



# Verzeichnungseinheiten und mehr

**Transformation von öffentlichen Daten im AIS nach RiC-O und Publikation als OGD-Set**

**(VSA Projektgruppe ENSEMEN, 25. März 2022)**

Oliver Schihin & Lambert Kansy, Staatsarchiv Basel-Stadt

# Überblick

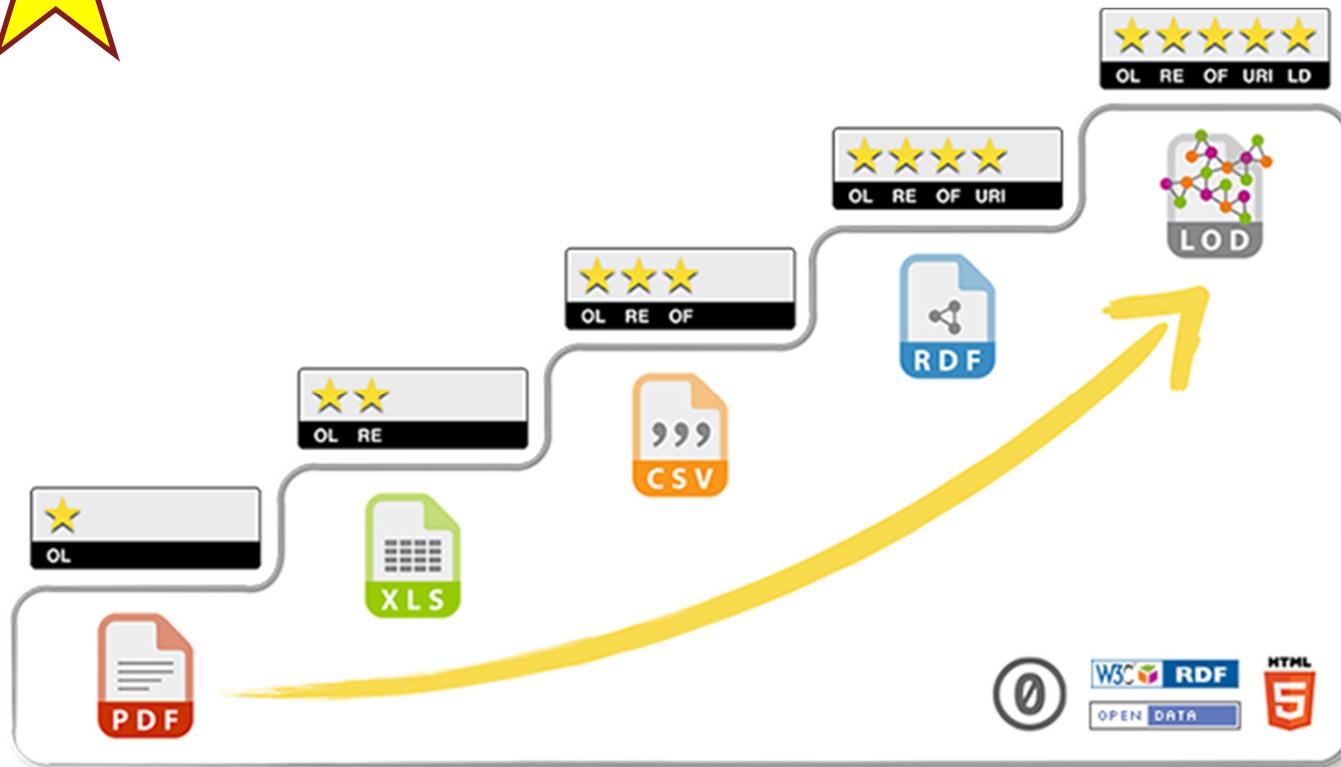
- Publikation der öffentlichen Inhalte des AIS scopeArchiv
- Linked Data-Service: Täglich aktualisierter und vollständiger Datensatz via eine öffentliche Schnittstelle (SPARQL)
- Umfasst Verzeichnungseinheiten, Deskriptoren, Ablieferungen und Partner
- Ontologie: Records in Context

## Kurzer Rückblick

Projekt ist eingebettet in die Open Government Data-Bestrebungen des Kantons

- 2017/18: Pilotprojekt OGD des Kantons Basel-Stadt (Staatskanzlei: StatA, AUE, Geodaten, StABS)
- 2019: Etablierung einer Fachstelle OGD beim Statistischen Amt
  - Datenkatalog [data.bs.ch](http://data.bs.ch)
  - Triple Store & Schnittstelle [Id.bs.ch](http://id.bs.ch)
- Fokus des Staatsarchiv: 4-Sterne-OGD / RDF / LD
- Archivseitig: Projekt ALOD (ab 2014)

# 5 ★ Open Data



# Service-Komponenten



## Datenquelle scopeArchiv

- Datenbankskripte (© by scope solutions)
- Täglich aktualisierte Materialized Views
- Datenschutzprüfung erfolgt bereits im Quellsystem (kein nachträgliches Filtern mehr nötig)
- Flache Datenstruktur (kundenspezifisch, kein Standard)

## ETL-Prozesse

- RDF-Pipeline und Mappings (open source, by Zazuko GmbH)
  - <https://github.com/Staatsarchiv-Basel-Stadt/RDF-Pipeline> (Pipeline)
  - <https://github.com/Staatsarchiv-Basel-Stadt/StABS-scope2RDF> (Mappings)
  - <https://github.com/Staatsarchiv-Basel-Stadt/stardog-docker> (stardog services)
- Sämtliche Anwendungen in Docker
- Wesentliche Prozessschritte
  - Extraktion der Daten aus den Views von scopeArchiv
  - Materialisierung anhand des Mapping in einen RDF-Graph in stardog
  - Datenkorrekturen und -verbesserungen mit JavaScript
  - Publikation des Graphen im öffentlichen Triple-Store des Kantons

## Publikation

- Öffentlicher Triple-Store mit Datenansicht und SPARQL-Schnittstelle
  - ld.bs.ch / ld.staatsarchiv.bs.ch
- 1 Datensatz (Named Graph <https://ld.staatsarchiv.bs.ch/graph/ais-metadata>)
- Ca. 9 Mio. Triples (Stand heute)
- Einträge auf kantonalem und nationalem Datenportal
  - data.bs.ch
  - opendata.swiss
- Go public im April 2022

# URIs & Klassen

@base <<https://ld.staatsarchiv.bs.ch/>>

URI	Klasse	AIS-Objekt
Record/{id}	rico:RecordResource (und Subklassen)	VE alle Stufen
Agent/{id}	rico:Agent skos:Concept	Partner und Personendeskriptoren
Place/{id}	rico:Place skos:Concept	Ortsdeskriptoren
Event/{id}	rico:Activity	Ablieferungen
Record/{id}/[Date Identifier] Extent]	rico:Date, rico:Identifier, rico:recordResourceExtent, etc.	Umfang, Signaturen / Identifier, Zugang, Datum, etc.
<b>Noch offen</b> Instantiation/{id}-i	rico:Instantiation	«Blätter» der Hierarchie

# StABS-Erweiterungen (zwei Beispiele)

PREFIX stabs-rst: <https://ld.staatsarchiv.bs.ch/vocabularies/RecordSetTypes#>

PREFIX stabs-idt: <https://ld.staatsarchiv.bs.ch/vocabularies/IdentifierTypes#>

(weitere sind ExtentTypes, ThesaurusTypes, RuleTypes, AccessType, etc.)

```
stabs-rst:Bestand a rico:RecordSetType , skos:Concept , owl:NamedIndividual ;
    rdfs:label "Bestand"@de ;
    rdfs:subClassOf ric-rst:Series ;
    skos:prefLabel "Bestand"@de ;
    skos:definition "Bestände werden bei Unterlagen von Dienststellen und juristischen Personen aufgrund bestehender Ordnungssysteme, physisch unterscheidbaren Ablagen, Zeitperioden oder nach den Aufgabengebieten und Produkten der Provenienzstelle gebildet. Bei natürlichen Personen und Familienarchiven können Bestände auch aufgrund von Generationen, Tätigkeitsgebieten oder Lebensabschnitten gebildet werden. (Def. sinngemäss nach StABS, Handbuch Erschliessung) "@de ;
    skos:inScheme <https://ld.staatsarchiv.bs.ch/vocabularies/recordSetTypes>
```

```
stabs-idt:Fruehere-Signatur a rico:IdentifierType, skos:Concept, owl:NamedIndividual ;
    rdfs:label "Frühere Signatur"@de ;
    skos:prefLabel "Frühere Signatur"@de ;
    skos:definition "Von der Provenienzstelle verwendete Signatur oder identifizierende Bezeichnung"@de ;
    skos:inScheme <https://ld.staatsarchiv.bs.ch/vocabularies/IdentifierTypes>
```

# Konversion

- Mapping (R2RML, RML, etc.): Grundlage für die Übertragung der Inhalte nach RiC-O
- Datenbearbeitung
  - Daten häufig invalide oder sehr unpräzise
  - Datenbearbeitung zwingend
    - Zahlenwerte (Laufmeter) → «1,3»
    - Datumswerte → «1539-1950 (ca.)»
    - Deskriptoreneinträge
      - «Verein der Lehrerinnen für Textilarbeit und Werken»
      - «Chevloh Ez Chajim»
      - «Reichert, Heinrich (1949-), Prof. Dr.»

# Schluss. Fragen?

\*\*\*

Danke für die Aufmerksamkeit

\*\*\*

[oliver.schihin@bs.ch](mailto:oliver.schihin@bs.ch)

[lambert.kansy@bs.ch](mailto:lambert.kansy@bs.ch)