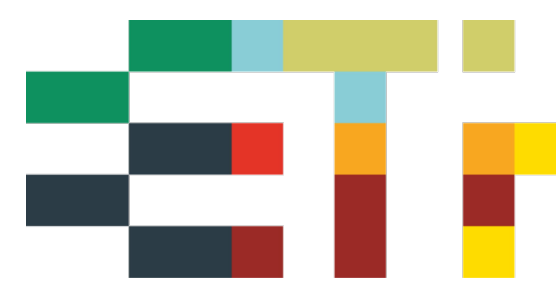




AKSW



# Base Platform for Knowledge Graphs with Free Software

Simon Bin, Claus Stadler, Norman Radtke, Kurt Junghanns,  
Sabine Gründer-Fahrer, Michael Martin

D2R2'23 @ ESWC 2023

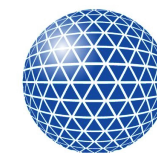
May 28th, 2023

Supported by



Federal Ministry  
for Economic Affairs  
and Climate Action

GA No. 01MK21007A



InfAI®  
Institute for Applied Informatics

# Project Vision

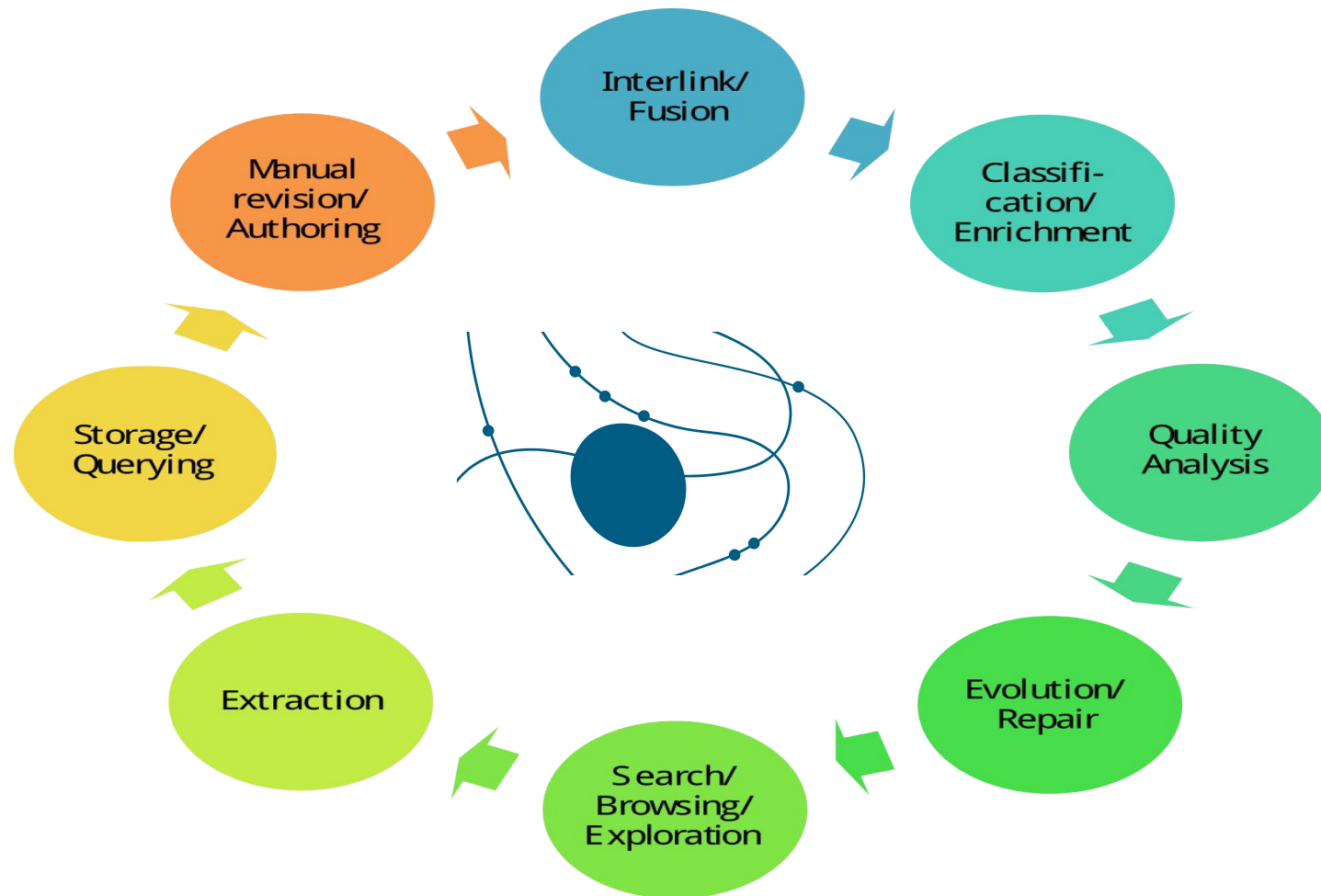


## “Cognitive Economy Intelligence Platform for resilience of economic ecosystems”

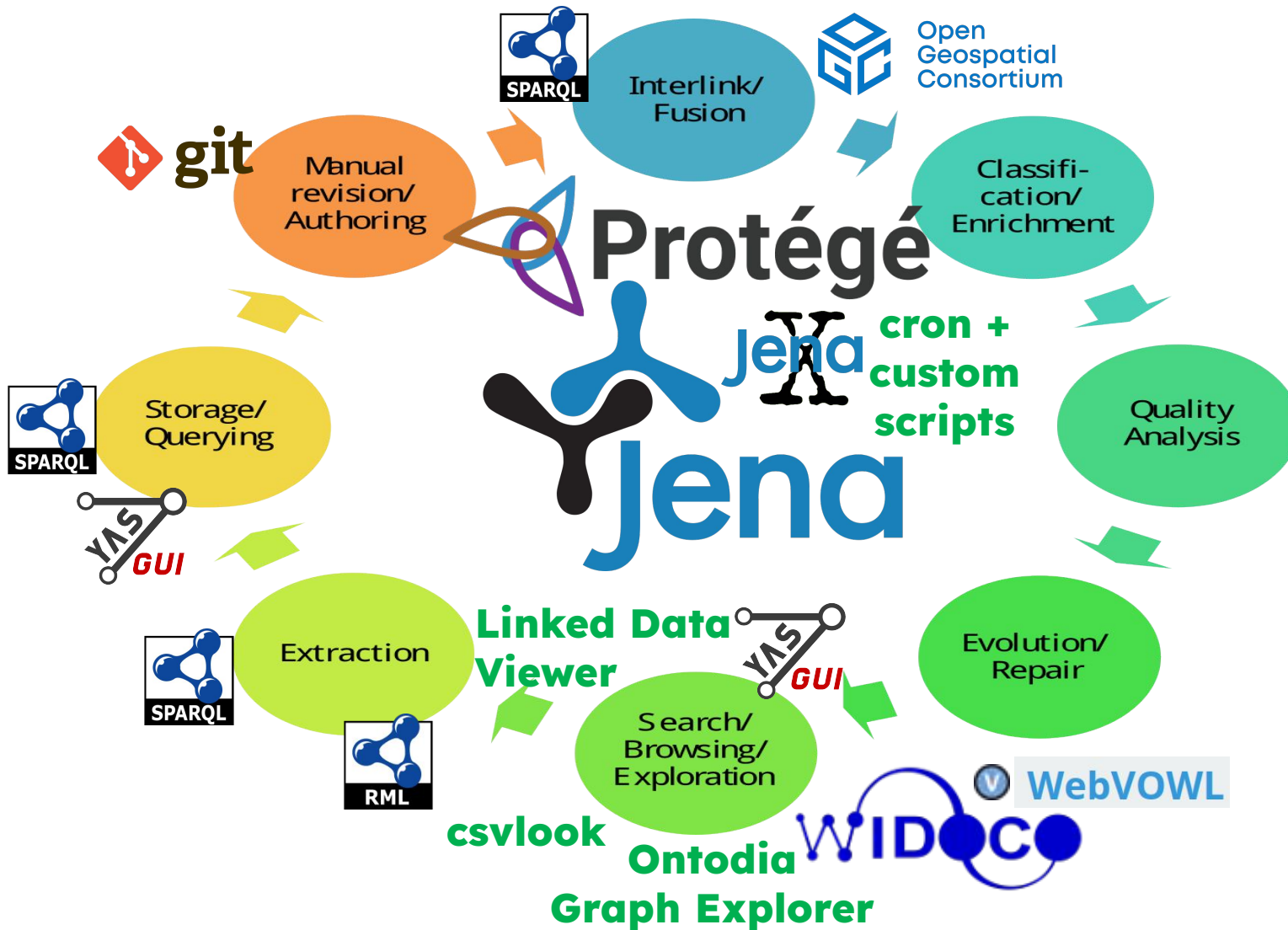
- Intelligent Platform for the integration, structuring, networking, analysis and evaluation of heterogeneous data
- cognitive modelling of data and flexibly configurable analysis tools
- qualitative, up-to-the-minute insights into facts, trends, interdependencies and forecasts for individual value networks
- Analyses for different regions, industries and company sizes as well as on a macroeconomic level
- Information transparency for companies and markets, especially in the SME sector

⇒ **Resilience of existing value chains**

# Linked Data Lifecycle



# Linked Data Lifecycle

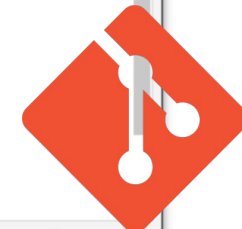


# Ontology creation

The screenshot displays the Protégé ontology editor interface. The main window title is "global (https://schema.coypu.org/global/2.3) : [CoyPu/coy-ontology/ontology/global/coy.ttl]". The breadcrumb navigation shows the path: Event > Disaster > TechnologicalDisaster > Industrial accident. The left sidebar shows a class hierarchy starting with owl:Thing, followed by Company, Event, Conflict, Disaster, NaturalDisaster, TechnologicalDisaster, and Industrial accident. The right pane shows the details for the selected class, "Industrial accident", including its annotations (rdf:type, rdfs:label, rdfs:comment, rdfs:seeAlso) and its description, which states it is a subclass of TechnologicalDisaster.



# Protégé

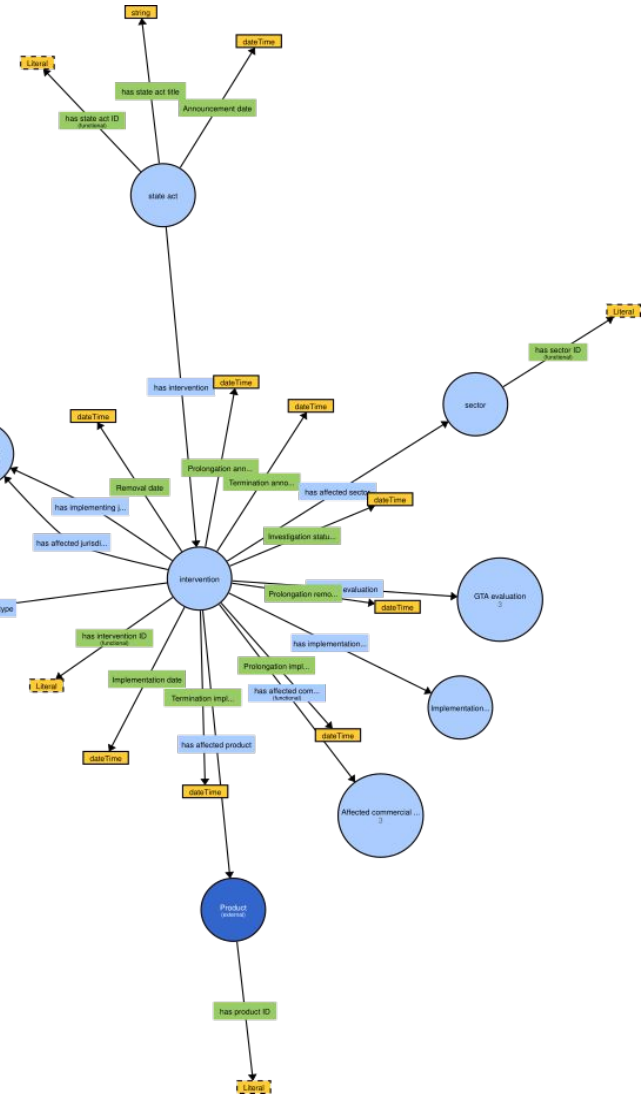


# git

# Ontology documentation



WebVOWL



Ontology Specification Draft

**Customer<sup>C</sup>** back to ToC or Class ToC

**IRI:** <https://schema.coypu.org/global#Customer>  
 An organization that sources something needed such as a product or service from another organization.

**Term status**  
 CoyPu graph: Not currently in use

**has super-classes**  
 Company<sup>C</sup>

**identity militias<sup>C</sup>** back to ToC or Class ToC

**IRI:** <https://schema.coypu.org/global#IdentityMilitias>

**Term status**  
 CoyPu graph: In use, distinct subjects: 7977

**Term rationale**  
 CoyPu graph: -- USED BUT NOT IN ONTOLOGY --

# Data catalogue



## Data Sources Overview

- CoyPu technical docs
- Data Sources main repository

## Graphs & Data

This is a list of Data Sources loaded into `data.coypu.org` on Fuseki.

### Static

Dataset	Named Graph	Domain
AIRPORTS	<a href="https://data.coypu.org/infrastructure/airports/">https://data.coypu.org/infrastructure/airports/</a>	Global airports
CITIES	<a href="https://data.coypu.org/cities/">https://data.coypu.org/cities/</a>	Global cities
CLIMATETRACE	<a href="https://data.coypu.org/infrastructure/climatetrace/">https://data.coypu.org/infrastructure/climatetrace/</a>	Natural resource extraction and processing
CORRUPTION	<a href="https://data.coypu.org/country-statistic/corruption/">https://data.coypu.org/country-statistic/corruption/</a>	Corruption perception index per country
COUNTRIES	<a href="https://data.coypu.org/countries/">https://data.coypu.org/countries/</a>	Counties and regions
DALICC	<a href="http://dalicc.net/licenselibrary/">http://dalicc.net/licenselibrary/</a>	License graph from DaliCC framework

```
:airports
  dct:license <https://metadata.coypu.org/licenses/CC-BY_v4> ;
  dcat:landingPage <https://datacatalog.worldbank.org/search/dataset/0038117/Global-Airports> ;
  t:graph <https://data.coypu.org/infrastructure/airports/> ;
  t:licenseNote "CC-BY-4.0 by Worldbank" ;
  t:licensePage <https://datacatalog.worldbank.org/search/dataset/0038117/Global-Airports> ;
  t:theme "Global airports" ;
  t:type "Dump"
```

(SPARQL + )  
**csvlook**  
( + Markdown renderer)

# Licences in the data catalogue



Talk by **Kurt Junghanns**: *Efficient Use of DALICC in Data Processing Pipelines with Fuzzy License Information.*

**Today 17:00–17:20 in Orpheas**





# RML Mapping Language



```
<map_stops_0> a rr:TriplesMap ;
```

```
rr:subjectMap [ a rr:SubjectMap ;  
  rr:template "http://example.org/stops/{stop_id}"  
];
```

```
rr:predicateObjectMap [ a rr:PredicateObjectMap ;  
  rr:predicateMap [ a rr:PredicateMap ;  
    rr:constant wgs84:long  
  ] ;  
  rr:objectMap [ a rr:ObjectMap ;  
    rml:reference "stop_lon" ;  
    rr:datatype xsd:double ;  
    rr:termType rr:Literal  
  ] ;  
];
```

```
rml:logicalSource [ a rml:LogicalSource ;  
  rml:referenceFormulation ql:CSV ;  
  rml:source "STOPS.csv"  
];
```

```
rr:predicateObjectMap [ a rr:PredicateObjectMap ;  
  rr:predicateMap [ a rr:PredicateMap ;  
    rr:constant <http://vocab.gtfs.org/terms#parentStation>  
  ] ;  
  rr:objectMap [ a rr:ObjectMap ;  
    rr:joinCondition [ rr:child "parent_station" ;  
      rr:parent "stop_id" ] ;  
    rr:parentTriplesMap <map_stops_0>  
  ] ;  
];
```

# Mapping data using Tarql and Javascript



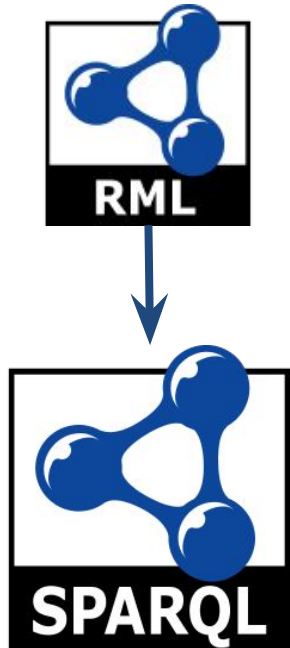
```
33
34   ?actor_1
35     rdfs:label ?actor1_label ;
36     rdf:type ?actor_type ;
37     .
38
39   ?actor_2
40     rdfs:label ?actor2_label ;
41     rdf:type ?actor2_type ;
42     .
43
44 }
45 WHERE {
46   # JS function to parse event date in format 'dd MMMM yyyy' to ISO format
47   #BIND(json:js("function(x) {return new Date(x).toISOString();}", ?event_date) AS ?date)
48   BIND(xsd:date(STRDT(REPLACE(js:toIsoDateString(?event_date), "Z", ""), xsd:dateTime)) AS ?date)
49
50   BIND(IRI(CONCAT(str(data:), ?event_id_cnty)) AS ?event)
51
52   # event subtype
53   BIND(IRI(CONCAT(str(acled:), REPLACE(js:toTitleCase(REPLACE(?sub_event_type, "-", " ")), " ", ""))) AS ?type)
54
55   # geo lon/lat
56   BIND(STRDT(?latitude, xsd:double) AS ?lat)
57   BIND(STRDT(?longitude, xsd:double) AS ?lon)
58
```

# Executing RML Mappings as parallel SPARQL queries



Talk by **Claus Stadler**: *Scaling RML and SPARQL-based Knowledge Graph Construction with Apache Spark.*

**Today 17:20–18:00 in Hermes**



# Mapping data using JSONL + GeoSPARQL



Open  
Geospatial  
Consortium

```
17 CONSTRUCT {
18   ?country
19   | geo:hasGeometry ?geometry_node .
20   ?geometry_node a geo:Geometry ;
21   | geo:asWKT ?geometry .
22 }
23 WHERE
24 {
25   {
26     SELECT
27     ?country
28     ?geometry_node
29     ?geometry
30     {
31       <env:INPUT> url:textLines ?text .
32       BIND(STRDT(?text, xsd:json) AS ?item)
33
34       BIND(json:path(?item, "$.properties") AS ?properties)
35
36       BIND(json:path(?properties, "$.ISO_CODE") AS ?iso3)
37       BIND(URI(concat("https://data.coypu.org/country/", ?iso3)) AS ?country)
38       BIND(URI(concat(str(?country), "/geometry/boundary")) AS ?geometry_node)
39
40       BIND(json:path(?item, "$.geometry") AS ?geometry_)
41       BIND(spatialF:transformDatatype(STRDT(str(?geometry_), geo:geoJSONLiteral), geo:wktLiteral) AS ?geometry)
42     }
43   }
44 }
```

The logo for Jena, featuring the word 'Jena' in a blue, sans-serif font with a large, stylized black 'X' overlaid on the letters 'e' and 'n'.

## Executing RML Mappings as parallel SPARQL queries



Talk by **Lars Michaelis**: *WikiEvents - A Novel Resource for NLP Downstream Tasks*.  
**Tomorrow 16:55–17:15 in Polymnia & Kalia**



# “Pipeline”

```
0 22 1-11,13-31 * 0,1,2,3,4,5,6 /data/coypu/tools/maintenance/live_reload_all_daily_and_recreate_tmp.sh
crontab [+] 1,1 All
cd

# run commodities fetcher&transformer
python /data/coypu/tools/commodities/main.py
python /data/coypu/tools/disasters/main.py
python /data/coypu/tools/wildfires/main.py

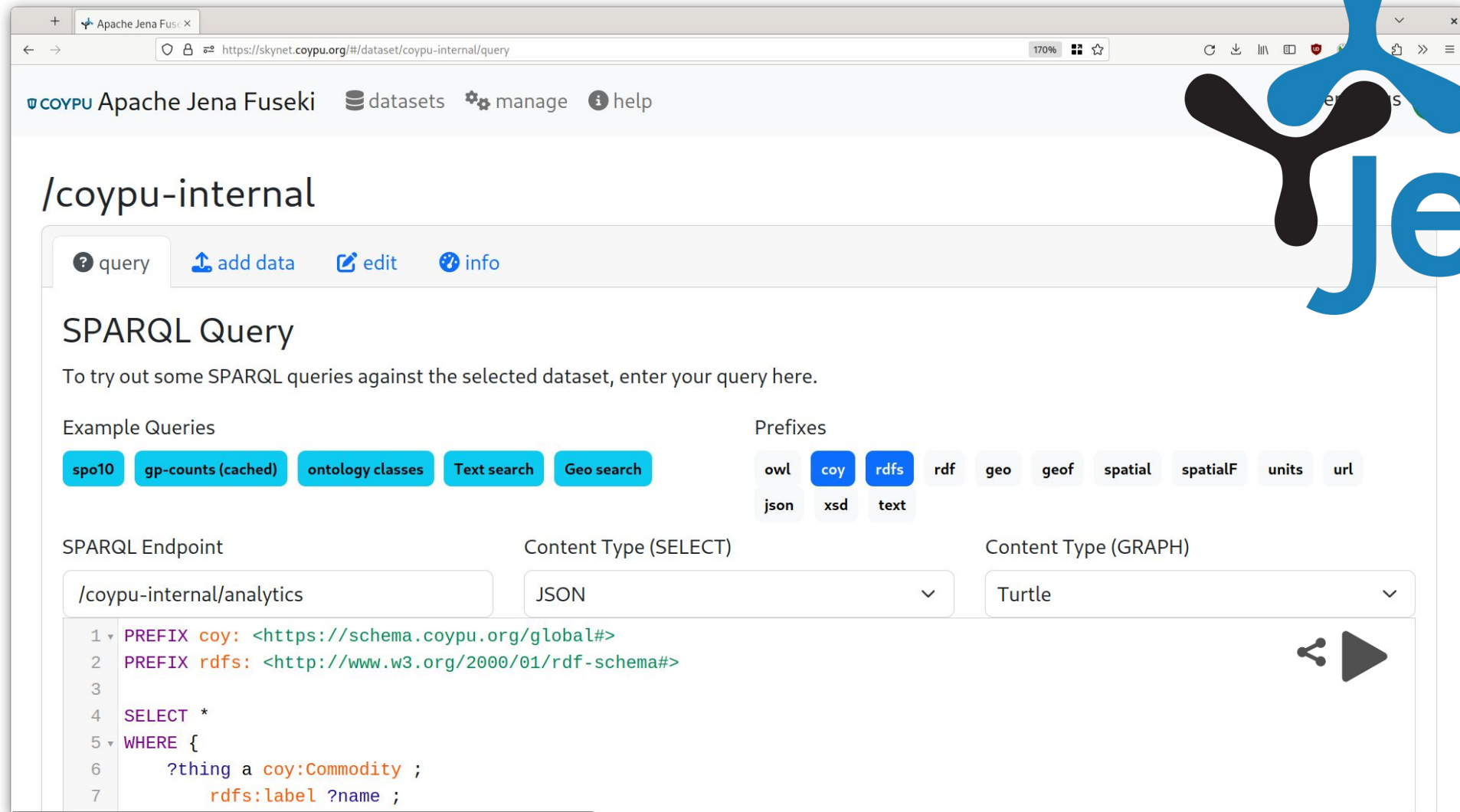
# extract daily wiki events
/data/coypu/tools/wikievents_loader/wikievents_loader_daily 2>&1 | tee /data/coypu/tools/log/wikievents_loader_daily.log

"$script_root/live_reload_all_daily_from_disk.sh"

/data/coypu/tools/maintenance/live_reload_all_daily_and_recreate_tmp.sh [+] 28,0-1 Bot
```

## cron + custom scripts

# Triple Store + SPARQL Endpoint



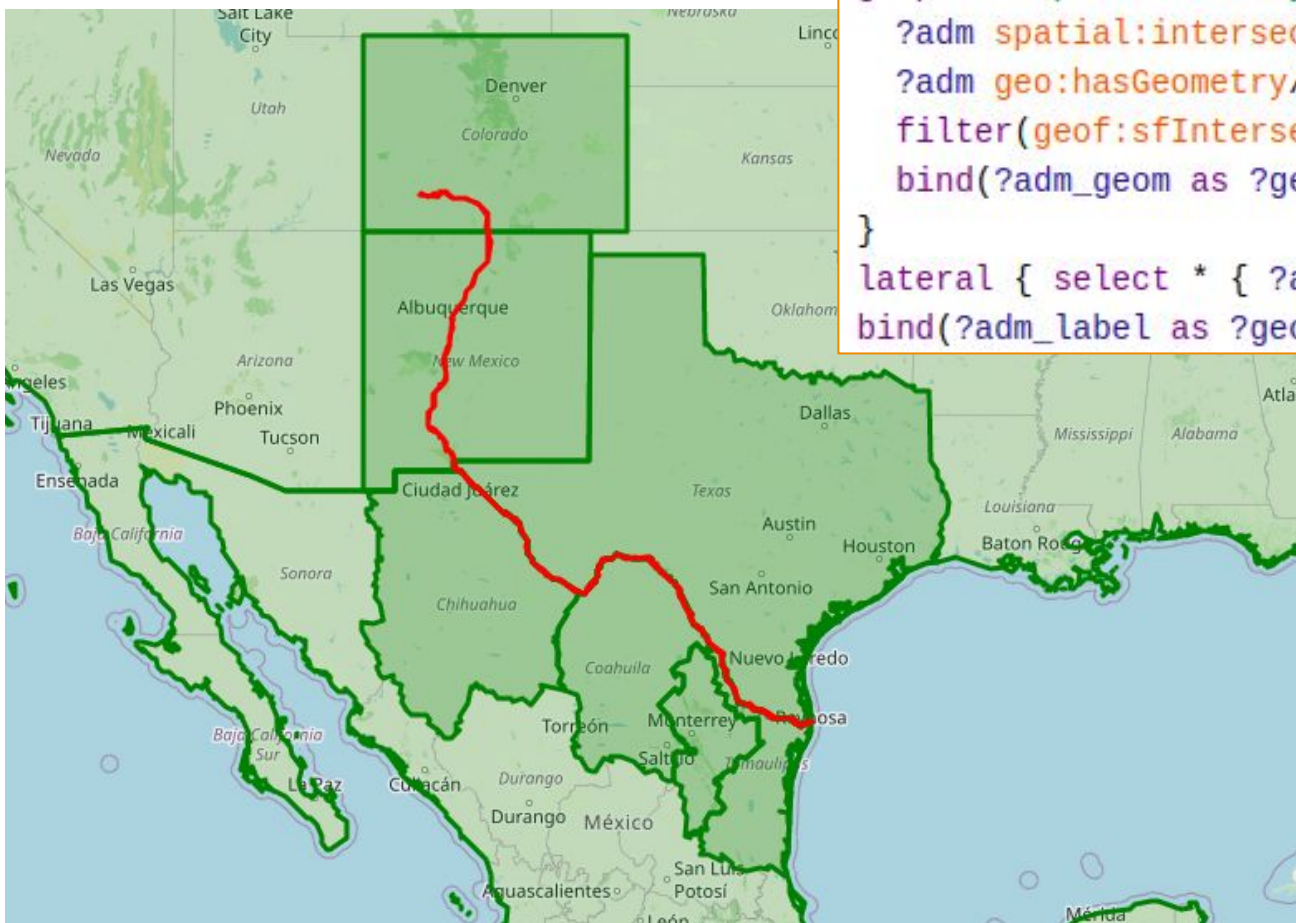
The screenshot shows the Apache Jena Fuseki web interface. The browser address bar displays `https://skynet.coypu.org/#/dataset/coypu-internal/query`. The page title is `COYPU Apache Jena Fuseki`. The main content area is titled `/coypu-internal` and features a `query` tab, `add data`, `edit`, and `info` buttons. The `SPARQL Query` section includes a text input field with the following query:

```
1 PREFIX coy: <https://schema.coypu.org/global#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3
4 SELECT *
5 WHERE {
6   ?thing a coy:Commodity ;
7   rdfs:label ?name ;
```

Below the query editor, there are sections for `Example Queries` (including `spo10`, `gp-counts (cached)`, `ontology classes`, `Text search`, and `Geo search`), `Prefixes` (including `owl`, `coy`, `rdfs`, `rdf`, `geo`, `geof`, `spatial`, `spatialF`, `units`, `url`, `json`, `xsd`, and `text`), and `Content Type` dropdowns for `(SELECT)` (set to `JSON`) and `(GRAPH)` (set to `Turtle`). A `SPARQL Endpoint` field contains `/coypu-internal/analytics`. A share icon and a play button are visible on the right side of the query editor.



# Combining geodata in the graph




```
graph <https://data.coypu.org/administrative-regions/> {  
  ?adm spatial:intersectBoxGeom(?flood_geom) .  
  ?adm geo:hasGeometry/geo:asWKT ?adm_geom_lit_ .  
  filter(geof:sfIntersects(?flood_geom, ?adm_geom_lit_))  
  bind(?adm_geom as ?geom) bind("green" as ?geomColor)  
}  
lateral { select * { ?adm rdfs:label ?adm_label } limit 1 }  
bind(?adm_label as ?geomTooltip) bind(?adm as ?geomLabel)
```



# Calculate inferences

## Germany



<https://data.coypu.org/country/DEU>

Resolve labels en  
 Calculate inferences  
[Local Browsing](#) | [Explore](#)

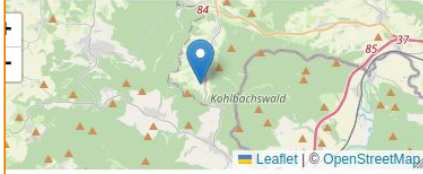
Country @en	
rdf: type	Country @en
geo: hasGeometry	<ul style="list-style-type: none"> <li>country/DEU/geometry</li> <li>country/DEU/geometry/boundary</li> </ul>
rdfs: label	Germany @en
owl: sameAs	Germany @en
skos: altLabel	<ul style="list-style-type: none"> <li>BR Deutschland @en</li> <li>Bundesrepublik Deutschland @en</li> <li>Deutschland @en</li> <li>Federal Republic of Germany @en</li> </ul>
has adaptive capability @en	29.62e0 (xsd: double)
coy: hasContinent	Europe @en
has coping capability @en	3.14e0 (xsd: double)
coy: hasCpiScore	Corruption Perception Score - Germany

```

CONSTRUCT {
  ?s ?p ?o .
} {
  SERVICE <sameAs+rdfs:> {
    # ...
  }
}
    
```



## Germany



<https://data.coypu.org/country/DEU>

Resolve labels en  
 Calculate inferences  
[Local Browsing](#) | [Explore](#) | [Link](#)

Linked Data  
Viewer

Country @en	
rdf: type	<ul style="list-style-type: none"> <li>Item</li> <li>xsd: string</li> <li>owl: Thing</li> <li>P3086</li> <li>Country @en</li> <li>Feature @en</li> <li>Location @en</li> <li>Region @en</li> <li>GlobalTradeAlertFeature @en</li> <li>Jurisdiction @en</li> <li>gn: WikipediaArticle</li> </ul>
dct: source	Germany
pagerank	65125.707 (xsd: float)
dateModified	<ul style="list-style-type: none"> <li>2023-02-28T12:56:13+00:00 (xsd: dateTime)</li> <li>2023-03-31T21:25:24+00:00 (xsd: dateTime)</li> </ul>

# Mapping of product sectors

[https://data.coypu.org/classification/hs\\_2012/382311](https://data.coypu.org/classification/hs_2012/382311)

a skos: Concept

**HS 2012**

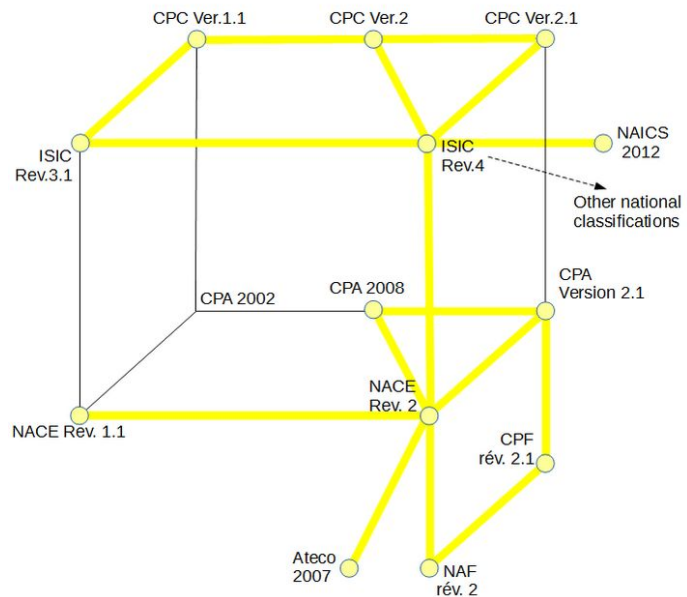
rdf: type

skos: Concept

skos: broader

Industrial monocarboxylic fatty acids; acid oils from refining; industrial fatty alcohols @en

skos: closeMatch



**Linked Data Viewer**

Industrial monocarboxylic fatty acids; acid oils from refining @en

rdf: type

skos: Concept

skos: broader

3412

**CPC Ver.2.1**

skos: closeMatch

2011

rdf: type

skos: Concept

skos: closeMatch

**ISIC Rev.4**

20.11

rdf: type

skos: Concept

**NACE Rev.2**

coreContentNote This class includes: -  
manufacture of  
liquefied or  
compressed inorganic  
industrial or medical

# Interactive exploration of the graph

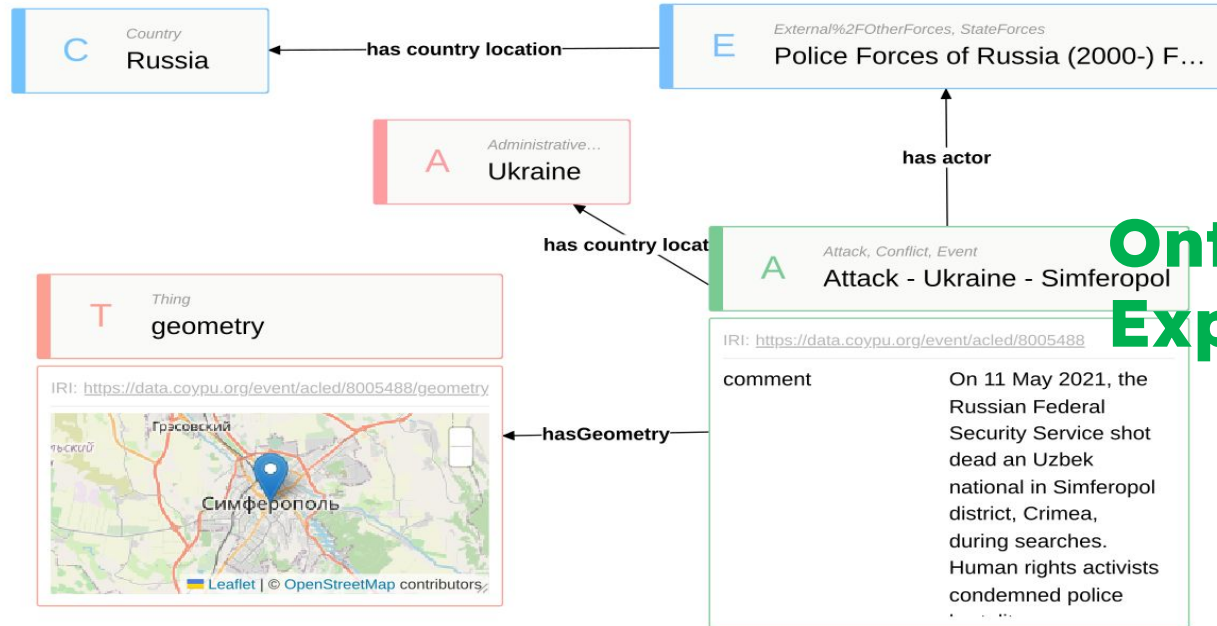
▼ Classes

- ConceptScheme
- MetaThing
- Thing
  - Company
  - Event
  - Feature
  - Industry
  - Infrastructure
  - Location
    - City
    - Continent
    - Country

▼ Instances

Has type **Country**

- Syria
- Thailand
- Saint Vincent and the Grenadines
- United States of America
- Slovenia
- Saint Lucia
- Sri Lanka
- Niger
- Iceland
- Philippines



Ontodia Graph Explorer



# Caching for fast retrieval of precalculated results

```
1 SELECT ?g (SUM(?c) AS ?count) WHERE {  
2   { SERVICE <cache:> { SELECT ("defaultGraph" AS ?g) ?p (COUNT(*) AS ?c) { ?s ?p ?o } GROUP BY ?p } }  
3   UNION  
4   { SERVICE <cache:> { SELECT ?g ?p (COUNT(*) AS ?c) { GRAPH ?g { ?s ?p ?o } } GROUP BY ?g ?p } }  
5 }  
6 GROUP BY ?g  
7 ORDER BY DESC(?count)
```



5 minutes →  
0.6 seconds

Table Response Gallery Chart Geo Geo-3D Geo events Markup Network Pivot Timeline 7 results in 0.597 seconds

Simple view  Ellipse  Filter query results Page size: 50

g	count
1 defaultGraph	"208467870"^^<http://www.w3.org/2001/XMLSchema#integer>
2 <urn:mvn.org.coypu.thirdparty.osm:osm-river-planet:2022-10-05.ttl.bz2>	"163310838"^^<http://www.w3.org/2001/XMLSchema#integer>
3 <urn:mvn.org.coypu.thirdparty.osm:osm-railway-planet:2022-10-05.ttl.bz2>	"43235550"^^<http://www.w3.org/2001/XMLSchema#integer>
4 <urn:mvn.org.coypu.thirdparty.osm:osm-railwaystation-planet:2022-10-05.ttl.bz2>	"1570048"^^<http://www.w3.org/2001/XMLSchema#integer>
5 <urn:mvn.org.coypu.thirdparty.gadm-level-0-2:gadm:4.1.ttl.gz>	"334776"^^<http://www.w3.org/2001/XMLSchema#integer>
6 <urn:mvn.org.coypu.thirdparty.msi.nga.mil:world-port-index:2022-10-05.ttl>	"187363"^^<http://www.w3.org/2001/XMLSchema#integer>
7 <urn:mvn.org.coypu.thirdparty.osm:osm-ports-planet:2022-10-05.ttl.bz2>	"63213"^^<http://www.w3.org/2001/XMLSchema#integer>

# Inline Javascript for API transformations



```
27 BIND(xml:path(?event_item, "//georss:polygon/text()") as ?polygon_points
28 BIND(json:jse("""
29   let arr = $0.split(' ')
30   let res=[]
31   for (var i = 0 ; i < arr.length - 1; i = i + 2) {
32     res.push(arr[i + 1] + ' ' + arr[i])
33   }
34   return res.join(', ')
35   """, ?polygon_points) AS ?value)
36 BIND("POLYGON((" + ?value + "))" as ?wkt)
37 BIND(strdt(?wkt, geo:wktLiteral) as ?geom)
38 }
39 LIMIT 100
40
```

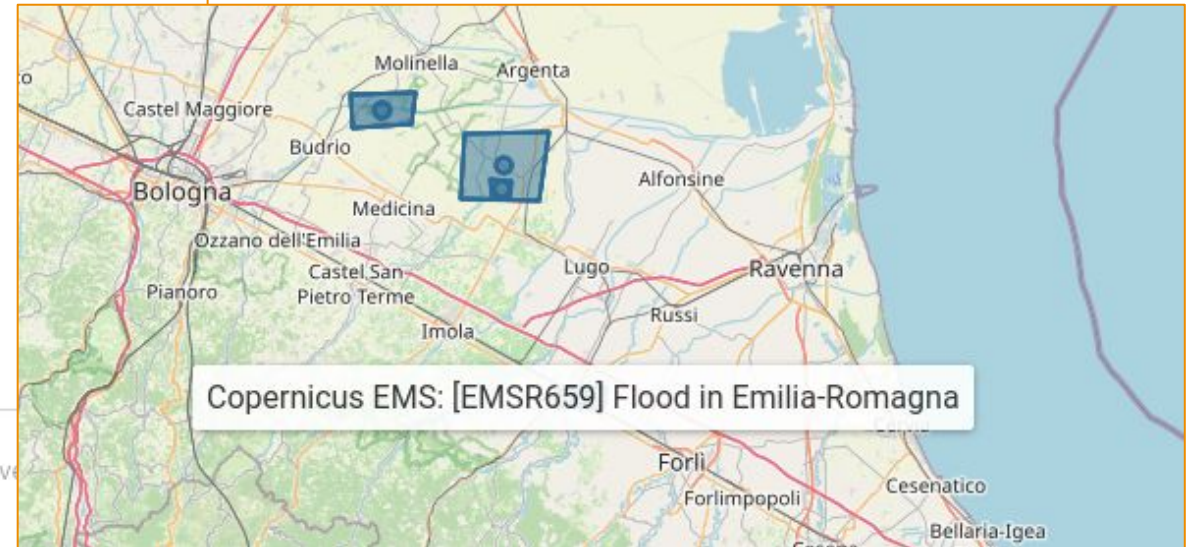


Table Response Gallery Chart Geo Geo-3D Geo ev

geom

```
1 "POLYGON((11.77606211830873 44.48668204095276, 11.80291361887779 44.4849722331955...
2 "POLYGON((11.87157062501555 44.27663992525394, 11.93386214668788 44.2763158019823...
3 "POLYGON((11.7330331630454 44.48405330348036, 11.84236179092796 44.48296196729031,...
```



AKSW



Thanks

[eswc.coypu.org](https://eswc.coypu.org)

Simon Bin

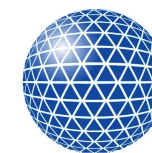
[sbin@informatik.uni-leipzig.de](mailto:sbin@informatik.uni-leipzig.de)

Supported by



Federal Ministry  
for Economic Affairs  
and Climate Action

GA No. 01MK21007A



InfAI®  
Institute for Applied Informatics

## How does the data end up in the graph?

- Ontology (if required, e.g. with Protégé), upload it into Gitlab
  - Pipeline will document the ontology on <https://schema.coypu.org/>
- Describe data with metadata: Licence, Source (RDF document is stored in Gitlab)
  - Pipeline will document data on <https://datasets.coypu.org/> (in tabular form) and <https://metadata.coypu.org/> (as RDF)
- [A1] Find/have Data/API, and map it (e.g. with Python, RML, Tarql, ...)
  - Put scripts/configurations into Gitlab, for documentation+reproducibility
- [A2] Upload data onto the server in the right folder (daily/weekly/static) and register it in dataset-mapping.csv
  - Pipeline will document file-sets on <https://metadata.coypu.org/> (as RDF)

## How does the data end up in the graph?

- [B] Create a periodic job comprising [A1] and [A2]
- (Currently) 1× per month the knowledge graph is created from all source data
- Daily updates are processed 1× per day
- Text-index for label search is built/refreshed
- Geo-index for GeoSPARQL queries is built/refreshed
- Statistics about the knowledge graph are calculated



# How to get data out of the graph

- Query with SPARQL
- Query with custom API (delegating to SPARQL internally)
- Query with prepared queries as CSV
- Interactive SPARQL results inspection using Triply Yasgui
- Redash-dashboards created on th basis of SPARQL queries
- Dereference and view RDF data using the CoyPu Linked Data Viewer at <https://data.coypu.org>
- Explore the knowledge graph using Ontodia Graph Explorer at <https://explore.skynet.coypu.org/coypu-internal>

# RhizomerEye



RhizomerEye About

🔍 Login en ▾

[datasets](#) / [got](#) / [dbo:FictionalCharacter](#) / resource

## Addam Marbrand

Noble FictionalCharacter

**allegiance:** [House Lannister](#)

**appearsIn:** [A Feast for Crows](#), [A Clash of Kings](#), [A Game of Thrones](#), [A Storm of Swords](#)

**bookIntroChapter:** 56

**genre:** Male

**name:** Addam Marbrand



# Linked Data Hub

Root / Copenhagen / Places /  
Hovedbiblioteket



Content

Properties

Map

Chart

Graph

## Hovedbiblioteket



Place

BACKLINKS

<b>Identifier</b>	stednavn_punkt.fid-361ad8c1_150a94a2aab_-83b
<b>Latitude</b>	55.68
<b>Longitude</b>	12.57
<b>Subject</b>	<a href="#">Bibliotek</a>
<b>Title</b>	Hovedbiblioteket



AKSW

Thanks

[eswc.coypu.org](https://eswc.coypu.org)

Simon Bin

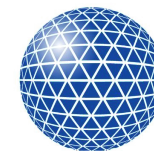
[sbin@informatik.uni-leipzig.de](mailto:sbin@informatik.uni-leipzig.de)

Supported by



Federal Ministry  
for Economic Affairs  
and Climate Action

GA No. 01MK21007A



InfAI®  
Institute for Applied Informatics

