Network of Terms

*Bringing links to your data!*

Enno Meijers – CTO Dutch Digital Heritage Network (NDE)

[enno.meijers@kb.nl](mailto:enno.meijers@kb.nl)  |  [https://mastodon.social/@ennomeijers](https://mastodon.social/@ennomeijers)

SWIB - 12 September 2023
Introduction to the NDE programme
The Dutch Digital Heritage Network (NDE)

The Dutch Digital Heritage Network (NDE) aims at increasing the social value of the cultural heritage information maintained by libraries, archives, museums and other cultural institutions.

The NDE strategy starts from the end user perspective and encourages institutions to provide digital heritage information that is more visible, usable and sustainable.

The NDE program is about building strong cross sector networks on the level of expertise and information. Linked Data is regarded as one of the enabling technologies.

https://netwerkdigitaal erfgoed.nl/activiteiten/nationale-strategie-digitaal-erfgoed/
Networks of cultural institutions

Archieven

Bibliotheken

Musea

Media en AV

Digital humanities

Design en digitale cultuur

Podiumkunst.net

Netwerk Archieven Design en Digitale Cultuur
Cross domain reference architecture (DERA)

Strategic goal

Increasing social value by putting the users first

Operational goals

- Users can freely navigate heritage information
- Users can determine the usefulness of heritage information
- Users can add information
- Heritage institutions work together

Principles

- Ensure that the authenticity of heritage information is clear
- Ensure that the availability of heritage information is clear
- Ensure that information is presented in a recognisable and user-friendly way
- Ensure an unambiguous description of heritage information
- Ensure that heritage information refers
- Ensure that heritage information is referable
- Respect the diversity of heritage information
- Provide distributed heritage information
Overview of the NDE program

Visible
Service providers
- Usage profiles
- Campagne & channel
- Rights & Usage

Usable
Infrastructure providers
- Data & Terminologie Sources (LOD)
- Registries
- Aggregators
- Knowledge Graph

Sustainable
Source providers
- PID
- Preservation Policy & Certification
- Cost model Preservation
- Index Preservation Services

OVERALL
- Supporting Network
- Training & Education
- Body of Knowledge
- Services Toolbox
- Service implementation & management
Infrastructure for digital heritage
Infrastructure for digital heritage information

Rethink the architecture:
• maximize the usability of data at the source
• refer to data instead of copying
• build portals as dynamic views based on a common, interconnected data layer
• minimize the intermediate layers
• improve visibility on the web in general

Apply:
• Linked Data / FAIR principles
• ‘web-centric’ technologies (HTTP, RDF, Web APIs)
• decentralized technologies where possible
Service platforms – generic and dynamic services

1. Register datasets with heritage information
2. Select datasets with heritage information
3. Retrieve heritage information
PROVINCIE GELDERLAND
Web Portals

Infrastructure

Data Sources

Services

Collection Management Systems

Dataset Register

Knowledge Graph (data summaries)

Service Platforms (Powered by KGs)

Network of Terms

"discover relevant datasets"

"improve the visibility on the web"

"smarter, more dynamic services"

"advertise your data"

"use things, not strings"

n = 2000+

n = 50+

n = 1500+
Network of Terms
Using Linked Data in a network...

A Pair of Leather Clogs (Van Gogh Museum)

Garden at Arles (Kunstmuseum)

National Library

Dataset Registry

Knowledge Graph

Network of Terms

schema:creator

rico:isCreatorOf

schema:creator

rico:isCreatorOf

schema:about

schema:subjectOf

Vincent van Gogh

https://data.rkd.nl/artists/32439
(RKD Artists)
Available sources for linking through the Network of Terms

National Institutes
- Cultural Heritage Agency (thesauri)
- National Library (thesauri, persons/org.)
- Sound & Vision Institute (thesauri, persons/org.)
- Music library of The Netherlands (thesauri, persons/org.)
- Performing Arts Network (thesauri, persons/org.)
- Netherlands Institute for Art History (persons)
- Nationaal Museum of World Cultures (thesauri)
- Indonesia Remembrance Centre (thesauri)
- LGBTI Heritage Organisation (thesauri)
- Second World War Documentation (thesauri, persons)

International Institutes
- Getty Research Institute (AAT)
- Geonames (NL, BE, DE)
- Wikidata (entities, persons, places, streets)
- European Union (EuroVoc)
- Henri van de Waal Foundation (Iconclass)

Regional Institutes
- Erfgoed Brabant (buildings)
- Gouda Time Machine (streets)
- Cultural Institutions Amsterdam (streets)

In numbers:
- Organisations: 18
- Individual terminology sources: 25
- Search queries: 40
Default direct approach

Use Case:
Find relevant Term URI(s) for concepts, persons, places,... when describing resources in my own system.

Challenges:
- harvest sources or query live apis?
- deal with multiple protocols
- deal with multiple data models
- complexity and maintenance issues

Collection Management System

Art & Architecture Thesaurus
RKDartists
Geonames

‘From string to thing’: “Van Gogh” → https://data.rkd.nl/artists/32439
Default platform approach

Some challenges:
- deal with multiple data models
- synchronisation with source data
- integration with other systems?
Network of Terms approach

Harmonisation of data models and protocols

Collection Management Systems

SKOS

Network of Terms

Art & Architecture Thesaurus (AAT)

RKDartists

Geonames

SKOS/XL, dcterms, schema.org, gvp

GraphQL

SPARQL
Searching shared thesauri, classification systems, and reference lists

for collection administrators | for administrators of terminology resources | for heritage software developers

Search terms: metal

Terminology sources: Art & Architecture Thesaurus - materialen (AAT - materialen)

Search

"metaal naar kwaliteit"

Copy URL | View at source

Broader term

metal
But it is really about the API...
Example of a SPARQL CONSTRUCT query used for transformation to SKOS

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX void: <http://rdfs.org/ns/void#>

CONSTRUCT {
  ?uri a skos:Concept ;
  skos:prefLabel ?rdfs_label ;
  skos:altLabel ?schema_name ;
  skos:altLabel ?schema_alternateName ;
  skos:scopeNote ?schema_description .
}
WHERE {
  ?uri foaf:isPrimaryTopicOf/void:inDataset <http://data.bibliotheeken.nl/id/dataset/stcn> ;
  schema:additionalType <http://www.productontology.org/id/Printer%28publishing%29> ; # Select printers.
  rdfs:label ?rdfs_label .

  VALUES ?predicate { rdfs:label schema:name schema:alternateName }

  OPTIONAL { ?uri schema:name ?schema_name }
  OPTIONAL { ?uri schema:alternateName ?schema_alternateName }
  OPTIONAL { ?uri schema:description ?schema_description }
}
LIMIT 1000
```

Map all types to skos:Concept for easy access

some “magic” for full text search depending on SPARQL-endpoint type
Reconciliation Service API

Reconciliation Service API v0.2
A protocol for data matching on the Web
Final Community Group Report 10 April 2023

This version:
https://www.w3.org/community/reports/reconciliation/CG-FINAL-specs-0.2-20230410/

Latest published version:
https://www.w3.org/community/reports/reconciliation/CG-FINAL-specs-0.2-20230410/

Latest editor’s draft:
https://reconciliation-api.github.io/specs/draft/

Editors:
Antonin Delpeuch (University of Oxford)
Adrian Pohl (Hochschulbibliothekszentrum NRW)
Fabian Steeg (Hochschulbibliothekszentrum NRW)
Thad Guidry Sr.
Osma Suominen (National Library of Finland)

Feedback:
GitHub reconciliation-api/specs (pull requests, new issue, open issues)
public-reconciliation@w3.org with subject line [specs-0.2] _ message topic _ (archives)

Copyright © 2023 the Contributors to the Reconciliation Service API v0.2 Specification, published by the Entity Reconciliation Community Group under the W3C Community Final Specification Agreement (PSA). A human-readable summary is available.

Abstract
This document describes the reconciliation service API, a protocol edited by the W3C Entity Reconciliation Community Group. It is intended as a comprehensive and definitive specification of
Add links to existing data using OpenRefine:

Every Network of Terms source gets a Reconciliation Service endpoint by default!
DIY – Network of Terms Tutorial

Configuring the catalog
Enno Meijers edited this page 3 weeks ago · 9 revisions

Introduction to the Catalog
The Catalog is the core component of the Network of Terms. Adding new terminology sources to the Network of Terms is only a matter of making changes to the Catalog configuration. No changes of the code itself are necessary. The Catalog only exists of files containing JSON-LD objects or SPARQL queries.

To make changes to the Catalog it is important to understand the structure and data model behind it.

The Catalog is part of the network-of-terms repository and is a separate package located in packages/network-of-terms-catalog. The Catalog has the following file structure:

```
* catalog/  
  * publishers.jsonld  # definition of organizations that publish data sources
  * datasets/  
    * <ds_name>.jsonld  # dataset definition for each data source
  * queries/  
    * search/  
      * <qname>.rq  # SPARQL queries for searching
    * lookup/  
      * <qname>.rq  # SPARQL queries for lookups
```

Data Model

So far we have found the following examples of that deploy full text search capabilities:

- **GraphDB - Lucene FTS plugin (deprecated)**

  Used in Getty sparql endpoint serving AAT, GTN and others:

  ```sparql
 PREFIX luc: <http://www.ontotext.com/connectors/lucene#>
  SELECT ?uri ?cntStr {
    ?cntUri luc:tempQuery
    luc:term: Brief, includes all terms (prefLabels and allLabels) and subject ID (default)
    luc:text: Full, includes all terms, qualifiers, subject ID, and scope notes.
  }
  ```

- **GraphDB - Lucene connector**

  If the Lucene connector is configured the available indexes can be listed in the following way:

  ```sparql
  PREFIX luc: <http://www.ontotext.com/connectors/lucene#>
  SELECT ?uri ?cntStr {
    ?cntUri luc:listConnectors ?cntStr .
  }
  ```

- **GraphDB - Elasticsearch connector**

  If the Elasticsearch connector is configured the available indexes indexes can be listed in the following way:

  ```sparql
  PREFIX elastic: <http://www.ontotext.com/connectors/elasticsearch#>
  SELECT ?uri ?cntStr {
    ?cntUri elastic:listConnectors ?cntStr .
  }
  ```

Searching by using one of the indexes is shown in this example from the SemOpenAlex endpoint:

```sparql
```

See for more info https://graphdb.ontotext.com/documentation/10.0/graphdb-connectors.html

- **Apache Jena Fuseki**

  Example for using the Lucene implementation with a Fuseki endpoint:

  ```sparql
  (?uri ?score) textquery ("<field1>...<fieldn> ?query 100")
  ```

Implementation
Integrations of Network of Terms API in other systems:

- Atlantis (DEVENTit)
- Memorix (Picturae)
- Axiell Collections (Axiell)
- Kleksi (Sofco)
- MuseumPlus (ZetCom - under discussion)
- Omeka-S (open source)
- Solid-CRS (open source - experimental)
- LDWizard - Heritage (open source - limited functionality)
- Mark Lindeman - VSCode Termennetwerk Extension (open source)

Source code (EUPL-licensed) available at
Usage over August '23
Takeaways:

- Federated querying of linked data works (for our use case)
- Comunica.dev is a powerful framework for federated querying over Linked Data
- Standardisation of full text search for SPARQL is needed
- Network-of-Terms is a generic tool, it could fit your use case too

=> give it a try and share your ideas with us!
Thank you!

Enno.meijers@kb.nl  |  https://mastodon.social/@ennomeijers

Learn more at:

netwerkdigitaalerfgoed.nl