Establishing Semantic Interoperability between HL7 v2.x and V3: A Communication Standards Ontology (CSO)

Abstract. Different communication standards in healthcare - esp. HL7 version 2.x [1] and v3 [2] - lack inter- and intra-family compatibility. Bridging between those to establish semantic interoperability using a formal ontology as a mediator in a mapping process [3] has demonstrated that both communication standards have in principle the same underlying architecture. This paper shortly analyses this structure in order to create a communication standards ontology (CSO) based on (basic) formal ontologies (BFO/FO) which is presented thereafter. The paper discusses problems which appeared during the development process and the established solution.

Keywords. HL7 v2.x, HL7 V3, communication standard, semantic interoperability, formal ontology, BFO.

Introduction

The increasing use of information exchange in healthcare for a vast variety of reasons (improvement of patient care, reduction of costs, patient safety, etc.) requires a data representation which enables the reuse of the transmitted data - the latter is commonly known as semantic interoperability. However, the most frequently used communication standards do not guarantee each others compatibility which makes it even harder to achieve the aforementioned goal.

In order to prepare compatible data exchange, knowledge about the used standards is essential. One way to express knowledge is the use of semantic web technologies which has lead to the Web Ontology Language (OWL) [4]. The efforts in creating a bridging (mapping) architecture [3] have revealed that the most popular communication standards in healthcare – HL7 v2.x and V3 – are incompatible and call for the aforementioned approach. This paper describes efforts and solutions to the problems of creating communication standards ontology for HL7 version 2.x and version 3 based on the basic formal ontology (BFO) [5, 6].

1. Materials and Methods

Assuming that the reader is aware of the different HL7 standards families [1, 2], they are not described in detail. Instead, short explanations about formal ontologies are provided. The complete work has been prepared by using the Generic Component Model (GCM) architecture framework [7]. Supporting information can be found in [8].

1.1. BFO: Basic Formal Ontology as a Foundation

In order to achieve the primary goal of the underlying work [3], the integration of a reference ontology bridging the different structures of those standards appears necessary. As a first approach when starting with the development process, a suitable reference ontology (RO) has to be created from scratch. Research on the Internet revealed an RO which seemed to be applicable and could be used instead: the Basic Formal Ontology (BFO) [5]. BFO itself is used as a foundation ontology for a set of other ontologies which are built for specific purposes. Here, the Advancing Clinico Genomic Trials on Cancer (ACGT) [9] ontology can be mentioned, which is used in [3] as the central mapping ontology.

In this paper, BFO plays the role of the reference ontology, which is used to align the basic concepts for communication standards with. Figure 1 demonstrates a rough overview which is explained later.

1.2. IHE Technical Frameworks

Another good source of information is the set of Technical Frameworks (TF) [10] IHE (Integrating the Healthcare Enterprise) is providing for set of different domains. IHE facilitates workflow specific integration profiles which are specified using the technical terminology as provided by the underlying (communication) standards. This leads to the problem that the same technical term, e.g. "required", is used in different TFs, but the meaning of it is taken from the corresponding communication standard. Hence, such a term has different semantics. However, even the experts are not always aware of this problem, so that it results in diverging interpretations and of course in incompatible implementations. IHE failed to harmonize its Technical Frameworks by providing a clear definition for mapping to the used standards [11]. The result of this examination is reflected in the CSO as well.

2. Results

The presented results are based on aforementioned conceptual work and represent the current status.

2.1. Ontology Structure

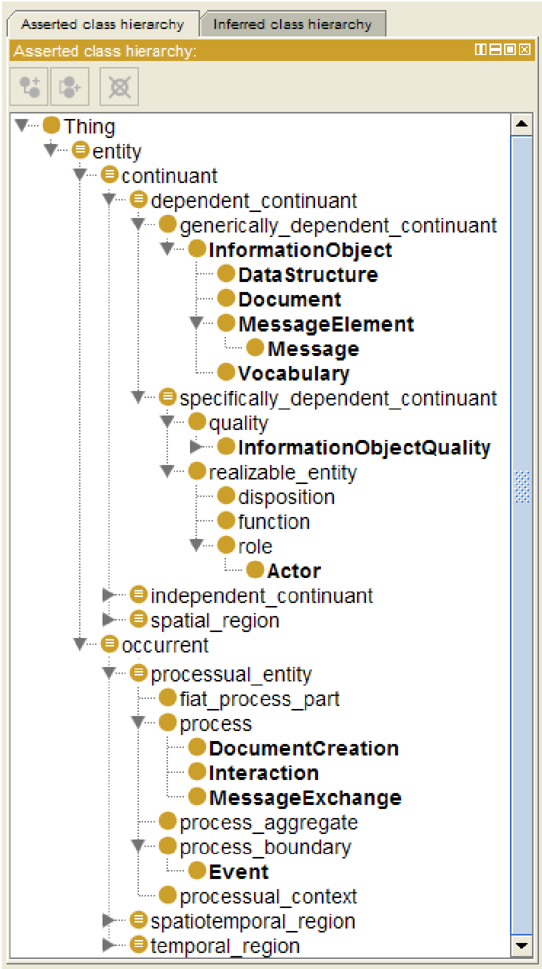

Figure 1. Ontology Structure embedded in BFO

Intensive work with and contributions to the analyzed communication standards during the past few years have revealed a deep insight into the architecture and the technical terminology used in the standard specifications. The IHE Whitepaper [11] lists the details.

The process of exchanging a message or creating a document is caused (triggered) by an event which leads to an interaction with a message between two or more actors to exchange this message (*MessageExchange*) or document (*DocumentCreation*). These dynamic aspects of the communication standards are shown in Figure 1. The last three concepts are different kinds of processes and represented as siblings in this hierarchy. BFO provides the appropriate parent concepts with *process* and *process_boundary*. The actors which are involved in this interaction fit in form of a behavioral role into *role* as a specialization of *realizable entity*.

The relations between those are represented in OWL as object relationships and explained later.

The concepts for the static aspects are aggregated below the common and newly introduced parent concept called *InformationObject* which is introduced below *generically_dependent_continuant*. The concepts *DataStructure*, *Vocabulary*, *MessageElement* and *Document* are represented as siblings. A *Message*, which is used in *Interaction*s, is a specialization of *MessageElement* because it is itself a message element, while it represents the top element of a partonomy of message elements at the same time: Message elements makes of use other message elements in form of substructures – or part-of relationships.. Therefore, a message inherits the same properties (features) while using additional constraints in combination with the object relationships allowing for being included in interactions.

Documents are also a specialization of *InformationObject* for the same reasons. Using object relationships, documents can reuse concepts out of the message element hierarchy for a structured representation of the necessary details as well. It is quite common in communication standards, that the same means are used.

The details of message elements vary between HL7 v2.x and V3. However, they fit below the same architecture:

v2.x:

- Component
- DataElement
- Field
- MessageStructure
- Segment

V3:

- CMET
- R-MIM
- RIM

Furthermore, the entry points for data types/structures and vocabulary items are located below *InformationObject* as well due to their nature when acting as a dependent continuant. But both require different specializations for the distinct communication standards as shown in Table 1.

Data structures are used to provide a hierarchy of part-of relationships, e.g. an address consists of address parts. This can either be represented as a specialization hierarchy within the same branch or within different branches [8]. In HL7 version 3 a separate hierarchy for data type parameters is necessary. One example thereof is “SET<T>” as a set of information using data type “T”.

Table 1. Specializations done in the standards

Concept	**Specialization**	**v2.x**	**V3**
Data Structure			
	individual datatypes with components	X	
	DataType		X
	DataTypeParameter		X
Vocabulary			
	Tables	X	
	table values	X	
	concept domains		X
	code systems		X
	value sets		X

In contrast to data types, vocabulary commonly provides different kinds of sets to reflect grouping and identification means to reference specific codes out of a controlled vocabulary [14, 15].

2.2. Object Relationships

As mentioned before, the different specialization hierarchies are linked together using object relationships. In principle, quite a lot of different relationships can be establish among the different concepts thereby assigning different names. BFO has a set of relations, but they do not represent a hierarchy. Instead, they are maintained as siblings without any relationship.

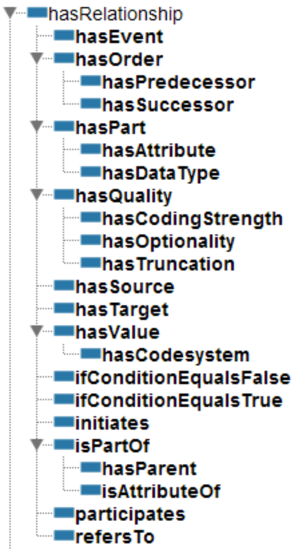

Figure 2. Object Relationships

But having in mind, that an automatic process should evaluate the relationships a hierarchical structure is necessary: It allows for separating and clarifying the semantics by using different names while still keeping the foundation for an automatic process in place. Figure 2 clarifies the result of this separation.

The availability of the source data as input for the generation process is restricted, so that some relationships are established from the "wrong side", but establishing the correct inverse relationship this disadvantage is eliminated.

2.3. Qualities

Communication standards make use of a set of concepts to specify behavioral, coding and other aspects. The complete detailed list is shown in Figure 3. The different concepts are maintained as siblings.

The detailed (i.e. specialized) concepts as extracted from the specific communication standards are linked by object properties (see above) from these concepts. The different communication standards only support a subset of it so that this hierarchy represents the common superset.

As introduced above, one aspect is the use of intelligent agents being able to interpret the represented knowledge automatically. It requires the awareness of the semantics in behind so that the differences while exploring a simplified programming logic can be considered.

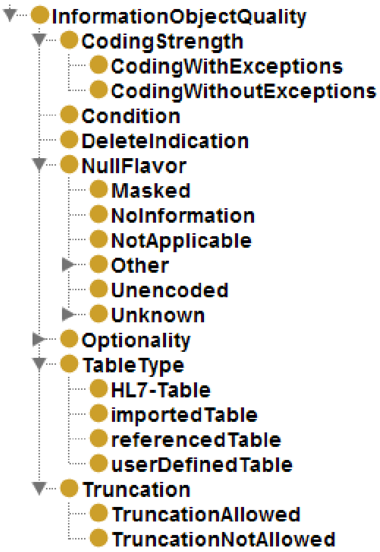

Figure 3. Information Object Qualities

In addition to the concepts mentioned in Figure 3, some further details must be considered which are described in the following.

To support the verification of message instances – this is not part of this work item – the ontology must additionally provide information about minimum, maximum and conformance length. These items are represented in OWL by using appropriate data properties.

The second important information is the possibility for repeating a message element. This information can be represented by the standard OWL properties minCardinality and maxCardinality.

Finally, the different table types may have associations with *CodingStrength*. For example, HL7-tables do not allow for extending the value sets which is represented by a link to *CodingWithoutExceptions*. In addition, all coded values – no matter what table type is used - should not allow for truncation.

2.4. Different Kinds of Optionality

Analyzing the IHE Technical Frameworks [10] based on the described analysis has lead to an abstract description of an application's architecture [11] including static and dynamic aspects of systems behavior. A central component in communication standards specifications is the definition of optionality, i.e. how an interface of an application has to work with certain elements.

A fundamental result of this analysis is the fact that the terms used in the specifications do not represent atomic concepts, i.e. they always combine different basic concepts (qualities). Unfortunately and not surprisingly, the different standards combine them differently.

The formalization of this analysis facilitating the web ontology language (OWL) in combination with namespaces leads to a conceptualization as shown in Figure 4.

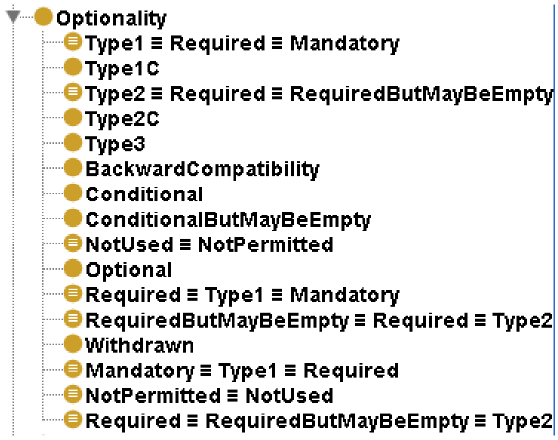

Figure 4. Different Kinds of Optionality

In this case, the OWL equality relationship can be used to represent identical concepts. One example thereof is the fact, that "DICOM type 1" is equal to "v2.x required" which is the same as "V3 mandatory". This relationship is shown as the first specialization in Figure 4. In the aforementioned figure namespaces are used to separate concepts with the same "name" – e.g. "required".

At least, this equality is true for concepts for real optionality. If null flavors must be taken into account, a separate equality will be established which must be represented by object relationships because different kind of behaviors are concerned.

3. Discussion

The *Mapping* concepts and their specializations as mentioned in Figure 1 are not really necessary in CSO, because their primary use is intended in a mapping process [3]. Therefore, it can be evacuated in a separated ontology on top of and importing CSO.

Another topic worth discussing is the necessity (and naming) of *InformationObjecctQuality*: In principle it can be left out. But on the other hand it allows for enclosing all the concepts representing specific qualities of information objects.

Due to the conceptualization and theoretical background of BFO, it is not allowed to have relationships among different qualities. Within communication standards, e.g. an HL7 table is always combined with using codes/table values of the coding strength *CodingWithoutExceptions*. It is possible to have the same representation in OWL without violating this constraint, but it requires an amount of additional definitions. Having in mind, that this knowledge representation is intended for intelligent agents, a simplification seems to be acceptable.

4. Conclusion/Recommendation

As demonstrated in [3], communication standards share the same underlying standards. A rough examination of DICOM [12] and xDT [13] supports this statement. Therefore, CSO should be brought forward to the OBO foundry for inclusion.

4.1. Acknowledgments

The authors are indebted to Stefan Schulz (University of Freiburg) and Mathias Brochhausen (IFOMIS) for kind cooperation and support.

5. References

[1] HL7 at http://www.hl7.org

[2] newest "HL7 Version 3 Ballot Material" at http://www.hl7.org/v3ballot/html/welcome/environment/index.htm

[3] Oemig F, Blobel B. Semantic Interoperability between Health Communication Standards through Formal Ontologies. Medical Informatics in a United and Healthy Europe. In: Adlassnig K-P, Blobel B, Mantas J, Masic I (Editors), Series "Studies in Health Technology and Informatics, Volume 150", Proceedings of the MIE 2009 - European Federation for Medical Informatics, IOS Press, ISBN: 978-1-60750-044-5, p.200-204, Sarajevo, Bosnia and Herzegovina, 30.8-2.9.2009, http://www.mie2009.org

[4] "OWL Web Ontology Language Overview" at http://www.w3.org/TR/owl-features/

[5] BFO: Basic Formal Ontology, http://www.ifomis.org/bfo

[6] OBO Foundry. http://www.obofoundry.org/cgi-bin/detail.cgi?id=bfo.

[7] Blobel B. Application of the Component Paradigm for Analysis and Design of Advanced Health System Architectures. International Journal of Medical Informatics 60 (3) (2000) 281-301

[8] Oemig F, Blobel B. An Ontology Architecture for HL7 V3: Pitfalls and Outcomes. Proceedings of the WC 2009, Munich, Germany, 10.9.2009, http://www.wc2009.org

[9] ACGT: Advancing Clinico Genomic Trials on Cancer, http://www.ifomis.org/wiki/acgt

[10] IHE. Integrating the Healthcare Enterprise. http://www.ihe.net

[11] Oemig F. IHE Internal Terminology – Whitepaper (Volume 0). 2009. http://groups.google.com/group/ihe-terminology/attach/7804505ee72e6ed9/IHE+Terminology+v07.zip?part=2

[12] DICOM. Digital Images and Communication in Medicine. http://www.rsna.org/Technology/DICOM/index.cfm

[13] xDT. 2001. http://www.kbv.de/ , http://www.zi-koeln.de/ und http://www.zi-berlin.de/ http://www.kbv.de/ita/4274.html

[14] Cimino J. Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. Methods Inf Med. 1998 Nov; 37(4-5):394-403.http://www.sahs.uth.tmc.edu/evbernstam/HI5300/Articles%20(Reading%20Materials)/cimino_desiderata-for-controlled-medical.pdf, last accessed 5.5.08

[15] Gruber T. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, vol.5, Issue 2, S. 199-200, 1993, ISSN: 1042-8143