

Metadatenprovenienz (einzelne Aussagen)

Wie die einzelnen Metadatenabgaben entstanden sind, soll auch in den RDF-Daten transparent sein. Deswegen beginnen wir ab 2019 mit einem mehrschrittigen Vorgehen, diese Informationen verfügbar zu machen:

- Modellierung der Metadatenprovenienz einzelner Aussagen in RDF ✓
- Identifizierung der Erzeugungsprozesse und Agenten ✓
- Bereitstellung eines Metadatenprovenienz-Dumps ✓ *Ab Januar 2019 laufend*
- Beschreibungen der Erzeugungsprozesse und Agenten verfügbar machen ⚠ *geplant für 1. Halbjahr 2019*
- Integration der Metadatenprovenienz in das Resolving der einzelnen Titel-Ressourcen ⚠ *geplant für 2019*
- ... Ausbau der Prozess- und Agentenbeschreibungen und des Detailgrads der Prozessidentifizierung ⚠ *2019 ff*

Die geplante Ausgestaltung und das Vorgehen sind flexibel. Wir möchten in allen Phasen gerne mit den Datennutzern und -nutzerinnen ins Gespräch kommen und deren Wünsche und Anregungen [aufnehmen](#) und [diskutieren](#).

Modellierung der Metadatenprovenienz einzelner Aussagen in RDF

Unser schrittweises Vorgehen beinhaltet, dass die Metadatenprovenienz zunächst nur für die Arten von Aussagen veröffentlicht wird,

- für die es Unterscheidungsbedarf gibt (*d.h. die von mindestens zwei verschiedenen Prozesse generiert werden*) und
- für die diese Unterscheidung anhand der Datenbasis möglich ist (*in diesem Punkt verbessert sich die Datenlage stetig*).

Derzeit sehen wir Anlass und Möglichkeit, Provenienz für RDF-Aussagen mit den Properties

- `dcterms:subject` und
- `dcterms:language`

zu veröffentlichen.

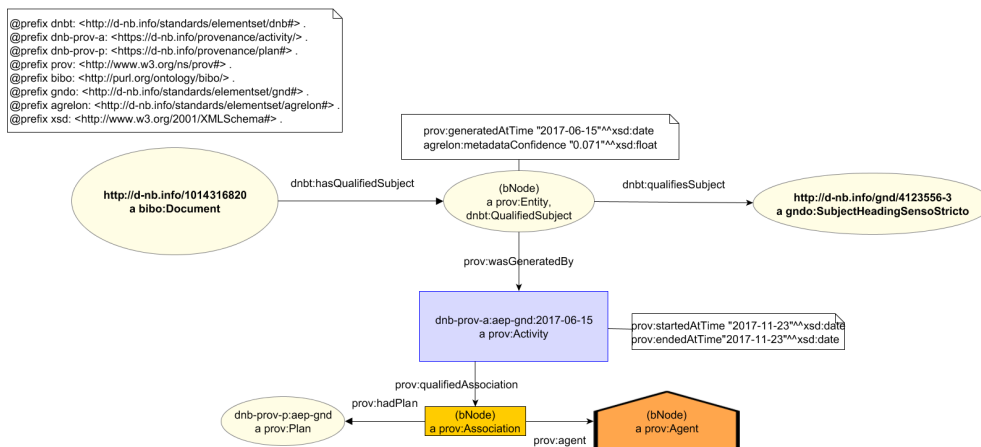
Die Modellierung erfolgt auf Grundlage des [PROV Data Model](#) (W3C Recommendation) und anhand der [PROV Ontology](#) (W3C Recommendation). Für die Anbindung der Provenienzinformatoren an die Titel-Ressourcen bedienen wir uns des [Qualified Relation patterns](#). Dazu haben wir in unserem RDF-Element Set [DNB Metadata Terms](#) die Klassen

- `dnbt:QualifiedLanguage`
- `dnbt:QualifiedSubject`

geschaffen, sowie die Properties

- `dnbt:hasQualifiedLanguage`
- `dnbt:qualifiesLanguage`
- `dnbt:hasQualifiedSubject`
- `dnbt:qualifiesSubject`.

Hier die grafische Darstellung eines Beispiels für "Titel wurde GND-Sachbegriff zugeordnet durch automatischen Prozess":



Dasselbe in RDF (Turtle)

```

@prefix dnbnt: <http://d-nb.info/standards/elementset/dnb#> .
@prefix dnb-prov-a: <https://d-nb.info/provenance/activity/> .
@prefix dnb-prov-p: <https://d-nb.info/provenance/plan#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix agrelon: <http://d-nb.info/standards/elementset/agrelon#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://d-nb.info/1014316820> a bibo:Document ;
  dnbnt:hasQualifiedSubject [
    a dnbnt:QualifiedSubject, prov:Entity ;
    dnbnt:qualifiesSubject <http://d-nb.info/gnd/4123556-3> ;
    prov:wasGeneratedBy dnb-prov-a:aep-gnd:2017-06-15 ;
    prov:generatedAtTime "2017-06-15"^^xsd:date ;
    agrelon:metadataConfidence "0.071"^^xsd:float
  ] .

dnb-prov-a:aep-gnd:2017-06-15 a prov:Activity ;
  prov:startedAtTime "2017-06-15"^^xsd:date ;
  prov:endedAtTime "2017-06-15"^^xsd:date ;
  prov:qualifiedAssociation [
    prov:hadPlan dnb-prov-p:aep-gnd ;
    prov:agent [
      a prov:Agent
    ]
  ] .

```

Identifizierung der Erzeugungsprozesse und Agenten

Die relevanten Entitäten werden durch URIs unter d-nb.info identifiziert:

- Im Bereich <https://d-nb.info/provenance/activity/> Entitäten des Typs **prov:Activity** (= die Erzeugung der einzelnen Aussage), z.B. <https://d-nb.info/provenance/activity/aep-gnd:2017-06-15>
- Im Bereich <https://d-nb.info/provenance/plan#> Entitäten des Typs **prov:Plan** (= der Prozess, der die Erzeugung der einzelnen Aussagen steuert), z.B. <https://d-nb.info/provenance/plan#aep-gnd>
- Im Bereich <https://d-nb.info/provenance/software#> Entitäten des Typs **prov:SoftwareAgent**

Bereitstellung von Metadatenprovenienz-Dumps

Die in der ersten Phase einzige Bereitstellungsform sind die 3-mal jährlich aktualisierten Metadatenprovenienz-Dumps. Sie werden zusätzlich zu den Titel-Dumps veröffentlicht und basieren auf demselben Datenstand. *Der Aussagenumfang der Titel-Dumps wird nicht verändert. Das heisst, dass sie weiterhin alle dcterms:subject- und dcterms:language-Aussagen ohne Provenienzinformation enthalten.*



Download der Dumps unter <https://data.dnb.de/opendata/>

Kleine Testdatensets für einen schnellen Überblick werden bereitgestellt unter <https://data.dnb.de/testdat/>

Verfügbar sind die Serialisierungen:

- RDF (Turtle)
- RDF/XML
- N-Triples
- HDT
- JSON-LD

Die Dateibenennung folgt dem Muster `dnb-all_ldsprov_JJJJMMTT.Dateieindung`, z.B. `dnb-all_ldsprov_20190113.ttl`

Beschreibungen der Erzeugungsprozesse und Agenten

Für Entitäten des Typs **prov:Activity** sind bereits RDF (Turtle)-Beschreibungen verfügbar. Die URIs lösen sie auf, z.B. <https://d-nb.info/provenance/activity/aep-gnd:2017-06-15>

Langfristig sollen auch eine HTML-Ansicht sowie weitere RDF-Serialisierungen verfügbar werden, sowie Content Negotiation implementiert werden. ⚠️ *Ein Zeitplan dafür liegt noch nicht vor*

Außerdem arbeiten wir daran, HTML- und RDF-Beschreibungen für **Entitäten des Typs `prov:Plan`** bereitzustellen. Zunächst lösen die URIs des Bereichs `https://d-nb.info/provenance/plan#` noch nicht auf.

Hier eine vorläufige grobe Einteilung der derzeit ausgewiesenen Erzeugungsprozesse:

Intellektuelle Erfassung
<code>https://d-nb.info/provenance/plan#i</code>
<code>https://d-nb.info/provenance/plan#ie-sg</code>
Automatische Erzeugung
<code>https://d-nb.info/provenance/plan#aep-gnd</code>
<code>https://d-nb.info/provenance/plan#aep-lc</code>
<code>https://d-nb.info/provenance/plan#aep-lcsh</code>
<code>https://d-nb.info/provenance/plan#aep-sg</code>
<code>https://d-nb.info/provenance/plan#m</code>
Übernahme aus Fremddaten
<code>https://d-nb.info/provenance/plan#oclc-fast</code>
<code>https://d-nb.info/provenance/plan#f</code>
Ableitung aus Datenmappings/Konkordanzen
<code>https://d-nb.info/provenance/plan#stw-gnd</code>

Auch **Entitäten des Typs `prov:Agent`** wird vorbereitet, fehlt derzeit jedoch noch gänzlich.

Integration der Metadatenprovenienz in das Resolving der einzelnen Titel-Ressourcen

Unser Ziel ist es, die Metadatenprovenienz langfristig auch auszuliefern, wenn die Beschreibung einer einzelnen Titel-Ressource über ihren URI angefordert wird. Z.B. `http://d-nb.info/1014316820`