Reusing Data of an EU-Wide EHR System for Clinical Trials: A Capabilities Analysis

Barbora KANKOVA[a,1] and Georg DUFTSCHMID[a]
[a] *Section of Medical Information Management, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria*

Abstract. Conduction of clinical trials may benefit from the reuse of EHR data. The upcoming EHR system eHDSI provides data exchange between European countries and thus represents an attractive data source for multinational trials. In this paper we analyze to what extent eHDSI could provide data relevant for trials with a focus on patient recruitment. Data elements identified in the EHR4CR project to be frequently used in trials were mapped to the HL7 templates of the eHDSI document types using the open source tool ART-DECOR. From the 149 EHR4CR data elements, 44 (30%) could be mapped to eHDSI document components. Despite this moderate coverage, eHDSI could still provide a substantial contribution for recruitment by an automatic pre-filtering process of large groups of potential trial candidates.

Keywords. electronic health records, clinical study, research design, health information exchange

1. Introduction

Data recorded in electronic health records (EHR) is primarily used in context of clinical care but also provides substantial potential for secondary use in clinical research [1, 2]. Reusing EHR data in the context of clinical trials was amongst others strived for by the Electronic Health Records for Clinical Research (EHR4CR) partnership [3].

The Smart Open Service for European Patients (epSOS) project, which was finalized in 2014, focused on development and evaluation of cross-border eHealth services for European countries, which would enable secure services for exchange of medical related data [4]. Within the eHealth Digital Service Infrastructure (eHDSI) the concepts for cross-border health data interchange developed in epSOS are currently being deployed across the EU member countries [5]. According to the definition of ISO/TR 20514, eHDSI can be considered a shared EHR system [6]. Taking into account the multinational patient cohort that will be covered by eHDSI, the latter could represent an attractive data source for reuse in the context of clinical trials.

Accordingly the goal of this work was to analyze to what extent eHDSI actually contains the data that are relevant for clinical trials. As our work is still ongoing, we present in this paper the first part of the results concerning the recruitment of trial patients.

[1] Corresponding Author: Barbora Kankova, Section of Medical Information Management, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria, E-Mail: n01400337@students.meduniwien.ac.at.

2. Method

Within the EHR4CR project, data elements were identified that are frequently used in the context of trial feasibility checking [7], for patient identification and recruitment [8], and for trial execution [9]. In order to analyze the availability of these data elements within eHDSI, we formally mapped them to the structures of the eHDSI document types using the tool ART-DECOR.

2.1. ART-DECOR

The open source tool ART-DECOR is used by various multidisciplinary stakeholders (e.g., healthcare institutions, regional/national networks, and IT service providers) in the field of health information exchange (HIE). Amongst other functionalities, it allows information needs to be specified as so-called concepts. They may then be mapped to the components of document types that allow the recording of the concepts. For the specification of document type structures, ART-DECOR supports the HL7 Clinical Document Architecture (CDA) templates standard [10]. ART-DECOR has been applied in more than 40 HIE projects in different countries, including Austria, where it is used in the context of the ELGA system [11].

2.2. Mapping the EHR4CR data elements to the eHDSI document types

We set up an ART-DECOR project named EHR4CR2eHDSI for the specification of our mappings [12]. The EHR4CR data elements were already specified as ART-DECOR concepts in an earlier work [13]. Within the corresponding ART-DECOR project [14] these concepts are grouped in the three data sets (i) “EHR4CR Data Inventory for trial feasibility” (95 concepts), (ii) “EHR4CR Data Inventory for patient identification and recruitment” (149 concepts), and (iii) “EHR4CR Data Inventory for trial execution” (94 concepts). Most concepts are associated with standardized codes (UMLS, LOINC, and/or SNOMED CT) that were specified in the EHR4CR project to unambiguously define the data elements’ semantics. We inherited all these concepts in our project EHR4CR2eHDSI as they represent the origin of our mappings.

Documents exchanged via eHDSI are represented in HL7 CDA format [15]. The structures of the eHDSI document types, which are based on the work done in epSOS, were defined as HL7 CDA templates. They are publicly accessible from an ART-DECOR project [16] that was set up by the Dutch National IT Institute of Health Care with support of HL7 International and HL7 Europe. In the course of our mapping we referenced those components of [16] that contain relevant targets for the EHR4CR data elements.

For the specification of the mappings we created so-called ART-DECOR “template associations”, which allow concepts to be formally linked to HL7 CDA template components within a GUI (compare figure 1). We applied three different approaches for finding semantically equivalent components within the epSOS templates for the different EHR4CR data elements:

- **Code search**: We used the codes associated with the EHR4CR data elements to find artefacts within [16] that are associated with the same code. Hereby we searched the fixed codes prescribed by the epSOS templates for the “code” element of the epSOS CDA components. Additionally, we searched the value

sets that were defined for epSOS CDA elements (e.g., “value” or “code”) to prescribe the allowed values. A hit in a value set lead us to the corresponding template that then formed the target of our mapping.

- **String-matching:** Referring to the names of the EHR4CR data elements, we tried to find suitable mapping targets in two different sets. On the one hand, we searched the concepts that were already defined and mapped to corresponding templates in the epSOS ART-DECOR project [16]. Several of the EHR4CR data elements already appeared within the epSOS concepts, which allowed us to reuse the corresponding template associations to the epSOS templates. On the other hand, we directly searched the epSOS templates for suitable pendants for the EHR4CR data elements. In this search we used text comparisons, i.e. whole or part string matches including synonyms.
- **Manual comparison of definitions:** Finally, we tried to find identical semantics based on the textual definitions of the EHR4CR data elements and the epSOS template components.

For the search process we relied on (i) the complete epSOS ART-DECOR project file that can be downloaded in XML format from https://art-decor.org/decor/services/RetrieveProject?prefix=epsos-&mode=compiled&download=true and on (ii) the epSOS Implementation Guideline that is a complete documentation of all ART-DECOR artefacts created for epSOS and can be downloaded in PDF format from https://art-decor.ehdsi.eu/publication/epSOS/ .

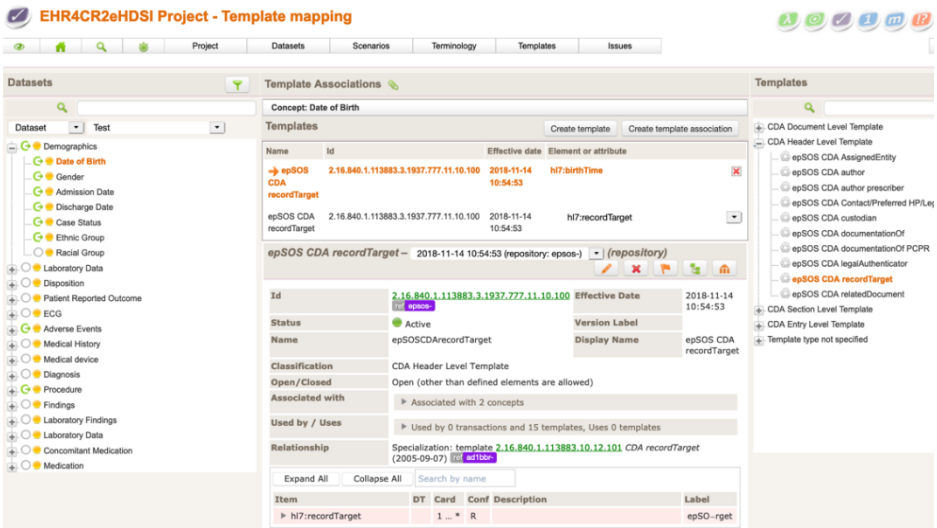

Figure 1. ART-DECOR template associations (center) were used to map the EHR4CR data elements (left) to corresponding components of the epSOS templates (right). The screenshot shows the mapping of the EHR4CR data element “Date of birth” to the component “hl7:birthTime” of the epSOS template “epSOS CDA recordTarget”.

3. Results

In the following we will present the results of mapping the EHR4CR concepts concerning patient identification and recruitment to the eHDSI document types. From the 149 EHR4CR concepts we were able to map 44 (30%) to corresponding epSOS template components. From these mappings, 42 (95%) refer to epSOS template components holding structured content, such as codes, dates, or quantities, whereas for 2 (5%) of the mappings the target holds unstructured content, i.e. text (compare Table 1).

Table 1. Results of mapping the EHR4CR data elements to epSOS templates.

Mapping type	Number of data elements	Percentage
Mapping to structured template component	42	28%
Mapping to unstructured template component	2	1%
No mapping possible	105	71%
Total	149	100%

From the 44 mappings, we found (i) 6 (14%) by searching for identical associated codes, (ii) 21 (48%) via string-matching, and (iii) 17 (39%) by manually comparing the definitions of the EHR4CR data elements and the epSOS templates (compare Table 2).

Table 2. Methods used for finding mapping targets.

Method	Number of data elements	Percentage
Code-bases search	6	14%
String matching	21	48%
Manual comparison of definitions	17	38%
Total	44	100%

4. Discussion

Recruitment of patients for clinical trials by reuse of routine data from EHR systems has been strived for by numerous projects [17, 18] and was shown to be effective and efficient [19]. Several papers also quantitatively analyzed to what extent their EHR system covers the data elements required for clinical trials.

Ateya and colleagues analyzed data elements that were used for patient recruitment within more than 200 primary care studies and compared them with data elements that they assumed to be typically present in an EHR system [20]. According to their assessment, a typical EHR system should be able to cover 74% of the trial-specific data elements in structured format.

Köpcke and colleagues extracted the recruitment-specific data elements from 15 deliberate clinical trials and examined their overlap with the data elements recorded within five hospital-based EHR systems [21]. They stated that, on average, 55% of the trial data elements were covered by the observed EHR systems. However, only for 35% of them data was actually recorded.

El Fadly and colleagues addressed a slightly different aspect of reusing EHR data for clinical trials by focusing on the pre-population of trial forms from EHR data elements [2]. They mapped the data elements of seven forms of a multi-center clinical trial to a hospital-based EHR system which allowed them to pre-populate 13.4% of the trial data elements.

The vast majority of earlier work concerning reuse of EHR data for recruitment of trial patients, including the before mentioned papers, focused on local EHR systems within a single health institution. In contrast to them we focus on a shared EHR system that aims to support HIE in the entire European Union. This has the obvious advantage that the covered patient cohort is much wider, which would be particularly helpful for patient recruitment in the context of multinational clinical trials. Besides, the eHDSI system is based on a standardized data model, which allows our mapping of the EHR4CR data elements to the eHDSI document types to be directly applied to all patients within the EU member states for whom data is recorded in eHDSI. In contrast, the trial data elements would have to be individually mapped to the content of each single local EHR system, as the latter typically employ proprietary data models. On the other hand, the content recorded in eHDSI is currently limited to patient summaries and medication data and therefore only constitutes a subset of the more detailed content that is typically stored in local EHR systems. This also explains eHDSI's rather low coverage of the EHR4CR data elements in comparison to the numbers reported in [2, 20, 21].

Reuse of data from a shared EHR system was addressed by an earlier work of us [22], where we analyzed the potential of the Austrian nation-wide EHR system ELGA for trial recruitment. We found that 61% of the EHR4CR data elements concerning patient identification and recruitment are considered in structured format within the current ELGA document types. An additional 5% could be recorded in ELGA in unstructured format. The main reason why ELGA covers more than twice as much recruitment-specific data elements than eHDSI is that it includes a laboratory report and more than half of EHR4CR's recruitment-specific data elements refer to laboratory data. None of these data elements referring to laboratory data could be mapped to eHDSI document types.

Strengths of our work are that our mappings are (i) publicly accessible and therefore fully reproducible via our ART-DECOR project EHR4CR2eHDSI [12], and they are (ii) formally specified within the corresponding project file in XML format, from where they can be electronically retrieved and easily transformed into a machine readable query [23].

A limitation of our work is that the mapping process requires a certain amount of interpretation in some cases, in particular when it comes to judging whether a certain specific data element can be expected to be present within a more generic unstructured text. As an example, we assumed that data element "expanded disability status scale (EDSS) score" could be recorded in element "text" of eHDSI template "Section Functional Status". Even though we discussed and tried to harmonize such unclear mappings, other scientists may come to different results. The problem is exacerbated by the fact that the documentation of the eHDSI templates in ART-DECOR is mostly rather Spartan.

Another limitation is that we only considered the theoretical availability of data elements within eHDSI by analyzing the eHDSI document type specifications. We did not have access to actual eHDSI documents and thus could not analyze to what extent that frequently optional data elements are in fact documented in practice.

Summarizing, our results show that with a mapping success rate of less than a third, eHDSI's coverage of the EHR4CR recruitment-specific data elements is moderate. However, even the possibility to have a third of a trial's inclusion/exclusion criteria checked in an automated way may drastically reduce the number of trial candidates and save a lot of time for manually checking the remaining criteria for the pre-filtered cohort. Besides, if eHDSI proves to be useful for cross-border patient care, it is realistic to assume that further document types may be added in the future. With an additional

laboratory report, eHDSI's current coverage of the EHR4CR recruitment specific data elements could be more than doubled immediately.

References

[1] Prokosch, H.U. and T. Ganslandt, Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med, 2009. **48**(1): p. 38-44.

[2] El Fadly, A., et al., Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform. J Biomed Inform, 2011. **44 Suppl 1**: p. S94-102.

[3] McCowan, C., et al., Using Electronic Health Records to Support Clinical Trials: A Report on Stakeholder Engagement for EHR4CR. Biomed Res Int, 2015. **2015**: p. 707891.

[4] European Commission. Cross-border health project epSOS: What has it achieved? 2014 January 15, 2020]; Available from: https://ec.europa.eu/digital-single-market/en/news/cross-border-health-project-epsos-what-has-it-achieved.

[5] European Commission. eHealth DSI Operations Home. 2020; Available from: https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHealth+DSI+Operations+Home.

[6] International Organization for Standardization, ISO/TR 20514:2005 Health informatics -- Electronic health record -- Definition, scope and context, I.O.f. Standardization, Editor. 2005.

[7] Doods, J., et al., A European inventory of common electronic health record data elements for clinical trial feasibility. Trials, 2014. **15**: p. 18.

[8] Doods, J., et al., A European inventory of data elements for patient recruitment. Stud Health Technol Inform, 2015. **210**: p. 506-10.

[9] Bruland, P., et al., Common data elements for secondary use of electronic health record data for clinical trial execution and serious adverse event reporting. BMC Med Res Methodol, 2016. **16**(1): p. 159.

[10] Heitmann, K.U., J. Curry, and L. Nelson, HL7 Templates Standard: Specification and Use of Reusable Information Constraint Templates, Release 1. 2014.

[11] Herbek, S., et al., The Electronic Health Record in Austria: a strong network between health care and patients. European Surgery, 2012(3): p. 155-163.

[12] Kankova, B., K.U. Heitmann, and G. Duftschmid. EHR4CR2eHDSI. 2020 January 28, 2020]; Available from: https://art-decor.org/art-decor/decor-project--ehr4cr2ehsdi-.

[13] Encheva, E., Reuse of data from the Austrian Nation-Wide EHR System ELGA for clinical trial feasibility checking, patient recruitment, and trial execution: A capability analysis. 2019, University of Vienna: Vienna.

[14] Encheva, E., et al. EHR4CR Data Inventory. 2019 January 16, 2020]; Available from: https://art-decor.org/art-decor/decor-project--ehr4cr-.

[15] Dolin, R.H., et al., HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc, 2006. **13**(1): p. 30-39.

[16] Nictiz. epSOS in ART-DECOR. 2020 January 16, 2020]; Available from: https://art-decor.ehdsi.eu/art-decor/decor-project--epsos-.

[17] Cuggia, M., P. Besana, and D. Glasspool, Comparing semi-automatic systems for recruitment of patients to clinical trials. Int J Med Inform, 2011. **80**(6): p. 371-88.

[18] Kopcke, F. and H.U. Prokosch, Employing computers for the recruitment into clinical trials: a comprehensive systematic review. J Med Internet Res, 2014. **16**(7): p. e161.

[19] Lai, Y.S. and J.D. Afseth, A review of the impact of utilising electronic medical records for clinical research recruitment. Clin Trials, 2019. **16**(2): p. 194-203.

[20] Ateya, M.B., B.C. Delaney, and S.M. Speedie, The value of structured data elements from electronic health records for identifying subjects for primary care clinical trials. BMC Med Inform Decis Mak, 2016. **16**: p. 1.

[21] Kopcke, F., et al., Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. BMC Med Inform Decis Mak, 2013. **13**: p. 37.

[22] Augustinov, G. and G. Duftschmid, Can the Austrian Nation-Wide EHR System Support the Recruitment of Trial Patients? Stud Health Technol Inform, 2019. **259**: p. 87-90.

[23] Ott, S., C. Rinner, and G. Duftschmid, Expressing Patient Selection Criteria Based on HL7 V3 Templates Within the Open-Source Tool ART-DECOR. Stud Health Technol Inform, 2019. **260**: p. 226-233.