APERITIF – Automatic Patient Recruiting for Clinical Trials Based on HL7 FHIR

Alexandra BANACH[a,1], Hannes ULRICH[a], Björn KROLL[a], Alexander KIEL[b], Josef INGENERF[a,c], and Ann-Kristin KOCK-SCHOPPENHAUER[a]

[a]*IT Center for Clinical Research, Lübeck, University of Lübeck, Germany*
[b]*LIFE – Leipzig Research Centre for Civilization Diseases, University of Leipzig, Germany*
[c]*Institute of Medical Informatics, University of Lübeck, Germany*

Abstract. Clinical trials are carried out to prove the safety and effectiveness of new interventions and therapies. As diseases and their causes continue to become more specific, so do inclusion and exclusion criteria for trials. Patient recruitment has always been a challenge, but with medical progress, it becomes increasingly difficult to achieve the necessary number of cases. In Germany, the Medical Informatics Initiative is planning to use the central application and registration office to conduct feasibility analyses at an early stage and thus to identify suitable project partners. This approach aims to technically adapt/integrate the envisioned infrastructure in such a way that it can be used for trial case number estimation for the planning of multicenter clinical trials. We have developed a fully automated solution called APERITIF that can identify the number of eligible patients based on free-text eligibility criteria, taking into account the MII core data set and based on the FHIR standard. The evaluation showed a precision of 62.64 % for inclusion criteria and a precision of 66.45 % for exclusion criteria.

Keywords. FHIR, CQL, clinical trials, patient recruitment, NLP

1. Introduction

Clinical trials are essential for verifying the safety and effectiveness of new therapies and pharmaceuticals. A sufficient number of suitable subjects is critical for the conduct of a study [1]. However, only a minority of clinical studies achieve their recruitment goal in the scheduled time [2]. Finding suitable participants who meet the increasingly specific eligibility criteria of a study is time-consuming and personnel-intensive for the local study team. Patient recruitment, especially in research on rare diseases, where the number of patients is limited, requires multi-center clinical trials to achieve the necessary case numbers. Identifying project partners providing suitable patients is an arduous initial task. The Medical Informatics Initiative (MII) develops the German *central application and registration office* (ZARS), a platform to perform feasibility queries and data retrievals. This platform will improve the patient recruitment process due to the eased identification of potential research collaborations by performing feasibility queries over all MII sites. The inquiries and the resulting data exchange are based on harmonized and consented

---

[1] Corresponding Author: Alexandra Banach, IT Center for Clinical Research, Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; E-mail: a.banach@uni-luebeck.de.

core data sets, the *MII Kerndatensätze* [3]. We propose APERITIF [4], an approach to streamline the design of multi-center trials based on the ZARS.

2. Methods

2.1. Core Data Set of MII

The project's main focus is to establish data integration centers at the local university hospitals in Germany to enable a federated data exchange among these centers. Therefore, the core data set of the MII was designed [3]. The whole data set contains several basic modules: Diagnoses, Procedures, Medication, Encounter, Observations, Demographics, and Person. Each module includes commonly required data elements for clinical care based on consented FHIR profiles.

2.2. FHIR, Ontoserver, and CQL

Digital healthcare requires standards such as Fast Healthcare Interoperability Resources (FHIR) to ensure interoperability. The basic idea of FHIR is to model concepts from the clinical domain with the goal to facilitate interoperability. Supporting FHIR, the Ontoserver provides services to integrate terminologies and code systems into digital healthcare solutions, e.g. performing code requests for medical concepts [5]. Using the medical terminology SNOMED CT, the Ontoserver is also able to process Expression Constraint Language (ECL) to request information about subconcepts, etc. As a query language for clinical contexts, Clinical Quality Language (CQL) [6] supports several data models including FHIR and provides a wide range of elements such as the integration of code systems, value sets, or complex data types such as codes [6].

3. Results

3.1. Data Analysis

In a first step, we analyzed 552 eligibility criteria from 27 randomly chosen studies. The identified concepts were assigned to different classes and groups oriented to the Semantic Network of Unified Medical Language System (UMLS), utilized by Luo et al. [7]. The most commonly classes that should be considered for implementation were "Disease, Symptom or Sign", "Therapy or Surgery", "Medication", "Diagnostic or Lab Result", "Age", and "Gender". Their proportion was about 84 % of all identified concepts. In the next step, the aim was to map corresponding data elements within the core data set of the MII. The core data set was able to cover about 99 % of our identified concepts. Missing data elements could be detected for concepts such as "Life Expectancy" or "Organ or Tissue Status".

3.2. Implementation of APERITIF

Since the developed framework for implementing automated case number estimation based on the MII ZARS is intended to achieve a maximum degree of automation, the

study criteria from national and international study registries are used. For studies from ClinicalTrials.gov, we implemented data acquisition via Representational State Transfer (REST) as well as importing data from a downloaded Extensible Markup Language (XML) file. All other study information has to be imported from a text file containing free-text eligibility criteria of a study in the English language.
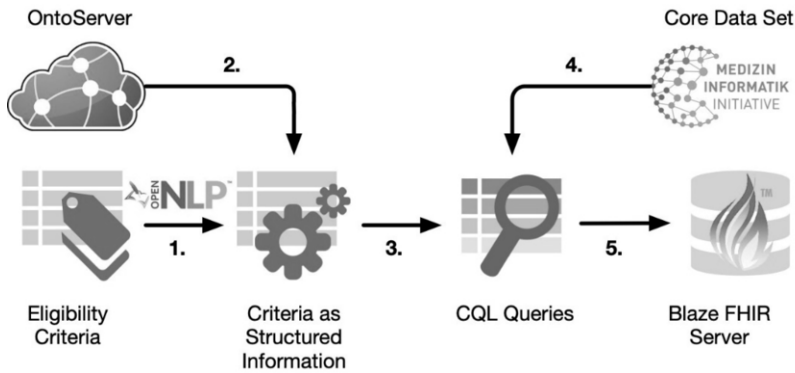
Figure 1. Schematic overview of APERITIF's architecture showing input and output formats.

Figure 1 illustrates the architecture of the program APERITIF. In a first step, the imported free-text eligibility criteria of clinical trials have to be converted into structured data using natural language processing (NLP) methods such as tokenization or, lemmatization provided by Apache OpenNLP [8]. We also integrated MetaMapLite (version 3.6.2rc5) [9], a UMLS based entity recognition including negation detection. Therefore, entities such as diagnoses, procedures, laboratory results, and medications can be categorized by their UMLS semantic types. In a second step, we collect the terms of the identified entities to perform a term search for SNOMED CT and Logical Observation Identifiers Names and Codes (LOINC) codes with the Ontoserver. Besides, subconcepts of single SNOMED CT concepts are extracted using ECL. For each criterion, the junctions "and" and "or" are extracted via regular expressions to connect several entities of a criterion in the query. For laboratory results, we also extract the laboratory value, the laboratory unit, and the comparator using regular expressions. The demographic data gender, minimum, and maximum age of a patient are extracted the same way. As the core data set has a FHIR specification, we chose CQL as query language supporting this standard and to overcome the limitations of FHIR search [10, 11]. With the help of the extracted information (step 3) and the corresponding data elements of the core data set (step 4), the CQL query is generated and finally sent to a Blaze instance of version 0.9.0 [12] (step 5), a FHIR server that can process CQL queries. The whole implementation is based on FHIR R4 and HAPI version 4.2.0 [13].

3.3. Qualitative Evaluation

We assessed 20 studies with 4 study centers of the campus Lübeck in Germany. All extracted information of a CQL query was translated in a comprehensible form (negations, identified entities, junctions of a criterion). The study nurses labeled the extracted criteria and the identified entity types (e.g. "diagnosis" for gastric cancer) as

"correct" or "false". 91 of 167 inclusion criteria contained medical information that was relevant for query generation. 57 of these 91 inclusion criteria were labeled as "correct" (62.64 %). 152 exclusion criteria contained medical concepts (total 261 exclusion criteria overall studies). 101 were labeled as "correct" (66.45 %). Major faults were underspecified entities, e.g. "excision" instead of "partial mesorectal excision" (31.46 %), wrong contexts, e.g. patients with MRI but diagnosis in this context was missing (22.48 %), and incorrect junctions, e.g. "and" instead of "or" (15.73 %). The extracted demographic data were correct for all tested studies. The identification of entity types resulted in a precision of 89.00 % for inclusion and 93.17 % for exclusion criteria.

3.4. Quantitative Evaluation

Adequate test data for the core data set is not yet available. Therefore, we manually generated test data for 20 studies to perform a quantitative evaluation of APERITIF. For each test patient of a study, a JavaScript Object Notation (JSON) file containing the required FHIR resources such as conditions or procedures was generated and uploaded using blazectl, a command line tool for Blaze. For each test study, at least one patient fulfilled the required information of the CQL query and one did not to check whether junctions and negations work correctly. Ten studies passed our instance data test scenario. Four queries resulted in a parsing exception and six queries contained contradictions.

4. Discussion

The qualitative evaluation showed that exclusion criteria performed better than inclusion criteria, which often contain more complex elements. During implementation and evaluation, MetaMapLite showed performance issues with complex criteria. The next step is to change the entity recognition to methods of the emerging field of machine learning [14]. One major impediment preventing the use of pre-trained networks is the missing support of UMLS, which is needed for the query generation in our approach. The new methods could allow the detection of temporal interrelations, which cannot be identified using MetaMapLite. Due to time constraints, the evaluation of missing information and extracted codes was out of the scope of this work.

The quantitative evaluation showed contradictions within the queries, which resulted in failing the expected test outcome for six studies. The reasons were errors in the automatic negation detection in MetaMapLite. To resolve the contradictions, the exclusion and inclusion criteria were compared and possible duplications were dropped from the query. This modification is less restrictive than removing codes from inclusion criteria since eligible patients would be excluded from recruitment. After the cropping step, only one contradiction remained for a study between entities of inclusion criteria. Upon queries with a high number of entities and corresponding codes, technical issues of the used FHIR store were encountered, especially by processing laboratory criteria. At this time, a different formulation of laboratory entities is technically not possible. It would require the implementation of further features in Blaze. Unfortunately, it was necessary to reduce the number of laboratory statements per entity to work around these issues, albeit information gets lost.

5. Conclusion

Our approach extracts medical and demographic data from free-text eligibility criteria of a study using NLP methods. The extracted information is mapped to the corresponding data elements of the core data set of the MII. Based on the processed information, a CQL-query is generated to represent the criteria. Based on free-text eligibility criteria, APERITIF enables estimating the number of potentially eligible patients for planning multi-center trials. With the chosen tools such as the Ontoserver for code extraction and CQL in combination with the Blaze server, we successfully implemented a promising application that can flexibly be adapted to possible central MII developments.

Acknowledgement

References

[1] Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. Fundamentals of Clinical Trials. Springer; 2015.

[2] McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et. al. What influences recruitment to randomized controlled trials? A Review of Trials Funded by Two UK Funding Agencies. Trials. 2006;7(1):9.

[3] Ammon D, Bietenbeck A, Boeker M, Ganslandt T, Heckmann S, Heitmann K, et al. Der Kerndatensatz Der Medizininformatik-Initiative: Interoperable Spezifikation Am Beispiel Der Laborbefunde Mittels LOINC Und FHIR. mdi: Forum der Medizin Dokumentation und Medizin Informatik. 2019;4(Mdi):113-117.

[4] ITCR-L, Uni Lübeck; 2021, Available at https://github.com/itcr-uni-luebeck/APERITIF, Accessed March 05, 2021.

[5] Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. Journal of biomedical semantics. 2018;9(1):24.

[6] International H. Clinical Quality Language (CQL); 2020, Available at https://cql.hl7.org/, Accessed Nov 11, 2020.

[7] Luo Z, Miotto R, Weng C. A Human-Computer Collaborative Approach to Identifying Common Data Elements in Clinical Trial Eligibility Criteria. Journal of biomedical informatics. 2013;46(1):33-39.

[8] Apache software foundation. Welcome to Apache OpenNLP. Available at https://opennlp.apache.org/, Accessed Nov 11,2020.

[9] Demner-Fushman D, Rogers WJ, Aronson AR. MetaMap Lite: an evaluation of a new Java implementation of MetaMap. Journal of the American Medical Informatics Association. 2017;24(4):841-844.

[10] Gulden C, Mate S, Prokosch HU, Kraus S. Investigating the Capabilities of FHIR Search for Clinical Trial Phenotyping. Studies in health technologies and informatics. 2018; 253:3.

[11] Kroll B, Schüttler C, Ulrich H, Kiel A, Deppenwiese N, Engels C, et al. Searching in FHIR® models-Researchers' expectations vs reality. Brussels, nov, 2020.

[12] Kiel A. Blaze Documentation; 2020. Available at https://github.com/samply/blaze, Accessed Nov 11, 2020.

[13] HAPI FHIR. A Free and Open Source Global Good:l Powering Interoperability Around the Word for 18 Years, 2020. Available at https://hapifhir.io/, Accessed Jan 01, 2021.

[14] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: A Pre-Trained Biomedical Language Representation Model for Biomedical Text Mining. Bioinformatics. 2020;36(4):1234-1240.