synaptica
Knowledge Organization & Discovery

SPARQL, You Taxonomy Star!
Defining, Designing, and Accessing Linked Data

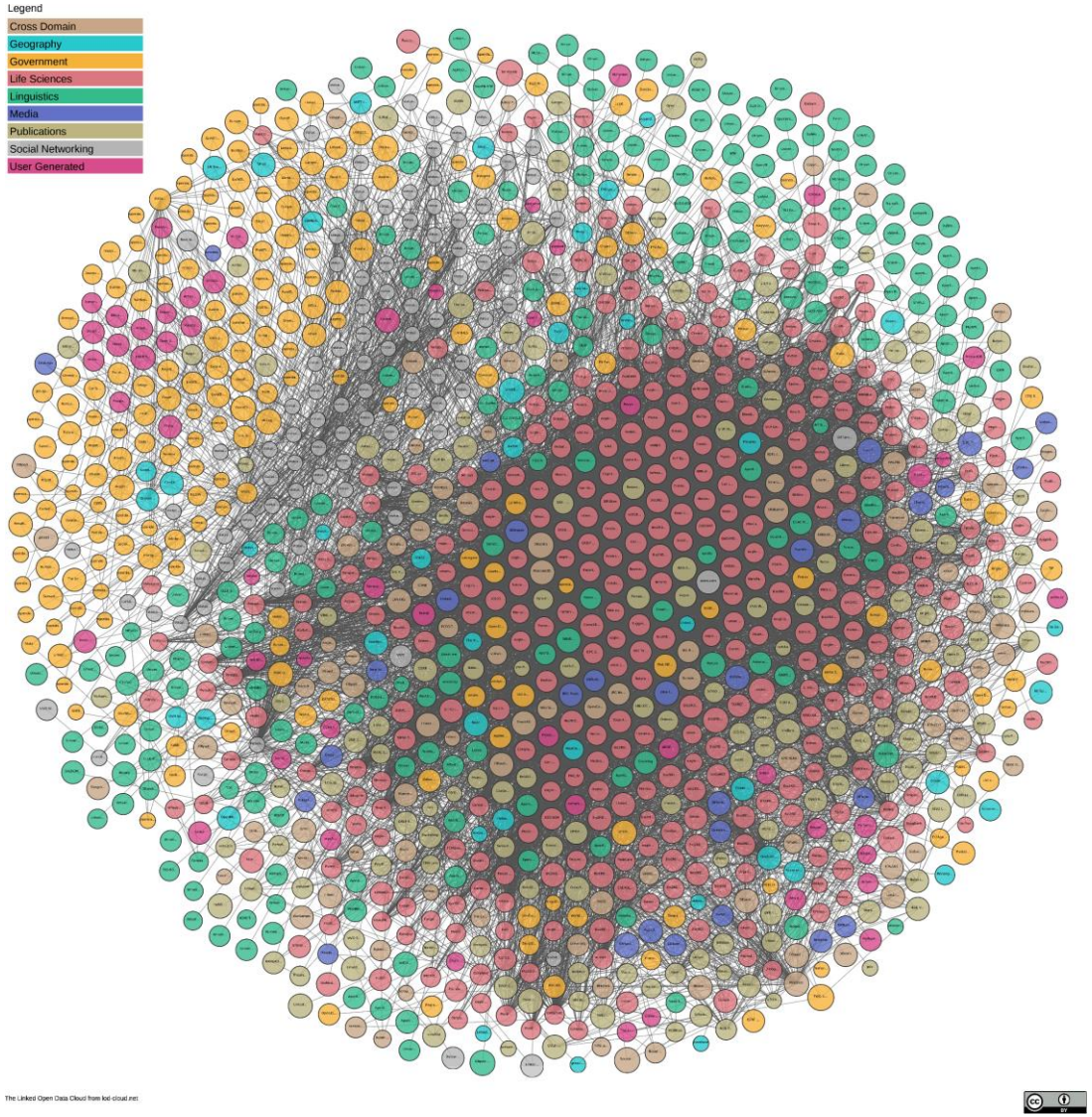
Monday November 5, 2018

Jim Sweeney, Senior Product Manager
Synaptica LLC

synaptica
Knowledge Organization & Discovery

The Semantic Web and Linked Data

- In 1991 Tim Berners-Lee described what he called the World Wide Web, which was meant to, “allow links to be made to any information anywhere.”[1]
- Linked Data, utilizing the Semantic Web and RDF Standards seeks to improve what was Web 1.0.
- With Linked Data protocols, one may send queries across the web to retrieve specific information from any source published in this format.
- This Image is of the Linked Open Data Cloud, which at last count had 1,224 submissions.



Linking Open Data cloud diagram 7-30-2018, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

[1] https://www.w3.org/People/Berners-Lee/1991/08/art-6484.txt

synaptica
Knowledge Organization & Discovery

The General WWW vs. Linked Data & Linked Open Data

- An essential feature of LD and LOD is that the sources must be *machine readable.*
- Unlike the rest of the WWW (the Web of Documents), each resource is composed of defined data elements (the Web of Data).
- Both LD and LOD use the same principles of applying RDF structure and using URIs to identify resources.

Linked Open Data is openly published and made available to any standard query

Linked Data uses the same principles, but limits access to specific audiences

synaptica
Knowledge Organization & Discovery

A Few Linked Open Data Sources

DBPedia
(WikiPedia LOD)

Getty Arts &
Architecture Thesaurus
(AAT)

GeoData
(Locations)

BioPortal
(Aggregated Life Sciences)

The Rules of Linked Open Data

★ Available on the web (whatever format) but with an open license - to be Open Data

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ Use a non-proprietary format (e.g. SKOS or OWL)

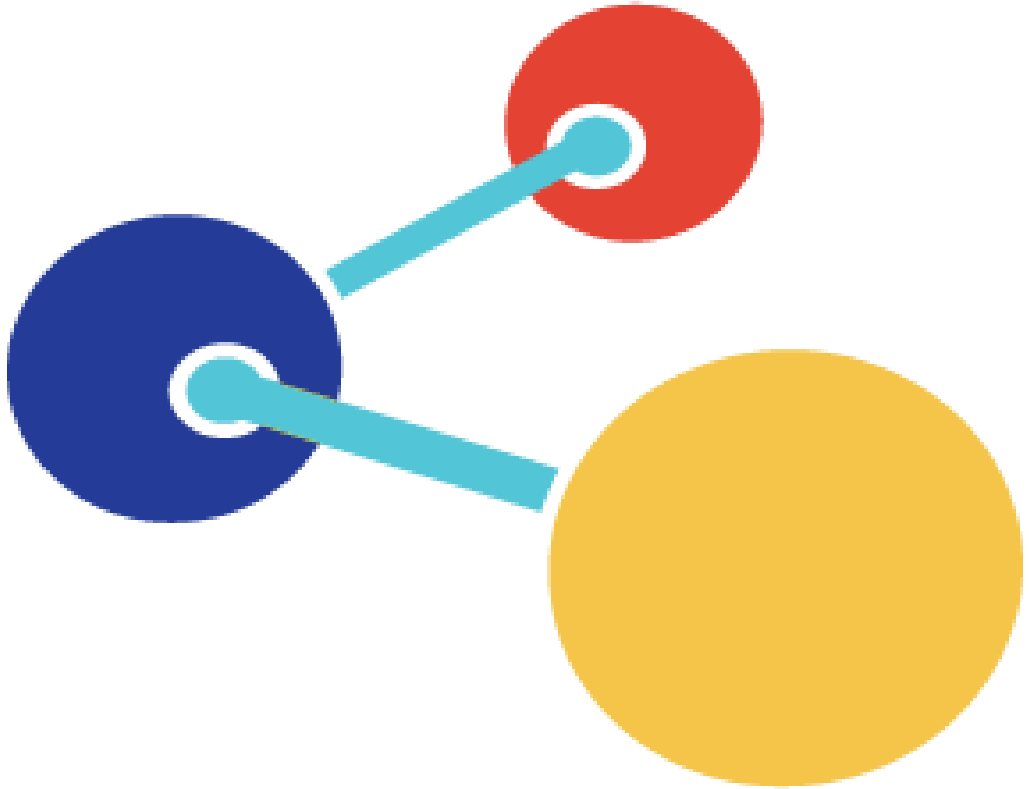
★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context (Linked Open Data Cloud)

synaptica
Knowledge Organization & Discovery

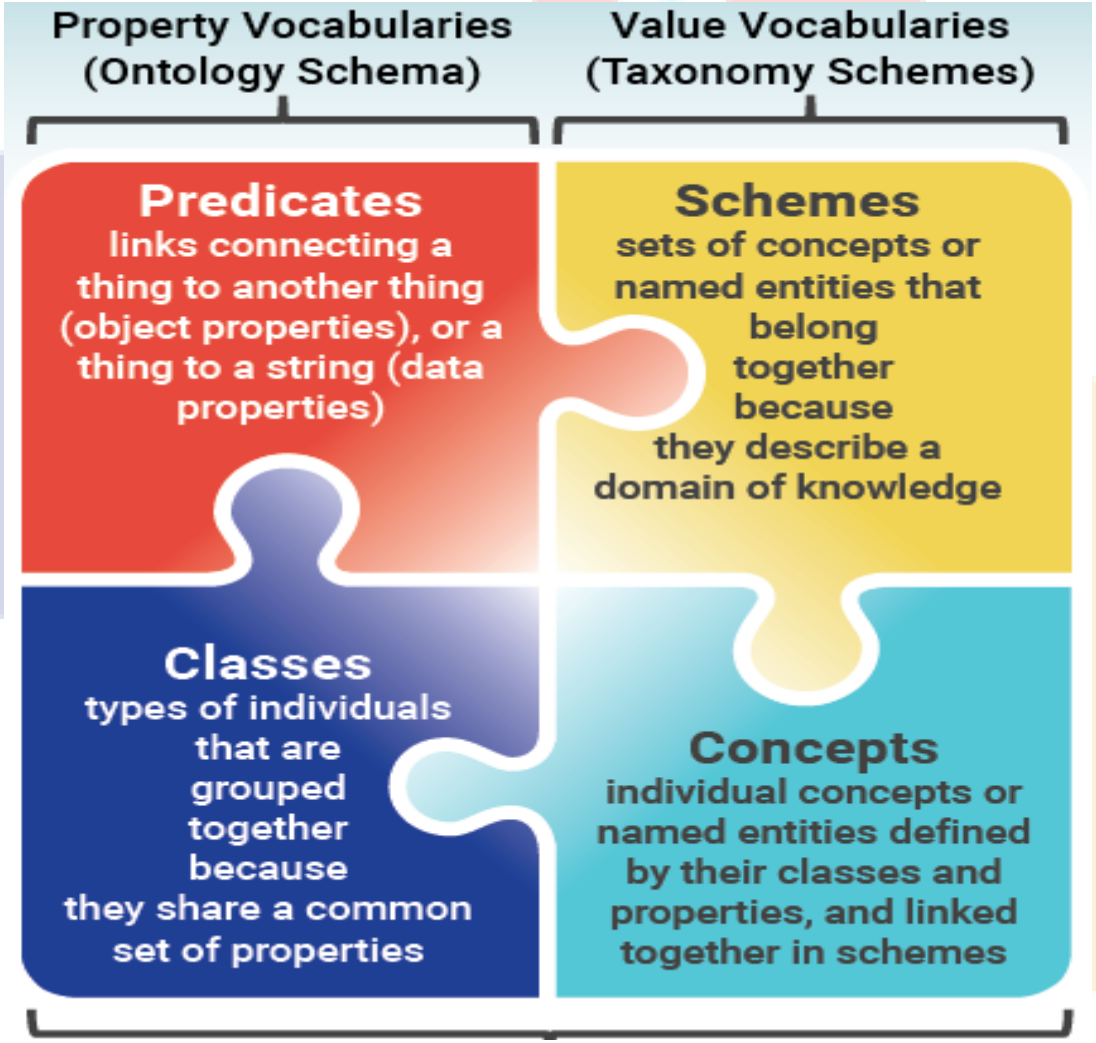
The Benefits of Linked Data

- Reduce costs and accelerate projects delivery by *reusing* Linked Open Data taxonomies and ontologies.
- Build smarter search and discovery applications by leveraging your built-in Object and Data Properties.
- Simplify systems integration work by using the open industry standards for data modelling and portability
- Take advantage of resources providing a broader set of knowledge and expertise



synaptica
Knowledge Organization & Discovery

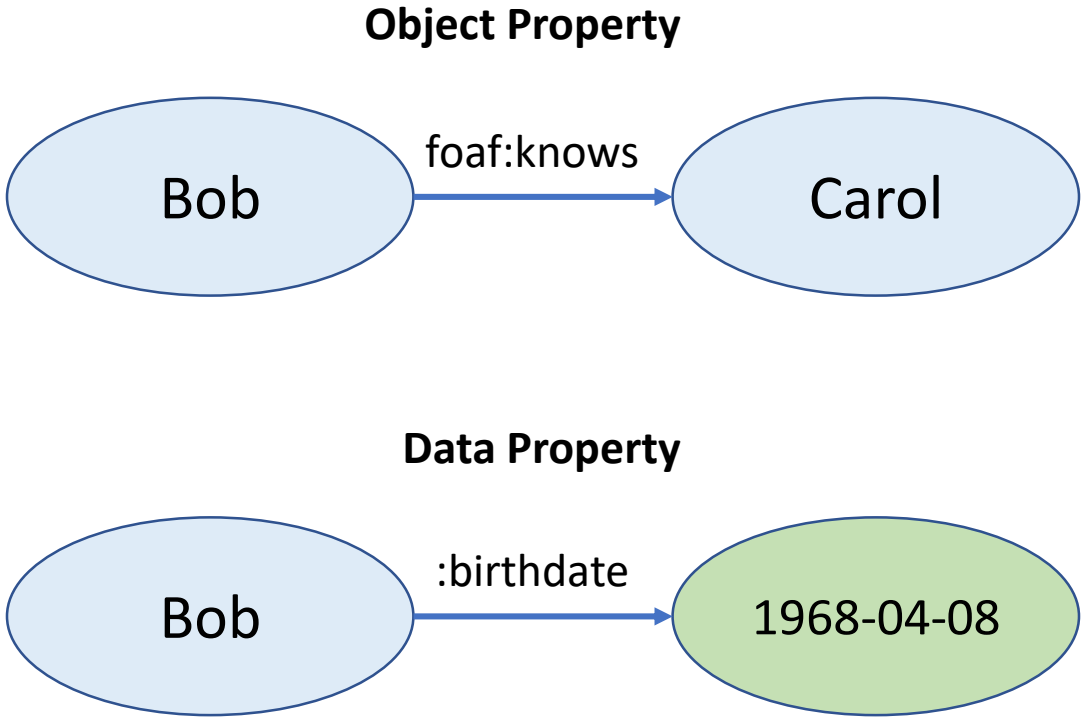
Property Vocabularies and Value Vocabularies



synaptica
Knowledge Organization & Discovery

RDF, Triples, and Graph Databases

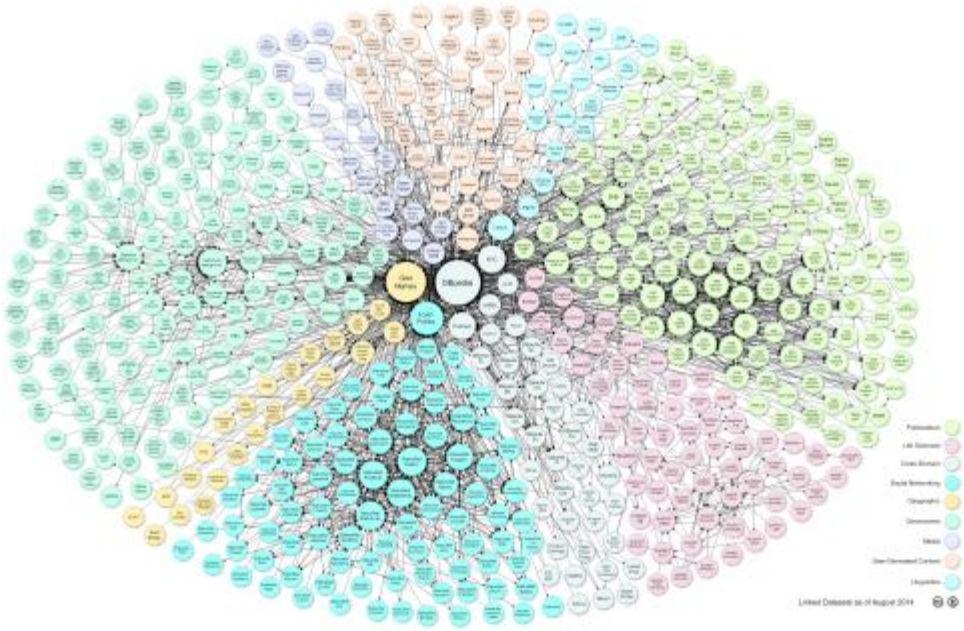
- RDF or Resource Description Framework provides the standards that we use to uniformly describe the relationships between nodes, as well as describe other attributes that we want to assign to our data as metadata.
- The standard establishes a *subject-predicate-object* unit called a *triple.*
- A triple relating a *thing* to a *thing* is called an Object Property
- A triple relating a *thing* to a *string* (a property or an attribute) is called a Data Property



synaptica
Knowledge Organization & Discovery

RDF, Triples, and Graph Databases

- Graph Databases, and especially RDF Triple Stores are well suited to storing this type of structured information.
- Entities within the Graph Database are connected via triples to all of their related Object and Data Properties.
- These data stores may contain millions of triples in large data sets.
- Data stored in this format is especially suited to be queried by SPARQL (and other means) to retrieve specified triples and return meaningful information.



synaptica
Knowledge Organization & Discovery

Designing for Linked Data Publication

- Utilize URI (Uniform Resource Identifier): a string of characters that unambiguously identifies a particular resource.

- Resolve to RDF Data:

  Turtle
  N-Triples
  N-Quads
  JSON-LD

  N3 or Notation3
  RDF/XML
  RDF/JSON

- Access to the entire database via SPARQL


```
{
  "https://graphite.synaptica.net/collection/0731825e7cea5a91dfd0299a209f4149" : {
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" : [
      {
        "value" : "http://www.w3.org/2004/02/skos/core#Collection",
        "type" : "uri"
      }
    ],
    "http://www.w3.org/2000/01/rdf-schema#label" : [
      {
        "value" : "Botanic Themes",
        "type" : "literal",
        "datatype" : "http://www.w3.org/2001/XMLSchema#string"
      }
    ],
    "http://www.w3.org/2004/02/skos/core#member" : [
      {
        "value" : "https://graphite.synaptica.net/concept/c7qrqi20mc486aoaokogq2qkac",
        "type" : "uri"
      }
    ]
  },
  "https://graphite.synaptica.net/concept/c7qrqi20mc486aoaokogq2qkac" : {
    "http://schema.synaptica.com/oasis#conceptStatus" : [
      {
        "value" : "Candidate",
        "type" : "literal",
        "datatype" : "http://www.w3.org/2001/XMLSchema#string"
      }
    ],
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#type" : [
```

synaptica
Knowledge Organization & Discovery

Design of an RDF Triple Store

Example of RDF XML:

```
<rdf:Description rdf:about="https://graphite.synaptica.net/concept/tcckl8a3j64o0i08g2kkoocaiq">

        <rdf:type rdf:resource="http://schema.synaptica.com/graphite#Concept"/>

        <rdf:type rdf:resource="https://graphite.synaptica.net/classes/j81vjzbqalpf"/>

        <skos:topConceptOf
        rdf:resource="https://graphite.synaptica.net/concept_scheme/jideyaq72g6dw"/>

        <skos:inScheme
        rdf:resource="https://graphite.synaptica.net/concept_scheme/jideyaq72g6dw"/>

        <rdfs:label xml:lang="en">Botany</rdfs:label>
```

synaptica
Knowledge Organization & Discovery

Design of an RDF Triple Store

Example of RDF XML:

```
<graphite:conceptStatus
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Candidate</graphite:conceptStatus>

<rdf:type rdf:resource="https://graphite.synaptica.net/classes/ja2plhld9lv2v"/>

<source xmlns=http://purl.org/dc/terms/
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">http://dbpedia.org/resource/Botany</source>

<skos:altLabel xml:lang="en">Plant science</skos:altLabel>
```

synaptica
Knowledge Organization & Discovery

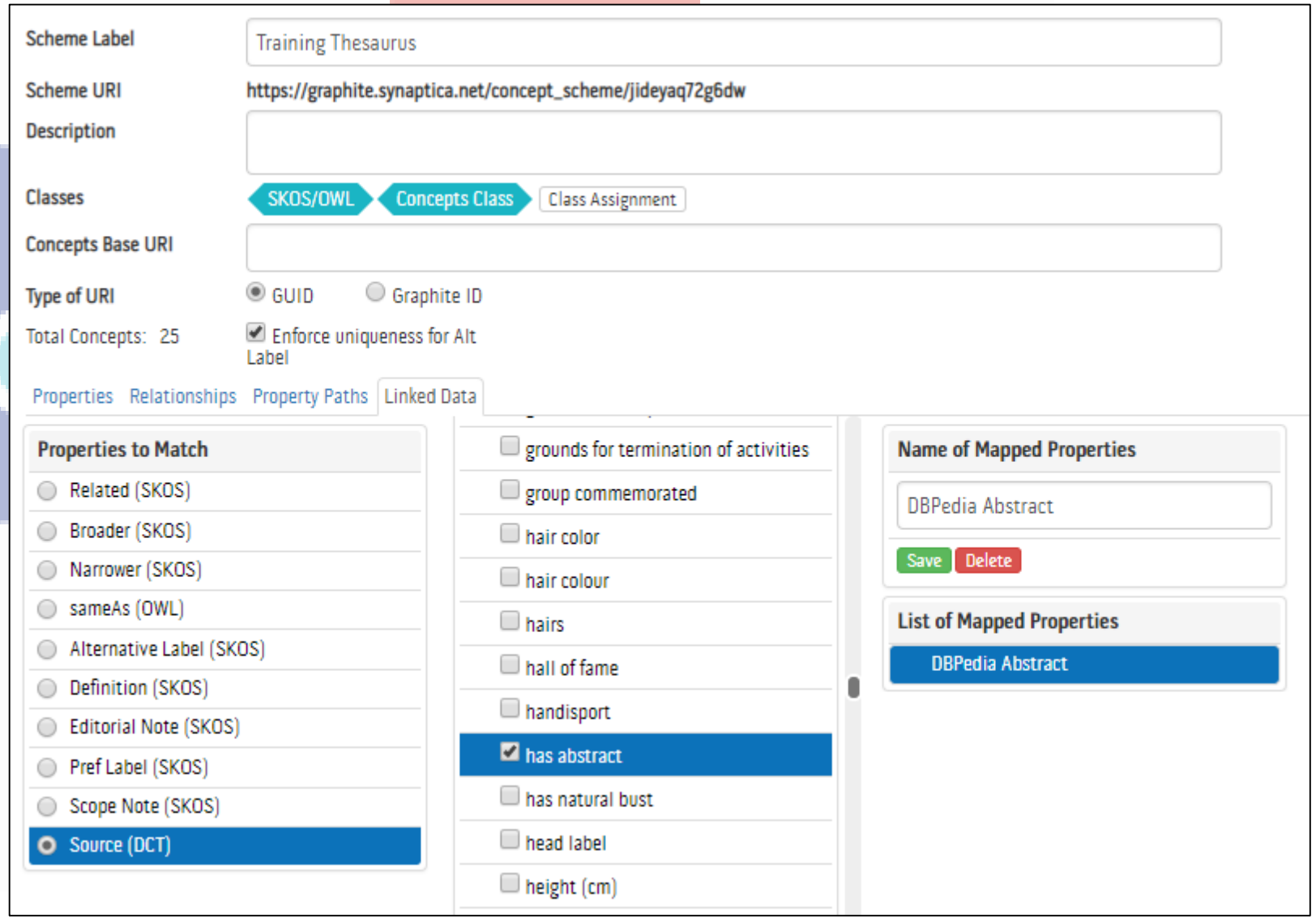
Design of an RDF Triple Store

Example of RDF XML:

```
        <skos:definition xml:lang="en">the scientific study of plants, including their physiology,
        structure, genetics, ecology, distribution, classification, and economic
        importance.</skos:definition>

        <skos:prefLabel xml:lang="en">Botany</skos:prefLabel>

        <skos:scopeNote xml:lang="en">http://dbpedia.org/resource/Botany</skos:scopeNote>

        <skos:related rdf:resource="https://graphite.synaptica.net/concept/ac1045494"/>

</rdf:Description>
```

Creating a SPARQL Query

Interface to design a SPARQL query:



synaptica
Knowledge Organization & Discovery

Querying an RDF Triple Store with SPARQL

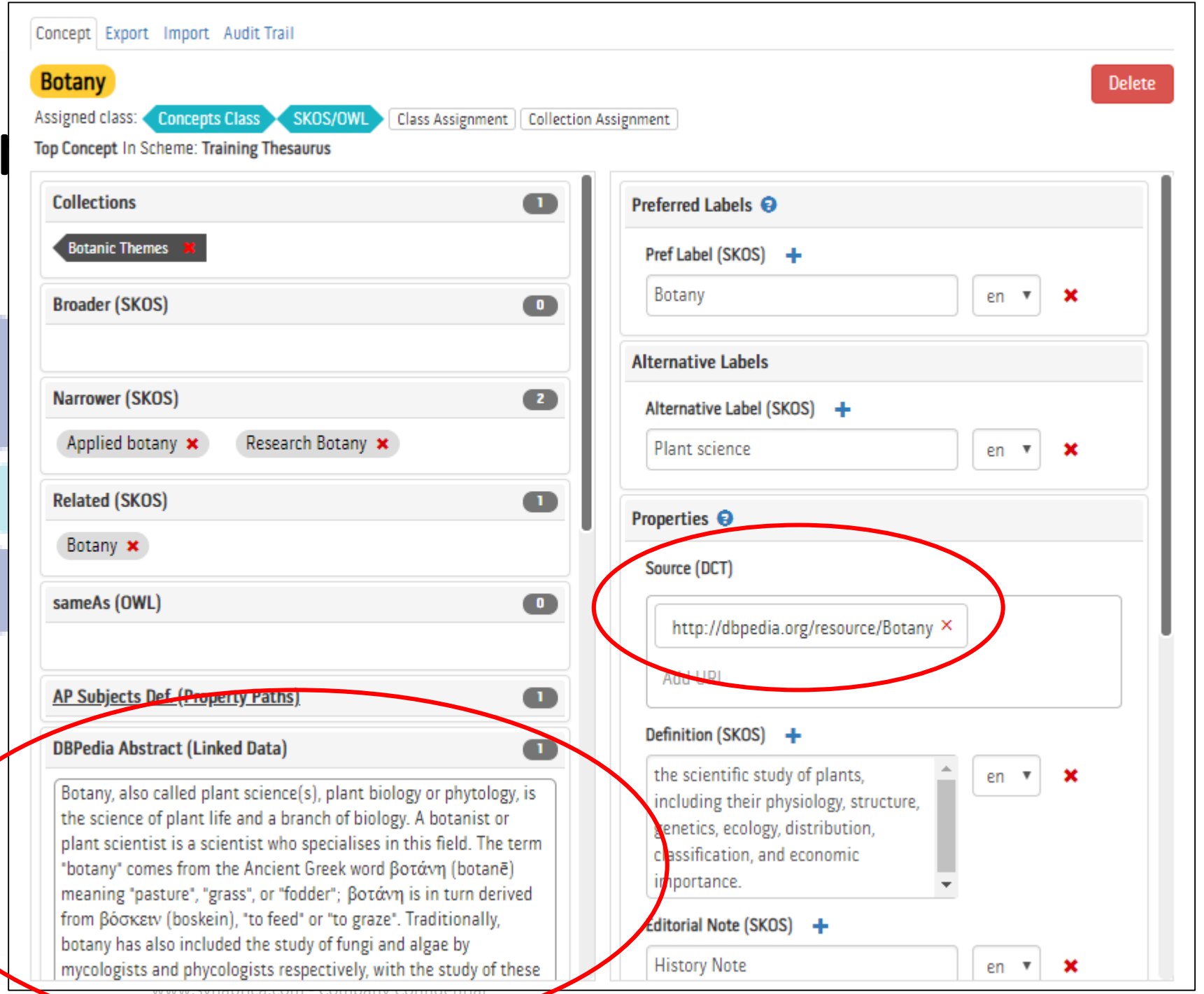
Example of a SPARQL query:

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbo: <http://dbpedia.org/ontology/>

        SELECT DISTINCT ?resource ?label WHERE {
                ?resource a skos:Concept .
                ?resource dbo:abstract ?label .
                #textfilter
        } LIMIT 100
```

synaptica
Knowledge Organization & Discovery

Querying ai

And the results of
that SPARQL query:



synaptica
Knowledge Organization & Discovery

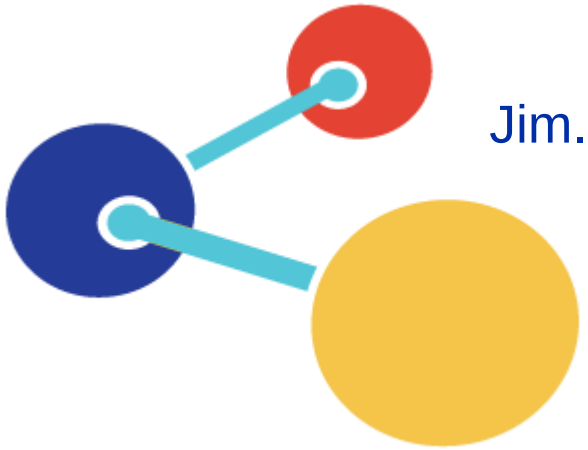
Final Thoughts

- Linked Open Data provides a way to jump-start taxonomy projects and reduce costs.
- LOD is a way to tap into external knowledge that can help answer your business questions.
- Using a Knowledge Organization System, one may construct complex queries to return highly relevant data and enable smart applications.

Thank You!

synaptica
Knowledge Organization & Discovery

Questions?



Jim.Sweeney@synaptica.com
www.synaptica.com