

Projektantrag

eingereicht von: KOST-Geschäftsstelle in Zusammenarbeit mit Lambert Kansy, StABS, und Martin Lüthi, StASG

geht an: Projektgruppe 17-034 und KOST-Steuerungsausschuss

Betreff: KOST Fortsetzungsprojekt 17-034 xlsadg Ontologie

Stand:

1 Ausgangslage

Im Rahmen des bentō-Projekts ist 2010 das Metadatenschema xlsadg zur Beschreibung von Descriptive Information nach ISAD(G) entstanden. Das XML Schema wurde damals in der Version xlsadg 1.6 veröffentlicht und hat seit dann verschiedene Anpassungen und Erweiterungen erfahren. Die Entwicklung von xlsadg ist auch Teil einer KOST-Initiative, die automatische Verzeichnung von Archivalien zu fördern. Ein Crosswalk zwischen eCH-0160 SIP-Metadaten und Descriptive Information nach xlsadg ist heute ein Teil von eCH-0160.

Das Stadtarchiv Zürich hat zusammen mit den Staatsarchiven Appenzell Ausserrhoden, Basel-Stadt, Bern, Luzern, St.Gallen und Thurgau und der KOST-Geschäftsstelle im KOST Projekt 17-034 xlsadg Änderungen, Anpassungen und Erweiterungen am Standard zusammengeführt und vereinheitlicht und darauf aufbauend eine neue Version xlsadg 2.1 spezifiziert: <https://kost-ceco.ch/cms/verzeichnungsschnittstellen-xisadg.html>

2 Projektziel

Das Projektziel des KOST-Fortsetzungsprojekt 17-034 «xlsadg Ontologie» ist es, das heutige [xlsadg Datadictionary](#) in eine Ontologie in OWL¹ umzuschreiben. Weil das keine an und für sich neue Aufgabe ist, soll das Projekt als Fortsetzungsprojekt mit der gleichen Nummer 17-034 weitergeführt werden.

Zum besseren Verständnis und weil der Begriff neu ist, ein kurzer Abriss darüber, was man sich unter einer Ontologie vorstellen soll:

Das *Semantic Web* basiert darauf, dass Wissen auch maschinenlesbar erfasst, verarbeitet, wiederverwendet und kommuniziert wird. Ontologien unterstützen bzw. sind die Grundlage dieser Aufgaben.

Eine Ontologie im Bereich des Semantic Web erfasst in formalisierter Weise und in digitaler Form die Struktur einer Domäne (eines Teilbereichs des Weltwissens), sie ist eine explizite Konzeptualisierung² dieser Domäne. Die Konzeptualisierung beschreibt das Wissen über die Domäne, nicht jedoch über den jeweiligen Sachverhalt in der Domäne. Mit anderen Worten, die Konzeptualisierung ändert sich nicht oder nur sehr selten, der

¹ OWL (Web Ontology Language) ist eine formale Sprache, die entwickelt wurde, um Wissen über Dinge und Beziehungen zwischen Dingen darzustellen. Sie ist Teil der W3C *Semantic Web Initiative*. <https://www.w3.org/OWL/>

² Ein Konzept für diesen Teilbereich des Wissens entwerfen

jeweilige Sachverhalt hingegen schon. Ontologie ist die formale Beschreibung dieser Konzeptualisierung unter Verwendung einer Modellierungssprache.

Eine Ontologie beschreibt eine Domäne, während eine Wissensbasis (basierend auf einer Ontologie) einen bestimmten Sachverhalt beschreibt.³ Jedes wissensbasierte System oder Agent hat seine eigene Wissensbasis, und nur was mit einer Ontologie ausgedrückt werden kann, kann in der Wissensbasis gespeichert und verwendet werden. Wenn ein Agent mit einem anderen Agenten kommunizieren möchte, verwendet er die Konstrukte aus einer Ontologie. Um in der Kommunikation verstanden zu werden, müssen Ontologien zwischen den Akteuren geteilt werden.

Im Rahmen der Standardentwicklung des Semantic Web wurden bereits früh Beschreibungssprachen (RDFs und OWL) entwickelt, um Ontologien in RDF abbilden zu können.

Genaugenommen handelt es sich dabei um ein Umschreiben⁴ des *Data Dictionaries*⁵ in eine Ontologie mithilfe der formalen Sprache OWL. Damit kann zwischen den in xlsadg vorliegenden Daten und dem als Ontologie vorliegenden *Data Dictionary* maschinell eine Verbindung hergestellt werden. Abfragen auf *Descriptive Information* in xlsadg sind dann mit den Mitteln von *Linked Data* im Sinne des *Semantic Webs* möglich. Dies ist auch deshalb von Relevanz, da eine Reihe von Archiven xlsadg für die Erfassung der archivischen deskriptiven Metadaten in AIPs verwenden.

3 Arbeitsschritte

Da in diesem Fortsetzungsprojekt die xlsadg Ontologie konzeptuell bereits im *Data Dictionary* definiert ist, braucht es keine Projektsitzungen zu inhaltlichen Fragen.

Der Technologietransfer, bzw. das Umschreiben des xlsadg *Data Dictionaries* in eine OWL Ontologie, wird am besten von einer Fachperson, konkret Claire Röthlisberger von der Geschäftsstelle vorgenommen. Martin Lüthi und Lambert Kansy haben sich bereit erklärt, sie in inhaltlichen und technischen Fragen zu unterstützen.

Die Projektgruppe übernimmt eine abschliessende Reviewfunktion.

4 Arbeitsaufwand und Budget

Um den Aufwand möglichst gering zu halten, soll das Projekt via Mail und Sharing über das bereits bestehende xlsadg Repository abgewickelt werden:

<https://github.com/KOST-CECO/xlsadg>.

Der initiale Aufwand bei der Geschäftsstelle besteht aus einem nicht unerheblichen Weiterbildungsanteil von etwa 10 Arbeitstagen, was die neue Technologie betrifft.

³ Konkret und vereinfacht beschreibt ISAD(G) im Sinne einer Ontologie Aufbau und Verhältnis von Verzeichnungseinheiten, die Wissensbasis hingegen ist die gesamte Architektur bzw. die Findmitteldaten eines Archivs. Agenten können diese Wissensbasis z.B. über ein *Query* austauschen und weiterverarbeiten.

⁴ Die Domäne *Descriptive Information* ist im xlsadg *Data Dictionary* bereits konzeptuell beschrieben.

⁵ Ein *Data Dictionary*, Datenwörterbuch oder Metadaten-Repository, ist eine Sammlung von Informationen über Daten, deren Bedeutung, Herkunft, Verwendung und Formate sowie deren Beziehungen zu anderen Daten. Diese Repository kann, muss aber nicht als Tabelle aufgebaut sein.

Dieser Aufwand kann aber, weil grundsätzlich notwendig, nicht vollständig diesem Projekt zugeordnet werden.

Für die Umsetzung xlsadg *Data Dictionaries* zu OWL Ontologie selber sind etwa 6 Arbeitstage zu veranschlagen.

Die Staatsarchive Basel-Stadt und St.Gallen müssen etwa je zwei Arbeitstage, wahrscheinlich jeweils mit einem Meeting für Technologiediskussionen, veranschlagen.

Bei den beteiligten Archiven wird mit einem halben Tag pro Archiv für die Review zu rechnen sein.

5 Zeitplan (Projektbeginn und -dauer)

Das Projekt soll möglich schnell gestartet und bis Ende 2018 abgeschlossen sein. Seitens StABS besteht ein akutes Bedürfnis nach einer xlsadg-Ontologie.

Nr.	Beschreibung	Termin	Verantwortlich	Status
1	Bewilligung durch Steuerungsausschuss	08.10.2018	Geschäftsstelle	P
2	Technologie Review und Einarbeitung in OWL	08.10.2018 – 31.10.2018	Geschäftsstelle CR	P
3	Vorgehen besprechen mit StABS und StASG	2 Termine Q4 2018	Geschäftsstelle StABS / StASG	P
4	OWL Ontologie erstellen	bis Ende Nov.2018	CR	P
5	Review durch Projektgruppe	Ende Q4 2018	Projektgruppe	P
6	Veröffentlichung	Ende Q4 2018	Geschäftsstelle	P
7	Statusbericht	Abschluss	Geschäftsstelle	P

Status: P = geplant, E = erledigt, A = in Arbeit