

Matterhorn RDF

VSA-Kickoff ENSEMEN, 1. März 2019

Alain Dubois, Tobias Wildi

Was ist Matterhorn RDF?

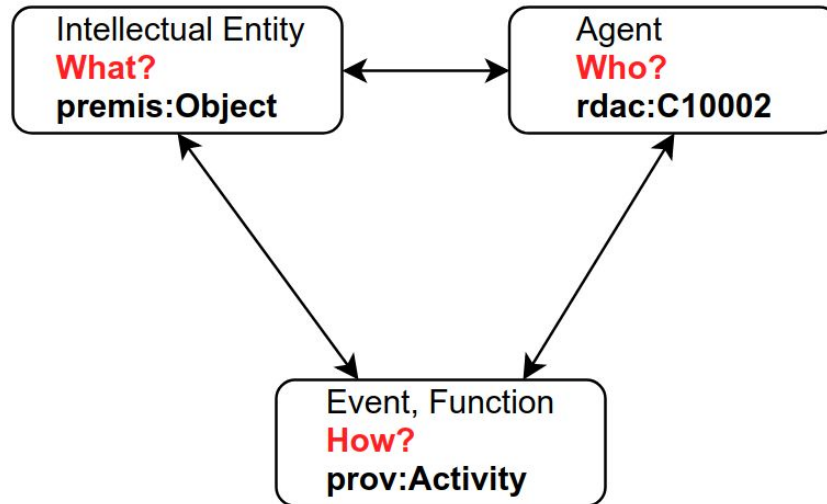
- Ein Datenmodell für archivische Metadaten, das
 - Die ICA-Standards und zukünftig das RiC-CM abbildet
 - “Preservation Description” und “Representation Information” vom OAIS-Informationsmodell abbildet
- Es ist somit ein Datenmodell, das beschreibende und technische Metadaten in einem einzigen Modell vereint.
- Basiert auf den Vorarbeiten für Matterhorn METS, einem METS-Profil, das 2012 bei der Library of Congress registriert wurde.

Verwendete Technologien

- Das Vorgängermodell “Matterhorn METS” basierte auf XML. Das führte in der technischen Umsetzung zu Problemen:
 - Tief verschachtelte XML-Strukturen zur Modellierung der Tektonik und Hierarchien
 - “Geschwätzige” Standards wie PREMIS und EAD erzeugen sehr grosse XML-Dateien, die nicht mehr verarbeitbar sind
 - Verlinkung zwischen den Standards, eindeutige Identifizierung von Ressourcen ist schwierig
- Entscheid für Umstieg auf semantische Technologien
 - Eindeutige Identifizierung jeder Ressource mit einer URI
 - Beziehungen zwischen Ressourcen können qualifiziert werden (“Kind von”, “verheiratet mit”)
 - Hierarchien können schlank modelliert werden (“has part”, “has child” o.ä.)
 - Nutzung externer Ressourcen und Nutzbarmachung eigener Ressourcen wird vereinfacht. Dies verändert die archivische Erschliessung fundamental: Übergang vom Ausfüllen von Textfeldern zur Verlinkung mit bestehenden Ressourcen in Wikidata, Vial, GND etc.

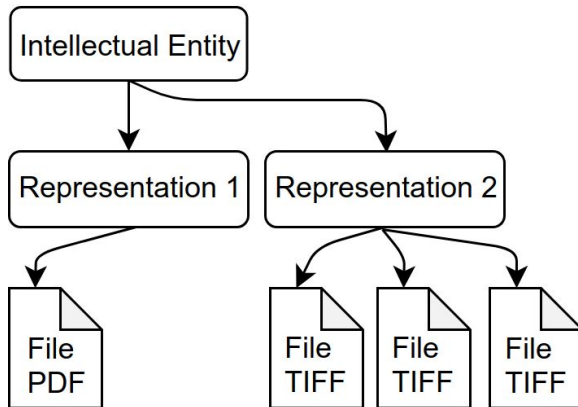
Matterhorn RDF ist ein “Best of”-Modell

- Matterhorn RDF basiert ausschliesslich auf bereits bestehenden Ontologien: PREMIS3, RDA, PROV, DC, DCTEMRS, Ebucore etc.



Repräsentations-Modell

Eine intellektuelle Einheit (z. B. ein Record) kann durch mehrere Repräsentationen abgebildet werden. Eine Repräsentation kann ein physisches Artefakt (Papier, Urkunde) umfassen oder eine oder mehrere Dateien.



Formalisierung mit SHACL

- Matterhorn RDF besitzt keine eigene Ontologie
- Die Formalisierung mit Hilfe der "Shape Constraint Language" des W3C. Vergleichbar mit der Schema-Sprache von XML
- Validierung von Triples gegen Matterhorn RDF wird dadurch möglich

```
sh:property
    sh:path                dc:title
    rdfs:label             "Title"@en
    rdfs:label             "Titel"@de
    rdfs:label             "Titre"@fr
    rdfs:comment           "ISAD 1.2"
    owl:sameAs           rico:title
    sh:datatype            xsd:string
    sh:minCount            1
    sh:maxCount            1
    sh:nodeKind            sh:Literal ;
]
```

Was machen wir gegenwärtig?

- Alle Arbeiten sind dokumentiert und verlinkt über <http://matterhorn.tools>
- Die Shapes sind in einem öffentlichen Git-Repo:
<https://bitbucket.org/docuteam/matterhorn>
- Die EGAD (insb. Florence Clavaud) kennt unsere Arbeit und wir tauschen uns aus.
- Wir präsentieren Matterhorn RDF am “Forum des Archivistes Français” am 4.4.2019 in der gleichen Session wie EGAD RiC präsentiert
- Wir reichen bis 19.3.2019 ein Paper für iPres 2019 in Amsterdam ein
- docuteam entwickelt die Umsetzung von Matterhorn RDF auf Fedora 4
- docuteam und Artefactual planen gemeinsam einen PoC für AtoM3, ein AiS und Repository basierend auf Matterhorn RDF

Wie weiter?

- A. **Going Fast:** Wir arbeiten weiterhin grundsätzlich zu zweit und kommen schnell voran. Wir stehen im Austausch mit der internationalen Community und engagieren uns in Normengremien (z. B. EGAD). Die Weiterentwicklung hängt von zwei Privatpersonen ab. Eine Standardisierung im Rahmen von eCH interessiert uns nicht.
- B. **Going Far:** Wir übergeben die Schweizer Standardisierung von Matterhorn RDF an die VSA-Gruppe und eCH. Das fördert die Nachhaltigkeit des Modells, verlangsamt aber die Entwicklung und verringert die Flexibilität. Wir verlieren als Zweierteam die Entscheidungshoheit, können die Arbeit aber auf mehrere Personen verteilen. Diese Personen müssen ein Commitment zur aktiven Mitarbeit abgeben und über das nötige Fachwissen verfügen.

