Open Biodiversity Knowledge Management System

Semantic Suite Running on top of the
Biodiversity Knowledge Graph

Viktor Senderov, Teodor Georgiev, Donat Agosti, Terry Catapano, Guido Sautter, Éamonn Ó Tuama, Nico Franz, Kiril Simov, Lyubomir Penev

TDWG 2016 Annual Conference
2016-12-05 02:30 PM – 02:45 PM
CTEC Auditorium

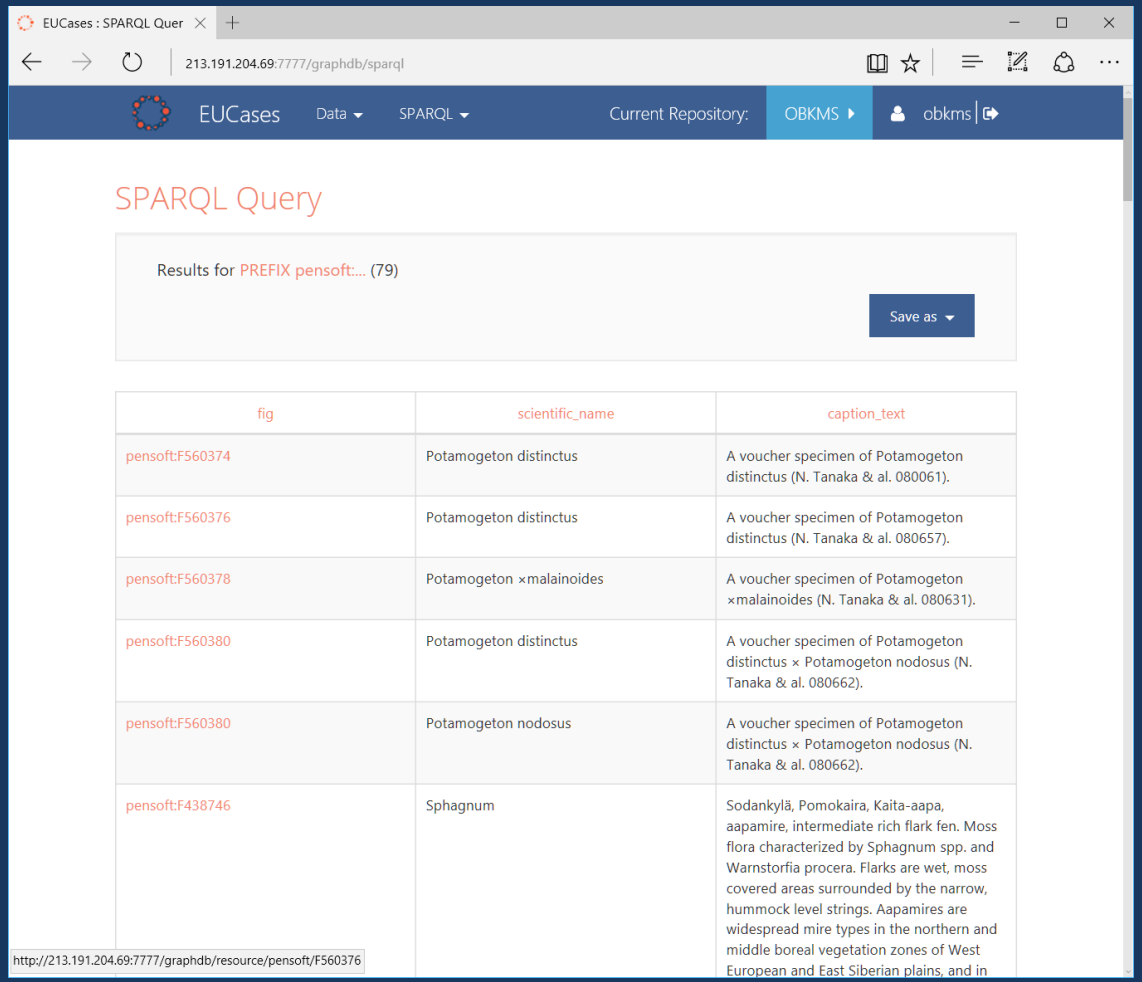
BIG4 Biosystematics, Informatics, Genomics

PLAZI TAKING CARE OF FREEDOM

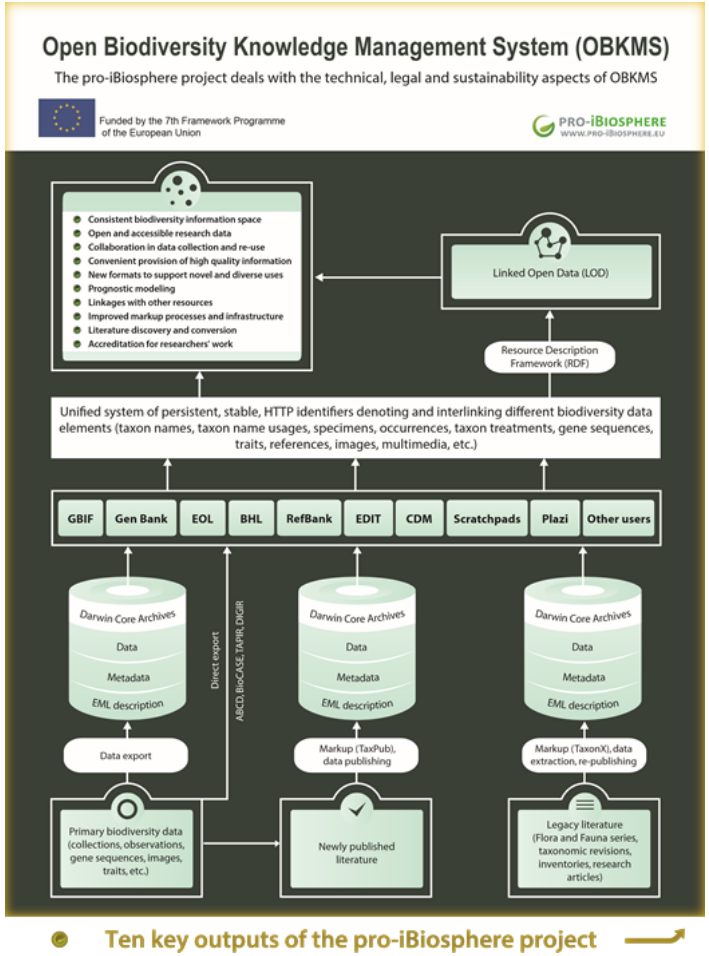
PENSOFT

BULGARIAN ACADEMY of SCIENCES 1869

I. What is OBKMS?



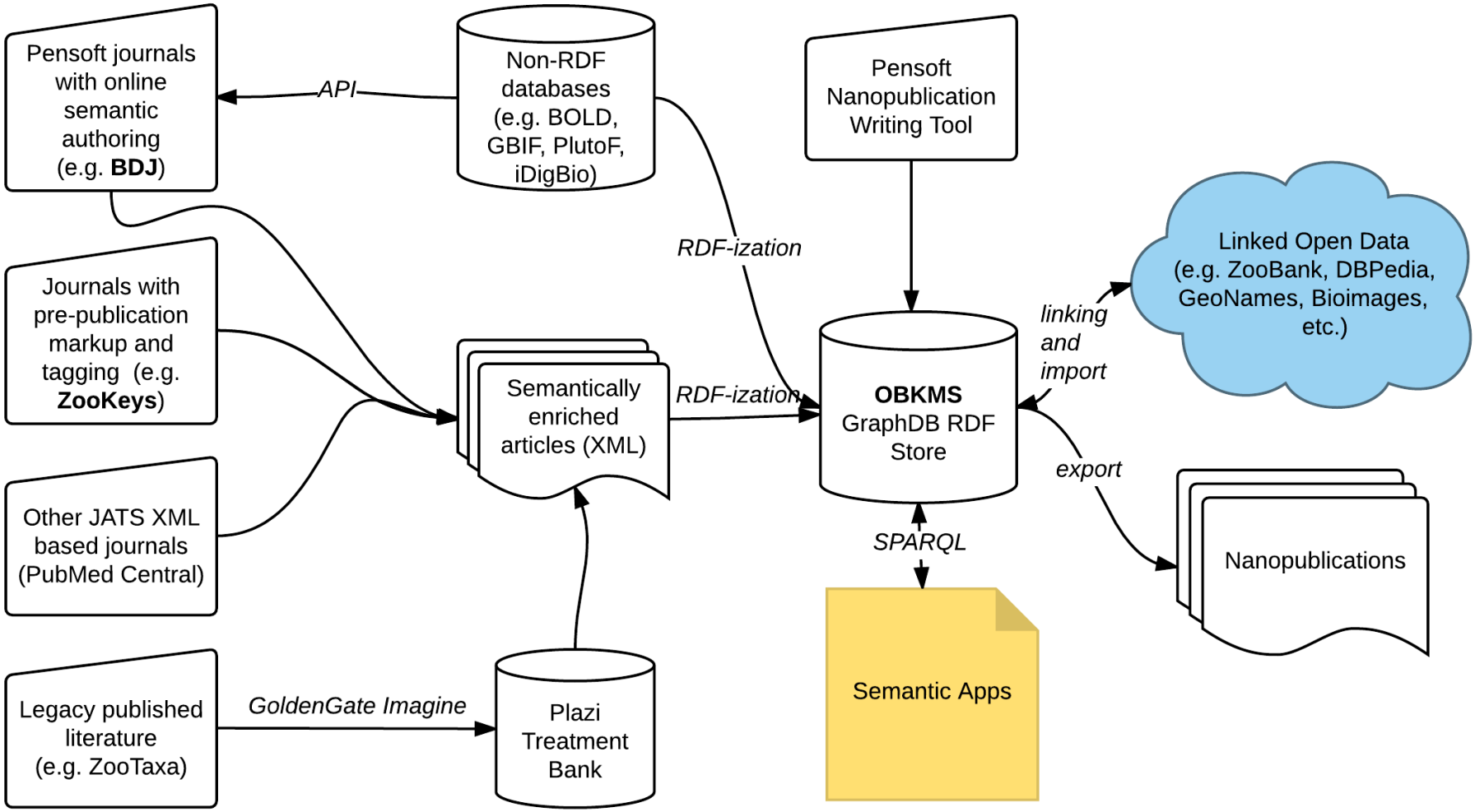
Vision



- Linked Open Data
  - Taxon names
  - Taxon name usages
  - Specimens
  - Occurrences
  - Taxon treatments
  - Gene sequences
  - Traits
  - References
  - Images
  - Multimedia
- Biodiversity Knowledge Graph

https://doi.org/10.5281/zenodo.191785

Software architecture



Rule-Based Inference

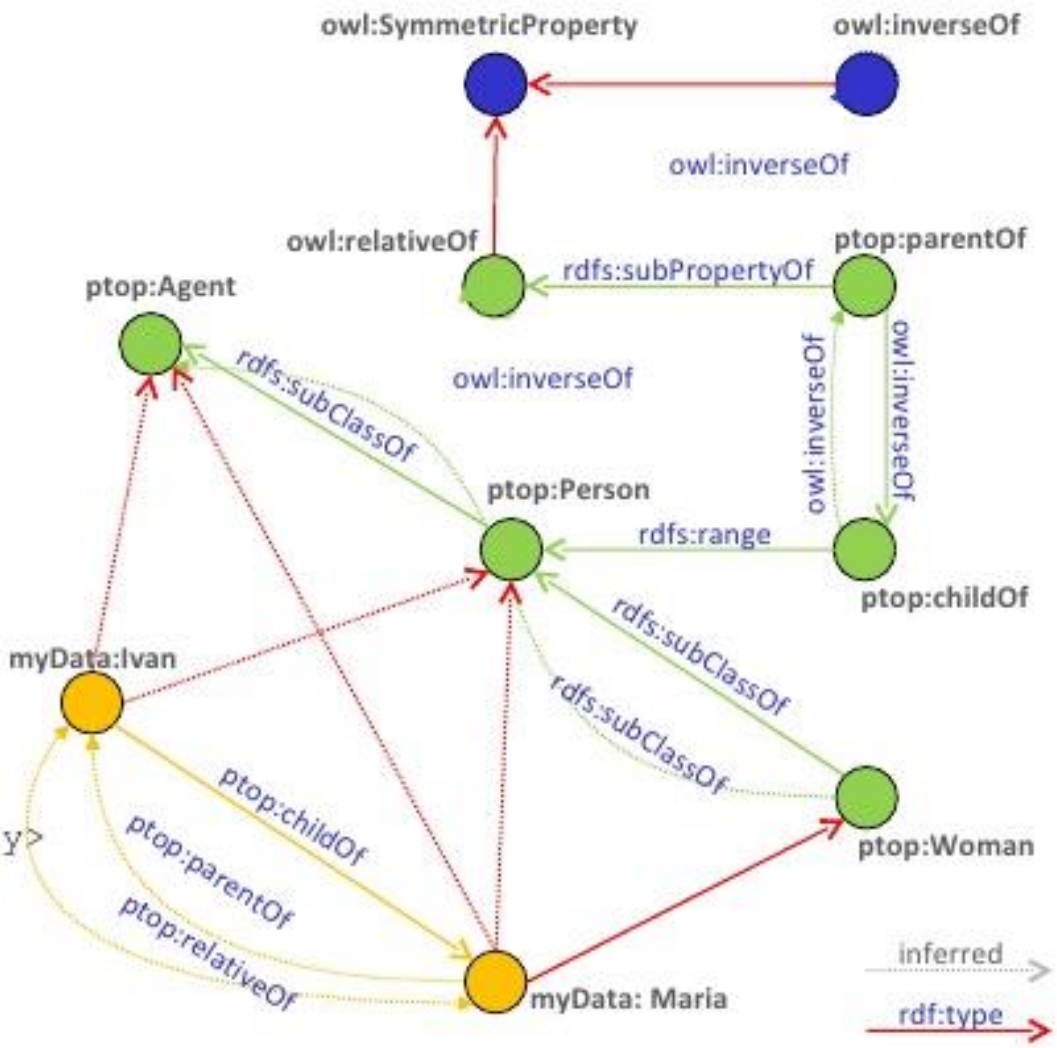
```
<C1,rdfs:subClassOf,C2>
<C2,rdfs:subClassOf,C3>
⇒ <C1,rdfs:subClassOf,C3>

<I,rdf:type,C1>
<C1,rdfs:subClassOf,C2>
⇒ <I,rdf:type,C2>

<I1,P1,I2>
<P1,rdfs:range,C2>
⇒ <I2,rdf:type,C2>

<P1,owl:inverseOf,P2>
<I1,P1,I2>
⇒  <I2,P2,I1>

<P1,rdf:type,owl:SymmetricProperty>
⇒  <P1,owl:inverseOf,P1>
```



Reason-able view of biodiversity data

http://www.slideshare.net/ontotext/two-reasonable-views-to-the-web-of-linked-data

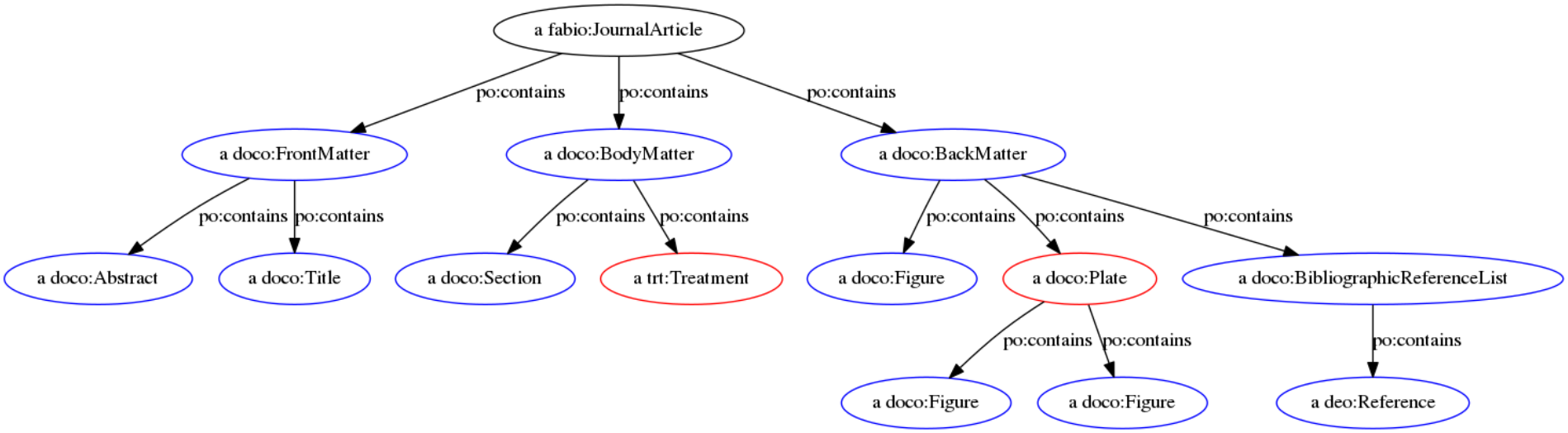
- Group selected datasets and ontologies in a **single semantic repository:** tractable reasoning
- OBKMS includes:
  - GBIF datasets (taxonomic backbone)
  - GeoNames country data
  - Bioimages Darwin-SW formatted data
  - Genetic data: Bio4j

II. Semantic Model of a Biodiversity Publication

Semantic model of a biodiversity publication: Key ontologies

- Semantic Publishing and Referencing Ontologies (SPAR)
  - FRBR-aligned Bibliographic Ontology (FaBiO)
  - Citation Counting and Context Characterization Ontology (C40)
  - Document Components Ontology (DOCO)
  - Publishing Roles Ontology (PRO)
- Treatment Ontology
- Darwin-SW

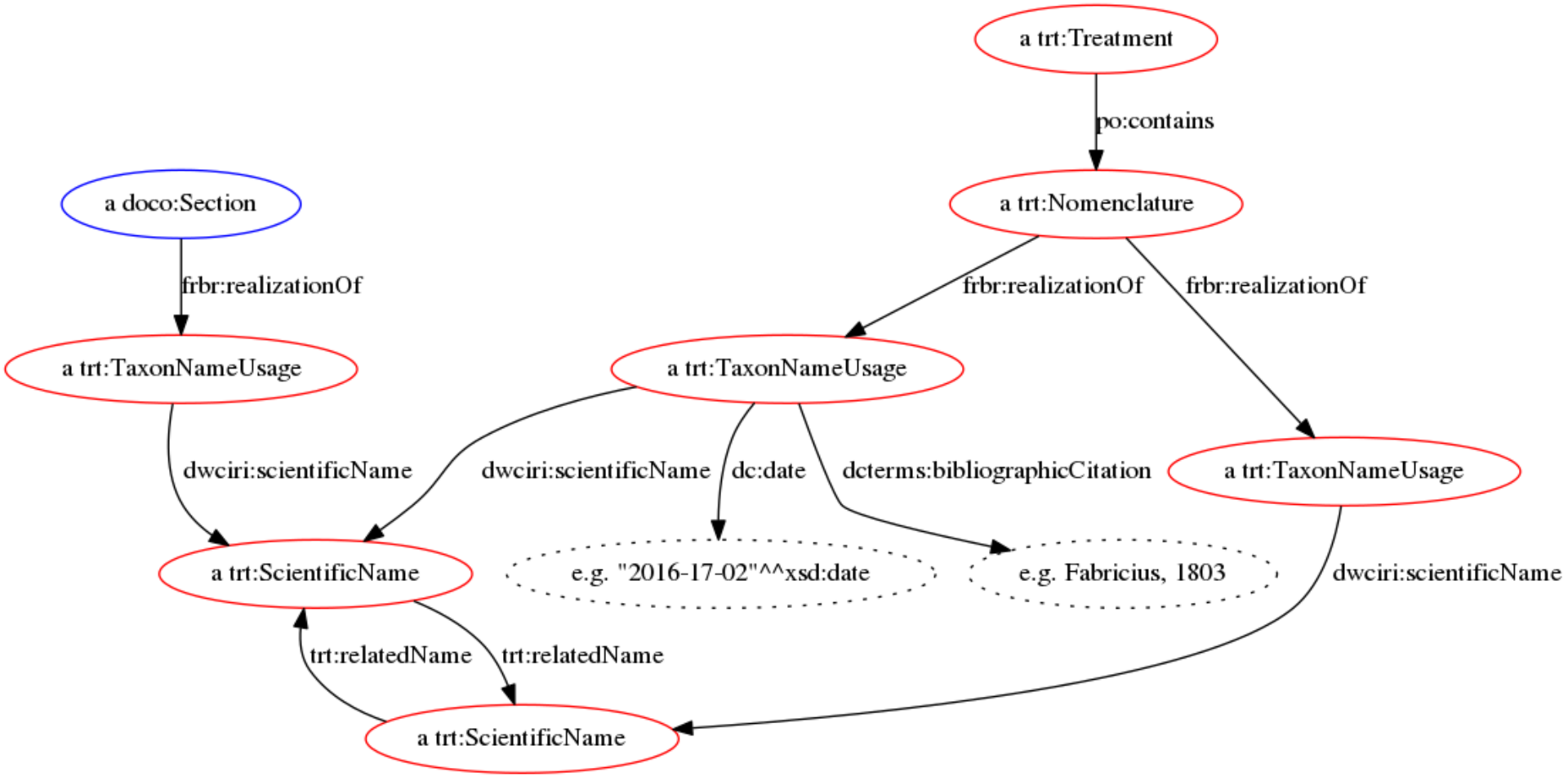
Semantic model of a biodiversity publication: Graph representation, Article structure



Legend:

- blue ovals: OBKMS nodes, representing XML elements
- red ovals: OBKMS nodes, representing XML elements, but not defined in SPAR

Semantic model of a biodiversity publication: Graph representation, Treatment



Properties of *taxon name usage (TNU)*

Property	Description
dwciri:scientificName	Links TNU to a trt:ScientificName
Type*	(Optional) E.g. “new sp.,“ new comb.,” etc.
dc:date	Gives the date of the taxon name usage as literal
dcterms:bibliographicCitation	(Optional) links the TNU to a bibliographic reference as literal
biro:references	(Optional) Links the TNU to a bibliographic reference
dwciri:toTaxon	(Optional) Links the TNU to a taxon (concept)
c4o:hasContent	Free text comments as literal

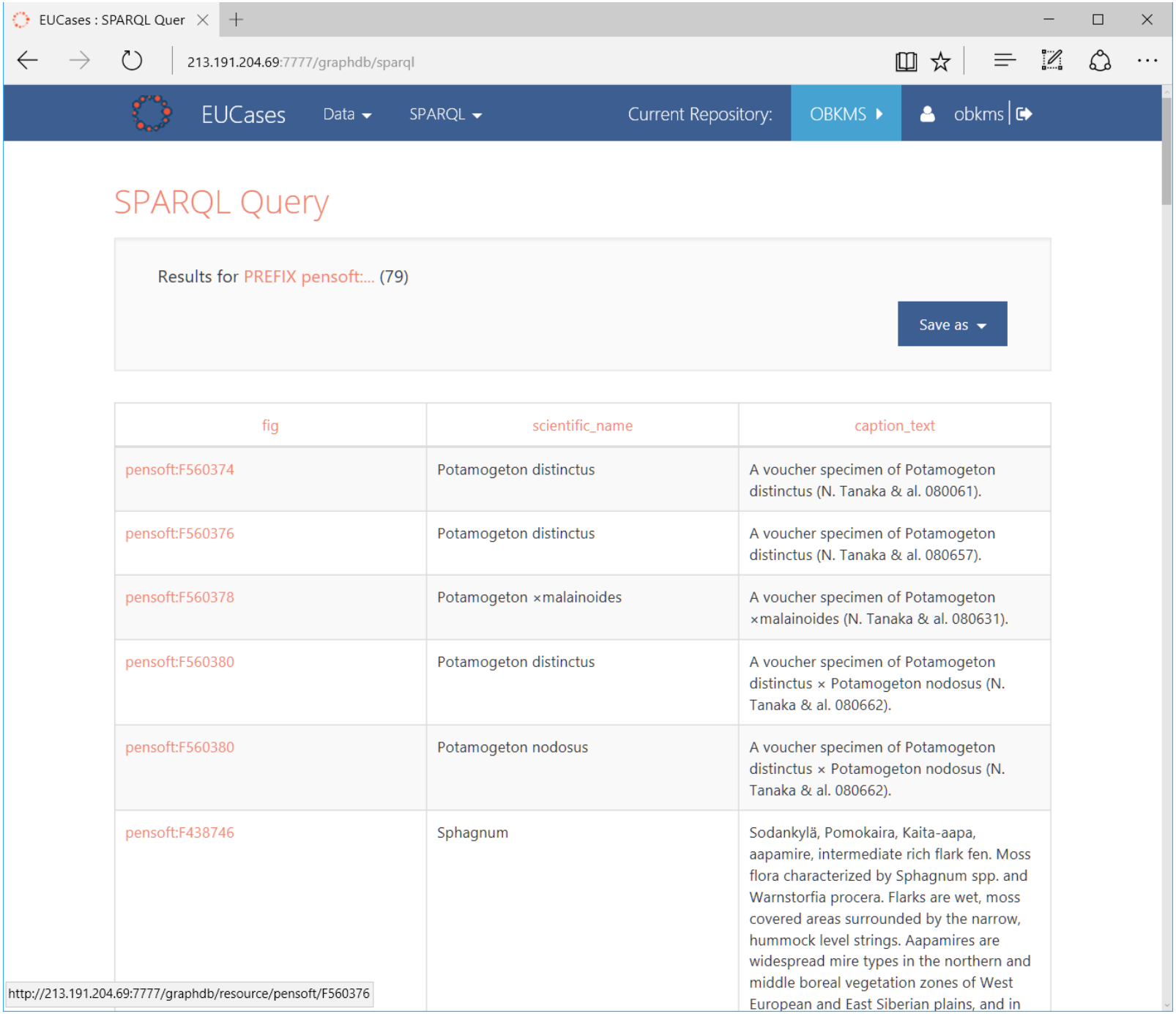
* Can be realized as sub-subproperty of dwciri:scientificName or as a separate type property

Example 1: Using the semantic model to understand the context of a *taxon name usage*

Task: *find all taxon name usages in figure captions, where the caption indicates that we dealing with a voucher or a type specimen!*

SPARQL

```
…
SELECT ?fig ?scientific_name ?caption_text
WHERE {
    ?name a                        trt:ScientificName ;
          skos:prefLabel           ?scientific_name .
    ?tnu  a                        trt:TaxonNameUsage;
          dwciri:scientificName    ?name .
    ?fig  frbr:realizationOf       ?tnu ;
          a                        doco:Figure ;
          po:contains              ?caption .
    ?caption c4o:hasContent ?caption_text.
    FILTER (regex ?caption_text, "voucher") || regex(?caption_text, "type"))
}
```


EUCases : SPARQL Quer
213.191.204.69:7777/graphdb/sparql
EUCases
Data
SPARQL
Current Repository:
OBKMS
obkms
SPARQL Query
Results for PREFIX pensoft:... (79)
Save as
fig
scientific_name
caption_text
pensoft:F560374
Potamogeton distinctus
A voucher specimen of Potamogeton distinctus (N. Tanaka & al. 080061).
pensoft:F560376
Potamogeton distinctus
A voucher specimen of Potamogeton distinctus (N. Tanaka & al. 080657).
pensoft:F560378
Potamogeton ×malainoides
A voucher specimen of Potamogeton ×malainoides (N. Tanaka & al. 080631).
pensoft:F560380
Potamogeton distinctus
A voucher specimen of Potamogeton distinctus × Potamogeton nodosus (N. Tanaka & al. 080662).
pensoft:F560380
Potamogeton nodosus
A voucher specimen of Potamogeton distinctus × Potamogeton nodosus (N. Tanaka & al. 080662).
pensoft:F438746
Sphagnum
Sodankylä, Pomokaira, Kaita-aapa, aapamire, intermediate rich flark fen. Moss flora characterized by Sphagnum spp. and Warnstorfia procera. Flarks are wet, moss covered areas surrounded by the narrow, hummock level strings. Aapamires are widespread mire types in the northern and middle boreal vegetation zones of West European and East Siberian plains, and in
http://213.191.204.69:7777/graphdb/resource/pensoft/F560376

Example 2: Using the semantic model to find related names

A rule for creating new relations:

```
PREFIX trt: <http://plazi.org/treatment#>
INSERT {
    ?a trt:relatedName ?b .
    ?b trt:relatedName ?a .
}
WHERE {
?a a trt:ScientificName .
?b a trt:ScientificName .
?c a trt:Nomenclature ;
   frbr:realizationOf
                  [ dwciri:scientifiCName ?a ];
                  [ dwciri:scientificName ?b ].
}

# if two names are mentioned in the same nomenclature
# section then they are related
```

Define related names as a transitive and reflexive property in OWL:

```
trt:relatedName rdf:type owl:ObjectProperty ,
                         owl:TransitiveProperty,
                         owl:ReflexiveProperty ;
                rdfs:domain trt:ScientificName ;
                rdfs:range trt:ScientificName ;
                rdfs:comment "Indicates that two
names are connected via a chain of nomenclatural
name usages." .
```

Example 2: Using the semantic model to find related names

R library enabling the taxonomist to browse OBKMS

```
Restarting R session...

> library(obkms)
> find_related_names("Harmonia manillana")
                                          id                          label    rank
1  pensoft:8c4f976c-1eac-4ea8-8ea0-ed8bf4b1e82e          Harmonia manillana species
2  pensoft:922d1e1e-25e1-4894-9bf4-cbb904a7a65c                Leis dunlopi species
3  pensoft:40f149be-f28c-4dfb-be31-077292c2359e             Caria manillana species
4  pensoft:3d6a8384-311e-4b40-b465-6746f4d2c85f             Leis atrocincta species
5  pensoft:283125dc-fd20-46fe-bb01-1630455d4226                  atrocincta variety
6  pensoft:130baaad-d5c4-4ac2-8ed5-4ac9201516dc               Neda paulinae species
7  pensoft:f92aaf7e-f6d0-454f-ab79-f93d4cd5248d              Caria paulinae species
8  pensoft:71899792-67ab-4753-95f5-82807fa1bc63            Leis cerasicolor species
9  pensoft:7e5fad4f-47b6-44b7-bd5b-351cae8cbf1d               Leis aterrima species
10 pensoft:db237747-571c-4e47-807a-4af33c38bf2d              Leis papuensis species
11 pensoft:ebb7e446-faaa-43a3-8d0a-7658bf3a519c Leis papuensis var. suffusa species
>
```

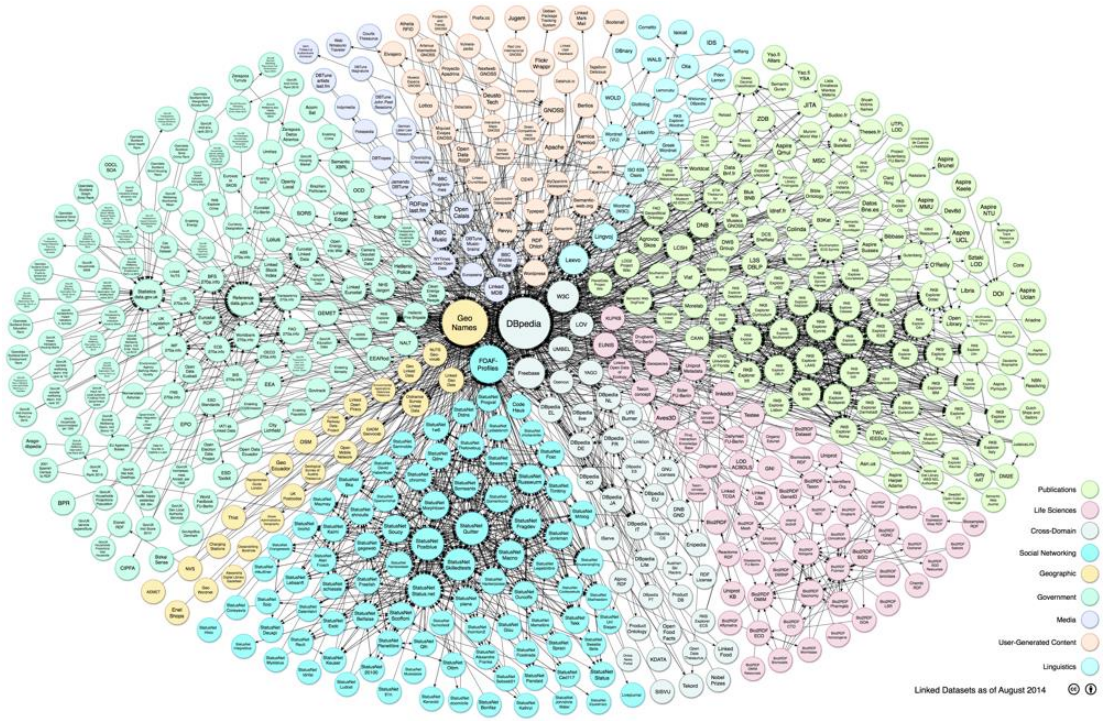
III. OBKMS as part of Linked Open Data

Crosslinking of identifiers (examples)

OBKMS Identifier Type	External Link
fabio:SubjectDiscipline	DBPedia
trt:ScientificName	ZooBank
dwc:Location	GeoNames
dsw:Occurrence	GBIF
fabio:JournalArticle	DOI
foaf:Agent	ORCID

OBKMS Linked Open Dataset

- Plain RDF
- Nanopublications



IV. Applications of OBKMS

OBKMS for collection managers and database aggregators

- Semantic web-app
  - Track published museum specimens
- Taxonomic API
  - Track new nomenclatural changes

OBKMS for scientists

Some ideas:

1. *Estimate the number of undescribed species per taxon by looking at the frequency of species description in that taxon. Also: generate stats on taxonomic activities, e.g., most studied/ published taxa.*
2. *Perform hidden topic analysis and create a recommendation algorithm based on top of that.*

Thank you!

Research funded under Marie Skłodowska-Curie BIG4 project, Grant agreement Nr. 642241

Please visit the demonstration of the OBKMS prototype and nanopublication presentation

Wednesday CTEC Auditorium

09:00 – 09:15 Talk 1028 and Talk 1012

02:00 PM Workshop 08

Friday CTEC Auditorium

09:00 – 09:15 Talk 1011