

Data Archiving and Networked Services

DANS

Decentralized research data infrastructure.

LLM and Data Repositories

CLARIAH 2nd Tech & Data Day

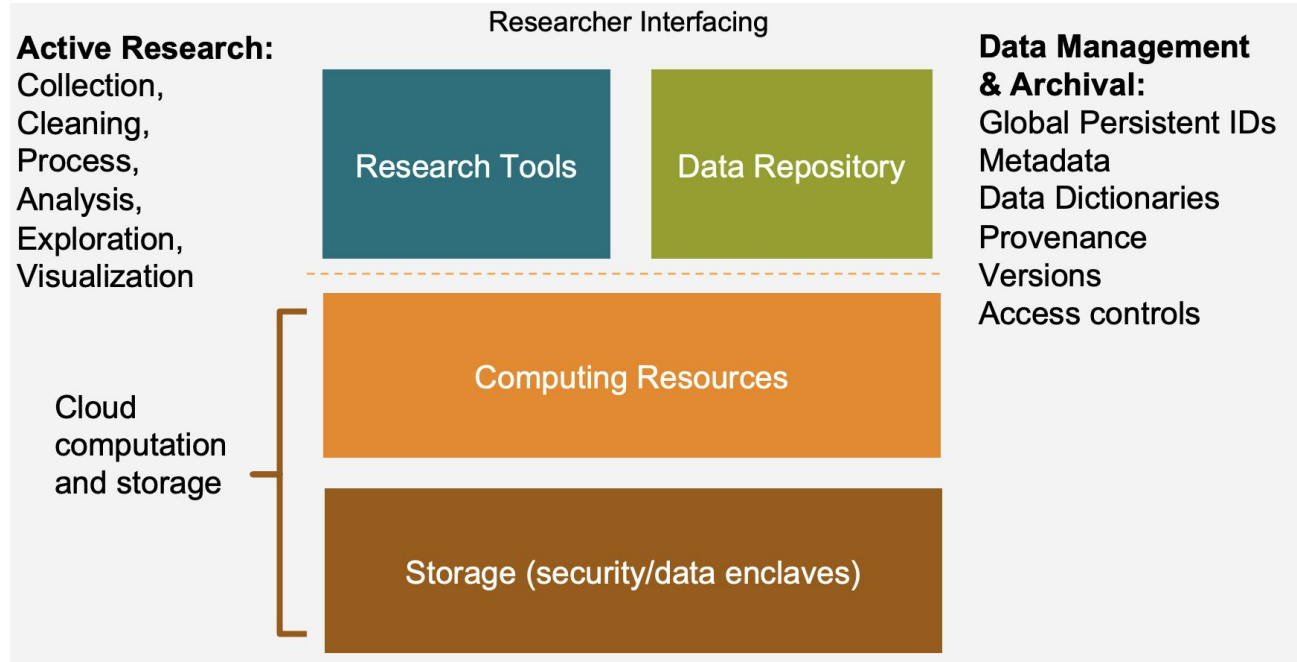
Slava Tykhonov, R&D
(DANS-KNAW, the Netherlands)

28 September 2023, The Hague

Building an Operating System for Open Science

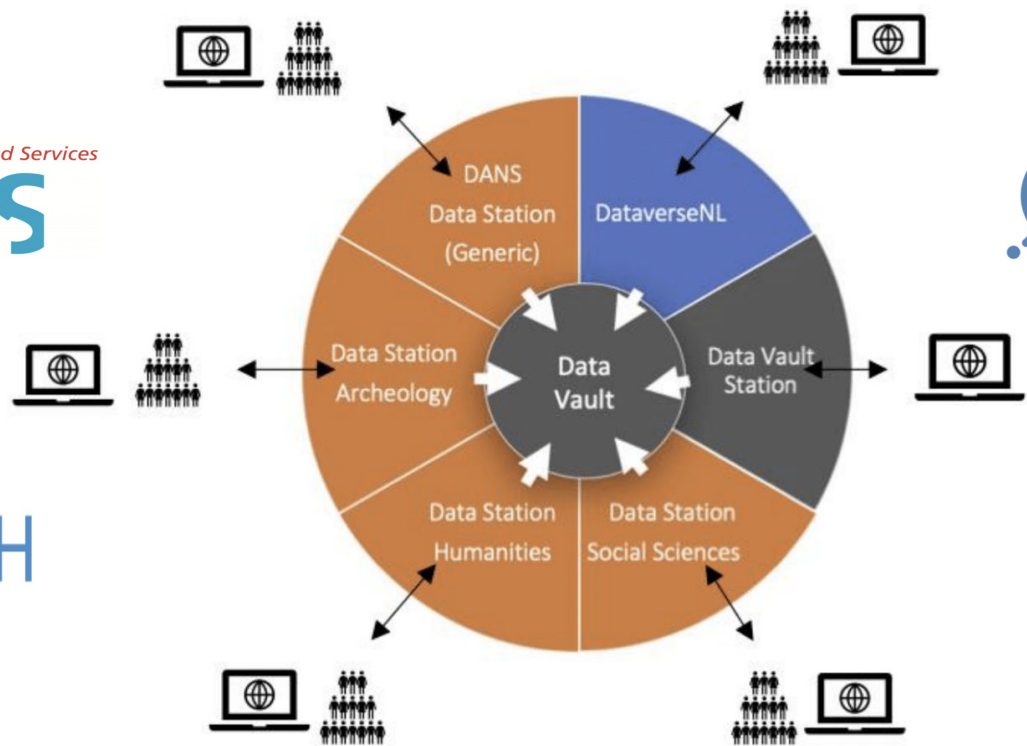
- Generic Common Research and Data Infrastructure should be distributed and robust enough to be scaled up and reused for any challenging tasks like cancer research etc
- Networked services built from Open Source components
- Data processed and published in FAIR way, the provenance information is the part of our Data Lake
- Data evaluation and credibility is the top priority, we're providing tools for the expert community for the verification of our datasets
- The transparency of data and services guarantees the reproducibility of all experiments and get bring new insights in the multidisciplinary research
- Infrastructure should enforce collaboration between people, bring together general public, researchers, citizen scientists, etc
- Infrastructure is free of charge, (meta)data is protected and licenced.

Looking for Commons



Merce Crosas, "[Harvard Data Commons](#)"

DANS Data Stations - Future Data Services



Dataverse is API based data platform and a key framework for Open Innovation!

Semantic interoperability on the infrastructure level

We envision a situation where thousands of Dataverse instances (due to EOSC) on the web can be simultaneously search for data and will form shared Data Lake.

The *old dream* of Federated search/Universal catalogue can only be realised if:

- (1) Crosswalks; mapping across different metadata schemes are implemented
- (2) In metadata schemes we seek for ways to enrich indexes with values from controlled vocabularies

Standard response (**centralized**) = standardisation and harmonisation = repository software, certain metadata standards, or certain controlled vocabularies

New response (**distributed**) = explore agile solutions (Proof of Concepts) which can be implemented by different communities (even smaller ones), so we keep variety and still enable integration in the Distributed Data Network by applying Linked Data technologies.

Dataverse Metadata Schema in KOS

#metadataBlock	name							
	citation							
#datasetField	name	allowmultiples	facetable	displayoncreat	required	parent	metadatablock_	termURI
	title	FALSE	FALSE	TRUE	TRUE		citation	http://purl.org/dc/terms/title
	subtitle	FALSE	FALSE	FALSE	FALSE		citation	
	alternativeTitle	FALSE	FALSE	FALSE	FALSE		citation	http://purl.org/dc/terms/alternative
	alternativeURL	FALSE	FALSE	FALSE	FALSE		citation	https://schema.org/distribution
	otherId	TRUE	FALSE	FALSE	FALSE		citation	
	otherIdAgency	FALSE	FALSE	FALSE	FALSE	otherId	citation	
	otherIdValue	FALSE	FALSE	FALSE	FALSE	otherId	citation	
	author	TRUE	FALSE	TRUE	TRUE		citation	http://purl.org/dc/terms/creator
	authorName	FALSE	TRUE	TRUE	TRUE	author	citation	
	authorAffiliation	FALSE	TRUE	TRUE	FALSE	author	citation	
	authorIdentifierScheme	FALSE	FALSE	TRUE	FALSE	author	citation	http://purl.org/spar/datacite/AgentIdentifierScheme
	authorIdentifier	FALSE	FALSE	TRUE	FALSE	author	citation	http://purl.org/spar/datacite/AgentIdentifier

Source: [Dataverse Metadata Schema](#)

Metadata Citation block as RDF

```
@prefix citation: <https://dataverse.org/schema/citation/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

citation: citation:accessToSources [ citation:schema_description "Level of documentation of the original sources." ;
    citation:schema_displayOrder 7.7e+01 ;
    citation:schema_fieldType "textbox" ;
    citation:schema_metadatablock_id "citation" ;
    citation:schema_name "accessToSources" ;
    citation:schema_title "Documentation and Access to Sources" ] ;
citation:alternativeTitle [ citation:schema_description "A title by which the work is commonly referred, or an abbreviation of the title." ;
    citation:schema_displayOrder 2e+00 ;
    citation:schema_fieldType "text" ;
    citation:schema_metadatablock_id "citation" ;
    citation:schema_name "alternativeTitle" ;
    citation:schema_title "Alternative Title" ] ;
citation:alternativeURL [ citation:schema_description "A URL where the dataset can be viewed, such as a personal or project website. " ;
    citation:schema_displayFormat "<a href=\"#VALUE\" target=\"_blank\">#VALUE</a>" ;
    citation:schema_displayOrder 3e+00 ;
    citation:schema_fieldType "url" ;
    citation:schema_metadatablock_id "citation" ;
    citation:schema_name "alternativeURL" ;
    citation:schema_title "Alternative URL" ;
    citation:schema_watermark "Enter full URL, starting with http://" ] ;
citation:author [ skos:broader citation:authorAffiliation,
    citation:authorIdentifier,
    citation:authorName ;
    citation:authorAffiliation [ citation:schema_advancedSearchField "True" ;
        citation:schema_description "The organization with which the author is affiliated." ;
        citation:schema_displayFormat "(#VALUE)" ;
        citation:schema_displayOrder 9e+00 ;
        citation:schema_displayoncreate "True" ;
        citation:schema_facetable "True" ;
        citation:schema_fieldType "text" ;
        citation:schema_metadatablock_id "citation" ;
        citation:schema_name "authorAffiliation" ;
        citation:schema_parent "author" ;
        citation:schema_title "Affiliation" ] ;
```

Source: [SEMAF client](#) (CLARIAH.nl project)

Relationships (predicates) in Dataverse metadata schema

Every predicate such as `skos:exactMatch` refers to the relationship between the subject (fieldname) and object (termURI):

```
citation:alternativeTitle skos:exactMatch "http://purl.org/dc/terms/alternative" .
citation:alternativeURL skos:exactMatch "https://schema.org/distribution" .
citation:dateOfDeposit skos:exactMatch "http://purl.org/dc/terms/dateSubmitted" .
citation:kindOfData skos:exactMatch "http://rdf-vocabulary.ddialliance.org/discovery#kindOfData" .
citation:otherReferences skos:exactMatch "http://purl.org/dc/terms/references" .
citation:relatedDatasets skos:exactMatch "http://purl.org/dc/terms/relation" .
citation:subject skos:exactMatch "http://purl.org/dc/terms/subject" .
citation:title skos:exactMatch "http://purl.org/dc/terms/title" .
citation:contributor skos:exactMatch "http://purl.org/dc/terms/contributor" ;
    skos:narrower citation:contributorName .

citation:authorAffiliation skos:altLabel "authorAffiliation" ;
    skos:broader citation:author ;
    skos:prefLabel "Affiliation" .

citation:authorIdentifier skos:altLabel "authorIdentifier" ;
    skos:broader citation:author ;
    skos:exactMatch "http://purl.org/spar/datacite/AgentIdentifier" ;
    skos:prefLabel "Identifier" .
```


Dataverse metadata export in OAI_ORE format are triples

```
dcterms:modified: "2023-08-17"
dcterms:creator: "Harvard Dataverse"
@type: "ore:ResourceMap"
▼ @id: "https://dataverse.harvard.edu/api/datasets/export?exporter=OAI_ORE&persistentId=https://doi.org/10.7910/DVN/63MIDP"
▼ ore:describes:
  ▼ author:
    citation:authorName: "Ravina, Mark"
    citation:authorAffiliation: "University of Texas at Austin"
    authorIdentifierScheme: "ORCID"
    authorIdentifier: "0000-0002-8726-7618"
  ▼ citation:keyword:
    citation:keywordValue: "Japanese geography"
  ▼ citation:dsDescription:
    ▼ citation:dsDescriptionValue: "File descriptions:\n\nVillage_level_calculations.R – calculates village-level metrics\n\nParcels_sample.txt – random sample of 37,295 parcels from 25,000 unique locations, a random sample of a complete data set of 65,201 unique locations\n\nShikoku_Voronoi_map.R – code to generate Voronoi map "Figure_11_Interactive_map_of_Iyo.html"\n\nShikoku_Voronoi_data.txt – data for Shikoku_Voronoi_map.R\n\nngadm40_JPN_shp – folder os shapefiles for Shikoku_Voronoi_map.R\n\nDomain_Simpson_complete.txt – complete domain-level data for logit calculations, based on all 65,201 locations and 97,553 parcels."
  ▼ citation:datasetContact:
    citation:datasetContactName: "Ravina, Mark"
    citation:datasetContactAffiliation: "University of Texas at Austin"
    citation:datasetContactEmail: "mr56267@eid.utexas.edu"
  ▼ title: "Replication Data for: Algorithmic Maps and the Political Geography of Early-modern Japan"
  dateOfDeposit: "2023-08-17"
  ▼ subject:
    0: "Arts and Humanities"
    1: "Social Sciences"
  citation:depositor: "Ravina, Mark"
  @id: "https://doi.org/10.7910/DVN/63MIDP"
```

Export Metadata ▼

- OAI_ORE
- DataCite
- OpenAIRE
- Schema.org JSON-LD
- DDI
- Dublin Core
- DDI HTML Codebook
- JSON



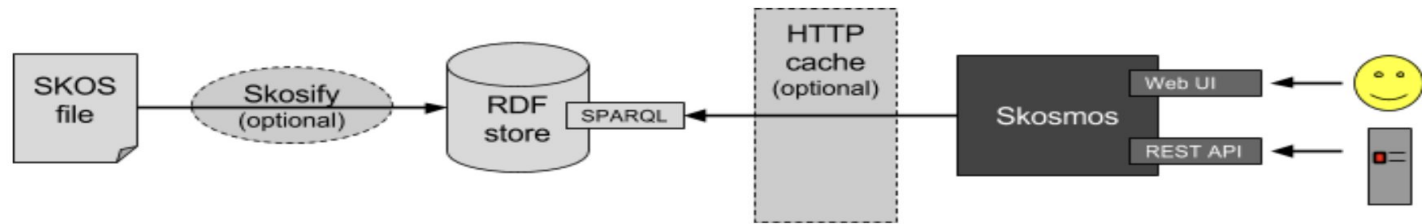
We need more controlled values, ORCID, Wikidata, ...!

Link: https://dataverse.harvard.edu/api/datasets/export?exporter=OAI_ORE&persistentId=doi%3A10.7910/DVN/63MIDP

External CV support as a key to FAIR interoperability

DANS has developed CV support as plugin solution in the Horizon 2020 funded SSHOC EU project, in the collaboration with GDCC and Harvard IQSS. This integration is based on Skosmos developed by National Library of Finland.

James D. Myers, & Vyacheslav Tykhonov. (2023). A Plug-in Approach to Controlled Vocabulary Support in Dataverse. <https://doi.org/10.5281/zenodo.8133723>



Skosmos architecture

ELLST thesaurus hosted by Skosmos

Key features:

- Skosmos is framework powered by **Jena Fuseki** triple store with **SPARQL**
- flexible API with search interface
- export in RDF
- Docker setup
- available as external component in DANS “Archive in a box”
- community support

The screenshot shows the Skosmos interface for the ELLST Thesaurus. The top navigation bar includes 'Vocabularies', 'About', 'Feedback', 'Help', and 'Interface language: English (United States)'. The main header is 'ELSST Thesaurus' with a search bar set to 'Content language English'. The left sidebar shows an alphabetical index and a list of terms, with 'RELIGIOUS EXPERIENCE' highlighted. The main content area displays the hierarchy 'RELIGION > RELIGIOUS EXPERIENCE' and provides details for the preferred term 'RELIGIOUS EXPERIENCE', including broader concepts (RELIGION), related concepts (MEDITATION, MYSTICISM), entry terms (ABJURATION OF FAITH, RELIGIOUS CONVERSION), identifier (urn:ddi:int.cessda.elsst:c9d00f39-4285-4911-852f-a2aaf20d3c8b:3), is a version of (elsst:c9d00f39-4285-4911-852f-a2aaf20d3c8b (elsst)), date issued (2022-09-20), has previous version (https://elsst.cessda.eu/id/2/c9d00f39-4285-4911-852f-a2aaf20d3c8b (elsst)), and translations in other languages (Czech: NÁBOŽENSKÁ ZKUŠENOST, Danish: RELIGIØS OPLEVELSE, Dutch: RELIGIEUZE ERVARING, Finnish: USKONNOLLISET KOKEMUKSET).

Source: [ODISSEI Skosmos](#)

Dataverse integration with Skosmos is term lookup based

Subject * ?

Keyword ?

Vocabulary ? Term ? + -

Vocabulary URL ?

Vocabulary ? Term ? + -

Vocabulary URL ?

Related Publication ?

Citation ?

- COVID-19
- COVID-19 pandemic
- COVID-19 pandemic in India
- Covi
- Covi
- Covi
- SARS-CoV-2

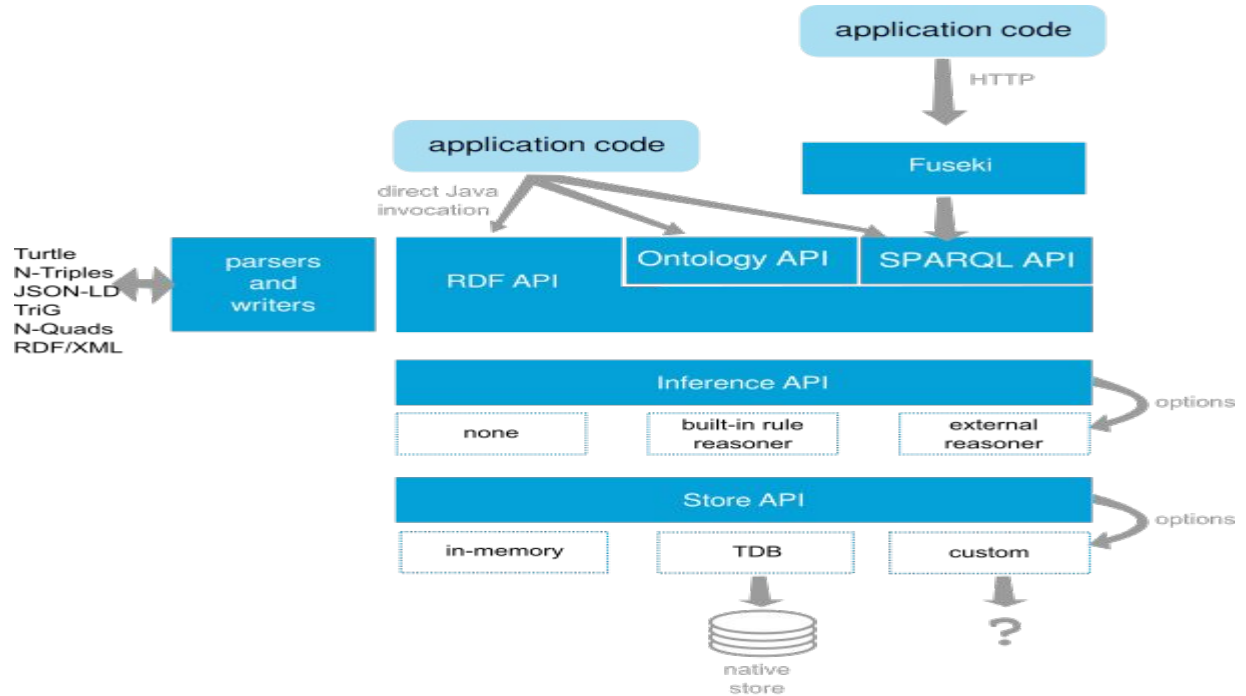
Resulting in TermURI

External controlled vocabularies support contributed by [SSHOC project](#) (data infrastructure for the EOSC)

Meet “built-in” Dataverse triple store: Jena Fuseki

Jena stores information as RDF triples in directed graphs, and allows your code to *add, remove, manipulate, store and publish* that information

Fuseki is a SPARQL server that provides REST-style API endpoint using the SPARQL protocol over HTTP



Querying Dataverse as Linked Open Data (ODISSEI project)

SPARQL query

To try out some SPARQL queries against the selected dataset, enter your query here.

EXAMPLE QUERIES

Selection of triples

Selection of classes

PREFIXES

rdf

rdfs

owl

xsd



SPARQL ENDPOINT

/odissei/query

CONTENT TYPE (SELECT)

JSON

CONTENT TYPE (GRAPH)

Turtle

```
5 PREFIX citation: <https://dataverse.org/schema/citation/>
6 PREFIX ore: <http://www.openarchives.org/ore/1.0/terms/>
7 PREFIX distributor: <https://dataverse.org/schema/citation/distributor#>
8 PREFIX variable: <https://portal.staging.odissei.nl/schema/variableInformation/variable#>
9 SELECT ?id ?title ?distributorname
10 WHERE {
11   ?id dcterms:title ?title;
12       dcterms:subject ?subject;
13       schema:datePublished ?date;
14       citation:Distributor ?distributor.
15   ?distributor distributor:Name ?distributorname.
16   ?id schema:datePublished "2010-01-01"
17 }
```

QUERY RESULTS



Table

Raw Response



Showing 1 to 1 of 1 entries

Search:

Show


50



entries

id	title	distributorname	
1	<doi:10.57934/0b01e41080216906>	"Minimumloonklasse van banen van werknemers"	"Centraal Bureau voor Statistiek"

Using grlc to manage SPARQL queries through github



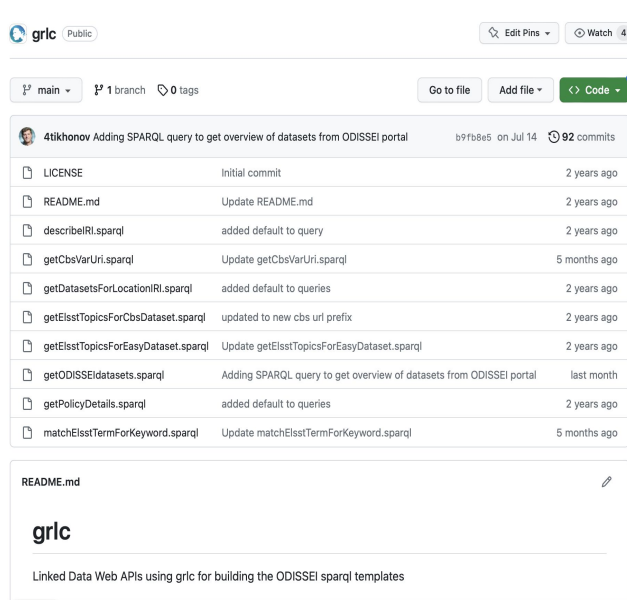
grlc makes all your [Linked Data](#) accessible to the Web by automatically converting your SPARQL queries into RESTful APIs. With (almost) no effort! Simply:

- 1 Create a GitHub repository, and store all your SPARQL queries in there (like in [this example](#)). If you don't have a GitHub account, [go get one](#). You can also just write down the username and the repository name of somebody else :-)
- 2 Go to the address bar of this page, and append `/api/github_username/repository_name` to it. So if I want the API derived from GitHub's username `foo` and repository `bar`, I append `/api/foo/bar/` to the domain name (`/api/foo/bar/api-docs` will work too). Now hit enter. Done!

[Show me an example SPARQL repo](#)

[Show me the equivalent API](#)

Take a look at the increasing number of [users](#) and [SPARQL repositories](#) on GitHub that are using **grlc** to generate APIs!



grlc Public

main 1 branch 0 tags

Go to file Add file Code

4tkhonov Adding SPARQL query to get overview of datasets from ODISEI portal b9fb8e5 on Jul 14 92 commits

LICENSE	Initial commit	2 years ago
README.md	Update README.md	2 years ago
describeIRI.sparql	added default to query	2 years ago
getCbsVarUri.sparql	Update getCbsVarUri.sparql	5 months ago
getDatasetsForLocationIRI.sparql	added default to queries	2 years ago
getElstTopicsForCbsDataset.sparql	updated to new cbs url prefix	2 years ago
getElstTopicsForEasyDataset.sparql	Update getElstTopicsForEasyDataset.sparql	2 years ago
getODISEIdatasets.sparql	Adding SPARQL query to get overview of datasets from ODISEI portal	last month
getPolicyDetails.sparql	added default to queries	2 years ago
matchElstTermForKeyword.sparql	Update matchElstTermForKeyword.sparql	5 months ago

README.md

grlc

Linked Data Web APIs using grlc for building the ODISEI sparql templates



grlc b9fb8e58e34f8a02d61b93cdc3b12a090a1bf042

[Base URL: `grlc.io/api-git/odissei-data/grlc/`]
[/api/odissei-data/grlc/swagger](#)

Linked Data Web APIs using grlc for building the ODISEI sparql templates
[odissei-data - Website](#)
[License](#)

default

- GET `/describeIRI`
- GET `/getCbsVarUri`
- GET `/getDatasetsForLocationIRI`
- GET `/getElstTopicsForCbsDataset`
- GET `/getElstTopicsForEasyDataset`
- GET `/getODISEIdatasets`
- GET `/getPolicyDetails`
- GET `/matchElstTermForKeyword`

Link: https://grlc.io/api/odissei-data/grlc#/default/get_getODISEIdatasets

grlc usage example in ODISSEI

SPARQL query stored in github:

GET /getODISSEIdatasets

```
#+ endpoint: http://fuseki.experimental.odissei.nl/odissei/sparql

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <https://schema.org/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX citation: <https://dataverse.org/schema/citation/>
PREFIX ore: <http://www.openarchives.org/ore/1.0/terms/>
SELECT ?id ?title ?description ?keyword
WHERE {
  ?id dcterms:title ?title;
      citation:Description ?descid;
      citation:Keyword ?keyid.
  ?descid ?descstype ?description.
  ?keyid ?k ?keyword
}
```

Response from SPARQL transformed by glibc in specified format:

Request URL

```
https://grlc.io/api-git/odissei-data/grlc/getODISSEIdatasets?endpoint=http%3A%2F%2Ffuseki.experimental.odissei.nl%2Fodissei%2Fsparql
```

Server response

Code Details

200

Response body

```
id,title,description,keyword
doi:10.17026/dans-xdu-8j3m,Green Xenophobia Project,<p>The experimental study is conducted to investigate the public opinion on how immigration leads to overpopulation and in turn might damage the sustainable development and environmental protection.</p>,Demography and population
```


Dataverse metadata semantic enrichment with Skosmos CVs

The screenshot shows the Dataverse ODISSEI Portal interface. At the top, the Dataverse logo and navigation links (Search, User Guide, Support, Sign Up, Log In) are visible. The main header includes the CBS (Centraal Bureau voor de Statistiek) logo and the text 'ODISSEI Portal >'. A search bar contains the term 'education' and a search button. Below the search bar, there are filters for 'Dataverses (0)' and 'Datasets (105)'. The 'Publication Year' filter shows results for 2021 (1), 2020 (1), 2019 (3), 2018 (5), and 2017 (6). The 'Keyword Term' filter shows 'Onderwijs (16)' and several related terms. The search results section displays '1 to 10 of 105 Results'. The first result is 'Adult Education Survey- Onderzoek onder volwassenen naar het volgen van opleidingen (lange, korte)' from Jan 1, 2011, by the Centraal Bureau voor de Statistiek. The second result is 'Scholieren en studenten in een kalenderjaar' from Jan 1, 1999, also by the Centraal Bureau voor de Statistiek. The search results do not include the keyword 'education' in the dataset metadata.

Multilingual (semantic) search: keyword 'education' isn't included in dataset metadata!

KG ecosystem: LOD Summarizer as “prompt” for SPARQL

did-summarizer

Linked Data summarizer driven by Decentralized Identifiers (DIDs)

Developed by DANS Labs, funded by [CLARIAH project](#).

The main aim of the Summarizer service is to gain an overview about which vocabularies are already used in CLARIAH, or might be useful in CLARIAH. The core is to build an analyzing pipeline containing data collection, vocabulary analysis, report. While concentrating on the automatised pipeline, we also indicate at which point where expert/manual curation is needed.

To bridge between different knowledge domains it is needed to find communalities, cross-walks, mappings between vocabularies. A prerequisite for this is to gain an overview what vocabularies exist (VOCABULARY part) and how they are used (DATASET part). Despite of many existing registries this problem is by no means solved, nor are there standard, off-the shelf solutions for gaining such an overview.

Decentralized identifiers (DIDs) are being used to create resolvable globally accessible unique and persistent identifiers to support various Linked Data tasks in FAIR way:

- assign DID to SPARQL query to make it persistent
- cache vocabulary concept content and relationships
- store and get statistics of usage for vocabulary concepts available in the time dimension, just like the Internet Archive
- assign unique DID to other services used in CLARIAH pipelines ...

<https://github.com/Dans-labs/did-summarizer>

```
curl -X 'GET' \  
'http://0.0.0.0:8001/summarizer?url=https%3A%2F%2Fraw.githubusercontent.com%2FAKSW%2Fdssn.rdf%2  
-H 'accept: application/json'
```

Response body:

```
{  
  "statements": {  
    "statements": 119,  
    "unique objects": 64,  
    "unique predicates": 25,  
    "unique subjects": 22  
  },  
  "prefixes": {  
    "http://www.w3.org/2000/01/rdf-schema#": "rdfs",  
    "http://www.w3.org/2002/07/owl#": "owl",  
    "http://www.w3.org/2004/02/skos/core#": "skos",  
    "http://xmlns.com/foaf/0.1/": "foaf",  
    "http://rdfs.org/sioc/ns#": "sioc",  
    "http://usefulinc.com/ns/doap#": "doap",  
    "http://www.w3.org/2003/06/sw-vocab-status/ns#": "vs",  
    "http://purl.org/dc/terms/": "dct",  
    "http://purl.org/net/dssn/": "dssn"  
  },  
  "stats": {  
    "dssn": 126,  
    "rdfs": 60,  
    "vs": 13,  
    "doap": 1,  
    "foaf": 20,  
    "owl": 15,  
    "dct": 1,  
    "sioc": 11,  
    "skos": 1  
  }  
}
```

KG ecosystem: controlled vocabulary recommender

Vocabulary Recommender Command-line interface (CLI) was developed by [Triply](#) and provides a recommendation interface which returns relevant Internationalized Resource Identifiers (IRIs) based on the search input. It works with SPARQL or Elasticsearch endpoints which contain relevant vocabulary datasets.

[DANS](#) has created API service out of it.

Usage example:

```
curl -X 'GET' \  
'http://0.0.0.0:8001/recommend?searchTerm=person&searchClass=class' \  
-H 'accept: application/json'
```

Response:

```
[  
  [  
    {  
      "searchTerm": "person",  
      "vocabs": [  
        "https://w3id.org/pnv#"  
      ],  
      "homogeneous": [  
        {  
          "iri": "https://w3id.org/pnv#Person",  
          "score": 1,  
          "vocabPrefix": "https://w3id.org/pnv#",  
          "vocabDomain": "https://w3id.org/pnv#",  
          "description": "A Person is a human being whose individual existence can somehow be doc",  
          "category": "class"  
        }  
      ],  
      "single": [  
        {  
          "iri": "http://xmlns.com/foaf/0.1/Person",  
          "score": 0.8853529652138054,  
          "vocabPrefix": "foaf",  
          "vocabDomain": "http://xmlns.com/foaf/0.1/",  
          "description": "A person.",  
          "category": "class"  
        }  
      ],  
    }  
  ],  
]
```

Knowledge Graphs as Supervisor for LLM

“You’ll see the real power of the usage of Large Language Models when you’ll be able to “talk” to data just like to your assistant, or even ask questions to your own documents”.

Anton Polishko, ex-Google, CoronaWhy

LLaMA: A foundational, 65-billion-parameter large language model

Querying on historical facts is OK

```
~ Tue Aug 22 2023 01:28 am
ollama run stable-beluga:70b
>>> Who was the 45th president in the US? Give response as RDF triples
  Subject: 45th President of the United States
  Predicate: is_a
  Object: Donald Trump

>>> Now list other 5 last presidents. Also as RDF triples
  Subject: 44th President of the United States
  Predicate: is_a
  Object: Barack Obama

  Subject: 43rd President of the United States
  Predicate: is_a
  Object: George W. Bush

  Subject: 42nd President of the United States
  Predicate: is_a
  Object: Bill Clinton

  Subject: 41st President of the United States
  Predicate: is_a
  Object: George H.W. Bush

  Subject: 40th President of the United States
```

Does know nothing about current events

```
curl http://172.28.112.1:8081/v1/chat/completions \
-H "Content-Type: application/json" \
-d '{
  "messages": [
    { "role": "user", "content": "### Who is winner of Eurovision in 2023? \n### Response:" }
  ],
  "temperature": 0.5,
  "max_tokens": -1,
  "stream": false
}'
{
  "choices": [
    {
      "finish_reason": "stop",
      "index": 0,
      "message": {
        "content": " I'm sorry, but the Eurovision Song Contest for 2023 has not taken place yet. The winner will be determined in the future when the competition occurs.",
        "role": "assistant"
      }
    }
  ],
  "created": 1692705863,
  "id": "chatcmpl",
  "model": "LLaMA_CPP",
  "object": "chat.completion",
  "truncated": false,
  "usage": {
    "completion_tokens": 38,
    "prompt_tokens": 52,
    "total_tokens": 90
  }
}
```

YouTube: <https://www.youtube.com/watch?v=m-phwvbbgXE>

MuseIT collections powered by Now.Museum

The screenshot displays the Dataverse web interface. At the top, the Dataverse logo is on the left, and navigation links for Search, User Guide, Support, Sign Up, and Log In are on the right. Below the header, a message states: "Test collection prepared by the MuseIT project using the Now.Museum technology." A search bar contains the text "Search this dataverse..." and a "Q" icon, with a link to "Advanced Search". On the left sidebar, there are filters for "Dataverses (0)", "Datasets (3,718)", and "Files (0)". Below these are filters for "Publication Year" (2023 (3,718)), "Author Name" (MuseIT (3,718)), "Subject" (Arts and Humanities (3,718)), and "Keyword Term" (media (1,954), song contest (863), eurovision song (784), Eurovision (562), grand final (473)). A "More..." link is at the bottom of the sidebar. The main content area shows "1 to 10 of 3,718 Results" and a "Sort" dropdown. Three search results are visible, each with a document icon, a title, a date, a snippet, and a DOI link. The first result is "esc_charts: Eurovision Charts@esc_charts-31mMost-streamed #Eurovision 2023 songs on..." dated Aug 22, 2023. The second is "Eurovision winner shows the power of music for nursing home residents" dated Aug 21, 2023. The third is "Duncan Laurence: Ο νικήτης της Eurovision 2019 παντρεύτηκε με τον επί χρόνια σύντροφό του" dated Aug 21, 2023.

- MuseIT is Horizon 2020 project to co-design and develop an inclusive multisensory platform with interactive technologies for people with disabilities to have enriched engagement with cultural assets and experiences.
- Now.Museum is Dataverse based hosting platform for MuseIT, CoronaWhy and other projects. In the collaboration with Yves Rozenholc, University Paris Cite.
- The technology powered by shared knowledge graph

Event/News registration in Now.Museum Dataverse

Eurovision 2023

(MuselT)

Root > MuselT > Eurovision 2023 >

Loreen calls Eurovision win 'surreal'

Version 1.0



MuselT, 2023, "Loreen calls Eurovision win 'surreal'", <https://doi.org/10.5072/FK2/JWI30C>, Root, V1

[Cite Dataset](#) ▾

[Learn about Data Citation Standards.](#)

Contact Owner

Share

Dataset Metrics 

0 Downloads 

Description

Loreen calls Eurovision win 'surreal'Eurovision winner Loreen has said she is still comprehending her 'surreal' victory'. The Swedish popstar claimed her second victory in Liverpool on Saturday with song Tattoo. She became the first woman to win the contest twice, having first won in 2012 with Euphoria.Popular VideosWatch MoreMore VideosHeadlinesSunday World NewsletterSign up for the latest news and updatesMore VideosDownload the Sunday World appNow download the free app for all the latest Sunday World News, Crime, Irish Showbiz and Sport. Available on Apple and Android devicesLatest News

Subject

Arts and Humanities

Keyword

loreen calls, calls eurovision, eurovision win, surreal'eurovision winner, winner loreen, swedish popstar, popstar claimed, second victory, song tattoo, contest twice, euphoria.popular videoswatch, videoswatch moremore, moremore videosheadlinesunday, videosheadlinesunday world, world newslettersign, latest news, updatesmore videosdownload, sunday world, world appnow, appnow download, free app, latest sunday, world news, irish showbiz, android deviceslatest, deviceslatest news, media

License/Data Use Agreement

[Custom Dataset Terms](#)

Assistant prompt in LLaMA

Prompt

```
curl http://172.28.112.1:8081/v1/chat/completions \  
-H "Content-Type: application/json" \  
-d '{  
  "messages": [  
    { "role": "user", "content": "### Instruction: Give me search terms helping to find out who was the Eurovision winner in 2023? Do not mention names of winners. \n### Give response in JSON format." }  
  ],  
  "temperature": 0.5,  
  "max_tokens": -1,  
  "stream": false  
'
```



Response

```
{  
  "choices": [  
    {  
      "finish_reason": "stop",  
      "index": 0,  
      "message": {  
        "content": " {\n      \"searchTerms\": [\n        \"Eurovision winner\",\n        \"2023\",\n        \"winner\"\n      ]\n}",  
        "role": "assistant"  
      }  
    }  
  ],  
  "created": 1692705207,  
  "id": "chatcmpl",  
  "model": "LLaMA_CPP",  
  "object": "chat.completion",  
  "truncated": false,  
  "usage": {  
    "completion_tokens": 37,  
    "prompt_tokens": 75,  
    "total_tokens": 112  
  }  
}
```


SPARQL query to find out who won Eurovision in 2023

PREFIXES

rdf rdfs owl xsd +

SPARQL ENDPOINT: /nowmuseum/sparql

CONTENT TYPE (SELECT): JSON

CONTENT TYPE (GRAPH): Turtle

```
5 PREFIX citation: <https://dataverse.org/schema/citation/>
6 PREFIX ore: <http://www.openarchives.org/ore/1.0/terms/>
7 PREFIX id: <https://doi.org/>
8 SELECT distinct ?title
9 WHERE {
10   ?id schema:distribution ?url;
11       dcterms:title ?title;
12       citation:dsDescription ?description;
13       citation:keyword ?keywordID.
14   ?description citation:dsDescriptionValue ?summary.
15   ?keywordID citation:keywordValue ?keyword.
16   filter contains(?title,"Eurovision")
17   filter contains(?keyword,"winner")
18   filter contains(?title,"2023")
}
```

QUERY RESULTS

Table Raw Response

Showing 1 to 1 of 1 entries

Search: Show 100 entries

	title
1	"Sweden wins Eurovision Song Contest 2023 with Tattoo by Loreen"

Showing 1 to 1 of 1 entries

Prompt to create SPARQL query with LLMs

```
Assuming you have an OAI-ORE repository with the following metadata in Dublin Core: <?xml version='1.0' encoding='UTF-8'?><metadata
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"
xmlns="http://dublincore.org/documents/dcmi-terms/"><dcterms:title>Taking the perspective of narrative characters</dcterms:title>
<dcterms:identifier>https://doi.org/10.34894/5KHVJV</dcterms:identifier><dcterms:creator>Vogels, Jorrig</dcterms:creator><dcterms:creator>Bimpikou,
Sofia</dcterms:creator><dcterms:creator>Kapelle, Owen</dcterms:creator><dcterms:creator>Maier, Emar</dcterms:creator>
<dcterms:publisher>DataverseNL</dcterms:publisher><dcterms:issued>2023-09-04</dcterms:issued><dcterms:modified>2023-09-04T07:54:54Z</dcterms:modified>
<dcterms:description>An ongoing debate in the interpretation of referring expressions concerns the degree to which listeners make use of perspective information during
referential processing. We aim to contribute to this debate by considering perspective shifting in narrative discourse. In a web-based mouse-tracking experiment in Dutch, we
investigated whether listeners automatically shift to a narrative character's perspective when resolving ambiguous referring expressions, and how different linguistic
perspective-shifting devices affect if and when listeners switch to another perspective. We compared perspective-neutral, direct, and free indirect discourse, manipulating what
objects are visible to the character. Our results do not show a clear effect of the perspective shifting devices on participants' eventual choice of referent, but our online mouse-
tracking data reveal processing differences that suggest that listeners are indeed sensitive to the conventional markers of perspective shift associated with direct and (to a
lesser degree) free indirect discourse.</dcterms:description><dcterms:subject>Arts and Humanities</dcterms:subject><dcterms:subject>perspective
shifting</dcterms:subject><dcterms:subject>narrative</dcterms:subject><dcterms:subject>free indirect discourse</dcterms:subject><dcterms:subject>mouse
tracking</dcterms:subject><dcterms:subject>reference resolution</dcterms:subject><dcterms:language>Dutch</dcterms:language>
<dcterms:language>English</dcterms:language><dcterms:isReferencedBy>Vogels, J., Bimpikou, S., Kapelle, O. E., and Maier, E. (2021). Taking the perspective of narrative
characters: a mouse-tracking study on the processing of ambiguous referring expressions in narrative discourse. Linguistics in the Netherlands 38(1), 144-162.
https://doi.org/10.1075/avt.00054.vog., doi, 10.1075/avt.00054.vog, https://doi.org/10.1075/avt.00054.vog</dcterms:isReferencedBy><dcterms:contributor>Elsenga,
Christina</dcterms:contributor><dcterms:dateSubmitted>2022-10-31</dcterms:dateSubmitted><dcterms:license>CC-BY-4.0</dcterms:license></metadata>. Give me
corresponding SPARQL query to get name, description, and creator of the item with identifier "https://doi.org/10.34894/5KHVJV":
```

Resulting SPARQL query from LLM

Sure! Here's a SPARQL query that should retrieve the name, description, and creator of the item with identifier "<https://doi.org/10.34894/5KHVJV>":

```
PREFIX dcterms: <http://purl.org/dc/terms/>
```

```
PREFIX dublincore: <http://dublincore.org/documents/dcmi-terms/>
```

```
SELECT ?name, ?description, ?creator
```

```
FROM <https://doi.org/10.34894/5KHVJV>
```

```
WHERE {
```

```
  ?item dcterms:title ["Taking the perspective of narrative characters"] .
```

```
  ?item dublincore:creator [ ?creator ] .
```

```
  ?item dcterms:description ["An ongoing debate in the interpretation of referring expressions concerns the degree to which listeners make use of |
```

```
}]
```

Here's how the query works:

1. The **PREFIX** statements define the prefixes for the Dublin Core (dublincore:) and DC Terms (dcterms:) vocabularies.
2. The **SELECT** statement retrieves three properties from the item with identifier "<https://doi.org/10.34894/5KHVJV>": name, description, and creator.
3. The **WHERE** clause filters the results to only include items that have a title property with value "Taking the perspective of narrative characters", and a creator property with at least one value (the `[]` syntax indicates a variable number of values).
4. The `dcterms:title` and `dublincore:creator` properties are used to retrieve the title and creator information, respectively.
5. The `dcterms:description` property is retrieved using the `[]` syntax to indicate that it is a variable number of values.

Note that this query assumes that the item with identifier "<https://doi.org/10.34894/5KHVJV>" has a title, creator, and description property, and that these properties are represented using the Dublin Core and DC Terms vocabularies. If your dataset uses different vocabularies or representations, you may need to modify the query according



Prompting ChatGPT with Dublin Core

V

Can you give me SPARQL query to get all titles and subject for records in Dublin Core format?



To retrieve all titles and subjects for records in Dublin Core format, you can use the following SPARQL query:

```
sparql Copy code
PREFIX dcterms: <http://purl.org/dc/terms/>

SELECT ?title ?subject
WHERE {
  ?resource a ?type ;
            dcterms:title ?title .
  ?resource dcterms:subject ?subject .
}
```

This query assumes that:

- The records are represented using Dublin Core (DC) terms.
- The titles are stored using `dcterms:title`.
- The subjects are stored using `dcterms:subject`.

You can execute this query against your RDF triple store or dataset to retrieve all titles and subjects for records in Dublin Core format. Adjust the prefixes and property names as needed to match your specific dataset schema.



SPARQL ENDPOINT: /odissel CONTENT TYPE (SELECT): JSON CONTENT TYPE (GRAPH): Turtle

```
1 PREFIX dcterms: <http://purl.org/dc/terms/>
2
3 SELECT ?title ?subject
4 WHERE {
5   ?resource a ?type ;
6             dcterms:title ?title .
7   ?resource dcterms:subject ?subject .
8 }
9
```

QUERY RESULTS

Table Raw Response

Showing 1 to 50 of 6,116 entries

title	subject
1 "Green Xenophobia Project"	"Social Sciences"
2 "Green Xenophobia Project"	"Social Sciences"
3 "An Instructional Manipulation Check (IMC) to prevent respondents from using a response style"	"Social Sciences"
4 "An Instructional Manipulation Check (IMC) to prevent respondents from using a response style"	"Social Sciences"
5 "Computer Adaptive Test (CAT)"	"Social Sciences"
6 "Computer Adaptive Test (CAT)"	"Social Sciences"
7 "DANS research programme 2018-2020"	"Social Sciences"
8 "DANS research programme 2018-2020"	"Social Sciences"
9 "DANS research programme 2018-2020"	"Arts and Humanities"
10 "DANS research programme 2018-2020"	"Arts and Humanities"
11 "DANS research programme 2018-2020"	"Business and Management"
12 "DANS research programme 2018-2020"	"Business and Management"
13 "DANS research programme 2018-2020"	"Law"

Questions?

Slava Tykhonov, R&D
(DANS-KNAW, the Netherlands)
vyacheslav.tykhonov@dans.knaw.nl