

Eine Ontologie für Kommunikations- Standards (CSO): Etablierung semantischer Interoperabilität zw. HL7 v2 und v3

Establishing Semantic Interoperability between HL7 v2.x and V3 using a Communications Standards Ontology

Frank OEMIG^{a,1} and Bernd BLOBEL^b

^a*Agfa HealthCare GmbH, Bonn*

^b*eHealth Competence Center, Regensburg*

Zusammenfassung. Den Kommunikations-Standards im Gesundheitswesen - bspw. HL7 Version 2.x [1] und v3 [2] - fehlt es an Kompatibilität innerhalb und zwischen den Familien. Um semantische Interoperabilität in der HL7-Standards-Welt zu etablieren, erfordert eine Kommunikations-Standards Ontologie (CSO) auf der Grundlage formaler Ontologien (BFO) als Vermittler in einem Mapping-Prozess [3], wobei beide Kommunikations-Standards der gleichen Architektur folgen müssen. Dieser Beitrag führt in die CSO ein.

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, Kommunikations-Standard, semantische Interoperabilität, formale Ontologie, BFO

¹ Frank Oemig, Agfa HealthCare GmbH, Konrad-Zuse-Platz 1-3, D-53227 Bonn, Frank.Oemig@agfa.com

1. Einleitung

Der zunehmende Informationsaustausch in der Gesundheitsversorgung aus einer Vielzahl von Gründen (Verbesserung der Patientenversorgung, Kostensenkung, Patientensicherheit, etc.) erfordert eine Informationsdarstellung, die eine Wiederverwendung der übermittelten Daten ermöglicht. Letzteres ist allgemein bekannt als semantische Interoperabilität. Kenntnisse über die verwendeten Standards sind dabei von wesentlicher Bedeutung, um einen kompatiblen Datenaustausch vorzubereiten.

Die am meisten genutzten Kommunikations-Standards im Gesundheitswesen – HL7 v2.x und V3 – sind zueinander inkompatibel und benötigen einen Mapping-Ansatz [3] auf Basis einer gemeinsamen Architektur.

Als Lösungsansatz dazu beschreibt dieser Beitrag die Kommunikations-Standards Ontologie (CSO) für HL7 Version 2.x und Version 3 auf Basis der formalen Ontologie (BFO) [5, 6]. Ein möglicher Weg zur Formalisierung von Wissen ist der Einsatz der Semantic Web-Technologien, die zu der Web Ontology Language (OWL) [4] geführt haben.

2. Material und Methoden

Die verschiedenen HL7-Familien (v2.x und V3, [1, 2]) werden in diesem Beitrag nicht im Detail beschrieben, sondern als bekannt vorausgesetzt: Die komplette Arbeit basiert auf dem Generic Component Model (GCM) Architecture Framework [7]. Weitere Informationen dazu finden sich in [8].

2.1. BFO: Die Basic Formal Ontology als Grundlage

Um das primäre Ziel der zugrunde liegenden Arbeit zu erreichen [3] ist die Integration einer Referenz-Ontologie zur Überbrückung der unterschiedlichen Strukturen der betreffenden Standards erforderlich. Mit dem Beginn des Entwicklungsprozesses müsste mit dem Aufbau einer geeigneten Referenz-Ontologie (RO) begonnen werden, der im ersten Ansatz dann von Grund auf neu erstellt wird. Recherchen im Internet führten zu einer RO, die verwendbar erschien: die Basic Formal Ontology (BFO) [5]. BFO wird selbst als Grundlage von einer Reihe von anderen Ontologien genutzt, die jeweils für bestimmte Zwecke aufgebaut werden. Hier wären die Advancing Clinico Genomic Trials on Cancer (ACGT) [9] Ontologie zu nennen, die in [3] als zentrale Mapping Ontologie verwendet wird.

In diesem Papier spielt die BFO die Rolle der Referenz-Ontologie, um daran die grundlegenden Konzepte der Kommunikationsstandards auszurichten. Abbildung 1 zeigt einen groben Überblick, der später weiter ausgeführt wird.

2.2. IHE Technical Frameworks

Eine andere gute Quelle für Informationen ist die Menge der Technical Frameworks (TF) [10] von IHE (Integrating the Healthcare Enterprise) für ein Reihe von verschiedenen Domains. IHE stellt workflow-spezifische Integrationsprofile bereit, die mit der technischen Terminologie spezifiziert werden, die durch die zugrunde liegenden (Kommunikations-)Standards vorgegeben sind. Dies führt zu dem Problem,

dass die gleichen technischen Begriffe, z. B. "required", in verschiedenen TFs verwendet werden, die Bedeutung wurde aber aus den entsprechenden Kommunikations-Standard übernommen. Daher hat ein solcher Begriff unterschiedliche Bedeutungen. Allerdings sind sich die Experten auch nicht immer dieses Problems bewusst, so dass es in unterschiedlichen Auslegungen und dann natürlich auch inkompatiblen Implementierungen mündet. IHE hat die Terminologie nicht harmonisiert, die in den TFs verwendet wird, und keine klare Definition in Form eines Mapping auf der verwendeten Standards [11] bereitgestellt. Das Ergebnis dieser Untersuchung spiegelt sich auch in der CSO wieder.

3. Ergebnisse

Die vorgestellten Ergebnisse basieren auf den vorgenannten konzeptionellen Arbeiten und repräsentieren den aktuellen Status.

3.1. Struktur der Ontologie

Intensive Arbeit mit und Beiträge zu den analysierten Kommunikations-Standards während der letzten Jahre haben einen tiefen Einblick in die Architektur und die technische Terminologie in der Standard-Spezifikationen offenbart. Das IHE Whitepaper zu den internen technischen Terminologien [11] listet dazu weitere Details auf.

Der Prozess zum Austausch einer Nachricht oder zu Erstellung eines Dokumentes wird ausgelöst durch ein Ereignis, das zu einer Interaktion zwischen zwei oder mehreren Akteuren mit dem Austausch dieser Nachricht (*MessageExchange*) oder des Dokumentes (*DocumentCreation*) führt. Diese dynamischen Aspekte der Kommunikations-Standards sind in Abbildung 1 unter *processual_entity* dargestellt. Die letzten drei Begriffe unter *process* sind verschiedene Arten von Prozessen und daher als Siblings in dieser Hierarchie vertreten. BFO bietet die geeigneten übergeordneten Konzepte mit *process* und *process_boundary*. Die Akteure, die in dieser Interaktion involviert sind, werden durch die Klasse *behavioral_role* als Spezialisierung einer *realizable_entity* repräsentiert.

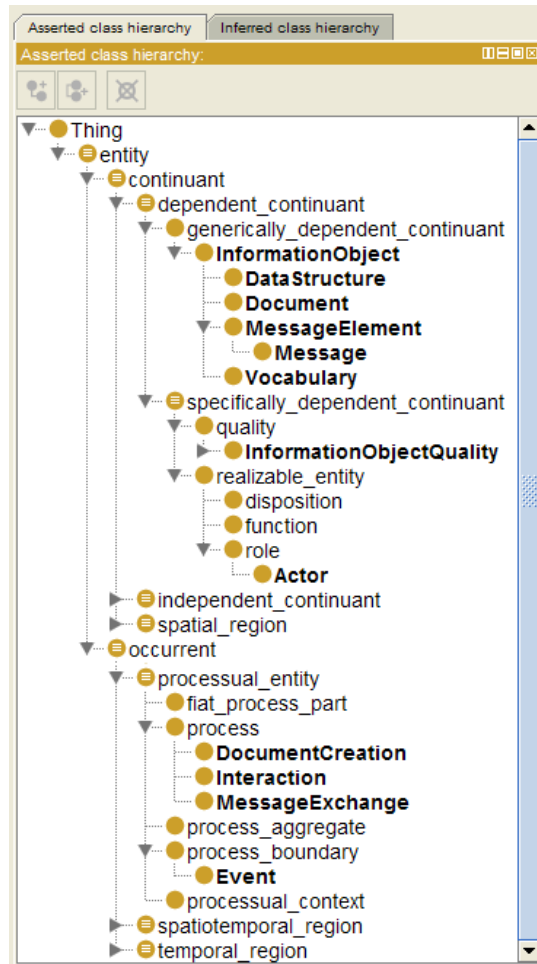


Abbildung 1. Ontology Structure embedded in BFO

Die Beziehungen zwischen diesen Klassen sind in OWL als Objekt-Beziehungen dargestellt und werden später erklärt.

Die Konzepte für die statischen Aspekte werden unter dem gemeinsamen, neu eingeführten Oberbegriff *InformationObject* eingeführt, das unter *generically_dependent_continuant* eingeordnet wird. Die Konzepte *DataStructure*, *Vocabulary*, *MessageElement* und *Document* werden als Siblings dargestellt. Eine *Message*, die in *Interactions* benutzt wird, ist eine Spezialisierung von *MessageElement* weil es selbst eine Nachrichtenelement ist, während es gleichzeitig das oberste Element einer Partonomie von Nachrichtenelementen ist: Nachrichtenelemente nutzen andere Nachrichtenelemente in Form von Substrukturen – oder Teilvon-Beziehungen. Daher erbt eine Nachricht die gleichen Eigenschaften (Features), während zusätzliche Einschränkungen dieser Spezialisierung in Kombination mit den Objekt-Beziehungen die Verwendung in Interaktionen ermöglichen.

Dokumente sind aus den gleichen Gründen, auch eine Spezialisierung von *InformationObject*. Mit Objekt-Beziehungen können Dokumente Konzepte aus der Hierarchie der Nachrichtenelemente für eine strukturierte Darstellung der

erforderlichen Details nutzen. In Kommunikations-Standards ist es durchaus üblich, dass die die gleichen Mittel genutzt werden.

Obwohl die individuellen Details zwischen HL7 v2.x und V3 variieren, können sie mit einem generischen Architektur-Framework harmonisch beschrieben werden. Nachfolgend sind die unterschiedlichen Spezialisierungen der Message Elements aus den beiden Kommunikationsstandards aufgelistet:

- | | |
|---|--|
| v2.x: | V3: |
| <ul style="list-style-type: none"> • DataElement • Field • MessageStructure • Segment | <ul style="list-style-type: none"> • CMET • R-MIM • RIM |

Die Einstiegspunkte für Datentypen/Strukturen und Vokabularen befinden sich auch unter *InformationObject* auf Grund ihrer Beschaffenheit als abhängiges Continuant. Aber beide erfordern wie in Tabelle 1 dargestellt unterschiedliche Spezialisierungen für die unterschiedlichen Kommunikationsstandards.

Datenstrukturen werden verwendet, um eine Hierarchie von Teil-von-Beziehungen aufzubauen, z.B. besteht eine Adresse aus Adressteilen. Dies kann entweder als eine Spezialisierungshierarchie innerhalb des gleichen Astes oder in verschiedenen Ästen vertreten sein [8]. In HL7 Version 3 wird auch eine separate Hierarchie für die Datentyp-Parameter notwendig. Ein Beispiel dafür ist "SET<T>" als eine Menge von Informationen unter Verwendung des Datentyps "T".

Tabelle 1. Spezialisierungen in den Standards

Konzept	Spezialisierung	v2.x	V3
Data Structure			
	individuelle Datentypen mit Komponenten	X	
	DataType		X
	DataTypeParameter		X
Vocabulary			
	Tables	X	
	table values	X	
	concept domains		X
	code systems		X
	value sets		X

Im Gegensatz zu den Datentypen stellen Vokabularen gemeinhin verschiedene Mengen dar, um bestimmte Codes aus einem kontrollierten Vokabular zu gruppieren und zu identifizieren [14, 15].

3.2. Object Relationships

Wie bereits erwähnt, werden die verschiedenen Spezialisierungshierarchien über Objekt-Beziehungen miteinander verknüpft. Im Prinzip können eine ganze Reihe von unterschiedlichen Beziehungen zwischen den verschiedenen Konzepten durch individuelle Ausprägungen unter Zuweisung von unterschiedlichen Namen geschaffen werden. BFO hat eine Reihe von solchen Beziehungen, aber sie stellen keine Hierarchie dar. Stattdessen werden sie wie Siblings ohne jede Beziehung zueinander verwaltet, so dass eine Nutzung schwierig wird.

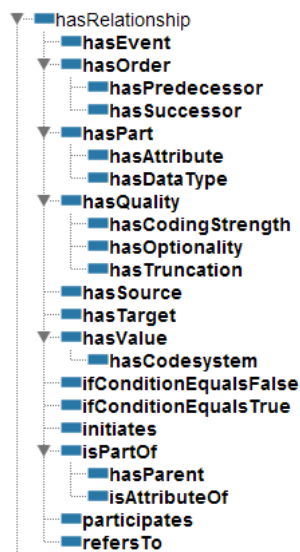


Abbildung 2. Object Relationships

Wenn aber ein automatischer Prozess die Beziehungen bewerten soll, ist eine hierarchische Struktur notwendig: Sie erlaubt die Trennung und die Klärung der Semantik durch die Verwendung unterschiedlicher Namen unter Beibehalt der Grundlage für einen automatischen Prozess. Abbildung 2 verdeutlicht das Ergebnis dieser Trennung.

Die Verfügbarkeit der Quell-Daten als Input für die Generierung ist eingeschränkt, so dass einige Beziehungen von der "falschen Seite" eingerichtet definiert werden. Dieser Nachteil wird jedoch mit der Definition der Umkehrbeziehung beseitigt.

3.3. Qualities

Kommunikationsstandards nutzen eine Reihe von Konzepten, um Verhaltens-, Codier- und andere Aspekte anzugeben. Die vollständige und detaillierte Liste ist in Abbildung 3 dargestellt. Die verschiedenen Konzepte werden als Siblings verwaltet.

Die detaillierten (d.h. spezialisierten) Konzepte, so wie sie aus den spezifischen Kommunikations-Standards extrahiert wurden, werden über ihre Objekt-Eigenschaften miteinander verbunden (siehe oben). Die unterschiedlichen Kommunikationsstandards unterstützen jeweils nur eine Teilmenge davon, so dass diese Hierarchie die gemeinsame Obermenge darstellt.

Wie oben erläutert, ist ein Aspekt der Einsatz von intelligenten Agenten, die in der Lage sind, das repräsentierte Wissen automatisch zu interpretieren. Die Programmierlogik muss die zugrunde liegende Semantik dann korrekt berücksichtigen.

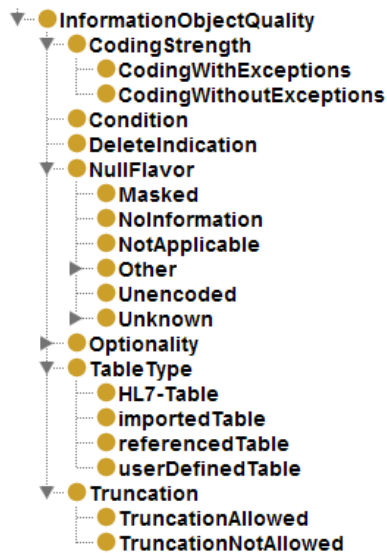


Abbildung 3. Information Object Qualities

Schließlich können die unterschiedlichen Tabellentypen Assoziationen mit *CodingStrength* haben. Zum Beispiel erlauben die HL7-Tabellen keine Erweiterung der Wertemenge, was sich durch eine Verknüpfung mit *CodingWithoutExceptions* darstellen läßt. Darüber hinaus dürfen alle codierten Werte - egal welcher Tabellentyp verwendet wird - nicht abgeschnitten (truncated) werden.

3.4. Verschiedene Ausprägungen der Optionalität

Die Analyse der IHE Technical Frameworks [10] hat zu einer abstrakten Beschreibung der Architektur einer Anwendung [11] einschließlich der statischen und dynamischen Aspekte des Systemverhaltens geführt. Ein zentraler Baustein in der Spezifikation von Kommunikationsstandards ist die Definition der Optionalität, d.h. wie die Schnittstelle einer Anwendung mit bestimmten Elementen umgeht.

Ein grundlegendes Ergebnis dieser Analyse ist die Tatsache, dass die Begriffe in den Spezifikationen keine atomaren Konzepten darstellen, d.h. sie können verschiedene grundlegende Konzepte (Qualitäten) miteinander kombinieren. Leider und nicht überraschend tun die verschiedenen Standards dies unterschiedlicher Art und Weise.

Die Formalisierung dieser Analyse mithilfe der Web Ontology Language (OWL) in Kombination mit Namespaces führt zu einer Konzeption, wie sie in Abbildung 4 dargestellt ist.

Zusätzlich zu den in Abbildung 3 erwähnten Konzepten müssen einige weitere Details berücksichtigt werden, die im Folgenden beschrieben werden.

Zur Unterstützung der Überprüfung der Nachrichteninstanzen muss die Ontologie zusätzlich Informationen über die minimale und maximale Länge sowie die Konformanzlänge bereitstellen. Diese Details werden in OWL durch die Verwendung geeigneter DataProperties realisiert.

Die zweite wichtige Information ist die Möglichkeit zur Wiederholung eines Nachrichtenelements. Dies kann mit den OWL-Standardeigenschaften *minCardinality* und *maxCardinality* umgesetzt werden.

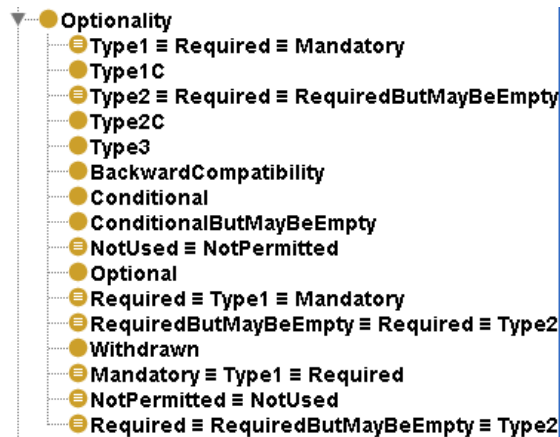


Abbildung 4. Verschiedene Ausprägungen der Optionalität

In diesem Fall kann die Gleichheitsbeziehung von OWL verwendet, um identische Begriffe zu identifizieren. Ein Beispiel hierfür ist die Tatsache, dass "DICOM Typ 1" mit "V2.x required" gleichgesetzt werden kann, was wiederum mit "V3 mandatory" identisch ist. Diese Beziehung ist als die erste Spezialisierung in Abbildung 4 dargestellt. In der vorgenannten Abbildung werden Namespaces genutzt, um verschiedene Begriffe mit dem gleichen "Namen" voneinander zu unterscheiden – z.B. "Required".

Zumindest stimmt diese Gleichheit für Konzepte für echte Optionalitäten. Für die verschiedenen Ausprägungen von „Null“ muss jedoch eine dedizierte Gleichheitsbeziehung geschaffen werden, weil verschiedene Verhaltensregeln betroffen sind.

4. Diskussion

Ein Punkt ist die Notwendigkeit (und Benennung) von *InformationObjectQuality*: Im Prinzip kann es weggelassen werden. Auf der anderen Seite erlaubt sie aber die Konzepte einzuschließen, die die Qualitäten der Informationsobjekte ausmachen.

Aufgrund der Konzeption und des theoretischen Hintergrundes der BFO ist es nicht gestattet, Beziehungen zwischen den verschiedenen Qualitäten zu etablieren. Innerhalb von Kommunikationsstandards kommt dies aber vor, so ist z. B. eine HL7-Tabelle immer mit Codes/Tabellenwerte der Coding Strength CodingWithoutExceptions kombiniert. Es ist möglich, dies in OWL ohne Verletzung dieser Einschränkung auszudrücken, aber es erfordert eine Reihe an zusätzlichen Definitionen. In Anbetracht des Umstandes, dass dieses Wissen für intelligente Agenten bestimmt ist, scheint eine Vereinfachung akzeptabel zu sein.

5. Schlußfolgerungen/Empfehlungen

Kommunikationsstandards weisen die gleichen Basisstrukturen auf. Eine grobe Prüfung von DICOM [12] und xDT [13] unterstützt diese Aussage. Daher sollte CSO der OBO-Foundry zur Aufnahme vorgelegt werden.

6. References

- [1] HL7 at <http://www.hl7.org>
- [2] neuestes "HL7 Version 3 Ballot Material",
<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"
<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](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