

13 Der epSOS Master Value Catalogue (MVC): Erfahrungen bei der Auswahl von Codesystemen

Christof Geßner und Jürgen Wehnert

Ein Ergebnis des epSOS-Projekts [epSOS, 2010] war die Vereinbarung von Regeln für die Codierung bestimmter medizinischer Konzepte auf Basis standardisierter internationaler Klassifikationen, Nomenklaturen und Ontologien. Zweck dieser Codierung war die maschinelle Unterstützung einer Übersetzung der Konzepte in den epSOS-Dokumenten ePrescription (eP) und Patient Summary (PS) von einer Quellsprache in eine Zielsprache. Eine Herausforderung bestand dabei in der Auswahl der Begriffssysteme für die Codierung sowie in einer sinnvollen Festlegung der zulässigen Wertebereiche für den gegebenen Inhalt und Verwendungszweck. Bei epSOS wurden dafür Wertelisten („Value Sets“) auf der Ebene einzelner Datenfelder festgelegt. Der vorliegende Artikel fasst Ergebnisse zusammen, die bereits zuvor im Rahmen des epSOS-Projekts in englischer Sprache publiziert wurden [Sonne Hansen, A. et al., 2013].

Angesichts der zunehmenden Bedeutung des institutionsübergreifenden Datenaustauschs im Gesundheitswesen werden Defizite in der semantischen Interoperabilität zwischen den beteiligten IT-Systemen zunehmend sichtbar. Eindeutige Bezeichnungen und Begriffssysteme werden immer wichtiger. Die Integration von Daten aus verschiedenen Quellsystemen stellt eine Herausforderung für die jeweiligen Zielsysteme dar.

Im Rahmen des epSOS-Projektes lag der Fokus auf grenzüberschreitenden Szenarien. Die Ausgangssituation waren also Kommunikationsszenarien, in denen die Partner unterschiedliche lokale Klassifikationen und Codesysteme

und zudem verschiedene Sprachen bzw. Sprachvarianten verwendeten. Um hier semantische Interoperabilität zu erreichen, musste die medizinische Information in einer Weise repräsentiert werden, die sowohl eine sprachliche Repräsentation in mehreren Quell- und Zielsprachen unterstützt, als auch eine standardisierte einheitliche Codierung ermöglicht, um den Übersetzungsprozess maschinell zu unterstützen.

Die vorgesehenen epSOS-Anwendungen „Patientenkurzakte“ (Patient Summary, PS) und „elektronisches Rezept“ (ePrescription, eP) mussten im gegebenen zeitlichen Rahmen des Projekts zur Einsatzreife in den angestrebten Pilotierungsszenarien gebracht werden. Bei der Umsetzung wurde daher eine Strategie verfolgt, die auf bestehenden Begriffssystemen aufsetzte, also zu direkt implementierbaren Ergebnissen führte, dabei aber andererseits auch moderne ontologiebasierte Ansätze und neuartige Vorgehensweisen bei der Abstimmung zwischen den Partnern betrachtete und bewertete.

Neben den genannten epSOS Dokumenten Patient Summary und ePrescription wurden im Projekt auch eDispensierung (eDispense eD) betrachtet, sowie in einer späteren Projektphase auch Health Care Encounter Report (HCER) und Medication Related Overview (MRO). Die für die zunächst pilotierten Dokumente (eP, PS) erarbeiteten Ergebnisse konnten mit nur geringen Anpassungen auch für die weiteren Dokumente wiederverwendet werden.

13.1 Codesysteme und Value Sets

Ist es möglich, alle Anforderungen durch ein einziges Begriffssystem zu erfüllen? Gibt es ein Codesystem für alle Zwecke? Diese Frage wurde zu Beginn des Projektes zwischen den Teilnehmern diskutiert. Man war sich einig: Das einzige Codesystem, das diese Anforderung potenziell erfüllen könnte ist SNOMED CT. Für einen Teil der medizinischen Information, etwa im Fall von Allergien oder Impfungen, war SNOMED CT in der Tat die einzige Möglichkeit, andere Terminologien sind nicht bekannt, nicht einmal mit geringerer Granularität.

Neben diesen medizinischen und semantischen Aspekten mussten hier aber auch organisatorische, politische, prozedurale und wirtschaftliche Fragen betrachtet werden. Die Nutzung von SNOMED CT ist nur den Ländern möglich, die eine entsprechende Lizenz erworben haben, indem sie z.B. Mitglied der IHTSDO geworden sind. Für mehrere teilnehmende Länder war das nicht der Fall, so dass es keine Rechtsgrundlage für eine Nutzung von SNOMED CT über das Projektende hinaus gab. Da in diesen Ländern SNOMED CT in der Fläche nicht verwendet wird, ist überdies eine – allerdings mit entsprechenden Risiken verbundene – Transcodierung der Inhalte erforderlich. Man einigte sich daher, nicht durchgehend SNOMED CT für alle Inhalte zu verwenden, sondern stattdessen für jedes einzelne Datenfeld ein jeweils geeignetes Codesystem zu suchen, dabei aber SNOMED CT als Referenzterminologie im Hintergrund zu nutzen.

13.2 Methodik

Für jedes Datenelement in den epSOS-Dokumenten für Patient Summary und ePrescription (PS, eP) sollte eine geeignete Terminologie identifiziert werden, die auf einem international standardisierten Codesystem beruht. Auf dieser Basis sollten dann geeignete Wertelisten (Value Sets) aufgestellt werden, die jeweils nur die Konzepte enthalten, die an dieser Stelle medizinisch sinnvoll sind. Später wurden die so vereinbarten Wertelisten dann als Basis für die Spezifikation der übrigen epSOS-Dokumente verwendet.

Die Teilnehmer des Projekts einigten sich auf HL7 CDA als technische Grundlage der Datenstrukturen und IHE-XDS für den Datentransport. Es wurde entschieden, dass jeweils zwei Repräsentationen der Inhalte übermittelt werden sollen:

1. Eine in CDA eingebettete PDF-Repräsentation der ursprünglichen menschenlesbaren Inhalte („narrative“) und
2. Eine Repräsentation in CDA, in der zusätzlich die codierten Informationen in entsprechenden XML-Elementen des „CDA Level 3“ enthalten sind. Dies ist relevant im Hinblick auf das Ziel der automatisierten Übersetzung von Konzeptbezeichnungen (Translation) und der Abbildung auf andere Codesysteme (Transcoding) für unterschiedliche Zielsysteme. Bei der Übermittlung wird das Quelldokument in mehreren Schritten durch übersetzte bzw. transcodierte Informationen ergänzt, die ursprünglichen Daten bleiben dabei erhalten und werden ebenfalls übertragen.

Beide Repräsentationen werden im CDA-Format übertragen und enthalten identische Kontextinformationen im CDA-Header.

Bei der Festlegung der Codesysteme und Value Sets wurde wie folgt vorgegangen: Zunächst wurden alle Datenelemente in Header und Body der in CDA abgebildeten Strukturen für PS bzw. eP untersucht. Für jedes Element wurde geprüft, ob es in existierenden Spezifikationen bereits entsprechende Inhalte mit äquivalentem Kontext gibt (z.B. IHE PCC CDA Content Modules). Falls es solch eine Übereinstimmung gab, wurde untersucht, ob die funktionalen Anforderungen übereinstimmen und ob das im Content Module verwendete Codesystem (oder Teile davon) im Hinblick auf medizinische Aspekte und auf Patientensicherheit geeignet ist.

Für den Fall, dass keine passende Übereinstimmung vorlag, hat das epSOS Semantic Core Expert Team Auswahl- und Entscheidungskriterien für die Auswahl von Codesystemen und Value Sets festgelegt[epSOS, 2010].

13.3 Auswahlkriterien für Codesysteme

- Internationale Verwendung: Von Vorteil ist hier die Erarbeitung durch Fachleute mit entsprechender Erfahrung in der Implementierung und Anwendung von Begriffssystemen und Terminologien. Meist liegen Leitlinien für Wartung und Lokalisierung vor.
- Verwendung in den teilnehmenden Ländern: Auf Basis von Umfragen wurde festgestellt, welche Codesysteme in den teilnehmenden Ländern bereits im Einsatz sind.
- Vorliegende Übersetzungen in verschiedene Sprachen: Dadurch wird der Aufwand für die Übersetzung drastisch reduziert. Vorliegende Übersetzungen sind üblicherweise bereits fachlich validiert worden, Anpassungen an neue Versionen finden regelmäßig statt.
- Definierter Prozess zur Pflege des Codesystems: Vorteilhaft ist eine regelmäßig erfolgende Pflege des Systems, mit geeigneten Verfahren für die Distribution neuer Versionen sowie für die Unterstützung der Nutzer.
- Mehrsprachigkeit, vorhandene Systeme oder Dienste zur Übersetzung in verschiedene Sprachen: In welchem Maße Übersetzungen und Transcodierungen („Mappings“) zwischen zwei unterschiedlichen Codesystemen bereits vorliegen bzw. methodisch unterstützt werden, hat immense Auswirkungen auf Kosten und Risiken bei der Nutzung.
- Kosten für Lizenzen, Implementierung und Pflege: Die Kosten für Nutzung, Implementierung und Wartung können erheblich sein, selbst wenn die meisten Codesysteme für die Phase der Forschung und Entwicklung bzw. Pilotierung kostenfrei zur Verfügung stehen.
- Einfachheit der Implementierung: Die Implementierung sollte möglichst einfach sein und auf bewährten Vorgehensweisen basieren, sowohl in Bezug auf die Syntax als auch auf die fachlichen Inhalte.

Nach der Identifikation geeigneter Codesysteme wurde für jedes Datenelement ein „Value Set“ festgelegt, das alle Konzepte umfasst, die als Ausprägungen des Datenelements in diesem Zusammenhang in Frage kommen. Es wurde entschieden, dass in einem Value Set nur Codes aus jeweils einem einzigen Codesystem verwendet werden sollen. Diese Designentscheidung stellte sich später in einigen Fällen als signifikante Einschränkung heraus.

13.4 Auswahlkriterien für Value Sets

- Das Codesystem wird von einer maßgeblichen Anzahl der Teilnehmer verwendet und der Umfang des Codesystems stimmt mit den gegebenen Anforderungen des Use Case überein: Das gesamte Codesystem wird als Value Set übernommen. (Beispiel: ATC für Wirkstoffe)
- Das Codesystem wird von einer maßgeblichen Anzahl der Teilnehmer verwendet und es sind Übersetzungen in alle Sprachen vorhanden. Das

Codesystem deckt die Anforderungen ab, enthält jedoch darüber hinaus zusätzliche Konzepte, die für das Datenelement im Rahmen der vorliegenden Anforderungen nicht gültig, nicht erforderlich oder nicht sinnvoll sind: Die inhaltlich passenden Konzepte werden ausgewählt. (Beispiele: aus den ISO 3166 Ländercodes werden nur die Codes der teilnehmenden Länder benötigt, EDQM Codes für Darreichungsform)

- Falls das Codesystem nicht von einer maßgeblichen Anzahl der Teilnehmer eingesetzt wird, ist eine Transcodierung erforderlich, ggf. liegen keine Übersetzungen vor: Es erfolgt eine sorgfältige fachliche Analyse (durch Ärzte, Apotheker), um die relevanten Konzepte zu identifizieren. (Beispiele: SNOMED-CT basierte Value Sets, ICD-10) mit anschließender Qualitätssicherung

13.5 Ausgewählte Codesysteme

45 unterschiedliche Value Sets basierend auf 24 Codesystemen wurden identifiziert. Damit konnten alle erforderlichen Informationen in den fünf genannten Dokumenten abgebildet werden. Einige Value Sets gelten für alle Dokumente, andere sind spezifisch für ein bestimmtes Dokument.

Die Datenelemente und zugehörigen Konzepte lassen sich vier Kategorien zuordnen:

- Konzepte für die grundlegenden technischen Anforderungen des Datenaustauschs:

Hier konnten die von HL7 CDA bzw. IHE XDS vorgegebenen Codes übernommen werden, lediglich die Codes und mehrsprachige Bezeichnungen der Dokumentabschnitte mussten für das Projekt ausgearbeitet und zwischen den Teilnehmern abgestimmt werden. Hierbei ergab sich die Frage nach der Zuständigkeit für die Weiterentwicklung und ggf. langfristige Nutzung dieser Inhalte.

DocumentCodes, SectionCodes, HL7 NullFlavors (nähere Angaben zu nicht enthaltenen Daten), IHEActionCode (structural attributes, Charakterisierung der Datenstrukturen, die die relevanten Daten enthalten)

- Konzepte zur Beschreibung des Kontexts:

Die für den Header in HL7 CDA bzw. in den IHE-Profilen vorgegebenen Codes erfüllen weitestgehend die funktionalen Anforderungen.

Patient, Heilberufler, Inhaltsangaben zum Dokument (Metadaten, CDA Header), Beschreibung der Rollen von Personen/Institutionen im Kontext des Dokuments. (name prefixes, suffixes; gender; confidentiality; contact information, communication data, nationality, language, professional function)

- Für ePrescription/eDispense spezifische Konzepte:

Hier werden Konzepte für verschiedene Aspekte der Arzneimittelverordnung und -abgabe beschrieben: Identifikation von Arzneimitteln, Wirkstoff, Darreichungsform, Anwendungsart, ggf. Substitution etc. (ATC

vs. SNOMED CT, EDQM vs. SNOMED CT). Entscheidungen fielen häufig auf Basis monetärer Argumente: Einige Länder sind nicht IHTSDO-Mitglieder, EDQM als europäisches Projekt mit wissenschaftliche fundierten Konzepten wurde gegen das eher historisch gewachsene Value Set von HL7 bevorzugt. Für Substitution wurde ein HL7 Codesystem identifiziert und als geeignet beurteilt.

- Für PS (patient summary) spezifische Konzepte:
Für die Repräsentation von Allergien wurden diverse Modelle diskutiert. Ein wichtiger Diskussionspunkt war z.B. die Unterscheidung zwischen Allergie und Intoleranz. Die getroffene Einigung hängt stark von der Wahl des konzeptionellen Modells für Allergien ab, für einige Metainformationen zu Allergien und nicht-arzneimittelinduzierte Allergien wurde SNOMED CT gewählt (Reaction Allergy, AdverseEventType, Allergen No drugs), für die arzneimittelbedingten Allergene hingegen ATC.

Bei Diagnosen hat die Verwendung von ICD-10 bzw. ICD-9 in praktisch allen teilnehmenden Ländern den Ausschlag gegeben, hier ist SNOMED CT nicht zu verwenden. Die Bereitstellung einer ICD-10 Übersetzung ergibt sich als Notwendigkeit für die Interoperabilität auch für die Länder, in denen ICD-9 im Einsatz ist. Die Existenz von (tlw. unvollständigen) SNOMED CT und ICD-9 Mappings zu ICD-10 erleichterte die Übersetzungen. ICPC wurde nicht verwendet, da es nicht weit verbreitet ist.

SNOMED CT wurde im Rahmen des epSOS-Piloten für zahlreiche andere medizinische Konzepte verwendet, für die sich keine andere geeignete und weit verbreitete Terminologie fand. Im Einzelnen sind das die Inhalte für Impfstoffe, Prozeduren, die Blutgruppe und Medizinprodukte, wie etwa Implantate.

Für Maßeinheiten wurde UCUM gewählt.

Für etliche Datenelemente gab es auch in den untersuchten IHE-Profilen bzw. in HL7 CDA schon Vorgaben, die unverändert übernommen wurden (Beispiele: LOINC für Blutdruck und Informationen zu Schwangerschaft; SNOMED CT für Severity, SocialHistory u.a.; HL7 z.B. für RoleCode, TimingEvent)

13.6 Empfehlungen/Erfahrungen/Lessons learnt

Kriterien für die Auswahl von Codesystemen und Value Sets sollten möglichst frühzeitig festgelegt werden, um sich später im Verlauf des Abstimmungsprozesses auf diese Kriterien beziehen zu können.

IP und Lizenzierung: Semantische Interoperabilität baut auf Arbeiten und Ergebnissen internationaler SDOs auf. Enge Abstimmung mit den SDOs sollte bereits vor Beginn der semantischen Arbeiten erfolgen. Frühzeitig sollten auch die Nutzungsbedingungen, IP-Rechte und Copyrights geklärt werden, einerseits um ein Projekt nicht zu verzögern, andererseits um sicherzustellen,

dass Ergebnisse auch nach Projektende genutzt werden können. Viele Entscheidungen innerhalb des epSOS-Projektes wären wahrscheinlich einfacher gewesen, wenn Vereinbarungen mit den SDOs über eine freie Nutzung nach Projektende hätten getroffen werden können.

Tooling: Die Zusammenstellung der Value Sets erfolgte zunächst über einfache Tabellen (Excel-spreadsheets). Angesichts des Umfangs der Inhalte führte dies schnell zu Fehlern und mangelhafter Dokumentation bzw. Nachvollziehbarkeit (Versionierung, Change Logs). Später wurde ein Terminologieserver mit geeigneten IT-Werkzeugen und Workflows für die Qualitätssicherung eingesetzt. Es wird empfohlen, ein entsprechendes Terminologie-Management-System zu verwenden und dabei insbesondere die Zugänglichkeit für Mediziner sicherzustellen. Alle Ergebnisse und Dokumentationen sollten in einem zentralen Repository abgelegt werden, das allen Projektbeteiligten zugänglich ist.

Austauschformate wie ClaML (Classification markup Language) und Services (IHE SVS) für den Import von Codesystemen und Value Sets sollten festgelegt werden, um Aufwand und Fehler bei der Konvertierung von Datenformaten zu vermeiden.

Metadaten für Value Sets, die den Kontext der Definition beschreiben, machen vorangegangene Diskussionen und Entscheidungen deutlich und erleichtern so die Wiederverwendung oder ggf. Anpassung existierender Value Sets.

Referenzen

- [epSOS, 2010] European Patients Smart Open Services (2010): Smart Open Services for European Patients – Work Package 3.5 – Semantic Services Definition D3.5.2.
- [Sonne Hansen, A. et al., 2013] epSOS (2013): The experience of selecting the code systems for the development of the epsos master value catalogue (MVC).