Navigating Healthcare Through Challenging Times
D. Hayn et al. (Eds.)

doi:10.3233/SHTI210087

Semiautomatic Recruitment of Trial Patients Using ELGA Data: Conceptual Design and Implementation of an IT Tool

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

Abstract. Reuse of EHR data can substantially improve the recruitment process of clinical trials. As shown earlier, Shared EHR systems are particularly attractive data sources. The goal of this work was to conceptually design and implement a user-friendly tool for semiautomatic trial recruitment using ELGA data. The tool applies a web-based client (Vue and Electron frameworks) – server (Django-Python and Java server, SQLite database) architecture. Trial eligibility criteria are expressed as XPaths. Access to ELGA documents is simulated using the eHealth Connector library and the IHE XDS Open eHealth Integration Platform framework. Usability was optimized in expert interviews with investigators of two active trials. First feedback based on synthesized ELGA test data indicates suitability for clinical end users. Further insights are expected from applying the tool to real ELGA data.

Keywords. Electronic Health Records, Clinical Trial Recruitment, Standards

1. Introduction

Randomized controlled trials provide a powerful research design for the evaluation of healthcare interventions and are widely recognized as the gold standard of clinical research [1]. Clinical trials enable to test the effectiveness of medications, medical devices, and health methods. For each trial, participants with a specific profile have to be recruited [2]. The profile is defined by eligibility criteria that may for example refer to health problems, gender, age, or treatment.

Unfortunately, many trials suffer from delays or are not completed at all [3–8]. One reason are the inefficient and time-demanding recruitment methods such as advertising through flyers or brochures. Another disadvantage of traditional methods are the high costs that may arise per patient [8].

Reuse of EHR data can improve trial recruitment [9], [10]. As we have shown, Shared EHR systems represent attractive data sources for this purpose [11], [12]. The Austrian national Shared EHR system ELGA [13] in particular holds structured data for more than 60% of eligibility criteria [11] that are commonly used in clinical trials according to an analysis of the EHR4CR project [14].

[1] Corresponding Author: Raik Müller, Section for Medical Information Management, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria, E-Mail: mail@raik-mueller.com

We are currently working on a project that aims to analyze to what extent trial recruitment at the Medical University of Vienna (MedUni) can benefit from reusing ELGA data. In this course we will examine whether the promising theoretical results of Augustinov and Duftschmid [11] can be confirmed in practice. The goal of the master thesis, which is part of this project and is presented in this paper, is to conceptually design and implement the underlying recruitment tool. Hereby, our primary goal was to achieve high usability by clinical end users.

2. Methods

2.1. Conception phase

Conception of our tool was based on the IEEE software development process [15]. In the analysis phase we aimed to determine the context and requirements of the planned tool [16]. It was carried out through analysis of literature and existing recruitment systems as well as a series of expert interviews with clinical end users. An initial set of candidate requirements derived from literature research and existing systems was represented within mockups of the future user interface, which served as starting points for feedback in the interviews and a following revision of requirements. Revised requirements and correspondingly adapted mockups were then validated in a second round of expert interviews. Particular efforts were made to achieve an intuitive visualization of the results of trial eligibility checks that should allow for easy comprehension by clinical end users.

2.2. System architecture

To prepare for an easy applicability, we chose a web-based client-server system architecture. This should avoid dependencies or restrictions on the end users' hardware solutions. To alleviate reuse of our work by other researchers, widely used open source technologies were selected that should simplify potential future enhancements [17]. The source code including installation description is published on GitHub [18].

The client provides the user interface for selecting the trial and the patients to be checked for eligibility. The analysis of the patients' documents is done by the server and the results are finally displayed by the client again.

For the implementation of the client the progressive JavaScript web framework Vue [19] was used, which allows the creation of single-page web applications. In addition, the Electron Framework was employed, which enables a cross-platform desktop application [20].

In the back-end structure, a Django-Python [21] server acts as the center for all task areas. It holds the REST interface implementation for communication with the client and processes the ELGA documents.

For communication with the IHE XDS environment, a Java server based on the open-source eHealth Connector (EHC) library [22] is used. EHC is based on international implementation guidelines and standards that promote a harmonized exchange of data and documents in the healthcare sector and can be integrated into a back-end. This also makes it possible to validate CDA documents for conformance to the ELGA templates. The Open eHealth Integration Platform framework is used to simulate an IHE XDS environment [23].

An SQLite database [24] is employed for storing the trial-specific metadata including the inclusion and exclusion criteria. The criteria are represented as XPaths that refer to those components of ELGA documents that hold the required data. Hereby, we relied on a mapping of data elements commonly referenced in trial eligibility criteria to ELGA documents [11]. A tool for an interactive specification of trial criteria referring to ELGA document components that can be exported as XPaths was presented in [25]. The XPaths originate from the template IDs of the HL7 V3 templates that define the structure of the respective ELGA document component. This allows accessing any ELGA document types. Currently, ELGA lab reports and medication documents seem particularly promising as most of their components are highly structured.

2.3. Data model of trial criteria

For the representation of trial criteria, a suitable data structure was developed (compare **Figure 1**). It distinguishes between inclusion and exclusion criteria and allows a criterion to be checked by multiple alternative conditions. As an example, the criterion "person has diabetes" may alternatively be checked via conditions "diabetes was diagnosed", "diabetes-specific medication was dispensed", and "diabetes-specific lab parameter exceeds a certain threshold".

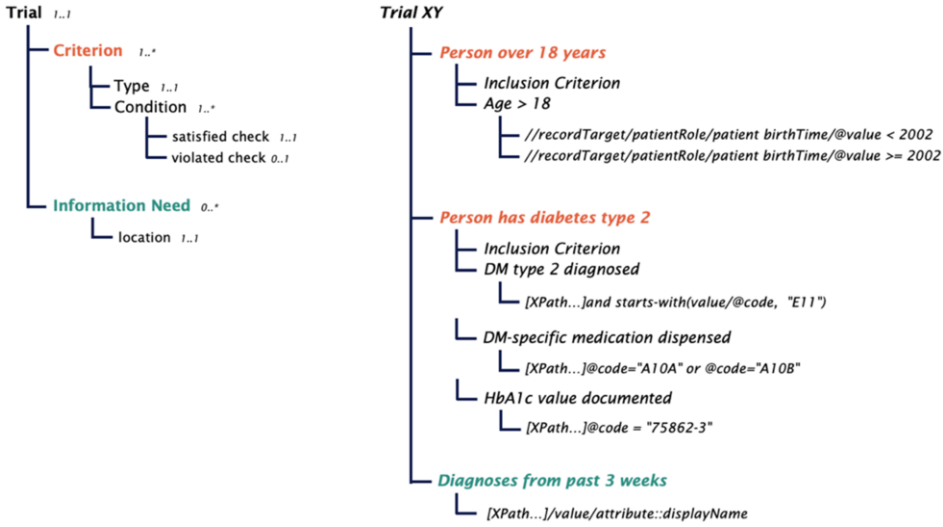

Figure 1 – Data model (left) and example instance (right) of trial criteria

In some cases, a condition may be clearly assessed to be violated based on existing data. As an example, condition "age > 18 years" can obviously be found to be satisfied or violated based on the documented birthdate of a person. Therefore, our data structure also covers checking for a conditions' violation. The latter is, however, optional as in most cases only the satisfaction of a condition will be assessable. As an example, condition "diabetes was diagnosed" cannot be rebutted just because no diabetes-specific diagnosis was recorded [26].

Further, it may sometimes be beneficial to provide a "fuzzy" alternative to strict condition checking. As an example, if condition "clinically-relevant disease in the last two weeks" is checked in a strict manner, a relevant disease recorded 15 days ago would

be ignored, even though it would probably be of interest for the trial investigator. Further, an explicit specification of a condition may sometimes be hard. As an example, a comprehensive formulation of "clinically-relevant disease" would require the listing of numerous ICD codes and would still bear the risk of missing one. Therefore, we included optional "information needs" in our data structure, in addition to criteria. In the present example, this would allow to query all recorded diseases of a person from e.g. the past three weeks and let the trial investigator decide whether they are clinically and temporally relevant.

2.4. Testing methodology

We plan to test our tool with two currently running trials of the MedUni. The trials have already been set up in our database. From those eligibility criteria that can be checked via ELGA data, we have translated about 90% to XPaths yet. We aim to analyze the practical usefulness of our recruitment tool by "re-evaluating" patients, who were already recruited for one of the two trials with conventional methods. These patients will serve as gold standard for eligible individuals. In order to measure the sensitivity of our recruitment tool, we will apply it on the ELGA documents of these patients and examine to what degree it indicates eligibility. As the trials have disjunctive study populations according to their inclusion/exclusion criteria, patients recruited for trial 1 can serve as gold standard for ineligible individuals for trial 2 and vice versa. This will allow us to analyze the specificity of our tool.

Only patients who provide written informed consent will be included in our analysis. Their ELGA documents will be downloaded in the Vienna General Hospital (AKH) information system and pseudonymized before being used in our project. The planned procedure was confirmed by the MedUni data protection officer to comply with the relevant legal regulations and received a positive vote by the MedUni ethics committee.

Currently we are working on the organizational steps of implementing an interface between the AKH hospital information system and a MedUni research platform, where the ELGA documents of trial patients will be stored for our project. As this rather bureaucratic procedure will not be completed within the limited timeframe of the master thesis presented here, we decided to focus in the thesis on the optimization of our recruitment tool's usability using test data.

Originating from the public ELGA test documents we synthesized fictive but valid documents holding data that are relevant for the eligibility criteria of our two trials. We composed our test documents in a way that should exhaust all visualization variants offered by our tool's user interface.

3. Results

As the result of the context analysis, a clear picture of the planned application environment of our recruitment tool within the MedUni clinics was achieved. In particular, the typical current recruitment procedure of trial patients was analyzed to prepare a suitable future integration of our tool.

3.1. Requirements to be covered

As the result of the requirements analysis, the following system functions were found to be essential:

- *Set up trial:* A new trial is set up in the system's database including the required trial metadata and the XPaths for the eligibility criteria and the information needs. This step is done by a technician in cooperation with the corresponding trial investigator.
- *View trial:* The clinical end user can look up all metadata of a selected trial. Further, results of selected patients from earlier eligibility checks for the trial are displayed.
- *Select patients to be checked for eligibility***:** The clinical end user selects the patients, who should be checked for trial eligibility. The patients' ELGA documents can either be requested and downloaded from an IHE XDS environment or they can be accessed at a local folder (where they might have been exported from a local EHR system).
- *View summary of check results***:** An intuitive overview of the eligibility check's results for all selected patients is presented to the clinical end user.
- *Inspect results in detail:* The results can be examined in detail for each patient. Hereby, a patient's results can be expanded to see which criteria are satisfied / violated or for which criteria no corresponding data are available in the patient's ELGA documents. Each criterion can further be expanded to show the results of the individual underlying conditions. For each satisfied / violated condition, the corresponding ELGA source data can be viewed directly within the embedding ELGA document.
- *Mark potential trial participants:* After going through the results of the eligibility checks, the clinical end user can mark particularly promising patients. They are saved in a shortlist that may be used later for establishing contact with the patients.

3.2. User interface design

The intuitive visualization of the results of an eligibility check proved to be a particular challenge. In particular, a balance had to be found between providing a coarse overview of the check's results for the complete selected patient cohort at a glance, and at the same time allowing a stepwise visual drilldown into each single patient to explore the respective ELGA data constellations that led to the satisfaction / violation of the individual eligibility criteria and the underlying conditions.

Figure 2 shows a screenshot of the results of checking the eligibility criteria of a diabetes-specific trial for a cohort of five fictive patients. In the header section the key trial metadata are displayed together with the trial's total number of eligibility criteria and the number of criteria for which ELGA covers the required data.

Beneath the header, the outcomes of the eligibility checks are separately displayed for inclusion and exclusion criteria. At the coarsest level, only the numbers of satisfied/violated/undecidable (due to missing data) criteria are shown for each patient. Patients are sorted according to the number of satisfied inclusion criteria. If one or more exclusion criteria are satisfied, the corresponding patients are moved to the bottom of the cohort and marked in red background color.

In order to comprehend the results of a particular patient, the corresponding row may be expanded to show the results of the individual criteria. Here, a traffic light color scheme is used to visualize whether a patient "has passed" a criterion (green dot), "has failed" on a criterion (red dot), or whether the criterion is undecidable due to missing data (gray dot). In this sense of "passing" a criterion, a satisfied *inclusion* criterion is shown as green dot, whereas a satisfied *exclusion* criterion is shown as red dot. If contradictory data are found for a patient (e.g., two blood glucose measurements within the period of interest, one above the criterion's threshold and one below), the criterion is displayed as a yellow dot.

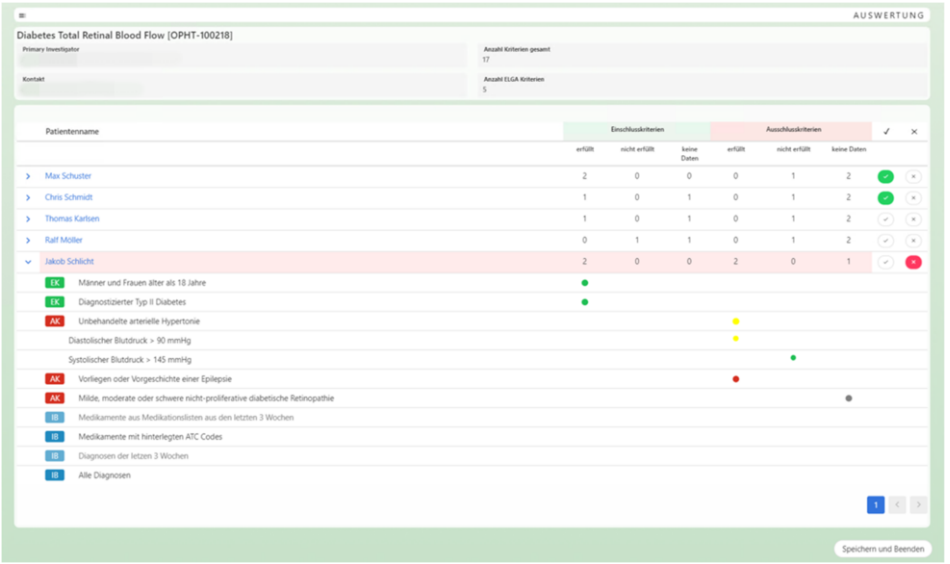

Figure 2 – Results screen of an eligibility check

If data was found for a criterion, the corresponding row may be further expanded to display the underlying conditions. The same color coding is used here to visualize the results. A condition with results can be clicked to open a pop-up window (Figure 3) that holds all source data found for the current patient that are relevant for the condition. A click on a particular value opens the embedding ELGA document and scrolls to the value within the document.

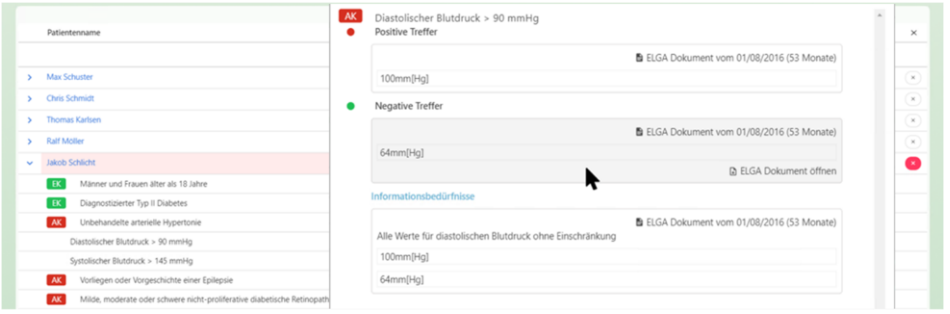

Figure 3 – For each condition the relevant ELGA source data can be retrieved

The checkbox in the right-most column can be used to manually mark those patients who were found to be most promising for a potential recruitment. The complete results of the executed eligibility check from selected patients can be saved with the trial for later processing.

4. Discussion

The present work follows a long tradition of implementing supportive tools for trial management at the Medical University of Vienna [27]. It extends these and other comparable activities for EHR-based trial recruitment [9], [28] by utilizing data of a national Shared EHR system.

We apply a rule-based approach for representing eligibility criteria and identifying suitable patients for a trial. While this is a rather straight-forward approach and knowledge bases exist that alleviate implementing these rules [29], achieving high sensitivity and specificity rates may require time-intensive fine-tuning. An alternative could be the identification of eligible patients by means of machine-learning approaches [30].

Our work is subject to several limitations. A system-immanent restriction is the fact that data available within ELGA will typically only allow a subset of a trial's eligibility criteria to be checked. However, even such kind of pre-filtering would allow the trial investigators to focus on the most promising candidates and thus in many cases entail a significant reduction of efforts in the recruitment process. We currently do not support natural language processing to locate data relevant for a trial's eligibility criteria within unstructured sections of ELGA documents.

The final goal of our project is to examine to what extent we can support patient recruitment for clinical trials at the MedUni by means of ELGA data. The master thesis presented here delivers an essential building block in this endeavor by developing a conceptual design and implementation of the underlying IT tool. The primary focus hereby was to achieve a high level of usability for clinical end users. Using the synthesized test documents, our tool was presented to the primary investigators of the two trials to gather feedback on its usability. Their preliminary feedback seems to indicate that we are on the right track in this regard.

References

[1] E. Hariton and J. J. Locascio, 'Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials', *BJOG: Int J Obstet Gy*, vol. 125, no. 13, pp. 1716–1716, Dec. 2018, doi: 10.1111/1471-0528.15199.

[2] C. A. Umscheid, D. J. Margolis, and C. E. Grossman, 'Key Concepts of Clinical Trials: A Narrative Review', *Postgrad Med*, vol. 123, no. 5, pp. 194–204, Sep. 2011, doi: 10.3810/pgm.2011.09.2475.

[3] G. Kolata, 'A Cancer Conundrum: Too Many Drug Trials, Too Few Patients (Published 2017)', *The New York Times*, Aug. 12, 2017.

[4] C. Halter, 'Liver Disease Researchers Struggle to Find Patients for Clinical Trials', *Hep*, Jan. 04, 2018. https://www.hepmag.com/article/liver-disease-researchers-struggle-find-patients-clinical-trials (accessed Oct. 14, 2020).

[5] K. Lopienski, 'Why do Recruitment Efforts Fail to Enroll Enough Participants?', *Forte*, Nov. 07, 2017. https://forteresearch.com/news/recruitment-efforts-fail-enroll-enough-patients/ (accessed Oct. 14, 2020).

[6] S. Feller, 'One in four cancer trials fails to enroll enough participants', *UPI*, Dec. 30, 2015. https://www.upi.com/Health_News/2015/12/30/One-in-four-cancer-trials-fails-to-enroll-enough-participants/2611451485504/ (accessed Feb. 21, 2020).

[7] A. M. McDonald *et al.*, 'What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies', *Trials*, vol. 7, no. 1, p. 9, Dec. 2006, doi: 10.1186/1745-6215-7-9.
[8] J. D. McAnulty, 'Bringing Patient Recruitment into Our Digital World', *Ther Innov Regul Sci*, vol. 43, no. 4, pp. 501–508, Jul. 2009, doi: 10.1177/009286150904300416.
[9] Y. S. Lai and J. D. Afseth, 'A review of the impact of utilising electronic medical records for clinical research recruitment', *Clinical Trials*, vol. 16, no. 2, pp. 194–203, Apr. 2019, doi: 10.1177/1740774519829709.
[10] S. Hussain, D. Ouagne, E. Sadou, T. Dart, and M.-C. Jaulent, 'EHR4CR: A semantic web based interoperability approach for reusing electronic healthcare records in protocol feasibility studies', *CEUR Workshop Proceedings*, p. 6, Jan. 2012.
[11] G. Augustinov and G. Duftschmid, 'Can the Austrian Nation-Wide EHR System Support the Recruitment of Trial Patients?', *Studies in Health Technology and Informatics*, pp. 87–90, 2019, doi: 10.3233/978-1-61499-961-4-87.
[12] B. Kankova and G. Duftschmid, 'Reusing Data of an EU-Wide EHR System for Clinical Trials: A Capabilities Analysis', *Stud Health Technol Inform*, vol. 271, pp. 17–22, Jun. 2020, doi: 10.3233/SHTI200069.
[13] S. Herbek *et al.*, 'The Electronic Health Record in Austria: a strong network between health care and patients', *Eur Surg*, vol. 44, no. 3, pp. 155–163, Jun. 2012, doi: 10.1007/s10353-012-0092-9.
[14] J. Doods, F. Botteri, M. Dugas, and F. Fritz, 'A European inventory of common electronic health record data elements for clinical trial feasibility', *Trials*, vol. 15, no. 1, p. 18, 2014, doi: 10.1186/1745-6215-15-18.
[15] A. Abran and J. W. Moore, *Guide to the Software Engineering Body of Knowledge*. IEEE Computer Society, 2004.
[16] U. Hammerschall and G. H. Beneken, *Software Requirements*. Pearson, 2013.
[17] 'Stack Overflow Developer Survey 2020', *Stack Overflow*. https://insights.stackoverflow.com/survey/2020/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2020 (accessed Jan. 22, 2021).
[18] R. Müller, *RecruitmentTool*. 2021.
[19] 'Vue.js'. https://vuejs.org/ (accessed Jan. 21, 2021).
[20] 'Electron'. https://www.electronjs.org/ (accessed Nov. 03, 2020).
[21] Django Software Foundation, 'The Web framework for perfectionists with deadlines | Django'. https://www.djangoproject.com/ (accessed Jan. 21, 2021).
[22] medshare GmbH, 'eHealth Connector Konzept zur Implementierung und API Spezifikation', Nov. 2013. [Online]. Available: https://www.yumpu.com/de/document/view/51786745/konzept-zur-implementierung-und-api-spezifikation-medshare/26.
[23] Open eHealth Foundation, 'IPF Open eHealth Integration Platform', *IPF Open eHealth Integration Platform*. https://oehf.github.io/ipf-docs/ (accessed Nov. 03, 2020).
[24] 'SQLite Home Page'. https://www.sqlite.org/index.html (accessed Jan. 21, 2021).
[25] S. Ott, 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*, pp. 226–233, 2019, doi: 10.3233/978-1-61499-971-3-226.
[26] O. Dameron, P. Besana, O. Zekri, A. Bourdé, A. Burgun, and M. Cuggia, 'OWL model of clinical trial eligibility criteria compatible with partially-known information', *J Biomed Sem*, vol. 4, no. 1, p. 17, 2013, doi: 10.1186/2041-1480-4-17.
[27] G. Duftschmid, W. Gall, E. Eigenbauer, and W. Dorda, 'Management of data from clinical trials using the ArchiMed system', *Med Inform Internet Med*, vol. 27, no. 2, pp. 85–98, Jun. 2002, doi: 10.1080/1463923021000014158.
[28] M. Cuggia, P. Besana, and D. Glasspool, 'Comparing semi-automatic systems for recruitment of patients to clinical trials', *International Journal of Medical Informatics*, vol. 80, no. 6, pp. 371–388, Jun. 2011.
[29] 'PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability - PubMed'. https://pubmed.ncbi.nlm.nih.gov/27026615/ (accessed Jan. 22, 2021).
[30] J. M. Banda, Y. Halpern, D. Sontag, and N. H. Shah, 'Electronic phenotyping with APHRODITE and the Observational Health Sciences and Informatics (OHDSI) data network', *AMIA Jt Summits Transl Sci Proc*, vol. 2017, pp. 48–57, 2017.