

kadaster



Lexi Rowland | 06-04-2022

GeoE3: Activity 2 Workshop

Activity 2 Deliverables

- **Semantic Search:** This task will **investigate existing national profiles of metadata** with the aim of subsequently using the profiles for the provision of content discovery and evaluation services.
- **Support for Search Engines:** This task will **develop mechanisms** for making data discovery services available through Internet search engines, like Google.

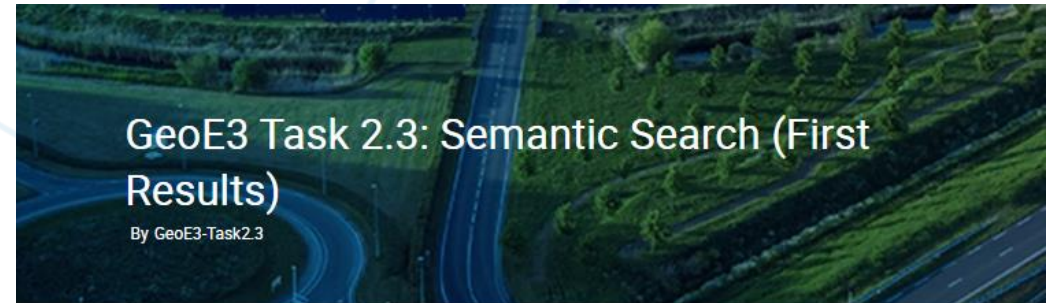
Semantic Search

- Transformation of European metadata records to linked data
- Manual enrichment of keyword lists with concepts from data models and themes using GEMET vocabulary
- Combination of records in a single dataset
- Implementation of ElasticSearch/SPARQL service
- POC provided as Data Story

Results:

1. Use of GEMET for keyword enrichment
2. Need for automation of enrichment step

Link: <https://data.pldn.nl/geoe3task23/-/stories/geoe3-semanticsearch-task2-3>



This data story serves to demonstrate the potential that enriching metadata with more extensive keywords lists has for improving the semantic searchability and findability of digital resources across Europe. The following are the first results of Task 2.3 of GeoE3 exploring the implementation of semantic searchability.

The following table provides an overview of all countries that include a keyword associated with 'address' in their metadata. These keywords have, in some cases, been added to existing metadata records based on a manual enrichment step for the purpose of highlighting the potential that metadata enrichment has for semantic searchability of datasets.

Table 1. A list of partner countries with keywords corresponding to the GEMET vocabulary value Address

[Go to dataset](#) [Try this query yourself](#)

asset_title	assets	keyword
1 Matrikkelen - Adresse	https://kartkatalog.geonorge.no/Metadata/uuid/f7df7a18-b30f-4745-bd64-d0863812350c	Adresse
2 DWV INSPIRE View Service for Addresses	https://inspire-wms.maanmittauslaitos.fi/inspire-wms/AD/ows	Addresses
3 DWV INSPIRE View Service for Addresses	https://inspire-wms.maanmittauslaitos.fi/inspire-wms/AD/ows	Osoitteet
4 DWV INSPIRE View Service for Addresses	https://inspire-wms.maanmittauslaitos.fi/inspire-wms/AD/ows	Addresses

Notes for Further Development

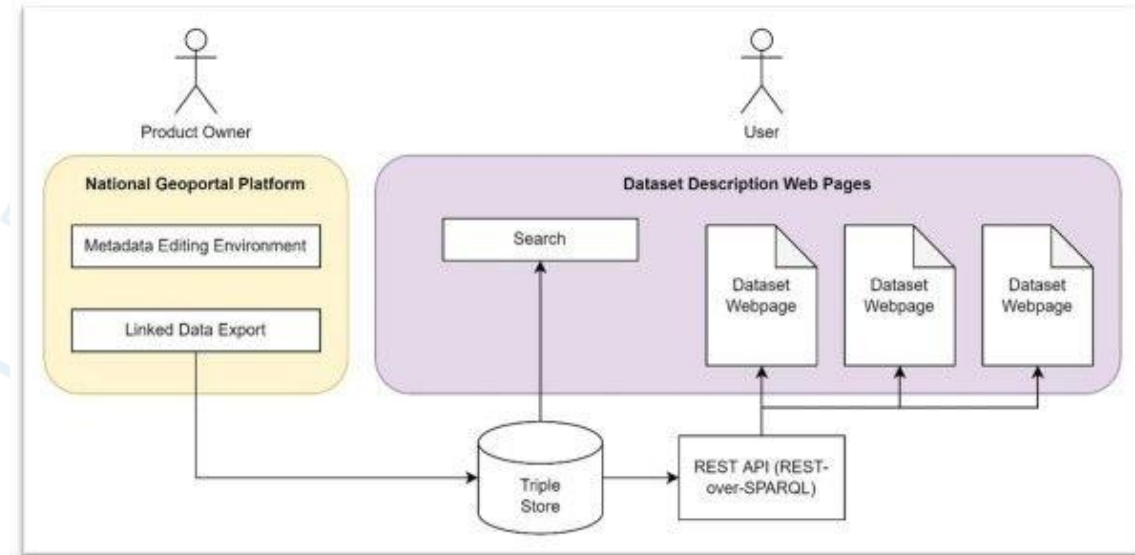
- Lack of consistency across metadata profiles making it difficult to ensure quality results (incomplete result set)
- Data quality issues (language tags, differing URIs)
- Case sensitivity of SPARQL queries
- Multi-lingual ElasticSearch?

Support for Search Engines

- Investigation into (Geo)DCAT-SDO mapping
- Implementation of ETL
- Rich Results Test
- Google Dataset Search

Results:

1. Tooling is available to support search engine findability
2. Not all tooling is available as open source
3. Validation of metadata records in RRT but indexing by search engines is still a barrier
4. Geonetwork tooling hinders the direct implementation of metadata records into NGR environment. This might be a similar factor for other partners.



Implementation Results (1)

Origin of visits to PDOK.nl

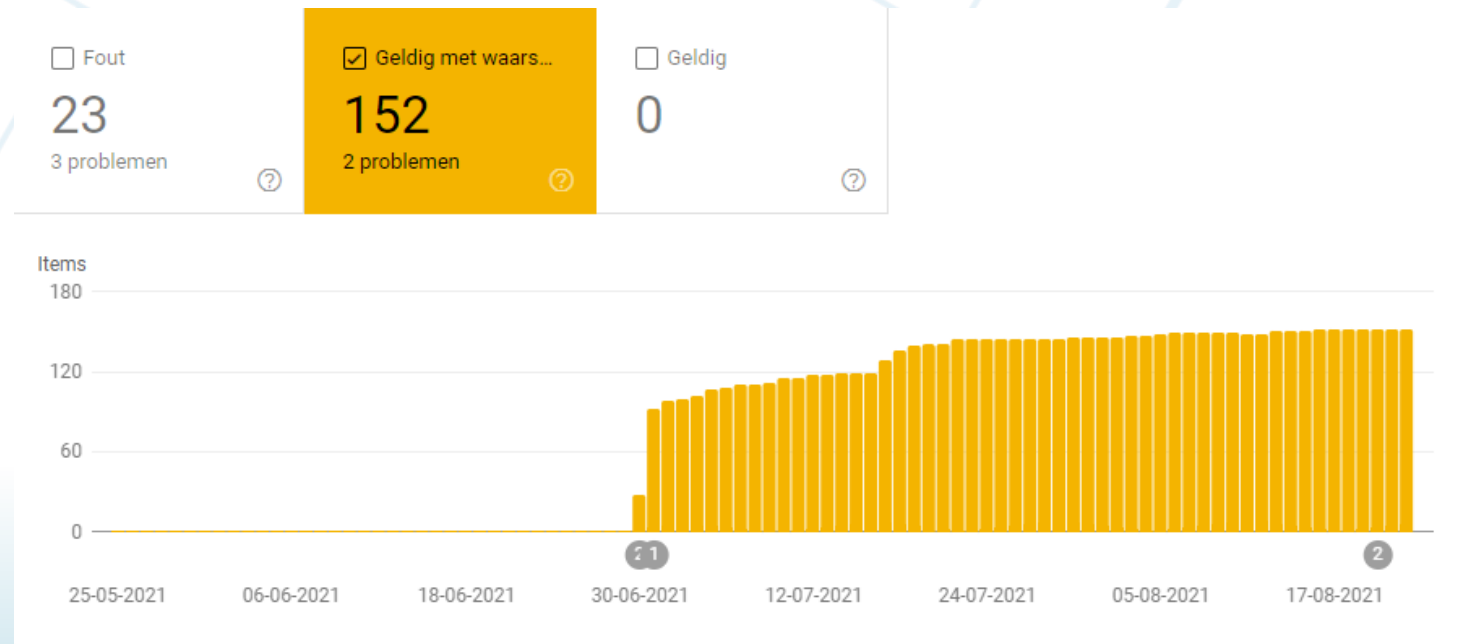
Bezoek herkomst PDOK				
Channel	Visitors	Vistisors %	Sessions	Sessions %
Search Engine	31.733	45,48%	43.470	41,47%
Direct Entry	27.916	40,01%	49.986	47,68%
Website	10.118	14,50%	11.374	10,85%
Totaal	69.767	100,00%	104.830	100,00%
1 april - 30 juni				
Channel	Visitors	Vistisors %	Sessions	Sessions %
Search Engine	30.290	47,19%	40.181	43,33%
Direct Entry	24.192	37,69%	41.696	44,97%
Website	9.708	15,12%	10.850	11,70%
Totaal	64.190	100,00%	92.727	100,00%
1 juli - 30 september				
Channel	Visitors	Vistisors %	Sessions	Sessions %
Search Engine	35.506	48,30%	49.420	44,77%
Direct Entry	28.652	38,97%	50.111	45,40%
Website	9.360	12,73%	10.849	9,83%
Totaal	73.518	100,00%	110.380	100,00%
1 oktober - 31 december				

Implementation Results (2)

For the PDOK website, we had the **Google Search Console** set up for the execution of this pilot.

- Insight into whether a website has structured data (the technology for improved findability) and what the quality of the improved findability is.
- The website also shows any warnings and errors that could affect findability.

As can be seen in the figure below, Google has been detecting in the metadata since 1 July that 152 datasets are available on PDOK. The pilot went live on June 28.



Implementation Results (3)

