doi:10.3233/SHTI190810

Semantically Annotated Metadata: Interconnecting Samply.MDR and MDM-Portal

Abishaa VENGADESWARAN[a,1] , Philipp NEUHAUS[b], Stefan HEGSELMANN[b], Holger STORF[a] and Dennis KADIOGLU[a]
[a] *Medical Informatics Group, University Hospital Frankfurt, Frankfurt am Main, Germany*
[b] *Institute of Medical Informatics, University of Münster, Münster, Germany*

Abstract. Interoperability is a growing demand in healthcare, caused by heterogeneous sources, which aggravate information transfer. The interoperability issues can be addressed by metadata repositories. These support to ensure syntactical interoperability, like compatible data formats or value ranges, however especially semantic interoperability is still challenging. Semantic annotation through standardized terminologies and classifications enables to foster semantic interoperability. This work aims to interconnect Samply.MDR and Portal of Medical Data Model (MDM-Portal) to allow facilitated semantic annotation with UMLS. Therefore, Samply.MDR was extended to store semantic information. While creating a data element, a request to MDM is send, which results in possible UMLS codes. The user can now adopt the most suitable code and select a link type between the code and the element itself. A successful enrichment of data elements with UMLS codes was shown by interconnecting Samply.MDR and MDM-Portal.

Keywords. Metadata, Interoperability, Semantic Annotation, Concept Association

1. Introduction

Nowadays shared data and shared benefits are keywords in the health sector, as large amounts of data are collected. However, these should be made more useful and meaningful. The Medical Informatics Initiative Germany funds four consortia by the German Federal Ministry for Education and Research (BMBF) in order to advance these problems [1]. The Medical Informatics in Research and Care in University Medicine (MIRACUM) consortium includes ten universities/university hospitals and one industrial partner as the largest of the consortia, in which the University Hospital Frankfurt is a partner site, and aims to share research and healthcare data within and especially across boundaries [2]. Sharing data comes along with many challenges. In hospitals, there are numerous information systems and therefore heterogeneous data sources, which makes exchanging data difficult and lavishly. The ability to transfer data is defined as interoperability [3]. Technical and syntactical interoperability includes

[1] Corresponding Author, Abishaa Vengadeswaran, Medical Informatics Group (MIG), University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; E-mail: vengadeswaran@med.uni-frankfurt.de.

transferring data and integrating services, like compatible interfaces, interconnection services or compatible data formats, whereas semantic interoperability deals with understanding the same meaning of the data. Because of nonuniform usage of words, there are many ambiguities and semantic heterogeneity [4], demanding continuously growing semantic interoperability [5].

Prerequisites for interoperability and exchanging data are capturing structured clinical data as well as using standardized terminologies to allow ontology-based data integration [6]. Today billing relevant documentation is already captured in a large structured manner using coding (ICD-10), whereas, in areas of treatment documentation, free text is often used to enter unstructured data.

A metadata repository supports the exchange and acquisition of structured data; however semantic annotation is still challenging. The Samply Metadata Repository (Samply.MDR), which is used in the MIRACUM consortium, lacks in supporting the user to reuse data elements and therefore avoid redundancies. This project aims firstly to extend Samply.MDR to enable semantic enrichment. Additionally, the precision of full text search for matching data elements can be improved by annotations.

2. State of the art

2.1. Samply Metadata Repository (Samply.MDR)

One key aspect to address interoperability issues is using metadata repositories (MDR). These could grant open access to metadata to increase interoperability. The Samply Metadata Repository (Samply.MDR) is an open source web application based on the standard ISO 11179-3. The development started in year 2011 to support sharing biomaterial collections and their annotation data. The development is still ongoing and it is already well established in numerous research projects, like Clinical Communication Platform (CCP-IT) of the German Cancer Consortium (DKTK) or Open Source Registry System for Rare Diseases (OSSE) [7]. The Samply.MDR supports to specify data elements in a standardized way. A data element is described by designation (name or label), definition (human readable description) and validation type (like data type, value ranges, etc.). For example, a data element in Samply.MDR for heart rate can be described as: designation = Heart Rate; definition = Heart Rate of a patient, measured by palpation on the wrist (radial pulse) and validation type = Integer.

2.2. Terminologies

Nonuniform usage of words results many ambiguities. A word can be used for different things (Homonym) or many words can be used for one thing (Synonym). Therefore, there is a semantic heterogeneity. Standard Terminologies addresses this issue and is fundamental for interoperability. However, for most health care providers using terminologies are unfamiliar and associated with additional work. Thus, user adoption in routine care is low. Moreover, there are more than one terminology, like Systematized Nomenclature of Medicine (SNOMED) [8] or Logical Observation Identifiers Names and Codes (LOINC) [9]. An important role in mapping between these terminologies plays the Unified Medical Language System (UMLS) [10], which contributes to the interoperability by identifying equivalent concepts across terminologies [6].

2.3. Portal of Medical Data Model (MDM)

The Portal of Medical Data Model (MDM-Portal) was developed at the Institute of Medical Informatics at the University of Münster. The portal aims to share data models in medicine for research and healthcare. A medical data model is considered as a description of data structures of information systems in medicine, like an electronic data capture form (eCRF) of a clinical trial, and is stored in a Operational Data Model (ODM) standard [11]. Moreover, the data elements are annotated with UMLS codes to address the issue of semantic heterogeneity.

3. Concept of interconnecting Samply.MDR with MDM-Portal

The Samply.MDR was extended to store concept associations for data elements. Therefore, the data model of MDR had to be extended. The Concept Association (including its attributes system, version, term, text), as specified in the currently actively developed standard ISO NP TS 21526 Healthcare Informatics – Metadata Repository Requirements, has been taken into account [12].

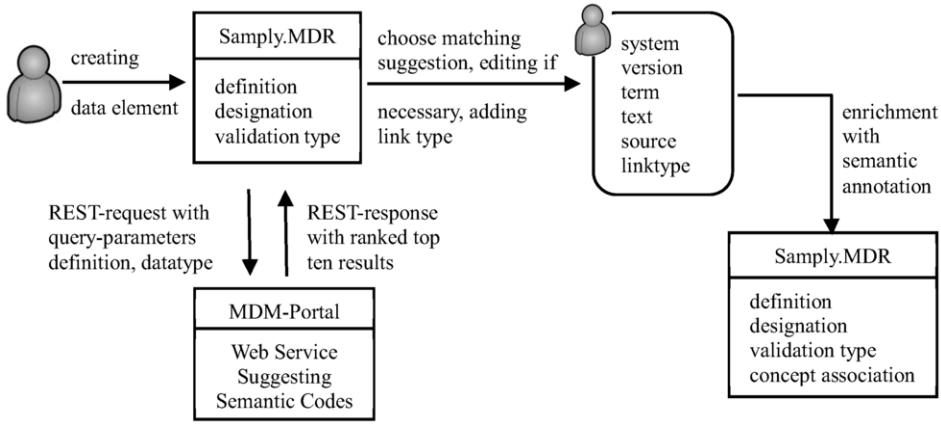

Figure 1. Conceptual process of semantically annotating data elements in Samply.MDR.

The process of creating a data element in Samply.MDR consists mainly of defining *definition*, *designation* and *validation type.* This is extended by one step for semantic enrichment by interconnecting the MDM-Portal,which is shown in Fig. 1. Hegselmann et al. developed a web service for suggesting semantic codes based on the MDM-Portal. It provides a REST-API and encodes the search result in JSON format. The output contains the top ten matching results, ranked by the number of occurrences in medical data models [13]. This web service for suggesting semantic codes of the MDM-Portal is used to find automatically matching UMLS codes through the given information (primarily designation). The matching code of the resulting response can be adopted. However, a 'link type' was attached to specify the link between a data element and the UMLS concept more precisely. For this purpose, the value set of ConceptMapEquivalence defined by the Fast Healthcare Interoperabilty Ressources

(FHIR) [14] was taken into account. Last but not least, to keep track of the provenance of the additional data retrieved from the MDM, 'source' was added to the list of attributes.

4. Implementation in Samply.MDR

First the backend of the Samply.MDR was expanded. A detailed overview of the added attributes can be seen in Table 1. The attributes system, version, term and text were adopted from ISO standard. For link type following codes can be selected: *equivalent*, *equal*, *wider*, *subsumes*, *narrower*, *specializes* and *inexact*. Codes like *unmatched* and *disjoint* from the FHIR value set were deliberately omitted since only suitable concept associations should be stored.

Table 1. Extended Samply.MDR attributes for the implementation of concept associations

Attribute	Datatype	Description
system	String	system referenced in the association
version	String	version or release of the system referenced, or the version of the term referenced at appropriate
term	String	the term identifier referenced by the association
text	String	the textual representation of the term
source	String	the source of the entry
linkType	EnumType	value of ConceptMapEquivalence for precise link type description between data element and concept

Within the UI, a wizard step was created, as shown in the following Fig. 2, which displays suggested concept codes (1). The user can select one of them, by pressing "Apply"-Button (2).

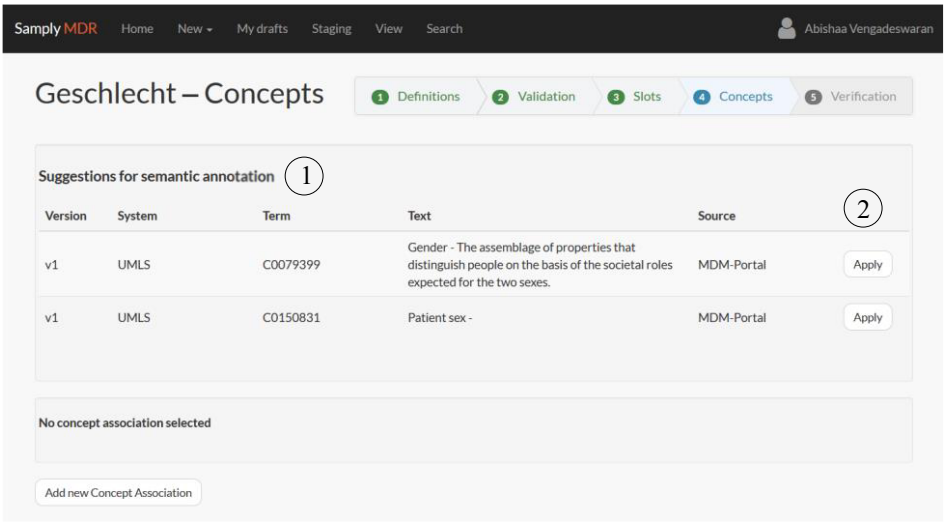

Figure 2. User interface of Concept Association in Samply.MDR – Suggestions of concept associations

Afterwards, the selected concept will be added to the list of "Selected semantic annotation",where a link type can be chosen from the given value set (3), as shown in Fig. 3.

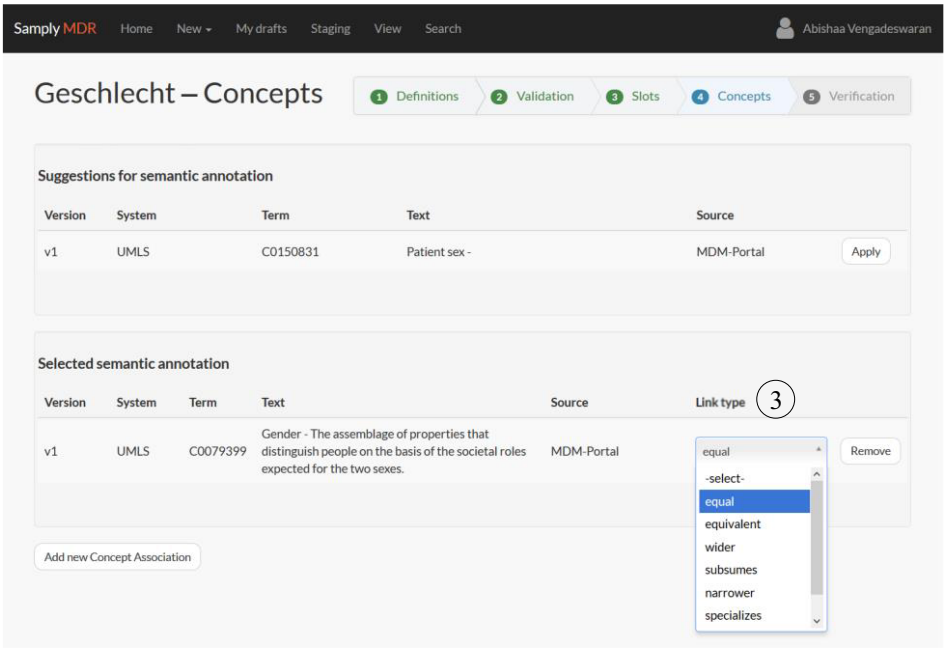

Figure 3. User interface of Concept Association in Samply.MDR – Adding Link type

Moreover, the user has the opportunity to define own concept association. After clicking "Add new Concept Association", input fields for attributes are given (Fig. 4). These can be used for example to establish individual ontologies (4). By clicking "Next"-Button the step of adding concept associations will be finished. A data element can also be edited after it has been released. During this process the same wizard step appears, so that concept assocations can be add, edit or delete, in the same way as proposed.

During the implementation of the prototype, emphasis was placed on facilitated extensibility, so that additional systems can easily be integrated. Currently, there are no restrictions to the handling of a concept association, however a basis for semantic annotation has been created.

5. Discussion

Semantic annotation increases semantic interoperability. The MDM-Portal addresses this issue and provides semantic annotation, which is curated by physicians and medical professionals. Hence, the quality of the annotation depends on the content of the portal [12]. The MDM facilitates the annotation process by suggesting UMLS codes, without additional user contribution. However, the validity or adjustment of the information, as well as contributing the LinkedType has to be done by the user, which requires some expertise, like understanding the meaning of the given link type. The major advantage of MDM-Portal for Samply.MDR is the ability of searching semantic annotation for

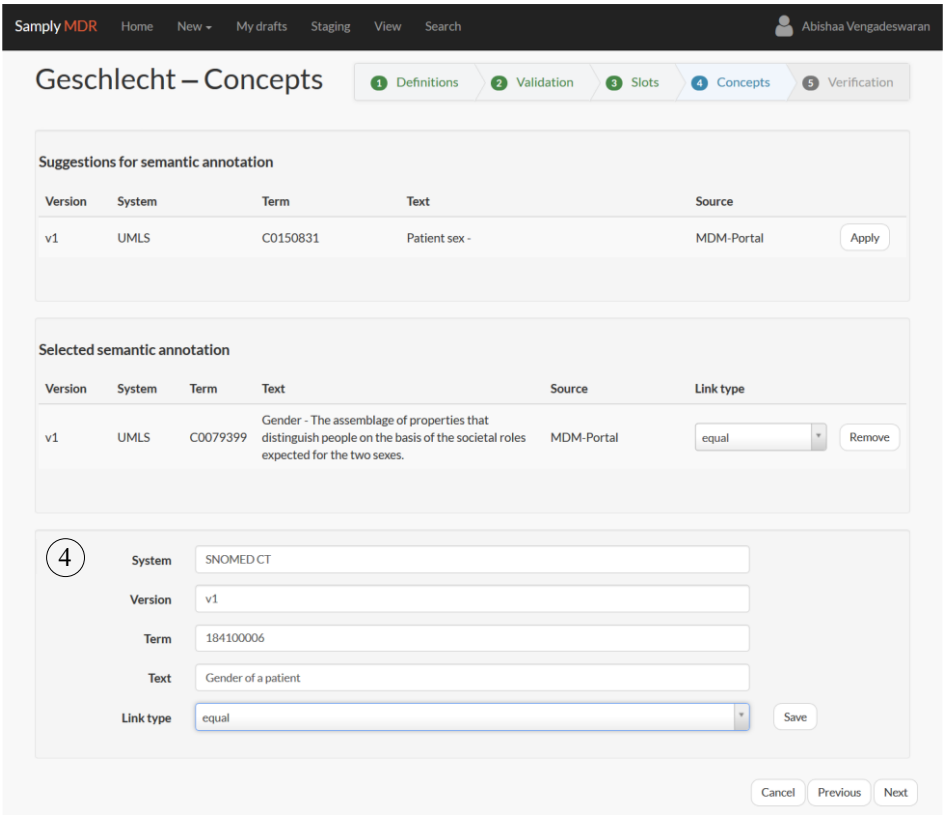

Figure 4. User interface of Concept Association in Samply.MDR – Adding own concept associations

especially German terms. However, if data elements are defined in English a query of the UMLS REST-API would be another possibilty.

Besides, another benefit of the MDM-Portal could be achieved by suggesting further data elements, which are used in combination with the created one. For example, if a data element body temperature is created, other vital parameters such as heart rate or blood pressure could be proposed to the user. Based on that, the user could finally decide, whether those proposals fit and could adopt them, instead of having to create those data elements by himself. Such an extension would foster harmonization of data elements in the long term and support the user during the process of creating data elements.

Furthermore, the mapping process within MIRACUM is supported. By adding semantic annotation, data scientist can easily find matching data elements with precise link type, which can be considered, when using the data element. However, Samply.MDR is used in numerous project, a mapping between the different data elements is easier.

This work enables enrichment of data elements with UMLS codes in a harmonized way, by interconnecting Samply.MDR with MDM-Portal. A first step towards semantic annotation has been taken and will be further expanded by linking additional systems. Finally, according to feedback and requirements, this protype will be improved continusouly.

Conflicts of Interest

The authors declare that there is no conflict of interest.

References

1. Medizin Informatik Initiative [cited 2019 Feb 18]. Available from: URL: https://www.medizininformatik-initiative.de/.

2. MIRACUM: Medical Informatics in Research and Care in University Medicine [cited 2019 Feb 18]. Available from: URL: https://www.miracum.org/.

3. 610.12-1990 IEEE Standard Glossary of Software Engineering Terminology. s.l.: IEEE / Institute of Electrical and Electronics Engineers Incorporated. Available from: URL: http://ieeexplore.ieee.org/servlet/opac?punumber=2238.

4. Kubicek H, Cimander R, Scholl HJ. Organizational Interoperability in E-Government. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011.

5. Garde S, Knaup P, Hovenga E, Heard S. Towards semantic interoperability for electronic health records. Methods Inf Med 2007; 46(3):332–43.

6. Park H-A, Hardiker N. Clinical Terminologies: A Solution for Semantic Interoperability. J Korean Soc Med Inform 2009; 15(1):1.

7. Kadioglu D, Breil B, Knell C, Lablans M, Mate S, Schlue D et al. Samply.MDR - A Metadata Repository and Its Application in Various Research Networks. Stud Health Technol Inform 2018; 253:50–4.

8. SNOMED International: Leading healthcare terminology, worldwide [cited 2019 Feb 18]. Available from: URL: http://www.snomed.org/.

9. LOINC: Version 2.65 [cited 2019 Feb 18]. Available from: URL: https://loinc.org/.

10. USA.gov. Unified Medical Language System® (UMLS®); 2016 [cited 2019 Feb 15]. Available from: URL: https://www.nlm.nih.gov/research/umls/.

11. CDISC. Operational Data Model (ODM)-XML [cited 2019 Feb 15]. Available from: URL: https://www.cdisc.org/standards/data-exchange/odm.

12. International Organization for Standardization TC 215 Health Informatics. ISO NP TS 21526 Healthcare Informatics – Metadata Repository Requirements; 2018 2018 Mar 12. Available from: URL: https://isotc.iso.org/livelink/livelink/open/tc215wg3.

13. Hegselmann S, Storck M, Geßner S, Neuhaus P, Varghese J, Dugas M. A Web Service to Suggest Semantic Codes Based on the MDM-Portal. Stud Health Technol Inform 2018; 253:35–9.

14. FHIR: FHIR R4; 2018 [cited 2019 Feb 19]. Available from: URL: https://www.hl7.org/fhir/valueset-concept-map-equivalence.html.