SHTI210547.

[6] Carus J, Nürnberg S, Ückert F, Schlüter C, Bartels S. Mapping Cancer Registry Data to the Episode Domain of the Observational Medical Outcomes Partnership Model (OMOP). Applied Sciences 2022;12:4010. doi:10.3390/app12084010.

[7] Haarbrandt B, Tute E, Marschollek M. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. J Biomed Inform 2016;63:277–94. doi:10.1016/j.jbi.2016.08.007.

[8] Li M, Cai H, Nan S, Li J, Lu X, Duan H. A Patient-Screening Tool for Clinical Research Based on Electronic Health Records Using OpenEHR: Development Study. JMIR Med Inform 2021;9:e33192. doi:10.2196/33192.

[9] Nicolaus H, Anywar M, Pazmino Pinto S, Ulrich H, Schladetzky J, Kock-Schoppenhauer A-K, et al. Integration of the Oncologic Base Data Set into the UKSH Medical Data Integration Center. Student Conference 2024, Medical Engineering Science, Medical Informatics, Biomedical Engineering and Auditory Technology, Lübeck: Infinite Science Publishing; 2024.

[10] Systems SG Ocean Health. Clinical Knowledge Manager - UKSH [Internet]; [cited 2024 Mar 11]. Available from: https://ckm.highmed.org/ckm/#showproject_1246.152.61

[11] HL7 International. Terminology Service - FHIR v4.0.1 [Internet]; 2019 [cited 2024 Mar 13]. Available from: https://www.hl7.org/fhir/R4/terminology-service.html

[12] Reimer N, Ulrich H, Busch H, Kock-Schoppenhauer A-K, Ingenerf J. openEHR Mapper – A Tool to Fuse Clinical and Genomic Data Using the openEHR Standard. In: Röhrig R, Beißbarth T, Brannath W, Prokosch H-U, Schmidtmann I, Stolpe S, editors. Studies in Health Technology and Informatics, IOS Press; 2021. doi:10.3233/SHTI210055.

[13] IHE International. IHE IT Infrastructure (ITI) Technical Framework: Cross-Enterprise Document Sharing (XDS.b) [Internet]; 2023 [cited 2024 Mar 13]. Available from: https://profiles.ihe.net/ITI/TF/Volume1/ch-10.html#10

[14] Reimer N, Schladetzky J. openehr-obds-code [Internet]; 2024 [cited 2024 Mar 11]. Available from: https://github.com/itcr-uni-luebeck/openehr-obds-code