

Projektabschlussbericht

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

geht an: Projektgruppe 17-034 und KOST-Steuerungsausschuss

Betreff: KOST Fortsetzungsprojekt 17-034 xlsadg Ontologie

1 Management Summary

Im Rahmen des bentö-Projekts ist 2010 das Metadatenschema xlsadg zur Beschreibung von Descriptive Information nach ISAD(G) entstanden. Das *Data Dictionary*¹ und das XML Schema wurden damals in der Version xlsadg 1.6 veröffentlicht und haben seitdem 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 zusammengeführt und vereinheitlicht sowie darauf aufbauend die neue Version xlsadg 2.1 von *Data Dictionary* und XML-Schema spezifiziert².

Das KOST-Fortsetzungsprojekt 17-034 «xlsadg Ontologie» hatte ursprünglich das Ziel, das heutige xlsadg *Data Dictionary* in eine Ontologie in OWL³ umzuschreiben. Wie sich im Verlauf der Arbeiten zeigte, war eine 1:1 Übertragung des *Data Dictionaries* in eine RDF⁴-Ontologie nicht möglich respektive nicht sinnvoll.

Das Projektteam hat daraufhin beschlossen, die Unterschiede im Abschlussbericht festzuhalten, die RDF-Ontologie als Betaversion zu publizieren und das Projekt abzuschliessen. Anhand der Rückmeldungen und Erfahrungen zur dieser Beta-RDF-Ontologie (xlsadgR v0.7.3) wird später entschieden, wie das weitere Vorgehen in Sachen xlsadg respektive xlsadgR aussehen soll.

¹ 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.

² <https://kost-ceco.ch/cms/verzeichnungsschnittstellen-xisadg.html>

³ 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/>

⁴ RDF (Resource Description Framework) ist ein W3C Standard und wurde zur Beschreibung von Metadaten konzipiert. <https://www.w3.org/RDF/>

2 Abweichungen zwischen xlsadg 2.1 und xlsadgR 0.7.3

2.1 Gründe für die Abweichungen

Die Annahme im Projektantrag, dass alles konzeptuell bereits im *Data Dictionary* definiert sei und es deshalb keine Projektsitzungen zu inhaltlichen Fragen brauche, war zu optimistisch.

Praktische Umsetzung des xlsadg *Data Dictionaries*

Beim Technologietransfer bzw. dem Umschreiben des xlsadg *Data Dictionaries* in eine OWL Ontologie durch Claire Röthlisberger tauchten Fragen auf, die in der Ontologie genauer definiert werden mussten, als es im *Data Dictionary* getan worden war. Zudem sind die xlsadg-Umsetzungen bereits in den beteiligten Staatsarchiven St. Gallen und Basel-Stadt unterschiedlich, insbesondere im Bereich der Repräsentationen respektive dem Abbilden von Archivobjekten in verschiedenen materiellen Ausprägungen. Es stellte sich schnell heraus, dass hier eine grosse Lücke im Konzept von xlsadg ist. Die technische Umsetzung des xlsadg *Data Dictionaries* in XSD ist ein Bestandteil von xlsadg 2.1, aber auch nicht in allen Punkten mit dem *Data Dictionary* übereinstimmend.

Aufbau des xlsadg *Data Dictionaries*

Das xlsadg *Data Dictionary* besteht genauer betrachtet aus drei unterschiedlichen Informationstypen; die Struktur, die Restriktionen sowie die Angaben zu Vererbung und Aggregation. Im Rahmen des Technologietransfers gewann das Projektteam folgende Erkenntnisse:

1. Für die Abbildung der Struktur ist OWL geeignet und so kann die Ontologie durch Maschinen verstanden werden.
2. Restriktionen können aufgrund der «Open-World» Annahme nicht in OWL abgebildet werden. Restriktionen in der Ontologie sind Shapes, und für diese Abbildung wurde SHACL⁵ verwendet. Die Shapes sind in einer eigenen Datei definiert.
3. Die Angaben zur Aggregation und Vererbung können nicht in die Ontologie einfließen, da dafür (noch) keine formale Sprache existiert.

xlsadgR ist entsprechend eine RDF-Zusammenstellung, welche die OWL *Ontology* und die SHACL *Shapes* beinhaltet; als Serialisierungsformat wurde Turtle⁶ verwendet.

Konzeptuelle Rückmeldungen zu xlsadg beim Review

Aus dem öffentlichen Workshop im Sommer 2019⁷ mit anschliessender Review-Runde kamen hauptsächlich konzeptuelle Rückmeldungen zu xlsadg respektive xlsadgR zurück. Diese flossen bislang lediglich in die Ontologie xlsadgR v0.7.3 ein.

⁵ SHACL (Shapes Constraint Language) ist eine W3C-Spezifikation zur Definition von Bedingungen bei RDF-Daten und deren Validierung. <https://www.w3.org/TR/shacl/>

⁶ Turtle (Terse RDF Triple Language) wird vom W3C zur Serialisierung von RDF-Graphen empfohlen. Zudem sind Turtle-Dateien (.ttl) kompakt und auch für Menschen lesbar. <https://www.w3.org/TR/turtle/>

⁷ https://kost-ceco.ch/cms/aid=759.html?newsdetail=20190515-148_linked-data-workshop

2.2 Die wichtigsten Abweichungen

In der Ontologie hat es zwei zentrale Klassen, die *ArchivalDescription* (Verzeichnungseinheit) und *MaterialExpression* (Archivobjekt):

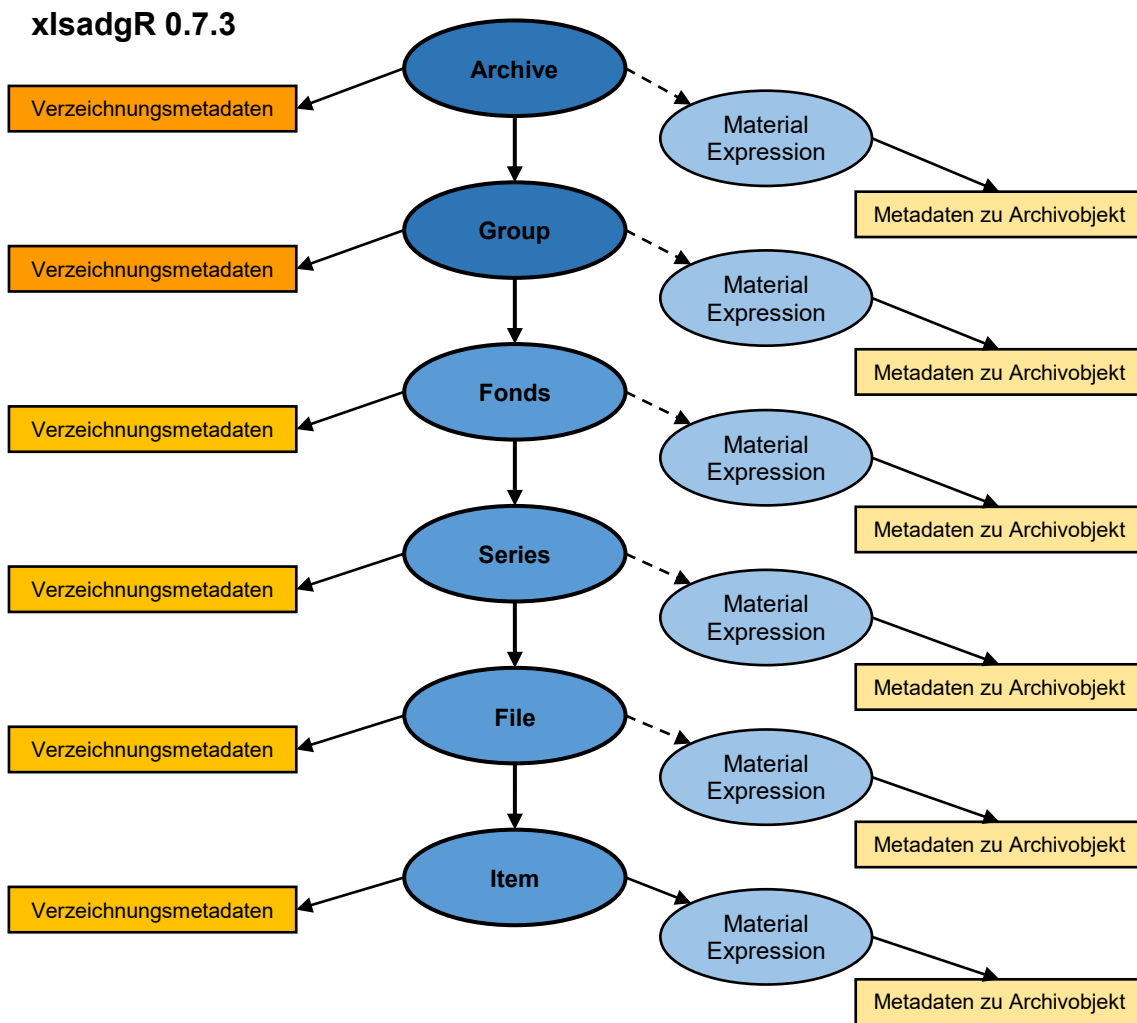
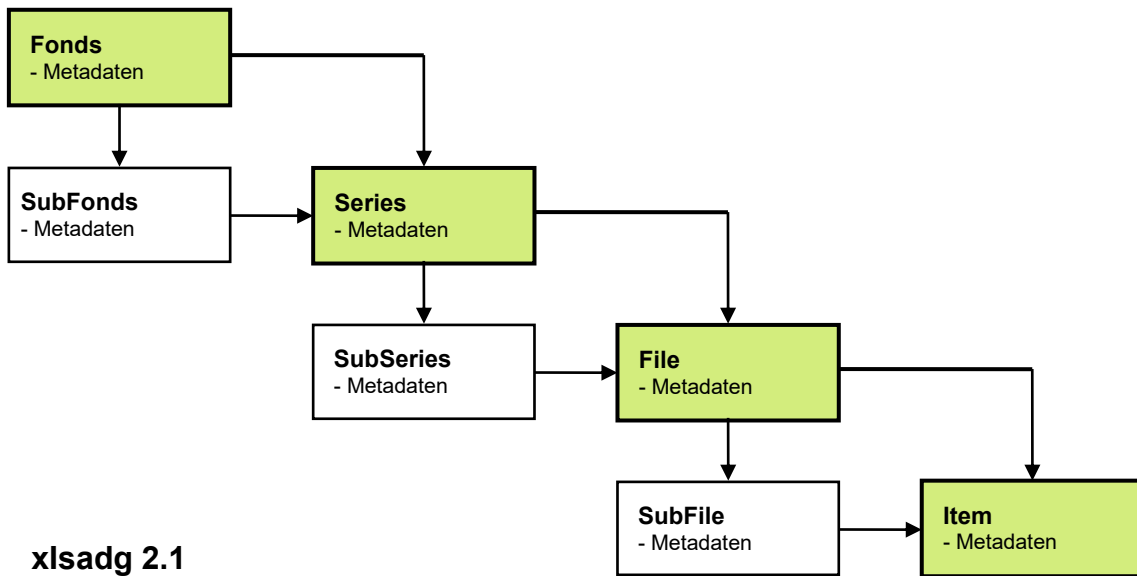
- ***ArchivalDescription*** enthält die wichtigsten Metadaten der Verzeichnungseinheit, welche unabhängig von der materiellen Ausprägung gleichbleibend sind. Dies ist zum Beispiel der Titel [1.2], die Provenienz [2.1] und der Inhalt [3.1]. Eine *ArchivalDescription* kann auf die verschiedenen *MaterialExpression* dieser Verzeichnungseinheit und auf andere *ArchivalDescription* verweisen.
- ***MaterialExpression*** enthält Metadaten des Archivobjektes und deren Besonderheiten. Die wichtigsten Metadaten sind unter anderem den Umfang [1.5a], die Archivalienart [1.5b] und die Zugangsbestimmungen [4.1]. Eine *MaterialExpression* kann auf eine oder mehrere andere *MaterialExpression* (Representation oder Version des vorliegenden Archivobjektes) verweisen.

Die Verzeichnungsstufen wurden oberhalb *Fonds* (Bestand) mit *Archive*⁸ (Archiv) und *Group*⁹ (Gruppe) erweitert. Im Gegenzug wurden die Teilstufen *SubFonds*, *SubSeries* und *SubFile* entfernt:

- Mit ***Archive*** und ***Group*** kann die Ontologie das ganze Archiv abbilden. Das bedeutet, dass in diesem Bereich keine zusätzliche Ontologie notwendig ist.
- Mit dem **Entfallen von *SubFonds*, *SubSeries* und *SubFile*** ist die Ontologie flexibler aufgebaut. Die Verzeichnungsstufen sind in der Ontologie kein Metadatum, sondern Subklassen von *ArchivalDescription*. Dies ist notwendig, damit die Hierarchiebildung mit SHACL validiert werden kann.

⁸ *Archive* ist die oberste Verzeichnungsstufe und soll in der Regel für die Abbildung von «Archiv», «Institution» und «Archivlandschaft» verwendet werden.

⁹ *Group* ist die Verzeichnungsstufe zwischen *Archive* und *Fonds*. Diese Zwischenstufe wird bis anhin meist als «Abteilung», «Departement» oder «Tektonikgruppe» bezeichnet.



Vereinfachte Verzeichnungsstufen und Metadaten

2.3 Übersicht der Abweichungen

Die Abweichungen werden hier aufgelistet und der Typ der Abweichung festgehalten. Es handelt sich um neue (Neu) oder gelöschte (Gelöscht) Elemente; Elemente, die neu Teil der materiellen Ausprägung (MaterialEx.) sind; Realisierung durch mehrere Wertebereiche (Datentypen) oder durch einen neuen Aufbau in Form einer neuen Klasse (Aufbau).

Nr	Bezeichnung	Neu	Gelöscht	MaterialEx.	Datentypen	Aufbau
1.1	Signatur + <i>hasReferenceCodeOld</i>	X			X	
1.3	Entstehungszeitraum / Laufzeit					X
1.4	Verzeichnungsstufe <i>ArchivalDescription</i> + <i>Archive</i> + <i>Group</i> = Bestand <i>Fonds</i> - Teilbestand = Serie <i>Series</i> - Teilserie = Dossier <i>File</i> - Teildossier = Dokument <i>Item</i> - Repräsentation	X X	X X			X X
1.5	Umfang (Menge und Abmessung)			X	X	
2.1	Name der Provenienzstelle					X
2.2	Verwaltungsgeschichte/Biografische Angaben					X
3.1	Form und Inhalt <i>hasScope</i> <i>hasContent</i>					x
4.1	Zugangsbestimmungen + <i>hasMetadataPrivacyProtection</i> + <i>hasIntellectualPropertyRight</i>	X X		X		x
4.4	Physische Beschaffenheit und tech. Anforderungen			X		
5.1	Aufbewahrungsort der Originale			X	X	
5.2	Kopien bzw. Reproduktionen			X	X	
5.3	Verwandte Verzeichnungseinheiten				X	
5.4	Veröffentlichungen				X	
Add	Referenz auf das AIP			X	X	
Add	Referenz auf Dateien			X	X	
Add	Referenz auf technische Metadaten			X	X	
ME	<i>MaterialExpression</i> <i>hasMaterialExpression</i>	X		X		
ME	<i>hasMaterialExpressionRepresentation</i>	X		X		
ME	<i>hasMaterialExpressionVersion</i>	X		X		
ME	<i>hasMaterialExpressionDate</i>	X		X		

3 Abschlussergebnisse

Es wurden folgende Ergebnisse erarbeitet:

- RDF-Ontologie xlsadgR v0.7.3: xlsadgR_Ontology.owl
- SHACL-Restriktionen zur Ontologie: xlsadgR_Shapes.ttl
- Grafische Übersicht der Ontologie: xlsadgR_0.7.3_VOWL.svg

Diese Abschlussergebnisse können hier heruntergeladen werden:

https://github.com/KOST-CECO/ontologies/releases/tag/xlsadgR_v0.7.3

Illustriert werden diese Dokumente mit diversen Beispielen.

3.1 xlsadgR_Ontology v0.7.3

Die Ontologie «xlsadgR_Ontology» wird auf GitHub publiziert und gepflegt:

https://github.com/KOST-CECO/ontologies/blob/master/xlsadgR_Ontology.owl

Die Beta-Version 0.7.3 ist vom 19.03.2020.

Die Ontologie wurde bewusst in mehreren Abschnitten aufgebaut, zuerst die ArchivalDescriptions mit ihren Metadaten und dann die MaterialExpressions. Innerhalb von ArchivalDescription und MaterialExpression wurde manuell anhand der ISAD(G)-Nummern sortiert und eingerückt. Dies soll zur besseren Lesbarkeit beitragen.

3.2 xlsadgR_Shapes v0.7.3

Die Restriktionen sind in «xlsadgR_Shapes» festgehalten und werden auf GitHub publiziert und gepflegt: https://github.com/KOST-CECO/ontologies/blob/master/xlsadgR_Shape.ttl

Die Beta-Version 0.7.3 ist vom 19.03.2020.

Mit «xlsadgR_Shapes» könne Einträge auch validiert werden¹⁰. Der Aufbau ist bewusst gleich gehalten wie derjenige der Ontologie.

3.3 Grafische Übersicht: VOWL

Für die grafische Übersicht wurde WebVOWL¹¹ verwendet. Damit der Nutzen noch grösser ist, wurde die erhaltene Darstellung mit weiteren Farben und Verweise auf Dublin Core ergänzt. Gleichzeitig wurde auch der Text in den Klassen wo nötig auf mehrere Zeilen aufgeteilt. Diese Übersicht einer Ontologie (VOWL¹²) ist wiederum auf GitHub (https://github.com/KOST-CECO/ontologies/releases/download/xlsadgR_v0.7.3/xlsadgR_0.7.3_VOWL.svg) und in der Beilage veröffentlicht.

¹⁰ Als Validator bietet sich der SHACL Playground an: <https://shacl.org/playground/>

¹¹ <http://www.visualdataweb.de/webvowl/>

¹² VOWL (Visual Notation for OWL Ontologies) <http://vowl.visualdataweb.org/v2/>

3.4 Beispiele: Ochs & Nesslau

Damit die Ontologie und deren Anwendung verständlicher ist, werden auf GitHub auch Beispiele publiziert. Diese sind im Ordner «/samples» ersichtlich:

<https://github.com/KOST-CECO/ontologies/blob/master/samples/stabs/ochs.ttl>

<https://github.com/KOST-CECO/ontologies/blob/master/samples/stasg/nesslau.ttl>

Für die Beispiele wurde abermals Turtle als Serialisierungsformat verwendet.

Das Beispiel Ochs zeigt auf, wie allgemein der Aufbau mehrere Metadatenaussagen aussehen könnte. Das Beispiel Nesslau zeigt die Besonderheiten der Materiellen Ausprägungen auf.

4 Nutzen für die KOST-Archive

Auch wenn alle Unterlagen nur in der Beta-Version 0.7.3 vorliegen, sind diese für die interessierten Archive und der KOST wichtig. Durch die Publikation kann die Ontologie für Prototypen sowie für die Sammlung von Erfahrung im Bereich *Linked Data* genutzt werden. Zudem konnte sich das Projektteam wichtige Erkenntnisse im Bereich *Ontologies*, *Linked Data* und RDF aneignen.

5 Ausblick

Die RDF-Ontologie «xlsadgR» beruht auf xlsadg respektive ISAD(G) und enthält hauptsächlich die ursprünglichen ISAD(G)-Metadaten. Das primäre Ziel – erste Erfahrungen im Bereich der Umsetzung eines *Data Dictionaries* als RDF-Ontologie zu sammeln – wurde erreicht. Das weitere Vorgehen in Sachen xlsadgR wird bewusst offengelassen. Zum heutigen Zeitpunkt kann nicht seriös abgeschätzt werden, wann und ob der Bedarf an einer finalen «xlsadgR» Ontologie mit Restriktionen gegeben ist.

Die ICA hat am 12.12.2019 «*Records in Contexts Ontology*» (RiC-O) in der Version 0.1 publiziert¹³. RiC-O ist ebenfalls eine RDF-Ontologie, beinhaltet aber ein Metadatenset auf der Grundlage des RiC-Datenmodells, welches im Verhältnis zu ISAD(G) eine erweiterte Beschreibung und Erschliessung archivischer Daten erlauben soll.

Der VSA hat 2018 eine Projektgruppe ENSEMEN berufen, die sich mit der Harmonisierung von RiC-O in der Schweiz befasst. Erste Ergebnisse liegen auch hier vor.

XisadgR und RiC-O sind nicht aufeinander abgestimmt und beide sind Pre-Releases.

6 Beilage

Die grafische Übersicht auf der folgenden Seite kann bei Bedarf aufgezoomt werden. Beim Ausdrucken empfiehlt sich das Format A1 oder A0.

¹³ <https://www.ica.org/en/records-in-contexts-ontology>