Distributed Medical Rule Engine (DMRE)-Project

Doctoral Consortium

Gerhard Kober

Tiani "Spirit" GmbH, Austria
gerhard.kober@tiani-spirit.com

Abstract. In medical-clinical informatics supporting medical research and medical routine is a core business. Through growing medical knowledge, medical doctors need to track all this new knowledge and activities. To support this high flow on specialized new information an IT-solution could be helpful. This solution takes care of the patient's data, as well as on existing and emerging medical knowledge and various medical processes. The main issue is combining these three entities to answer medical questions for a specific patient for a certain use-case. The intention is to find a generic methodology for the combination by using semantic technologies, FHIR as medical-data-standard and ontology-mappings.

Keywords: FHIR · medical guidelines · ontology · medical data · semantic technologies

1 Introduction

Medical-clinical research and the associated data analysis are of immanent importance in improving and ensuring patient care [12]. Out of this research, medical guidelines are developed, which describe a standardized, structured procedure for assessing and treating a patient within a special clinical picture. These guidelines are considered as necessary decision-making aids, ensuring a high quality of care. This leads to a general common understanding for every physician how to provide healthcare to a patient for a specific medical issue.

From a more generic point of view such a guideline consists of the entities *processes*, *medical knowledge* and *data*. The process describes, in a step-by-step approach which actions must be set. Knowledge is represented in the form of ontologies. Data as the third component is given for each patient. In further processes, decisions are strongly based on every single data-element. These patient-related entities interact with each other to facilitate decision-making.

The over all topic is on linking medical guidelines, medical data and, medical knowledge. The main problem is within combining these entities in a formal way, as well from a technical point of view. Currently, there is no standardized, formal way to achieve interactions in between *processes*, *medical knowledge* and *data*, to achieve the purpose of decision support.

2 Background

2.1 Data-Definition

Fast Healthcare Interoperability Resources (FHIR): Fast Healthcare Interoperability Resources (FHIR) is a standard framework defined by Health Level 7 (HL7). The core of FHIR contains so called "resources" and "APIs" as primary components [5]. Resources are information models defining data elements including constraints and relationships in between. This means FHIR describes the data-structure, the data-flow and the connection between different FHIR-resources. The aim of the FHIR Standard is [7] to enable the exchange of healthcare-related information, by reusing existing standards as HL7-Version 2 and Clinical Document Architecture (CDA). There is a focus on implementing fast and easy solutions, but also taking web standards like XML, JSON, HTTP, OAuth, etc. into account [9]. One more goal of this standard is to ensure the interoperability. There exists also the option to represent FHIR-resources as RDF-graphs [8]. This option is mainly used to bridge information between the data which is generated during operations and formal knowledge processing systems.

2.2 Problem in the field

Medical workflows are getting more relevant through medical studies and standardizing medical processes [1]. There is a variety of medical guidelines, which are interesting for the different medical disciplines and medical application areas. For example, there is a guideline for chronic cardiac insufficiency [3] which is a common disease for persons older than 60 years. The presented medical data exists in a standardized form (FHIR), but without any relation to one another. As an example: Elderly patients are taking medication to lower their pulse-rates. In the technical representation of the FHIR-resources, this link between the pulse-rate and the influence of the medication is missing. Medical knowledge is available in the form of different ontologies. Currently, the combination of medical guidelines, medical data, and the associated ontology is not available. Even the semantics of the different medical resources stored as FHIR-observations to each other is missing. Representing the different relationships is of high importance for making decisions. Furthermore, the issue of highly distributed medical-data-stores is given. This means medical data from one single patient can be stored either in the same physical location, but also in multiple different locations. A problem is the generation of a complete view of the patient's data and all relations between the data. One more issue regarding FHIR is, that the FHIR-data-model is not intended for semantic queries since it is built upon records, rather than stating facts (e.g. "The Patient A has a heart attack")[8].

In a more generic way, the problem-field is within the combination of process-knowledge, ontologies, and a certain sort of data. The issues which arise during the combination are about the integration of data and ontologies to find the semantics of the single data, but also to find all available data for a given representation in the ontology. The medical guideline needs to be applied to a concrete data-set for a patient, but also the ontology for a step in the guideline is relevant. For example, finding the blood-pressure for the patient. The rule-engine first needs to retrieve the "coding" for blood-pressure in the Ontology. One more problem is the question, which guideline to apply for a given set of data (e.g. which guideline take care of high blood pressure?) Also the question – "has the patient a medical issue" should be answered, related to the given ontology, the known medical guidelines, and the patient's data in order to react properly on common but also on rare diseases.

3 Related Work

This section describes work, done by others which are related to the field of the theses, and to find open tasks.

The work 'Structurally Mapping healthcare data to HL7-FHIR through ontology alignment' [11] analyses the possibility of a transformation of any healthcare data. Since the argumentation is about different healthcare-formats the incoming data needs to be analyzed in order to transform via an ontology-alignment to FHIR-resources in terms of structure. There are referenced a knowledge base to take care of the ontologies and relationships, and a structure mapping library. The structure translator provides a mechanism for setting the correct FHIR-format and put the values in the given FHIR-version-3 form. There is also a way to get a complete overview of the resources and the relationship among each. This solution also stores the combined information in a triple-store, but has a flip side when thinking on interoperability (since FHIR is meant to exchange observations between different healthcare-actors), and creating views "on the fly", with already existing data. Even a change of patient-information might cause issues in terms of the need to change the complete patient's information in every stored entity. There is also a need to find the ontology within an FHIR-resource, but this is not done yet.

In [14] the approach for modelling validating FHIR-Profiles is using semantic web shape expressions (ShEx). For this purpose, the FHIR-resources, which are represented either in JSON-format or XML-format are transformed. The transformation which needs to be done is from FHIR-JSON to RDF. This is done using an implementation by the HL7-Community. Somehow this function is not present anymore, there is an implementation missing. In [14] the conversion of a single resource mentioned, but not the combination of multiple resources. Such a conversation will be needed in a productive environment, where a combination of different resources is done to generate an RDF-graph over the resources. In some tests done with an old version, it occurred that the transformer is not able to convert resource-bundles, but just single FHIR-resources.

The idea in [4] is to take healthcare data (in different formats like HL7-V2-Message, or FHIR-resources) as input, encode it to RDF and store it into a triplestore database. The described query-stage then returns the matching results. Therefore SPARQL is being used. This engine acts simultaneously as a collector and a bus for delivering messages to sources, as well as a real-time analytic platform. This means the 'semantic enrichment machine' takes the FHIR-resources, transforms it to RDF, and stores them. This has the issue that existing FHIR-Data can not be used for further transactions because it is not available in the triple-store.

A medical guideline can technically be represented in Guideline Interchange Format version 3 (GLIF3) [13]. There are dependencies on the HL7-Arden-Syntax [2] and Medical Logic Modules (MLM). GLIF3 models a clinical guideline as a flowchart. The goal of GLIF3 was to find a representation format for sharable computer-interpretable clinical practice guidelines. The Arden syntax itself is a knowledge representation standard for medical knowledge and allows queries to existing data. The definitions on the MLMs are strict, the formulation on the 'rules' need to be exact. This means it takes an action on each input (e.g. if heart-rate is higher than 100 bpm (beats per minute), alert the doctor), and it does not build up an information-chain for taking decisions. In this case a 'longer' time period for evaluation would be helpful, to avoid not needed alerts. Even every single 'medical knowledge' needs to be built up by a person, and there is not a possibility to rely on existing information. Because of the high integration of Arden-Syntax to GLIF3, and a needed possibility to build semantic graphs over medical information and medical knowledge, GLIF3 is not a usable option.

To summarize these works: there are different transformers (from proprietary format to FHIR, from FHIR to RDF), and also the approach for storing in RDF or creating some FHIR-like-resources. These transformations are needed, to find a common 'language' for traversing an RDF-graph, for finding the needed information. It is also needed to have 'standard-FHIR-resources' (which are not enriched for a special use case) and to use them for better patient treatments. All found approaches are nonworking mechanisms for resolving ontologies within FHIR-resources. FHIR itself provides an ontology [6]. It defines the formal structure of the individual FHIR-resources, but not in the medical context.

4 Research questions and hypothesis

The over-all question is how to find the semantics of medical data, and how semantic technologies can be applied to medical data and medical guidelines and medical knowledge. Including the question of how to apply medical guidelines on withing the semantic context to answer medical questions during a medical assessment.

I assume that there is a generic methodology for automated processes in rule-engines to be combined with medical data. This allows answers to the questions which arise during the execution of a medical guideline.

More specific questions are about the different components:

1. Is there a system-architecture which can build up a solution to satisfy the needs of the over-all solution?
   The hypothesis is that there can be created an architecture which allows to operate a system to cover the needs for a rule-based medical decision support system.
2. How to transform medical guidelines (technically) in a way that a rule-engine can apply these guidelines and execute the different workflows?
   The expectation is that a rule-engine is the right tool to perform a medical guideline in general. It is also expected that it coordinates and orchestrates different tasks. Moreover, it can take decisions based on the input.
3. How to extend rule-engines in a way to deal with 1.) medical data and 2.) medical knowledge?
   The extension of a rule-engine is possible for the special case of medical applications, to handle this specific data also in a semantic context.
4. How are the relations between medical data/guidelines and medical knowledge described?
   The belief is that there is a semantic relationship among all these components (medical data, medical knowledge and medical guidelines), which need to be "merged" together, to reach the goal of semantic interoperability.
5. How to do semantic queries towards a medical-data-store (in this special case an FHIR-store)?
   Since there is an RDF specification for FHIR-resources, the premise is there is a technical option for the FHIR-JSON-to-RDF conversion.

5 Research Design and methods

As research methodology, Design Science is chosen, because it provides specific guidelines for the creation and evaluation of IT artifacts. The following guidelines should be implemented in the course of the work:[10]:

- Design as an Artifact
- Problem Relevance
- Design Evaluation
- Research Contributions
- Research Rigor
- Design as a search Process
- Communication of Research

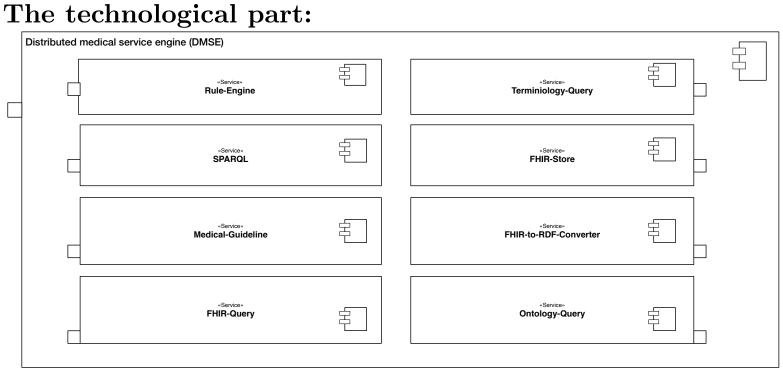
These guidelines will be used throughout the development of the thesis. An artifact will be developed, and the problem relevance will be described in the problem statement. The evaluation will be done in each single design artifact as well as in the overall solution. The overall evaluation will be done on a medical use case for a patient having a heart attack, expecting the solution to find that the patient has a heart attack, and providing the appropriate medical guideline to the doctor.

6 Proposed solution

The solution is partitioned in different parts which all will fit together in an over-all-solution for the problem.

The architectural part: from an architectural point of view it is a a centralized as well as a decentralized approach. The services can be applied on different servers, to do the needed calculations and transformations. Then it is possible to hold all information at the point of origin but just the needed information is then submitted to a central service, so that there is no need for huge data-transfers. Nevertheless the important (reduced) results are then centrally available and ready for processing. If there is a need for processing the data in the central environment the services there can also be used. There are different services needed, which can be executed during run-time. The dependencies on the services are loosely coupled, just the rule-engine itself, as the central workflow-component is a 'must' for the other services. For example, if there is no need for a FHIR-Query-Service, there is no need to activate it during the execution of the rule-engine's workflow. The rule-engine itself takes care of the over-all communication and coordination as well as the results from the different services. Decentralized services are specialized ones to act as an interface towards the different FHIR-Stores, and on the other hand, are they an interface for the central component. Since knowledge can be widely distributed a decentralized approach is needed to resolve all the available information for a use-case. The rule-engine as the core-service to evaluate the results and take decisions should be a centralized one.

The technological part:



As shown in the image above there are different services available on the solution to provide a certain function. The services are:

1. Rule-engine
2. FHIR-Query
3. SPARQL
4. Terminology-Query-Service
5. Medical-guideline-service
6. Ontology-query-service

7. FHIR-Store-Service
8. FHIR-to-RDF-Converter.

The services are able to interact with each other, provide information on requests, decide during the workflow, process incoming data or read medical guidelines.

1. The Rule-Engine: the rule-engine is the central component in the system architecture. It covers the main workflow during different medical processes like inserting new Observations, but also the medical guideline. The rule-engine also takes care of the execution of the different medical guidelines. To fulfill this task the rule-engine has to strongly interact with the other services.
2. FHIR-Query-Service: the FHIR-Query service allows the solution to perform queries towards an FHIR-Store to provide the patient's relevant information to the rule-engine. The query can either be done by the standard Hl7-FHIR-Services or by an intermediate layer (the FHIR-to-RDF-Converter), to get results in RDF-TTL-Format.
3. FHIR-Store-Service: The FHIR-Store service, takes care of all information that should be stored into an FHIR-Store. This service decides which is the correct FHIR-Store, and due to the interception, this service can also enrich data with other relevant information (e.g. intake of medications and the effect on the pulse-rate).
4. SPARQL-Service: this service allows the rule-engine to ask SPARQL-queries towards the result of the FHIR-Query-Service or the FHIR-to-RDF- Converter. These services hold its data in RDF-Format, and so SPARQL-Queries can be asked.
5. Terminology-Query-Service: The Terminology-Query-Service is relevant for semantic interoperability. By using this service, it is possible to use consistent terminologies in the distributed environment. This applies for the storage-component as well as to the query-component [15].
6. Medical-guideline-service: this is a very central service since it needs to cover the execution of the guideline. This service holds information on the rules.
7. Medical-knowledge-query-service: This service returns results on medical questions. For example: what is the meaning of a certain value in a dataset? What does the result mean in the over-all-context?
8. FHIR-to-RDF-Converter: This service cares on the correct transformation from FHIR-JSON-Objects to RDF-TTL, to allow SPARQL-Queries on the result-set. This service converts FHIR-resource-bundles and single resources as well.

Workflows in the system architecture: in the distributed system the workflow is important. The workflow is defined by two things: 1.) a general path/decision-part for the execution-environment and 2.) by the rule-engine which takes content from the medical-guideline. Besides the rule-engine for the workflow described earlier, the general path describes the workflow through the services in different network-endpoints.

7 Conclusion

Concluding there is the idea of a system-architecture, which needs to be proved and implemented, as well as the technological part, to cover the needs for a support system, and to resolve the relevant information for the medical doctor as well as for the patient. The architecture describes a centralized and decentralized approach at the same time, by having certain services located in the edges of the needed (medical) network, but the core-components on a central place to make decisions and guide trough the designated workflow for a medical use-case.

References

1. A. Webb, L.G.: Oxford textbook of critical care. Oxford University Press (2016)
2. Arden Syntax v2.10 (Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10), https://www.hl7.org/implement/standards/product_brief.cfm?product_id=372, accessed: 2020-03-29
3. Programm für nationale Versorgungsleitlinien: Chronische Herzinsuffizienz, https://www.leitlinien.de/nvl/html/nvl-chronische-herzinsuffizienz/3-auflage/kapitel-1, accessed: 2020-03-29
4. Cotter, D., Bumgardner, V.K.C.: Semantic Enrichment of Streaming Healthcare Data (12 2019), http://arxiv.org/abs/1912.00423
5. HL7 FHIR: Architect's Introduction, https://www.hl7.org/fhir/overview-arch.html, accessed: 2020-03-29
6. FHIR OWL Ontology, https://w3c.github.io/hcls-fhir-rdf/spec/ontology.html, accessed: 2020-03-29
7. HL7 FHIR: Summary, https://www.hl7.org/fhir/overview-clinical.html, accessed: 2020-03-29
8. HL7 FHIR: RDF, https://www.hl7.org/fhir/rdf.html, accessed: 2020-03-29
9. HL7 FHIR: Summary, https://www.hl7.org/fhir/summary.html, accessed: 2020-03-29
10. Hevner, A., S.March, J.Park: Design Science in Information Systems Research. MIS Quarterly Vol 28 No. 1 (2004)
11. Kiourtis, A., Mavrogiorgou, A., Menychtas, A., Maglogiannis, I., Kyriazis, D.: Structurally Mapping Healthcare Data to HL7 FHIR through Ontology Alignment. Journal of Medical Systems **43**(3) (2019). https://doi.org/10.1007/s10916-019-1183-y
12. Ollenschlaeger, G.: Leitlinien und Evidenzbasierte Medizin in Deutschland. Springer - Zeitschrift fuer Gerontologie und Geriatrie (2000)
13. Peleg, M., Boxwala, A.A., Ogunyemi, O., Zeng, Q., Tu, S., Lacson, R., Bernstam, E., Ash, N., Mork, P., Ohno-Machado, L., Shortliffe, E.H., Greenes, R.A.: GLIF3: the evolution of a guideline representation format. Proceedings. AMIA Symposium pp. 645–9 (2000), http://www.ncbi.nlm.nih.gov/pubmed/11079963 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2243832
14. Solbrig, H.R., Prud'hommeaux, e.a.: Modeling and validating HL7 FHIR profiles using semantic web Shape Expressions (ShEx). Journal of Biomedical Informatics **67**, 90–100 (2017). https://doi.org/10.1016/j.jbi.2017.02.009, http://dx.doi.org/10.1016/j.jbi.2017.02.009
15. HL7 Austria: Terminologieserver, https://wiki.hl7.at/index.php?title=TS:Terminologieserver, accessed: 2020-03-29