



KERNGRUPPE ENSEMEN

14. Sitzung, 19.08.2020



TEILNEHMENDE

Anwesend

Michele Merzaghi

Martin Kaiser

Krystyna Ohnesorge

Lamber Kansy

Niklaus Stettler

Tobias Wildi

Abwesend:

Jean-Daniel Zeller

Jean-Luc Cochard



THEMEN

1. Protokoll 13. Sitzung der Kerngruppe ENSEMEN
2. Information an den Teilnehmenden der Projektgruppe
3. Stand der Arbeiten
4. Termine für die nächsten Sitzungen:
 - Besuch des VSA-Vorstands
 - Organisation von «Face-to-Face» Sitzungen
5. Varia



PROTOKOLL DER 13. SITZUNG

Entscheid

Das Protokoll der letzten Sitzung wird genehmigt und zur Veröffentlichung auf der VSA-Webseite freigegeben.



INFORMATIONEN KERNGRUPPE - 1

Information aus der Sitzung der Präsidenten der AGs mit dem Vorstand:

- Aufgrund der Corona-Situation hat die Projektleitung die Verschiebung des 5. Moduls «Linked Data für Archive – die Chance nutzen. Ein Erfahrungsbericht aus der Arbeit der VSA-Projektgruppe ENSEMEN» vom 20.11.2020 im Zyklus «Archivpraxis Schweiz» beim Bildungsausschuss beantragt. Der Termin in 2021 ist noch nicht festgelegt worden.
- Die AG RiC von EGAD nimmt nach der Sommerpause ab 35. KWO ihre Arbeit wieder auf. Wegen der Corona-Situation wird die RiC-Version 1.0 voraussichtlich nicht vor Ende 2020 fertig gestellt.



INFORMATIONEN KERNGRUPPE - 2

Information aus der Sitzung der Präsidenten der AGs mit dem Vorstand (ff):

- Innerhalb der AG RiC von EGAD sollten «Relations» standardisiert und QS für die bisherigen Arbeiten durchgeführt werden, u.a. Erfahrungen aus dem Projekt zum «Mapping SQL zu RiC» und Rückmeldungen aus dem Memoriav-Projekt berücksichtigt.
- Zum Thema RiC wurde im Juli 2020 ein Webinar organisiert. Die AG Normen und Standards prüft die Durchführung weiteren «Webinars», so dass wir beim Thema RiC / ENSEMEN mit den VSA-Mitgliedern im Kontakt bleiben.



ABSTIMMUNG DER STRUKTUR DES EINLEITUNGSDOKUMENTS

Kapitel *Struktur der Arbeit* wird ebenfalls an die aktuelle Planung angepasst:

- *Thesepapiere* zu den einzelnen Entitäten mit RIC abstimmen.
- Innerhalb RIC ein *kontrolliertes Vokabular* für die Schweiz entwerfen (Tobias)
- *Experimentelle Umsetzung* von RIC in den vorhandenen Systemen und Lösungen



PLANUNG ARBEITEN ENSEMEN - 1

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (geplant bis Ende September):

1. lindas-data.ch: Tabelle -> RDF (Jean-Luc, Status: offen)
2. Transfer des ALOD Schemas nach RIC (Jean-Luc; sample vorhanden; Status: offen)
3. Graph-DB (Niklaus + Jean-Luc, Status: offen):
Anstelle der Graph-DB wird eine andere Umgebung als Graph-DB gewählt.
4. Matterhorn METS -> RIC (Tobias, Status: offen):
Aufgrund der Corona-Situation werden die Arbeiten später starten und Ende 2020 abschliessen.
5. Thesauri/Autoritätsdaten nach RIC exportieren (Staus: offen)
6. Migration relationaler Datenbank -> RDF/RIC (Lambert):
Aufgrund der Corona-Situation werden die Arbeiten zum Mapping der «Scope-DB zu RiC» im StA BS erst im 1. Q'2021 durchgeführt.

PLANUNG ARBEITEN ENSEMEN - 7

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (ff):

8. Migrieren relationaler Datenbank eines bestehenden AIS → RDF/RiC

(Tobias), in drei Schritten:

- Konzeptuelle Arbeit: Festlegen des Mappings der einzelnen Datenbankfelder auf RiC.
- Mit Hilfe des Expressive RDF Mapper (XRM) von Zazuko (<https://github.com/zazuko/expressive-rdf-mapper>) wird das Mapping in Form einer DSL (Domain Specific Language) beschrieben. XRM erzeugt dann ein r2rml-File.
- Ontop (<https://github.com/ontop/ontop>) führt die r2rml-Umwandlung dann durch, bzw. materialisiert das r2rml-Mapping basierend auf einer konkreten Datenbank.

PLANUNG ARBEITEN ENSEMEN - 7

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (ff):

- Das Pilotprojekt mit eigenen / internen AIS wird gegenwärtig von Docuteam durchgeführt.
- Die Daten und Resultate sind im Anhang-1 zum Protokoll dokumentiert.

Schlussbemerkung (relationale DB → RiC)

- Wenn das Mapping aus einer Scope- oder CMI-Datenbank gemacht werden soll, dann können wohl grosse Teile der konzeptionelle Arbeit aus dem Pilotprojekt übernommen werden. Der 2. Arbeitsschritt (Mapping) muss allerdings praktisch neu geschrieben werden, weil dort die SQL-Befehle drin stehen.



PLANUNG ARBEITEN ENSEMEN - 2

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (ff):

6. eCH-0160/xlsadg → RDF/XML RIC (Martin):

Die Konvertierung von Metadaten aus eCH-0160 SIP (Datei: metadata.xml) nach RIC-O ist gelungen.

- Zuerst wurden die Metadaten mithilfe von «KOST Konverter - Descriptive Information 2.1» (https://kost-ceco.ch/isadg/app_beta/) in eine EAD Datei umgewandelt (in der Beta Version 129 des Konverters sind Anpassungen an den EAD Dialekt der «Archives nationales de France» gemacht worden).
- Anschliessend musste die EAD Schemadeklaration in «RIS_St.Gallen.xml» durch eine `<!DOCTYPE ead SYSTEM "ead.dtd">` Deklaration ersetzt werden.
- Im nächsten Schritt erfolgt die Konvertierung zu RIC-O mit dem Konverter «ricoconverter-1.0.0» der Archives nationales de France (<https://github.com/ArchivesNationalesFR/rico-converter>).

PLANUNG ARBEITEN ENSEMEN - 3

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (ff):

- Dann wurde `xml:base=http://data.archives-nationales.culture.gouv.fr/">` durch eine neue Base URI `xml:base=http://bar.admin.ch/arelda/v4/">` ersetzt und die seeAlso-Links auf die «Archives Nationales»-URL entfernt.
- Die Daten und Resultate sind im Anhang-2 zum Protokoll dokumentiert.

Schlussbemerkung (eCH-0160 → RiC):

- Der Lösungsweg «eCH-0160 → EAD → RIC-O» mit dem «ricoconverter-1.0.0» ist nicht ganz überzeugend. Die einzelnen Schritte lassen sich dabei nicht wirklich nachvollziehen. Eine direkte Konvertierung von «Arelde/eCH-0160 → RIC-O» ohne Umweg über EAD wäre wahrscheinlich besser.

PLANUNG ARBEITEN ENSEMEN - 4

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (ff):

Fragen / Diskussion (eCH-0160 → RiC):

- Wie kann eine Transformation von XML (z.B. EAD oder ARELDA) in «RIC-O XML/RDF» formal überprüft werden, so wie dies XML mit schema.xsd möglich ist?
 - Grundsätzlich kann die XML/RDF oder Turtle Syntax formal mit einem entsprechenden Validator prüfen (z.B. <https://www.w3.org/RDF/Validator/> oder <https://www.easyrbeidf.org/converter>).
 - Die Konformität zur RIC-Ontologie kann aber wegen der sog. «Open-world assumption» ohne SHACL- Erweiterung zu RIC-O nicht überprüft werden.

PLANUNG ARBEITEN ENSEMEN - 5

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (ff):

Fragen / Diskussion (eCH-0160 → RiC):

- Wie können die RDF-Daten nach der Transformation «EAD XML → RIC-O RDF» visualisiert und damit die Korrektheit der Transformation verifiziert werden?
- Für XML stehen relativ ausgereifte Tools zur Verfügung, um nicht nur die Struktur, bzw. das Datenmodell, sondern auch die Daten zu betrachten (XML Spy, Oxygen etc.).
- Wie kann man die RDF Daten betrachten und mit dem Input (Original-Daten) vergleichen?
- Für die Ontologie selber gibt es recht gute Visualisierungstools (Protégé oder b.de/webvowl/)

PLANUNG ARBEITEN ENSEMEN - 6

Experimentelle Umsetzung von RIC in den vorhandenen Systemen und Lösungen (ff):

Fragen / Diskussion (eCH-0160 → RiC):

- Die Tools für die RDF-Datenvisualisierung sind relativ bescheiden in ihren Fähigkeiten, z.B. <http://www.ldf.fi/service/rdf-grapher>
- Versteht man den Output der Transformation als Triplestore und damit als Datenbank, stellt sich genauso die Frage nach der Visualisierung. Für Datenbanken gibt es eine Menge von Tools, welche nicht nur die Struktur der Datenbank sondern auch die Daten selber sichtbar machen können (TOAD, HeidiSQL, DBeaver etc.)
- Gibt es so etwa auch für Triplestore? Oder ist die einzige Sicht auf die Daten im Store über SPARQL Abfragen möglich?



TERMINE FÜR DIE NÄCHSTEN SITZUNGEN & VARIA

- Termine für die nächsten Sitzungen:
 1. Nächste Sitzung findet am 25.09 statt
 2. Ein Termin für den Besuch des VSA-Vorstands in 2020 wird nach der Doodle-Umfrage festgelegt, die noch im August stattfinden wird.
 3. Bis auf weiteres finden virtuelle Sitzungen statt.
Ein «Face-to-Face» Meeting wird erst beim Vorliegen der praktischen Ergebnisse im BAR organisiert.
- Keine weiteren Varia