How to Design & Build Semantic Applications with Linked Data

#DCMI17

Dave Clarke
CEO Synaptica

asis&t
https://www.asist.org

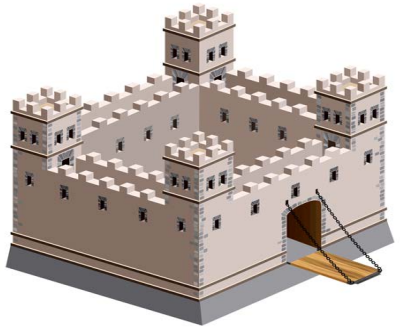
METADATA Dublin Core Metadata Initiative INNOVATION
http://dublincore.org/

www.synaptica.com

1. **Linked Data**: what is it and why is it good for you (10 mins)
2. **Our Use-case**: building an art-history educational application (5 mins)
3. **Design**: a data architecture based on ontologies (10 mins)
4. **Open Data**: adopting and reusing open data resources (5 mins)
5. **User Experience**: end-to-end UX examples (20 mins)
6. **Graph Databases**: types and performance considerations (5 mins)
7. **Discussion**: Q&A (10 mins)

Linked Data:
what is it and why is it good for you

Before Linked Data: Information was locked inside proprietary databases, each of which used custom database schema and every database record was accessed by identifiers that were only unique and intelligible within the system in where they resided.



The guardians of these data ‘fortresses’ didn’t often like sharing data. If it had to be done at all then data was extracted and delivered under lock and key with cryptic instructions how to use it.

After Linked Data: resources (concepts, names, database records, etc.) have global Unique Resource Identifiers (URIs) that are accessible on the web and intelligible to anyone on the planet.

MeSH (Medical Subject Headings):
Internal Medicine
http://id.nlm.nih.gov/mesh/D007388

Getty Art & Architecture Thesaurus:
Renaissance
http://vocab.getty.edu/aat/300021140

Library of Congress Name Authority:
Barack Obama
http://id.loc.gov/authorities/names/n94112934

GeoNames:
London
http://www.geonames.org/2643743

European Environment Agency:
Inland surface waters
http://eunis.eea.europa.eu/habitats/58

WordNet Lexical Database:
finance
http://wordnet-rdf.princeton.edu/wn31/101136358-n

World Wide ~~Web~~ Database

Before Linked Data: web hyperlinks were mere sign posts pointing to other web pages.

After Linked Data: links become semantic, i.e. they express the specific reason why two entities are related.

foaf:Topic **owl:sameAs** **prov:wasInfluencedBy**

These semantically expressive links, called predicates, assert factual statements and support machine reasoning.
They also have their own URIs identifying their place in ontological schema.

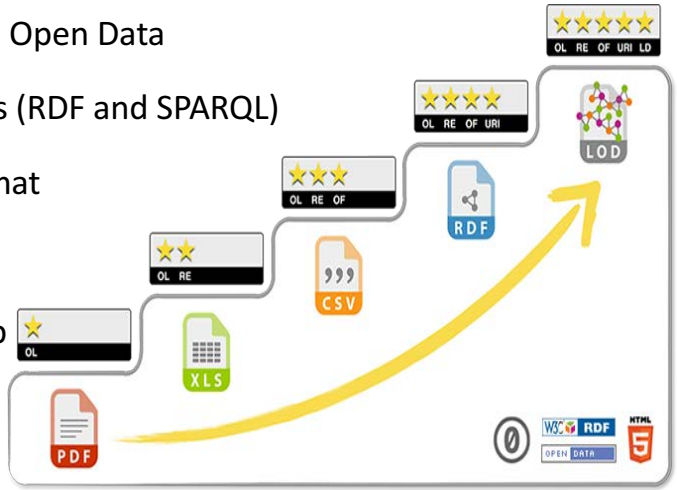
5☆ Links to other Linked Open Data

4☆ W3C Open standards (RDF and SPARQL)

3☆ non-proprietary format

2☆ machine-readable

1☆ available on the Web



Costs and benefits of Five Star Linked Open Data

As a consumer:

- You can discover more (related) data while consuming the data.
- You can directly learn about the data schema.
- You now have to deal with broken data links, just like 404 errors in web pages.
- Presenting data from an arbitrary link as fact is as risky as letting people include content from any website in your pages. Caution, trust and common sense are all still necessary.

As a publisher:

- You make your data discoverable.
- You increase the value of your data.
- Your own organisation will gain the same benefits from the links as the consumers.
- You'll need to invest resources to link your data to other data on the Web.
- You may need to repair broken or incorrect links.

More information on Costs, Benefits and Examples at: http://5stardata.info/en/

all Linked Data is *capable* of being shared

open is optional

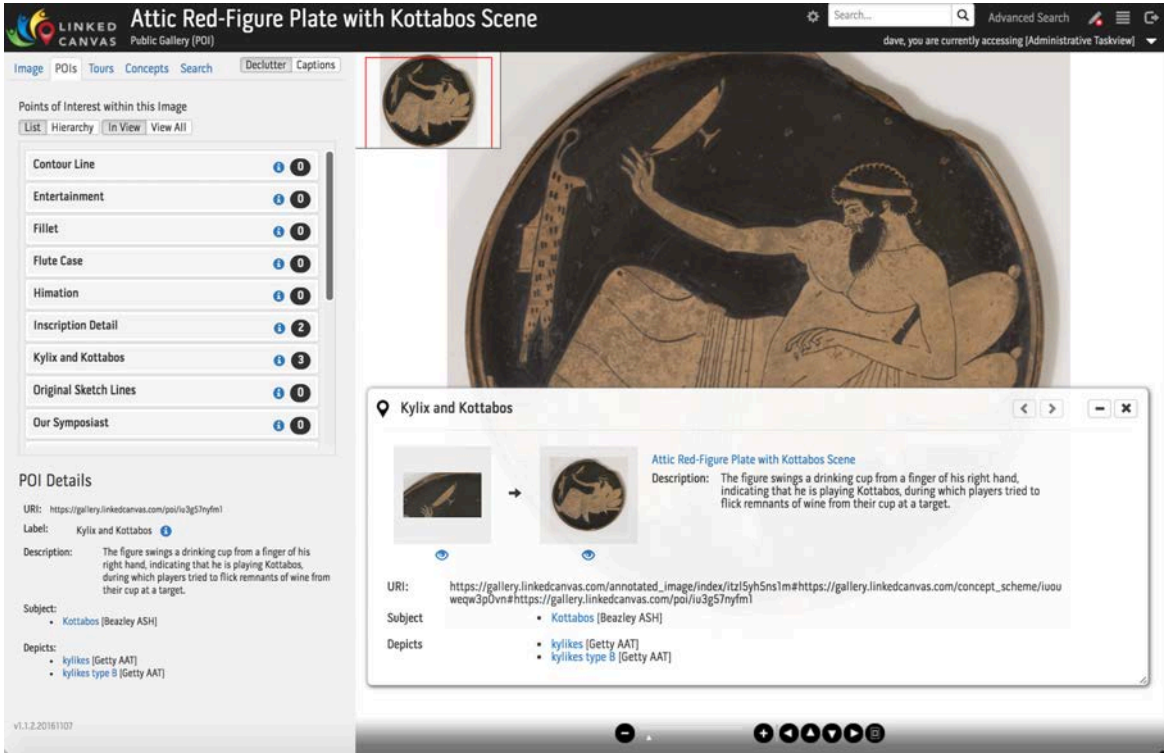
but Linked Data can also reside behind the firewall

Our Use-case:
building an art-history educational application

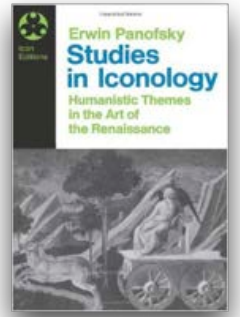
Art history is the academic study of the history and development of painting, sculpture, and the other visual arts. It usually involves a close analysis of visual images as well as texts.

Our use-case project set out to build an interactive educational application to support the close analysis of art images by multiple users, and to capture their analysis in the form of annotations, semantic indexing and multi-media commentary.

The entire application was designed and built using Linked Data.



Erwin Panofsky's 1939 seminal work *Studies in Iconology* describes a three-stage approach to the iconographic analysis of art images



(1) Natural analysis: Describe the visual evidence	**Representational**	**Image Annotation**
(2) Conventional analysis: Interpret the symbolism, allusions, and ideas behind the visual details	**Conceptual**	**Semantic Indexing**
(3) Intrinsic analysis: Explore and the wider cultural and historic context	**Contextual**	**Knowledge Graphs**

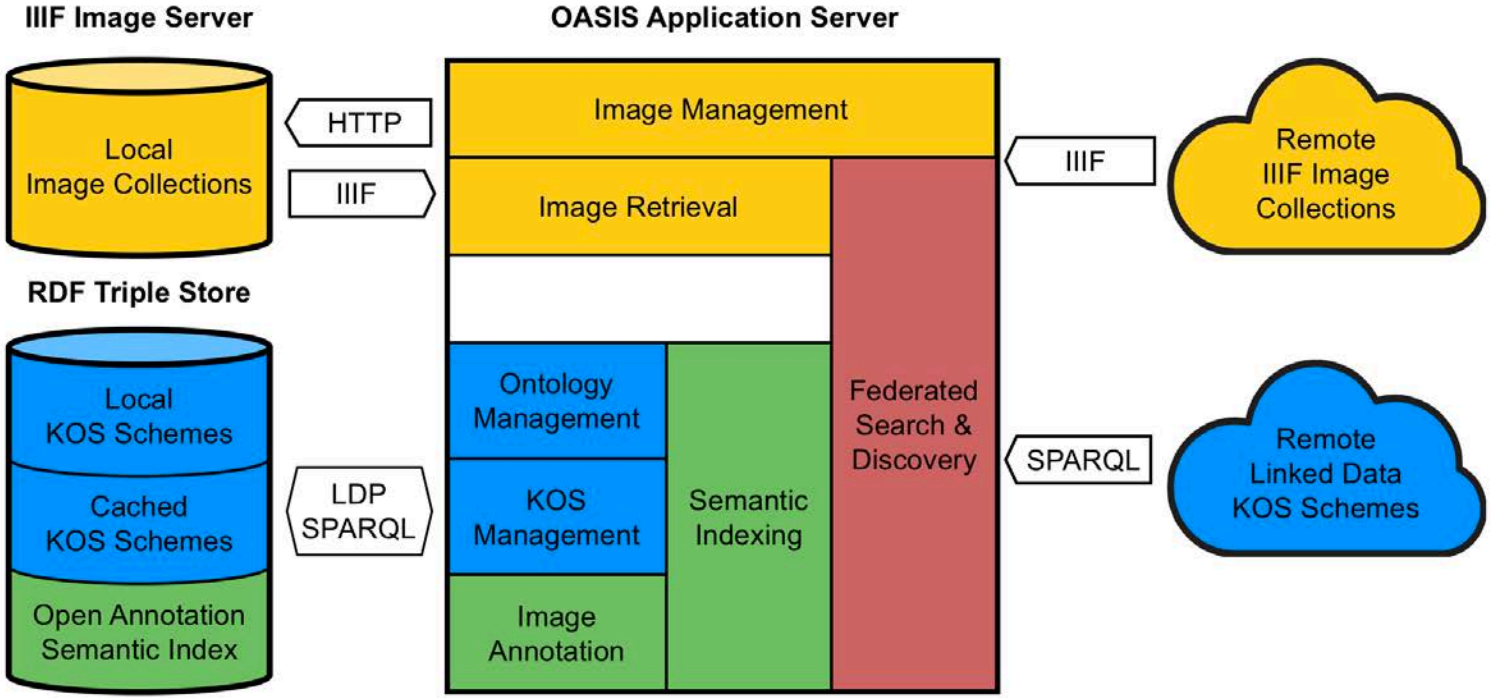
User Requirements

1. Upload and store high-resolution images
2. Catalog images with metadata
3. Pan-and-zoom inside images and define points and areas of interest
4. Annotate points of interest with labels and descriptions
5. Index images and points of interest using Linked Open Data taxonomies
6. Create guide-tours with audio-visual commentary
7. Search and browse across image collections and inside individual images

Design:
a data architecture based on ontologies

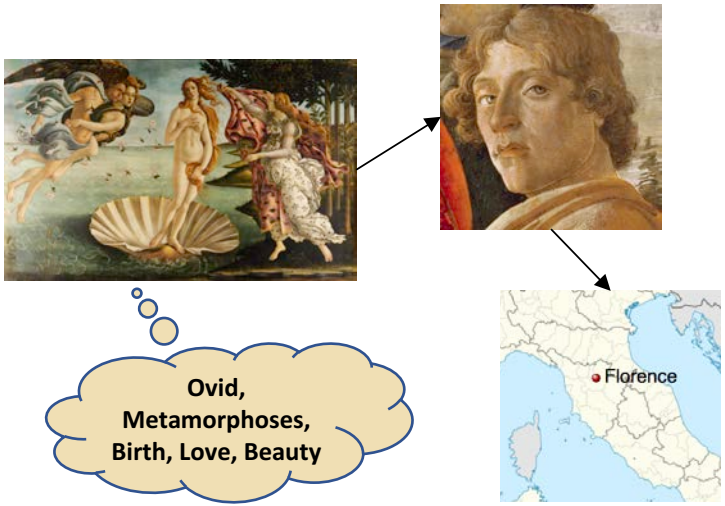
Design: General Principles

1. Manage all metadata and annotations using W3C Linked Data standards
2. Store Linked Data in an RDF triple store
3. Access images using the IIIF open standard image API
4. Annotate images using the W3C Open Annotation (now Web Annotation) Data Model
5. Adopt and reuse Linked Open Data ontologies
6. Adopt and reuse Linked Open Data taxonomies and reference resources



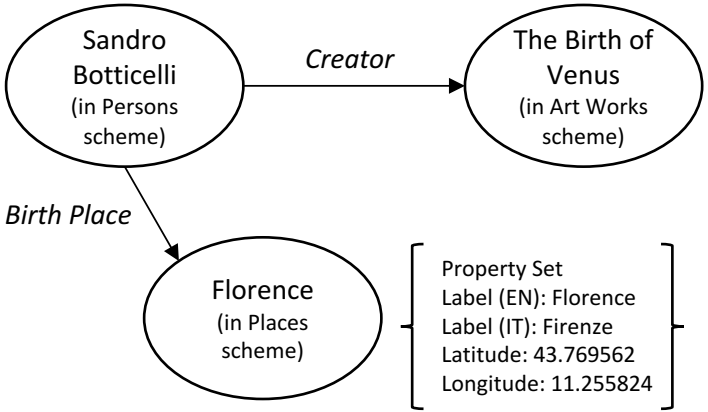
KOS	Knowledge Organization System	https://en.wikipedia.org/wiki/Knowledge_Organization_Systems
IIIF	International Image Interoperability Framework	http://iiif.io
LDP	Linked Data Platform	http://www.w3.org/TR/2015/REC-ldp-20150226/
OA	Open Annotation Data Model	http://www.openannotation.org/spec/core/
SPARQL	SPARQL Query Language for RDF	http://www.w3.org/TR/rdf-sparql-query/

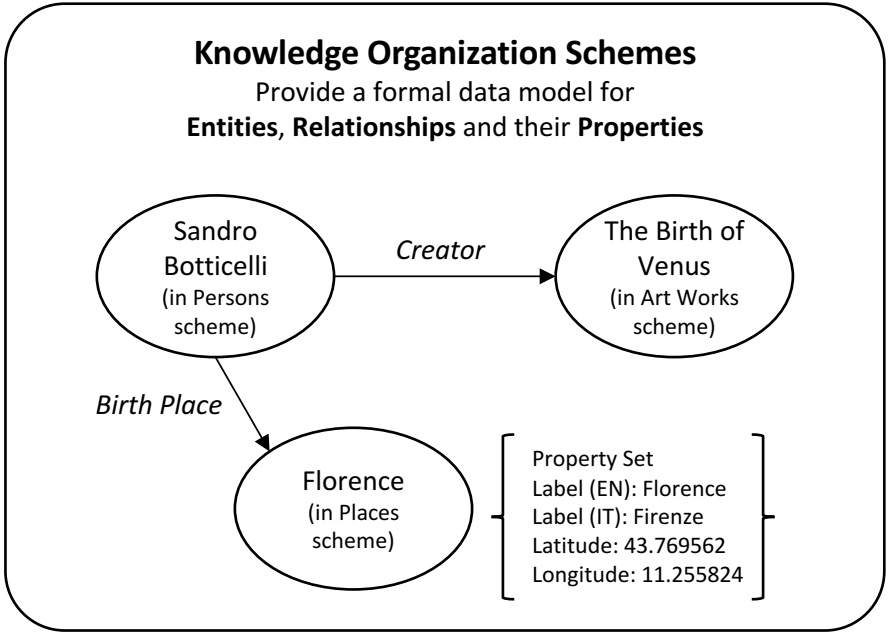
Data modelling begins by taking an inventory of the different types of real-world entities and abstract ideas that an information system needs to capture. In our use-case this includes *Art Works… Places… People… Concepts… and Allusions*



Knowledge Organization Schemes

Provide a formal data model for
Entities, **Relationships** and their **Properties**

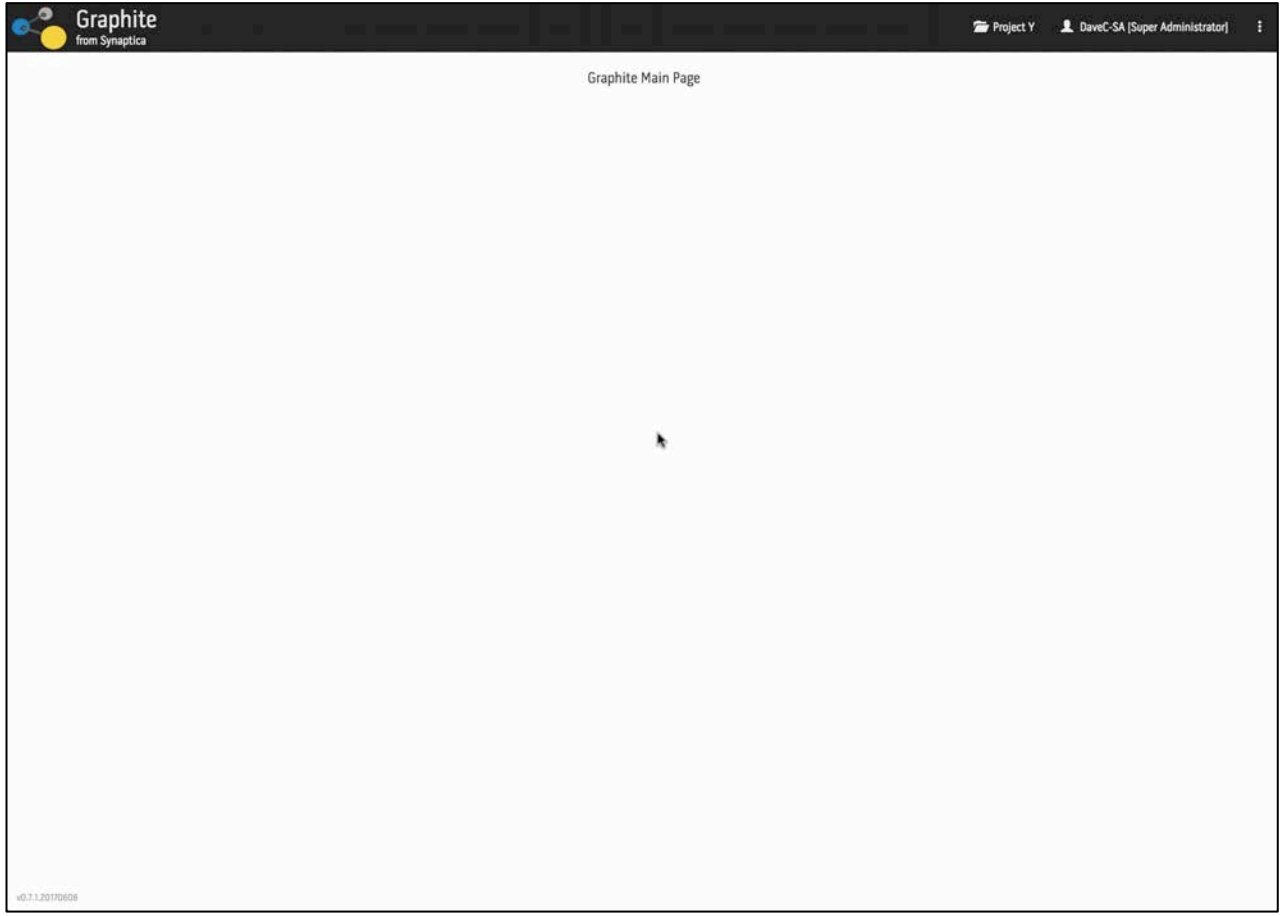




- Entities are identified by URIs
- Entities link to their literal properties via property types
- Property types are identified by URIs
- Entities link to other entities via relationship types
- Relationship types are identified by URIs

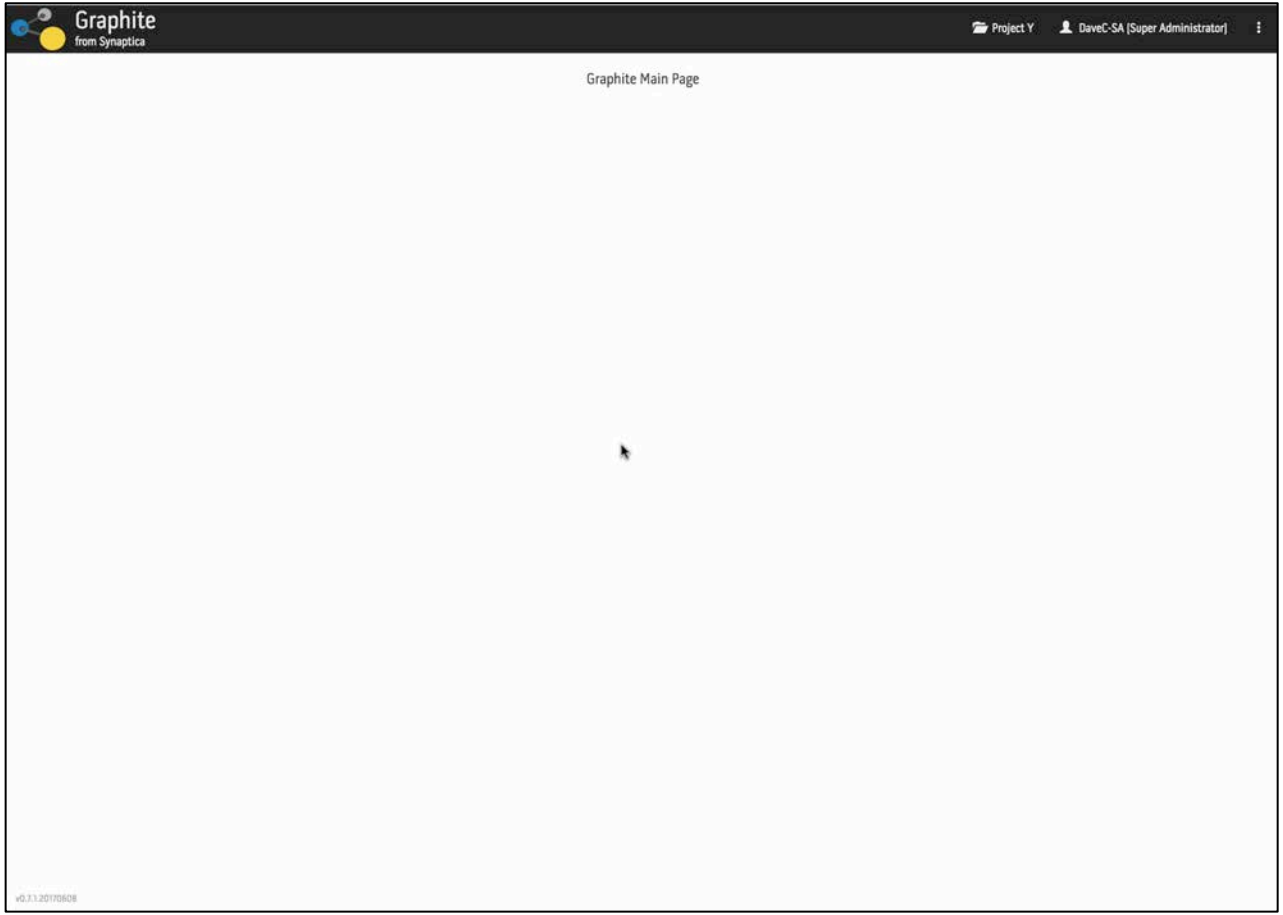
Linked Open Data ontologies provide a ready-made library of reusable property types and relationship types

- By adopting Linked Open Data predicates project **lead times may be reduced** and one's data becomes immediately **intelligible to others**



Mini Screencast: Ontology Management

- A library of open data ontologies can be adopted and managed
- Custom ontologies can also be created as and when needed
- Specific Properties and Relationship predicates can be selected
- Predicates include the URI from the external open data source, or custom minted URIs



Mini Screencast: KOS Scheme Management

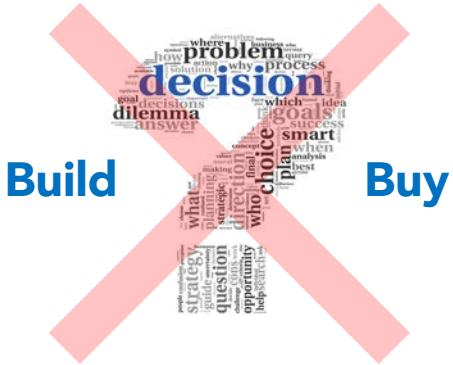
- A library of fully customized concept schemes can be managed
- Schemes are designed by adopting property and relationship types from ontologies
- Cardinality, label uniqueness and other attributes can also be specified
- Inter-scheme permissions control relationship building logic and rules

Open Data:
adopting and reusing open data resources

Build

decision
problem
process
goals
dilemma
decisions
answer
goal
solution
why
how
where
alternatives
business
query
which
idea
success
smart
when
plan
choice
direction
who
what
planning
strategic
final
best
analysis
concept
making
moral
values
strategy
question
opportunity
search
guide
cons
help
people
confusion
uncertainty
factors
challenge
criteria
individual
ask
between
perception
decide
work

Buy



New mantra for taxonomy projects:

ADOPT first

ADAPT second

CREATE third

Trusted authorities

Many subject domains

Millions of concepts

Open data sources

Standard electronic formats

Live query endpoints

and/or file downloads

Linked Open Data

Indexing & Classification Concept Schemes

Examples used for Linked Canvas Project

LCNAF
- 9.5M primary resources
- 80M relationships & properties

LCSH
- 419K primary resources
- 3.9M relationships & properties

Getty AAT
- 42K primary resources
- 14.7M relationships & properties

IconClass
- 40K primary resources
- 3.4M relationships & properties

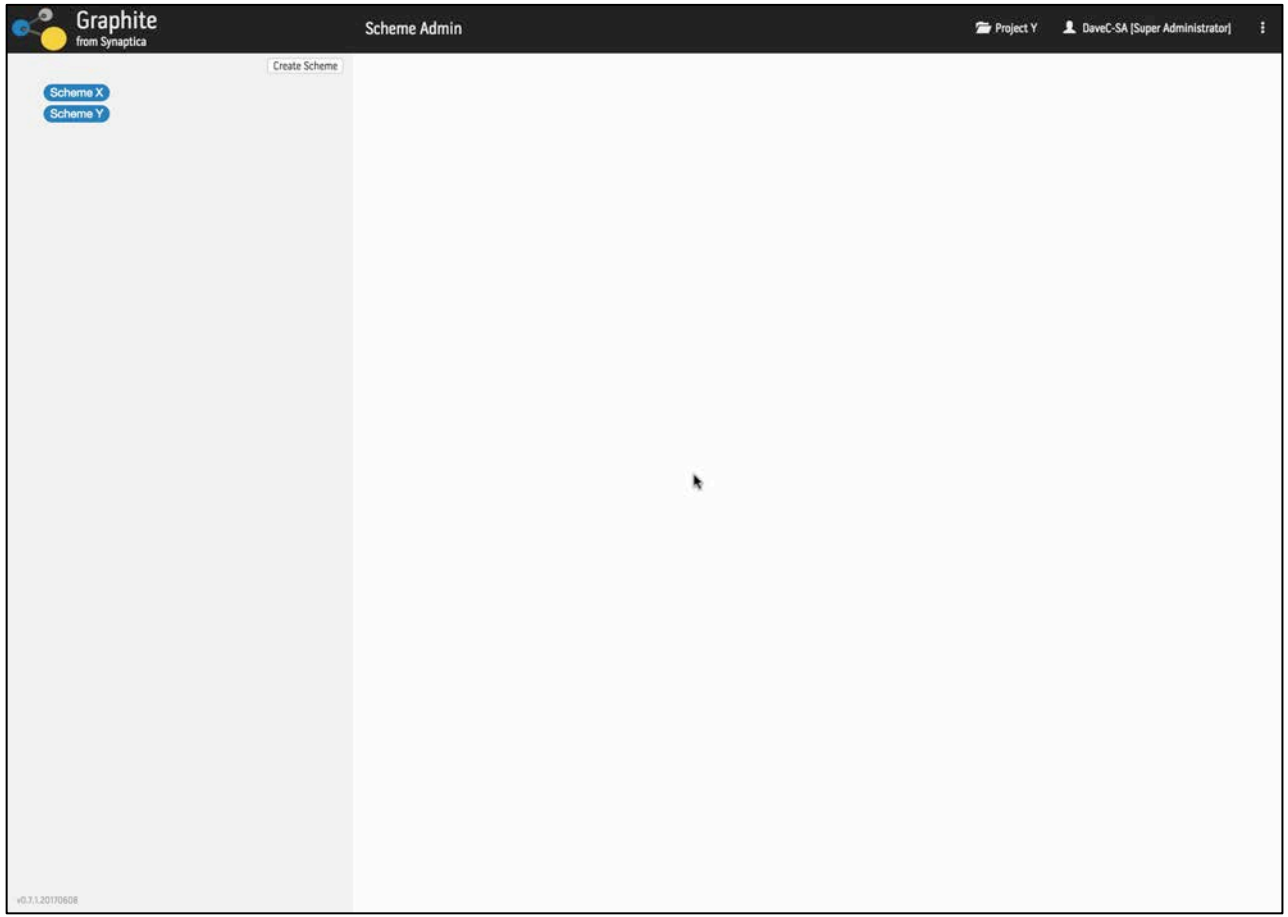
DBpedia
- 1.3M primary resources
- 31M relationships & properties

Totals
- ***11.3M*** *primary semantic resources*
- ***133M*** *relationships & properties*

Corporate & Enterprise Taxonomies

Products & Services
Commodities
Finance
Legal & Regulatory

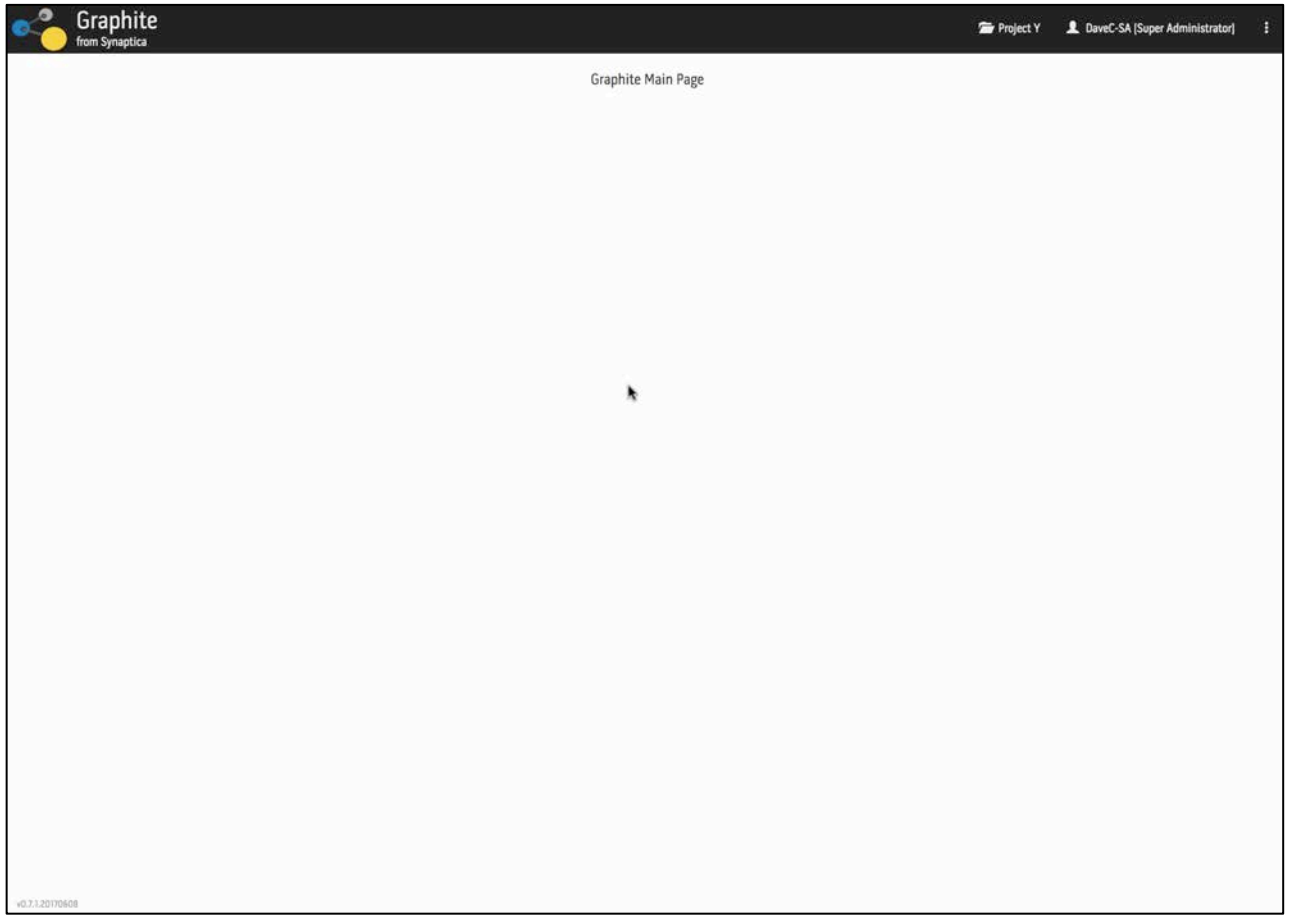
STEMs: Science, Technology, Engineering & Mathematics
HCLS: Health Care & Life Sciences
Cultural Heritage
News Media
Geospatial
Person Names



Mini Screencast: Design New Concept Scheme

- Create a new KOS called Scheme Z
- Base design on SKOS ontology
- Enable prefLabel, altLabel and Scope Notes properties
- Enable SKOS Broader, Narrower and Related relationships

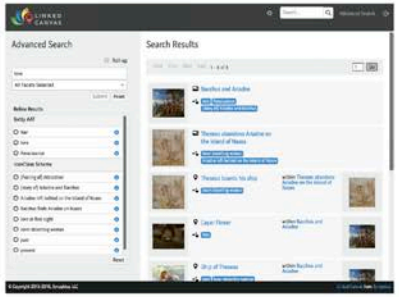
- Finally… enable the owl:sameAs relationship to support crosswalk mapping from Scheme Z to Scheme Y



Mini Screencast: Create Concepts and Crosswalk

- Create a concept tree animals > mammals > canines in Scheme Z
- Browse in Scheme Y then drag concept ‘dogs’ and drop on concept ‘canines’ in Scheme Z using the owl:sameAs relationship

User Experiences: end-to-end UX examples

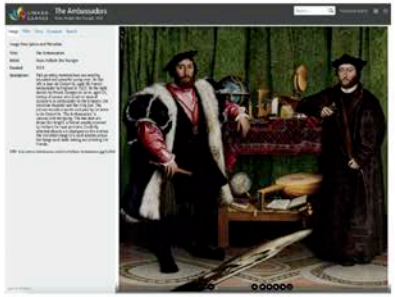


Search inside images or across image collections

LINKED CANVAS delivers advanced search functionality that is easy to use. Search identifies images across collections as well as searching deep inside images to reveal specific points of interest. Search works conceptually to identify images based on subjects, themes and symbols as well as descriptive text.

Arrange POIs into a hierarchical table of contents

LINKED CANVAS allows POIs to be explored as an alphabetical list or as an expandable hierarchy. This allows the picture space and the scenes and figurative details in a composition to be arranged and navigated like the table of contents of a book.

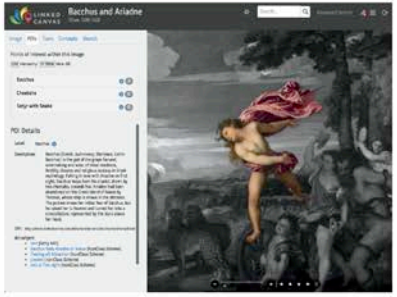


Upload and catalog images

LINKED CANVAS allows users to design their own image catalogs with an extensible set of metadata. It is also possible to import data from external catalogs and integrate the system with collections management systems. Image catalog records may be subject indexed using internal taxonomies or Linked Open Data controlled vocabularies.

Explore images conceptually

With LINKED CANVAS images can be explored conceptually. As the user pans and zooms around an image, a tag-cloud dynamically updates to reveal the ideas behind the image. Selecting a concept provides explanatory notes as well as links to other visual details on the image and to other images.

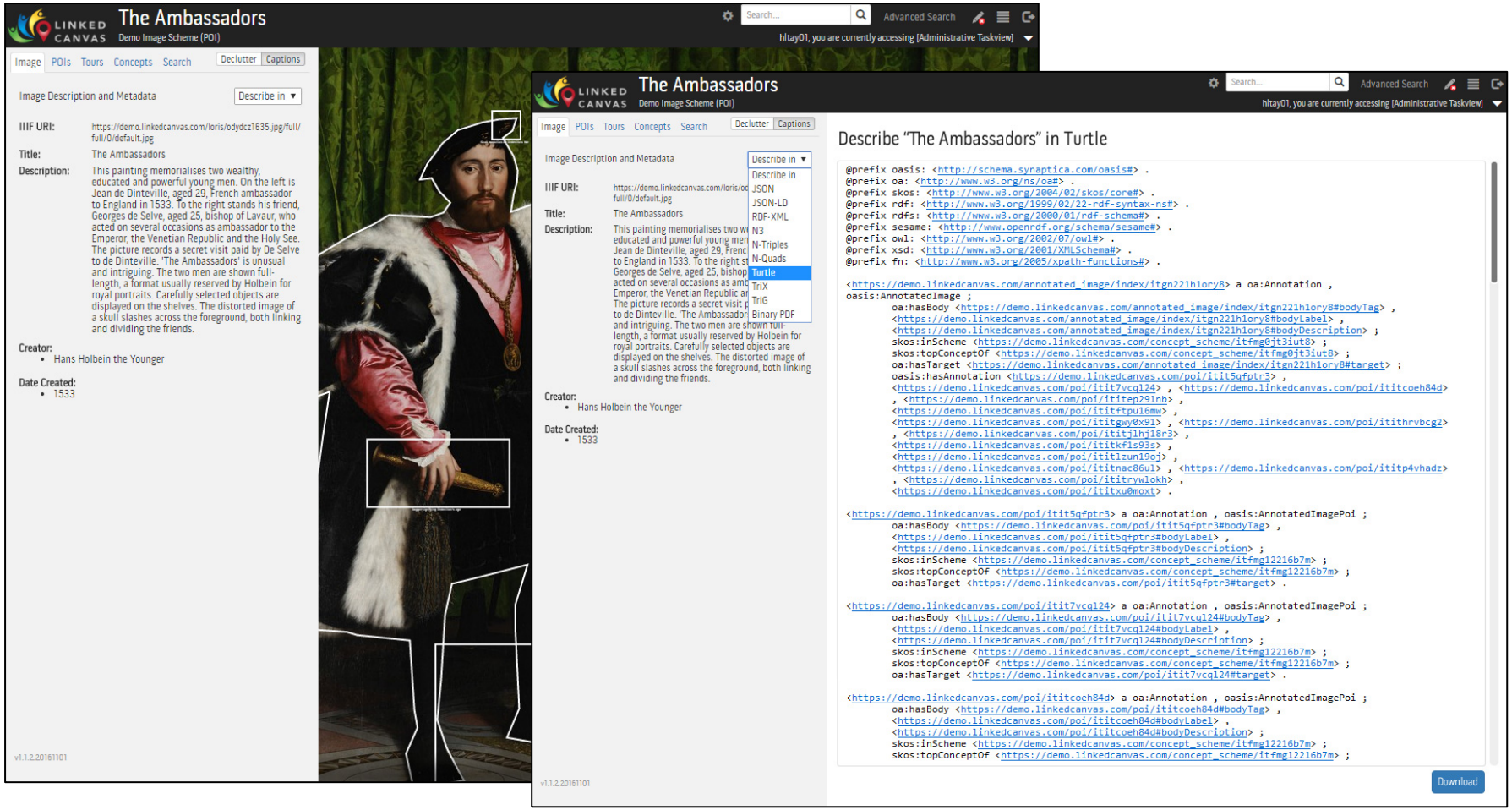


Define points of interest (POIs)

With LINKED CANVAS users can pan and zoom inside an image to define points of interest. These can be created quickly using drop-pin or rectangle-frame markers, or the outlines of specific figures can be traced to create a visually stunning way to highlight specific visual details in the context of their surroundings. POIs can then be labeled, annotated and subject indexed.

Create guided tours with audio visual narration

LINKED CANVAS lets users build time-sequenced guided tours with embedded audio and video narration. Multiple guided tours can be created to describe different aspects of an image, such as subject matter versus form and technique. With guided tours users can explore artworks using 'playlist' style controls.

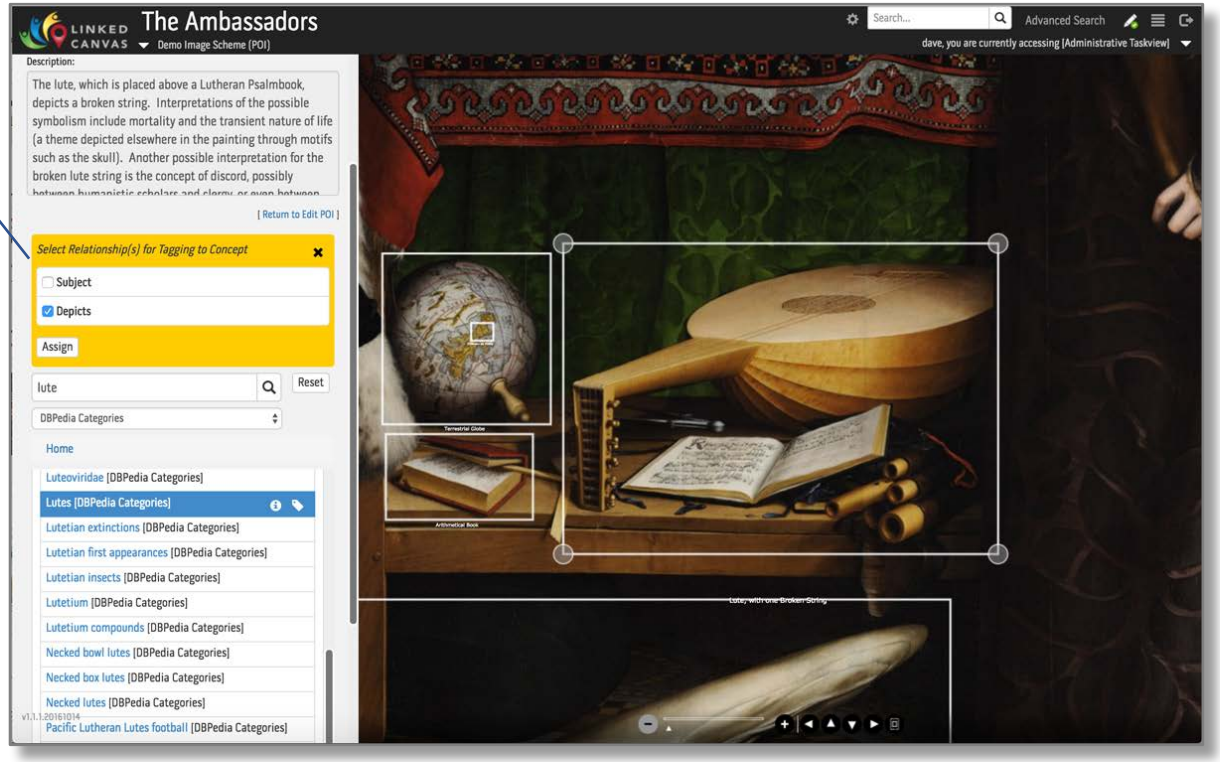

LINKED CANVAS
The Ambassadors
Demo Image Scheme (POI)
Search...
Advanced Search
hltay01, you are currently accessing [Administrative Taskview]
Image
POIs
Tours
Concepts
Search
Declutter
Captions
Image Description and Metadata
Describe in
IIIF URI:
https://demo.linkedcanvas.com/loris/odydcz1635.jpg/full/full/0/default.jpg
Title:
The Ambassadors
Description:
This painting memorialises two wealthy, educated and powerful young men. On the left is Jean de Dinteville, aged 29, French ambassador to England in 1533. To the right stands his friend, Georges de Selve, aged 25, bishop of Lavaur, who acted on several occasions as ambassador to the Emperor, the Venetian Republic and the Holy See. The picture records a secret visit paid by De Selve to de Dinteville. 'The Ambassadors' is unusual and intriguing. The two men are shown full-length, a format usually reserved by Holbein for royal portraits. Carefully selected objects are displayed on the shelves. The distorted image of a skull slashes across the foreground, both linking and dividing the friends.
Creator:
Hans Holbein the Younger
Date Created:
1533
v1.1.2.20161101
Describe in
JSON
JSON-LD
RDF-XML
N3
N-Triples
N-Quads
Turtle
TriX
TriG
Binary PDF
Describe "The Ambassadors" in Turtle
@prefix oasis: <http://schema.synaptica.com/oasis#> .
@prefix oa: <http://www.w3.org/ns/oa#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix sesame: <http://www.openrdf.org/schema/sesame#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix fn: <http://www.w3.org/2005/xpath-functions#> .
<https://demo.linkedcanvas.com/annotated_image/index/itgn221h1ory8> a oa:Annotation ,
oasis:AnnotatedImage ;
oa:hasBody <https://demo.linkedcanvas.com/annotated_image/index/itgn221h1ory8#bodyTag> ,
<https://demo.linkedcanvas.com/annotated_image/index/itgn221h1ory8#bodyLabel> ,
<https://demo.linkedcanvas.com/annotated_image/index/itgn221h1ory8#bodyDescription> ;
skos:inScheme <https://demo.linkedcanvas.com/concept_scheme/itfmg0jt3iut8> ;
skos:topConceptOf <https://demo.linkedcanvas.com/concept_scheme/itfmg0jt3iut8> ;
oa:hasTarget <https://demo.linkedcanvas.com/annotated_image/index/itgn221h1ory8#target> ;
oasis:hasAnnotation <https://demo.linkedcanvas.com/poi/itit5qfptr3> ,
<https://demo.linkedcanvas.com/poi/itit7vcql24> , <https://demo.linkedcanvas.com/poi/itit coeh84d>
, <https://demo.linkedcanvas.com/poi/itiep291nb> ,
<https://demo.linkedcanvas.com/poi/itiftpu16mw> ,
<https://demo.linkedcanvas.com/poi/itigwy0x91> , <https://demo.linkedcanvas.com/poi/itithrvbcg2>
, <https://demo.linkedcanvas.com/poi/itijhj18r3> ,
<https://demo.linkedcanvas.com/poi/itikf1s93s> ,
<https://demo.linkedcanvas.com/poi/itit1zun19oj> ,
<https://demo.linkedcanvas.com/poi/itit nac86u1> , <https://demo.linkedcanvas.com/poi/itit p4vhadz>
, <https://demo.linkedcanvas.com/poi/itit rywlokh> ,
<https://demo.linkedcanvas.com/poi/itit xu0moxt> .
<https://demo.linkedcanvas.com/poi/itit5qfptr3> a oa:Annotation , oasis:AnnotatedImagePoi ;
oa:hasBody <https://demo.linkedcanvas.com/poi/itit5qfptr3#bodyTag> ,
<https://demo.linkedcanvas.com/poi/itit5qfptr3#bodyLabel> ,
<https://demo.linkedcanvas.com/poi/itit5qfptr3#bodyDescription> ;
skos:inScheme <https://demo.linkedcanvas.com/concept_scheme/itfmg12216b7m> ;
skos:topConceptOf <https://demo.linkedcanvas.com/concept_scheme/itfmg12216b7m> ;
oa:hasTarget <https://demo.linkedcanvas.com/poi/itit5qfptr3#target> .
<https://demo.linkedcanvas.com/poi/itit7vcql24> a oa:Annotation , oasis:AnnotatedImagePoi ;
oa:hasBody <https://demo.linkedcanvas.com/poi/itit7vcql24#bodyTag> ,
<https://demo.linkedcanvas.com/poi/itit7vcql24#bodyLabel> ,
<https://demo.linkedcanvas.com/poi/itit7vcql24#bodyDescription> ;
skos:inScheme <https://demo.linkedcanvas.com/concept_scheme/itfmg12216b7m> ;
skos:topConceptOf <https://demo.linkedcanvas.com/concept_scheme/itfmg12216b7m> ;
oa:hasTarget <https://demo.linkedcanvas.com/poi/itit7vcql24#target> .
<https://demo.linkedcanvas.com/poi/itit coeh84d> a oa:Annotation , oasis:AnnotatedImagePoi ;
oa:hasBody <https://demo.linkedcanvas.com/poi/itit coeh84d#bodyTag> ,
<https://demo.linkedcanvas.com/poi/itit coeh84d#bodyLabel> ,
<https://demo.linkedcanvas.com/poi/itit coeh84d#bodyDescription> ;
skos:inScheme <https://demo.linkedcanvas.com/concept_scheme/itfmg12216b7m> ;
skos:topConceptOf <https://demo.linkedcanvas.com/concept_scheme/itfmg12216b7m> ;
Download

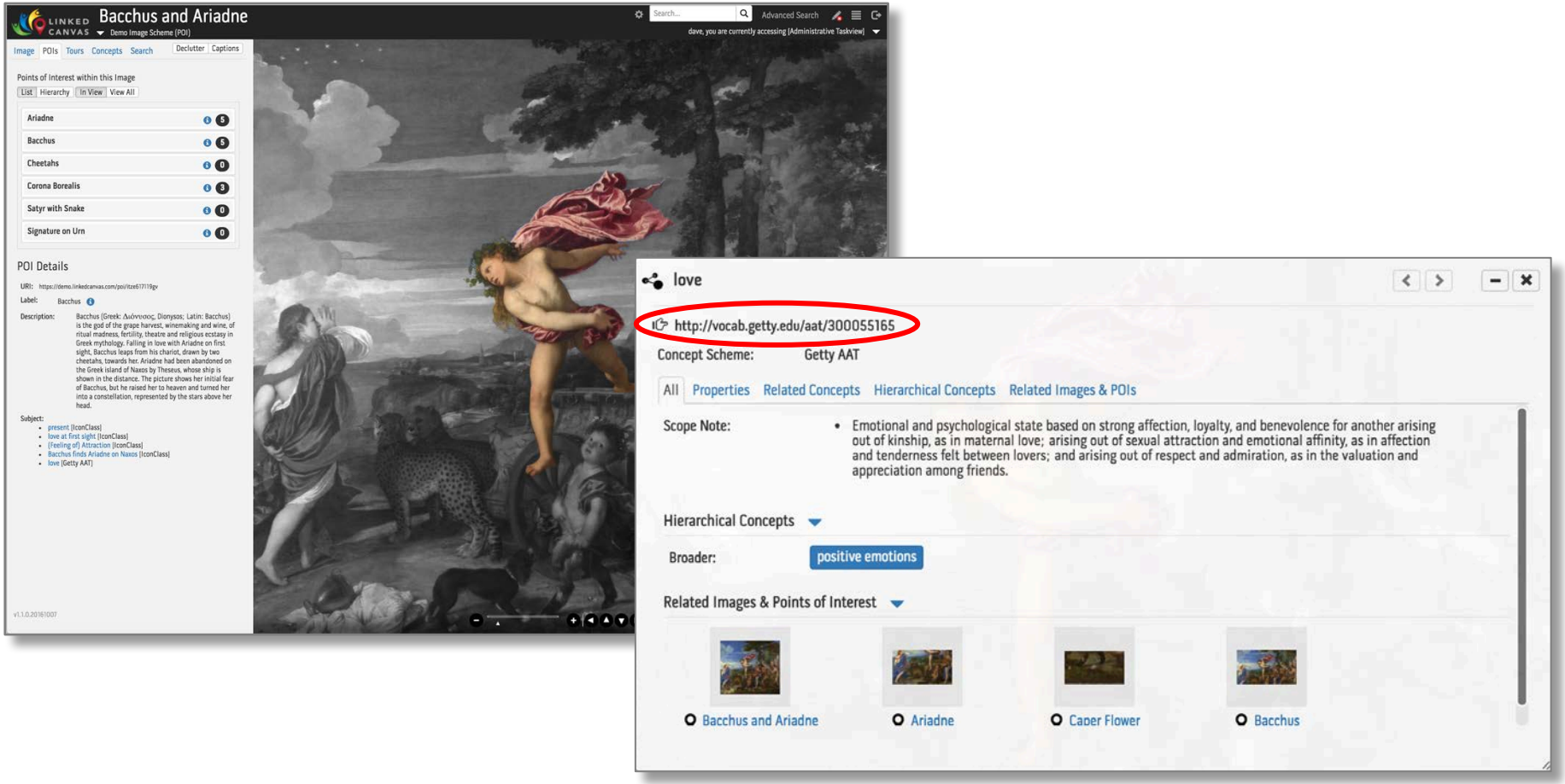
Process of indexing a visual detail inside a painting to the DBpedia category resource for ***Lutes***

http://dbpedia.org/resource/Category:Lutes

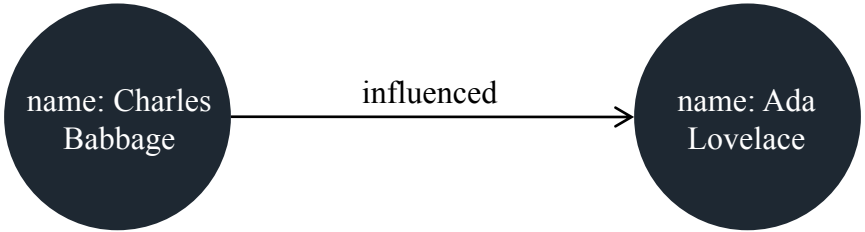
via the FOAF predicate
Depicts

foaf:depicts




LINKED CANVAS
Bacchus and Ariadne
Demo Image Scheme (POI)
Search...
Advanced Search
dave, you are currently accessing [Administrative Taskview]
Image
POIs
Tours
Concepts
Search
Declutter
Captions
Points of Interest within this Image
List
Hierarchy
In View
View All
Ariadne 5
Bacchus 5
Cheetahs 0
Corona Borealis 3
Satyr with Snake 0
Signature on Urn 0
POI Details
URI: https://demo.linkedcanvas.com/poi/itze617119gv
Label: Bacchus
Description: Bacchus (Greek: Διόνυσος, Dionysos; Latin: Bacchus) is the god of the grape harvest, winemaking and wine, of ritual madness, fertility, theatre and religious ecstasy in Greek mythology. Falling in love with Ariadne on first sight, Bacchus leaps from his chariot, drawn by two cheetahs, towards her. Ariadne had been abandoned on the Greek island of Naxos by Theseus, whose ship is shown in the distance. The picture shows her initial fear of Bacchus, but he raised her to heaven and turned her into a constellation, represented by the stars above her head.
Subject:
present [IconClass]
love at first sight [IconClass]
(Feeling of) Attraction [IconClass]
Bacchus finds Ariadne on Naxos [IconClass]
love [Getty AAT]
v1.1.0.20161007
love
http://vocab.getty.edu/aat/300055165
Concept Scheme: Getty AAT
All
Properties
Related Concepts
Hierarchical Concepts
Related Images & POIs
Scope Note:
Emotional and psychological state based on strong affection, loyalty, and benevolence for another arising out of kinship, as in maternal love; arising out of sexual attraction and emotional affinity, as in affection and tenderness felt between lovers; and arising out of respect and admiration, as in the valuation and appreciation among friends.
Hierarchical Concepts
Broader:
positive emotions
Related Images & Points of Interest
Bacchus and Ariadne
Ariadne
Caper Flower
Bacchus

Graph Databases: types and performance considerations

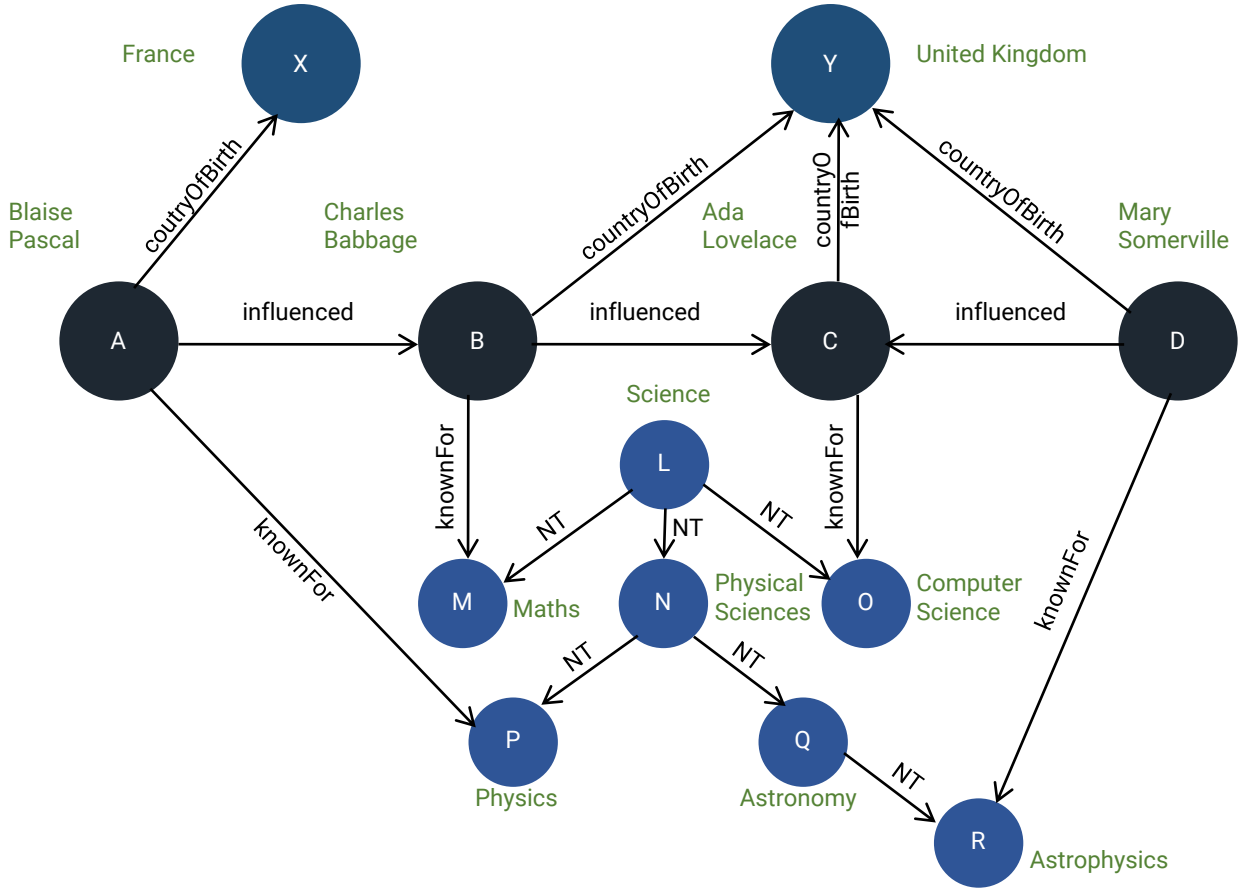


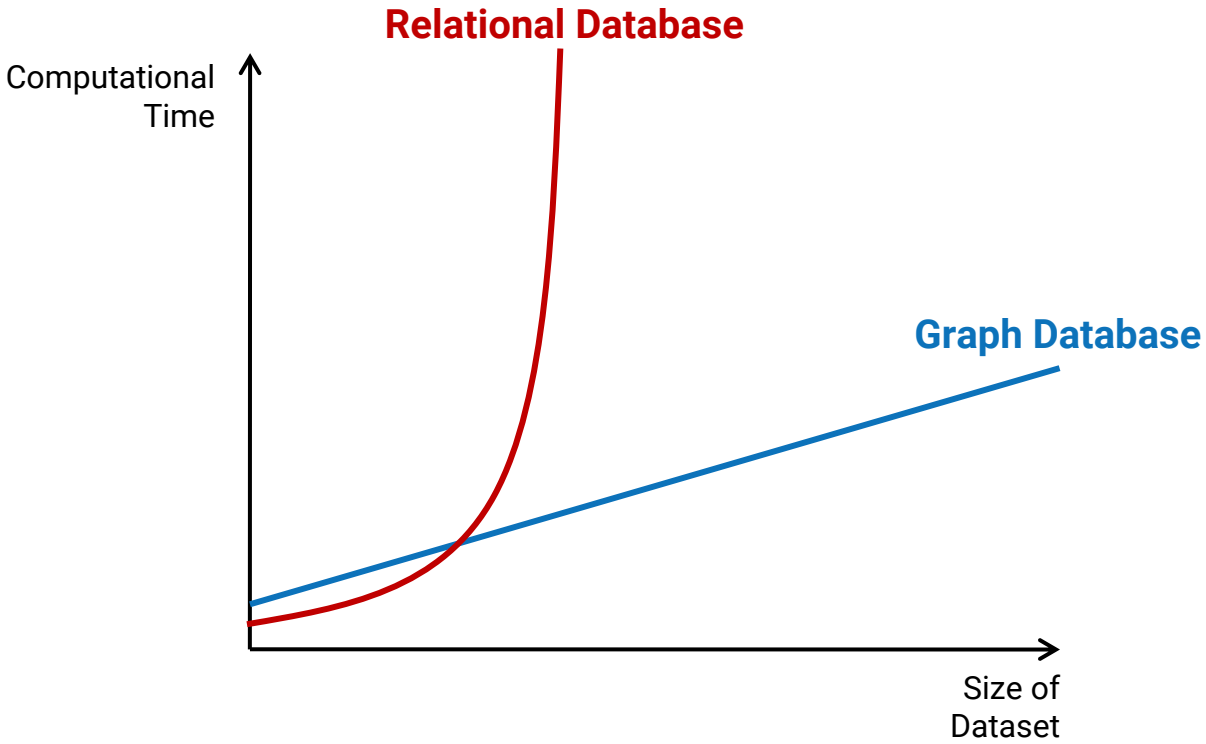
Subject	Predicate	Object
<http://id.loc.gov/authorities/names/n50031102>	prov:influenced	<http://id.loc.gov/authorities/names/n78030997>
<http://id.loc.gov/authorities/names/n50031102>	foaf:name	”Charles Babbage"
<http://id.loc.gov/authorities/names/n78030997>	foaf:name	“Ada Lovelace”

GeoNames

LC Name Authority File

LC Subject Headings




Relational Database
Computational Time
Graph Database
Size of Dataset

Graph databases don't naturally perform well with free text search… and remote SPARQL Endpoints may or may not be responsive how can these problems be overcome?

Query:

```
1  select ?Subject ?Term ?Parents ?Descr ?ScopeNote ?Type (coalesce(?Type1,?Type2) as ?ExtraType) {
2    ?Subject luc:term "vessel AND fish*"; a ?typ.
3    ?typ rdfs:subClassOf gvp:Subject; rdfs:label ?Type.
4    filter (?typ != gvp:Subject)
5    optional {?Subject gvp:placeTypePreferred [gvp:prefLabelGVP [xl:literalForm ?Type1]]}
6    optional {?Subject gvp:agentTypePreferred [gvp:prefLabelGVP [xl:literalForm ?Type2]]}
7    optional {?Subject gvp:prefLabelGVP [xl:literalForm ?Term]}
8    optional {?Subject gvp:parentStringAbbrev ?Parents}
9    optional {?Subject foaf:focus/gvp:biographyPreferred/schema:description ?Descr}
10   optional {?Subject skos:scopeNote [dct:language gvp_lang:en; rdf:value ?ScopeNote]}}
```

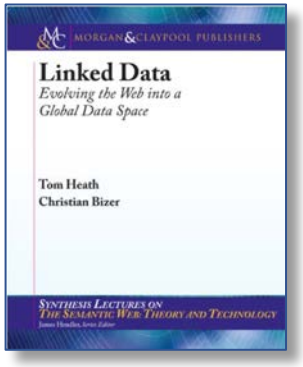
Some Common Options

- Download and cache remote LoD data sources
- Index literal property fields using search tools like Apache Solr
- Use triple store databases with built-in indexing like GraphDB
- Create 'named graphs' that can be used to sub-select from the master graph database

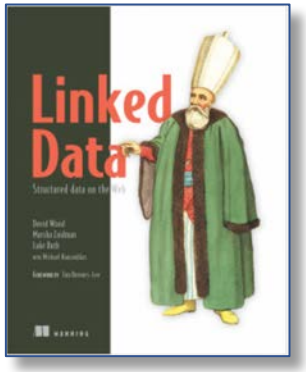
In the Linked Canvas system with > 11 million entities and > 130 million properties and relationships we execute simple searches sub-second and multi-word wildcard phrases in 1-5 seconds

To RDF or not to RDF?

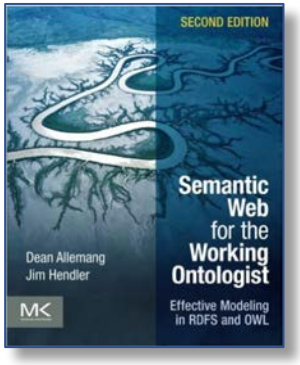
http://www.zdnet.com/article/graph-databases-and-rdf-its-a-family-affair/



Heath & Bizer
Morgan & Claypool (pub.)



Ruth, Wood & Zaidman
Manning (pub.)



Allemang & Hendler
Morgan Klaufman (pub.)

LINKED CANVAS

Home Design Features Videos Engage About

Tutorial Video Series

Synaptica Tech Talks - Data Modeling with Graphs

This tutorial explores data modelling by comparing and contrasting traditional relational databases with property graph database and RDF Linked Data graph databases.

Linked Canvas 60 Second Tutorials: 01 Uploading Images

This tutorial demonstrates how images may be uploaded and managed within Linked Canvas.

Linked Canvas 60 Second Tutorials: 02 Cataloging Images

This tutorial demonstrates how images may be cataloged within Linked Canvas. Linked Canvas supports an extensible set of image-level metadata.

Showcase Video Series

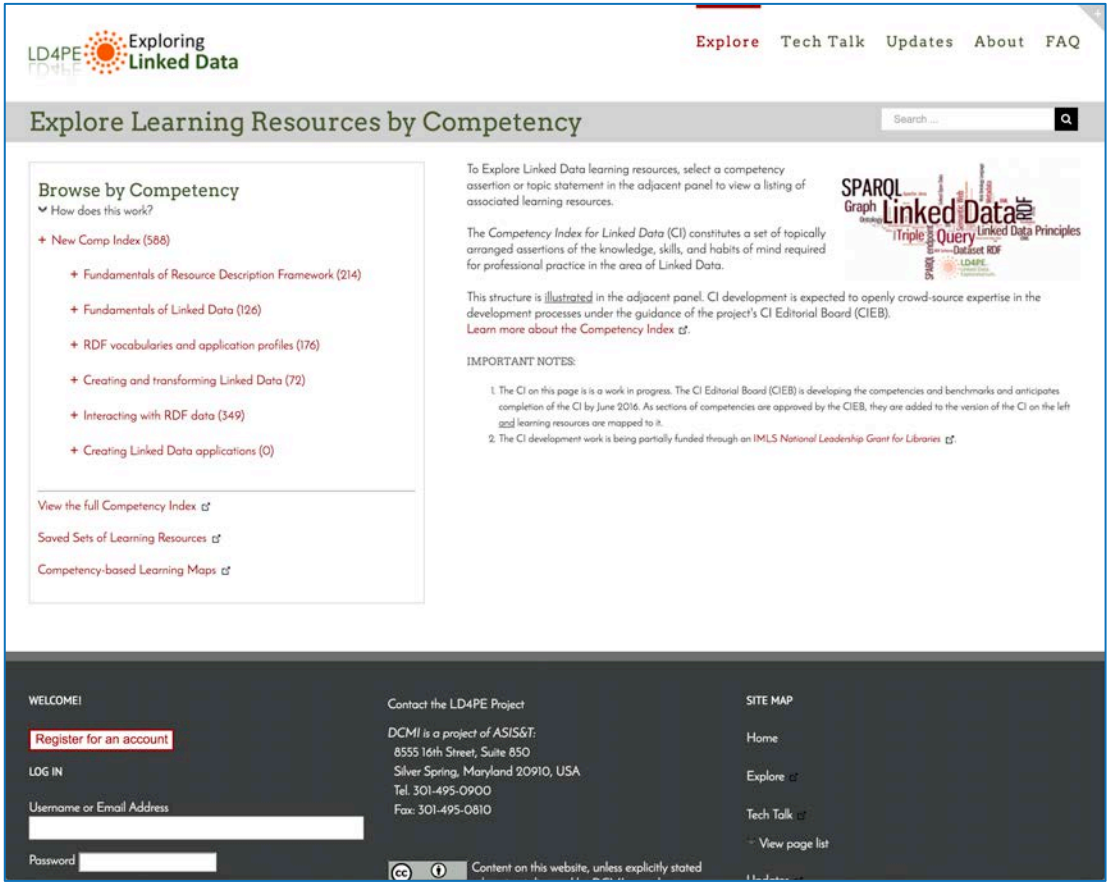
Bryn Mawr College - The World of the Symposium

Take a guided tour of a 5th C. BCE Attic Red-Figure Plate from the Bryn Mawr College special collection, and discover insights into The World of the Symposium. The annotations and subject indexing (using the Beazley Archive Subject Headings) were prepared by Bryn Mawr College as a cross-faculty exercise.

View this image live in Linked Canvas

www.linkedcanvas.org

http://explore.dublincore.net/explore-learning-resources-by-competency/



Thank You Dublin Core!

Thank You

David Clarke
CEO

dave.clarke@synaptica.com

www.synaptica.com

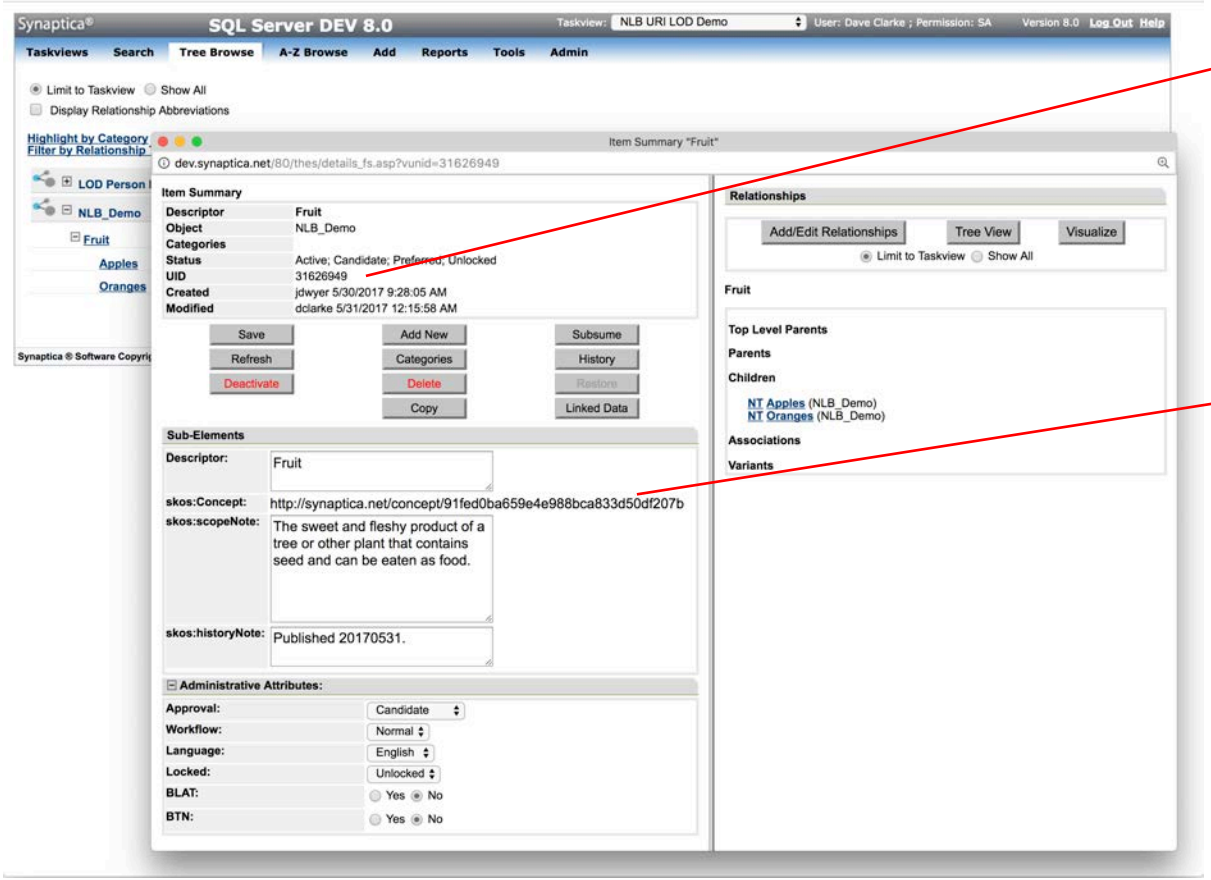
www.linkedcanvas.org

Reserve Slides

Managing HTTP-URIs & Linked Data Mapping Within Relational KOS Management Systems

Minting HTTP-URIs and Managing GUIDs

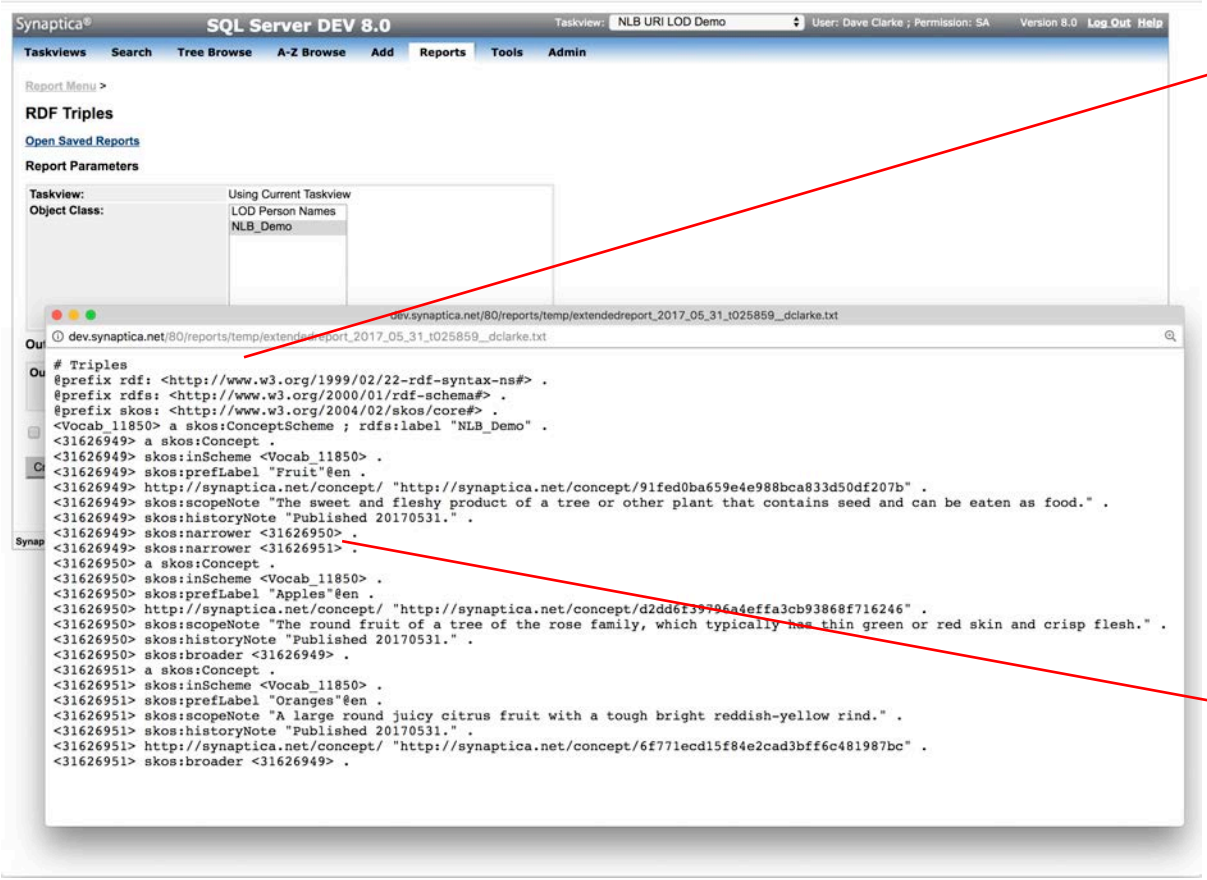
Example of a Term with HTTP-URI and GUID



All terms in Synaptica always have a system-generated unique ID (UID). It is a sequence number unique to each term in the database instance. It is used as the primary ID for terms internally and in imports/exports.

Synaptica also supports the creation of HTTP-URIs. A URI is composed of an HTTP-prefix followed by a unique ID. Any HTTP prefix can be configured in the Admin. tools, thereby supporting custom namespaces. The unique ID has three options: (i) the Synaptica UID; (ii) a true GUID (random 32 character alphanumeric); or the term descriptor. A descriptor is more user friendly than an ID, but is generally discouraged because it necessarily favours a single natural language and because descriptor terminology may change over time. The GUID is longer and more cumbersome than a UID, but GUIDs are generally preferred by the publishing industry as they are universally unique.

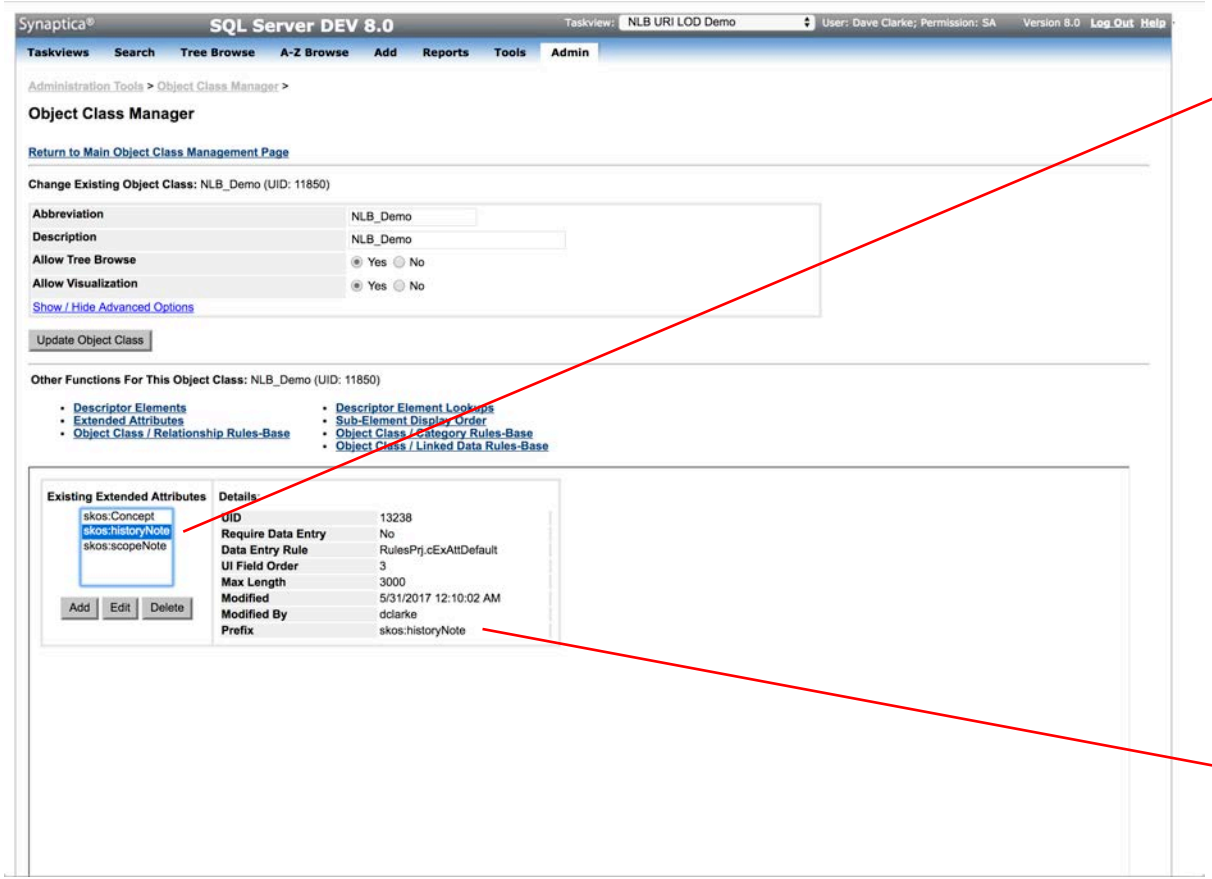
Example of Triples Output



The optional LOD module which will be released with V8.0 supports the export of triples. These can leverage namespace prefixes and HTTP-URIs for terms and predicates (relationship types and property types).

In this example Synaptica UIDs are used as Subjects and Objects in the triples, but in the final version to be released term HTTP-URIs could be used to directly represent Subjects and Objects.

Configuring a concept scheme to use HTTP-URI-based properties

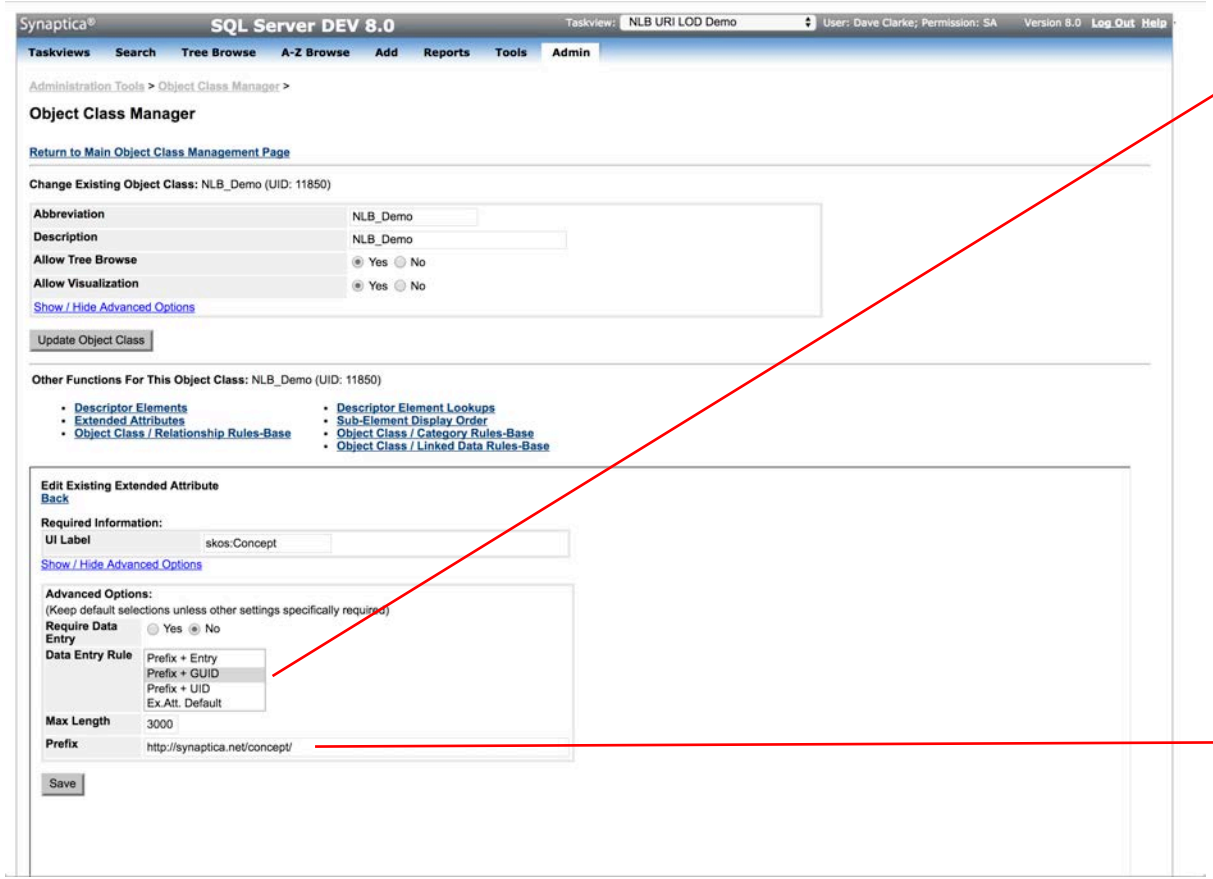


HTTP-URIs for terms and properties can be configured in the Object Class manager.

In this example the scheme 'NLB_Demo' has been configured with three HTTP-URI properties: skos:Concept; skos:scopeNote; and skos:historyNote.

Note the use of RDF namespace prefixes, which will also be illustrated on a subsequent screen.

Configuring a concept scheme to use HTTP-URI-based properties

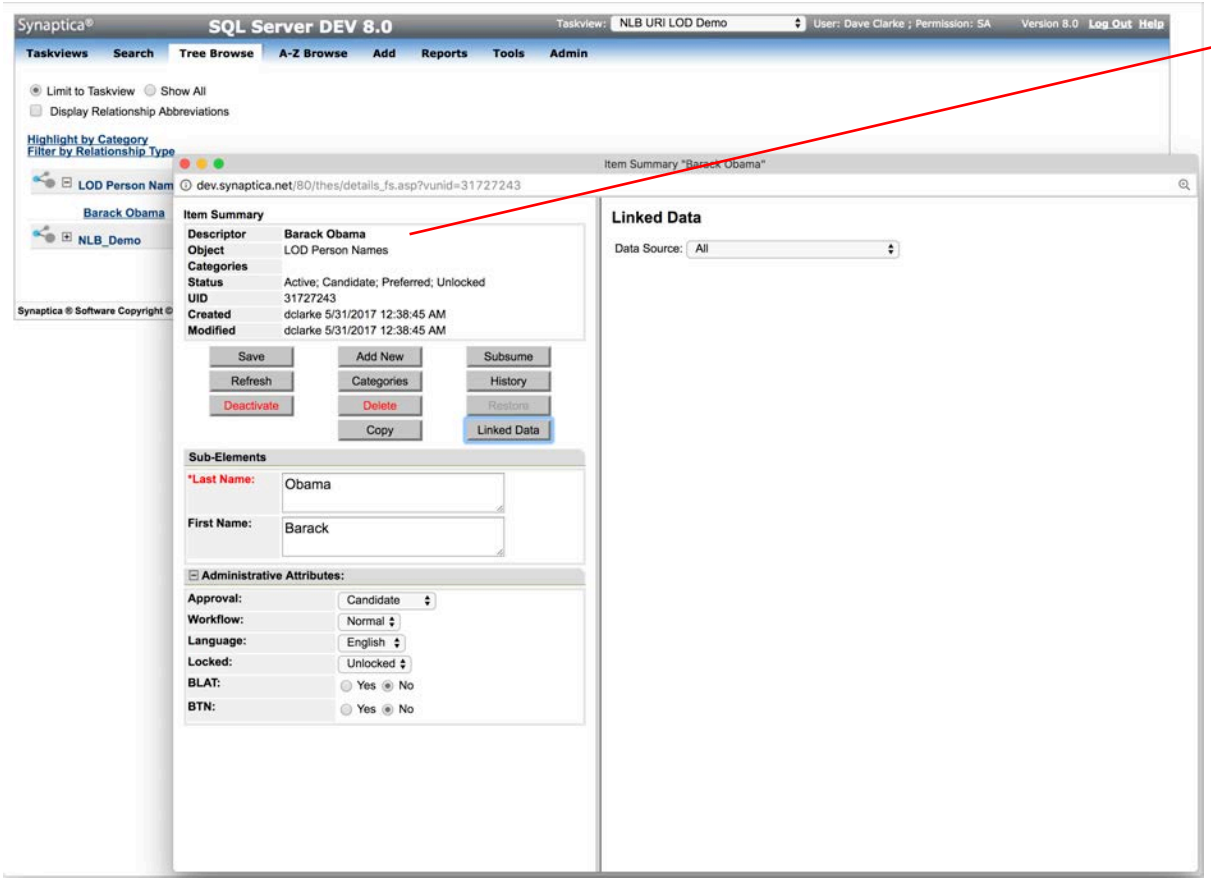


The HTTP-URI property for terms (concepts) uses the skos:Concept predicate to identify the URI and the data entry rule has been set to use an HTTP Prefix value + a GUID.

In this example a full HTTP prefix has been specified for concepts, but we could have used a namespace prefix if specified in the RDF Namespace Manager.

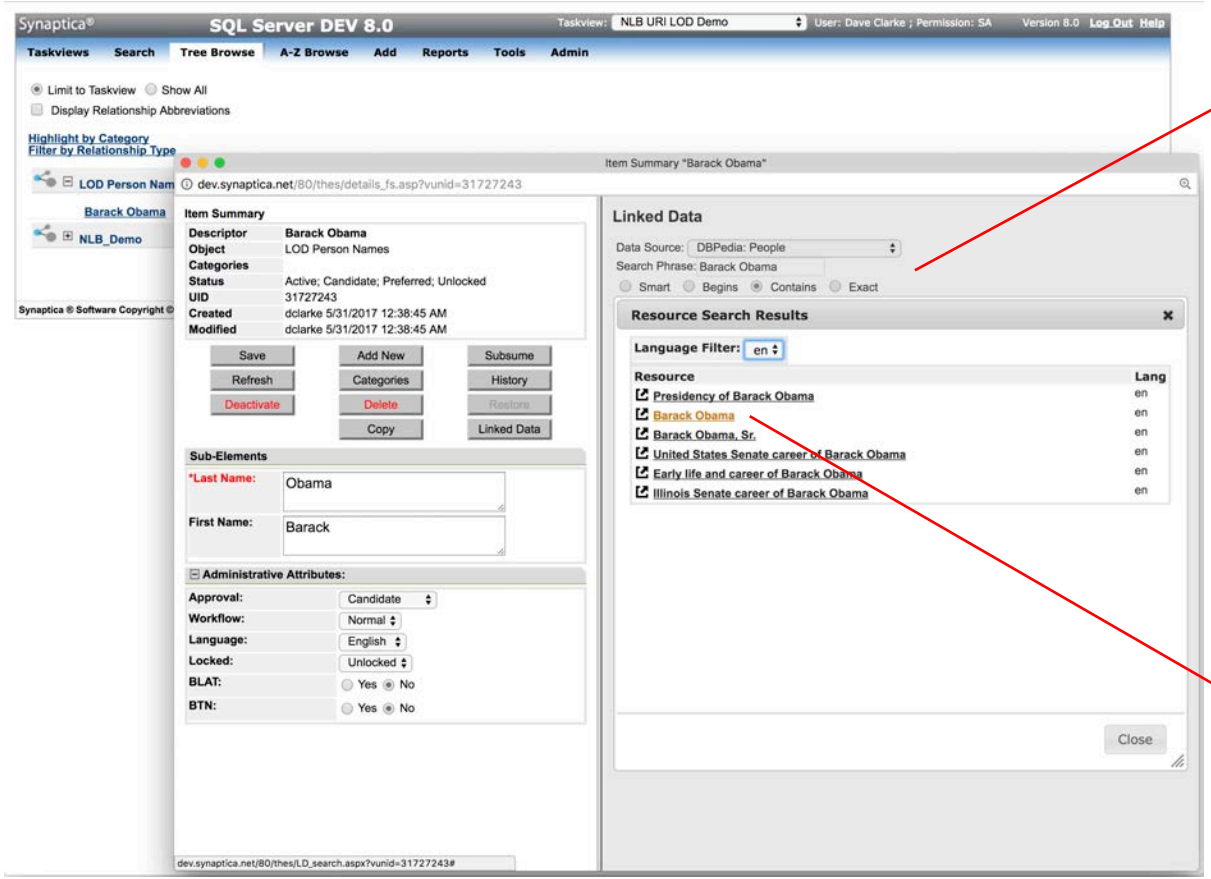
Mapping Internal Concepts to External Linked Open Data

Example of a person name record before LoD mapping



This named entity only has two fields of information in the internal taxonomy.

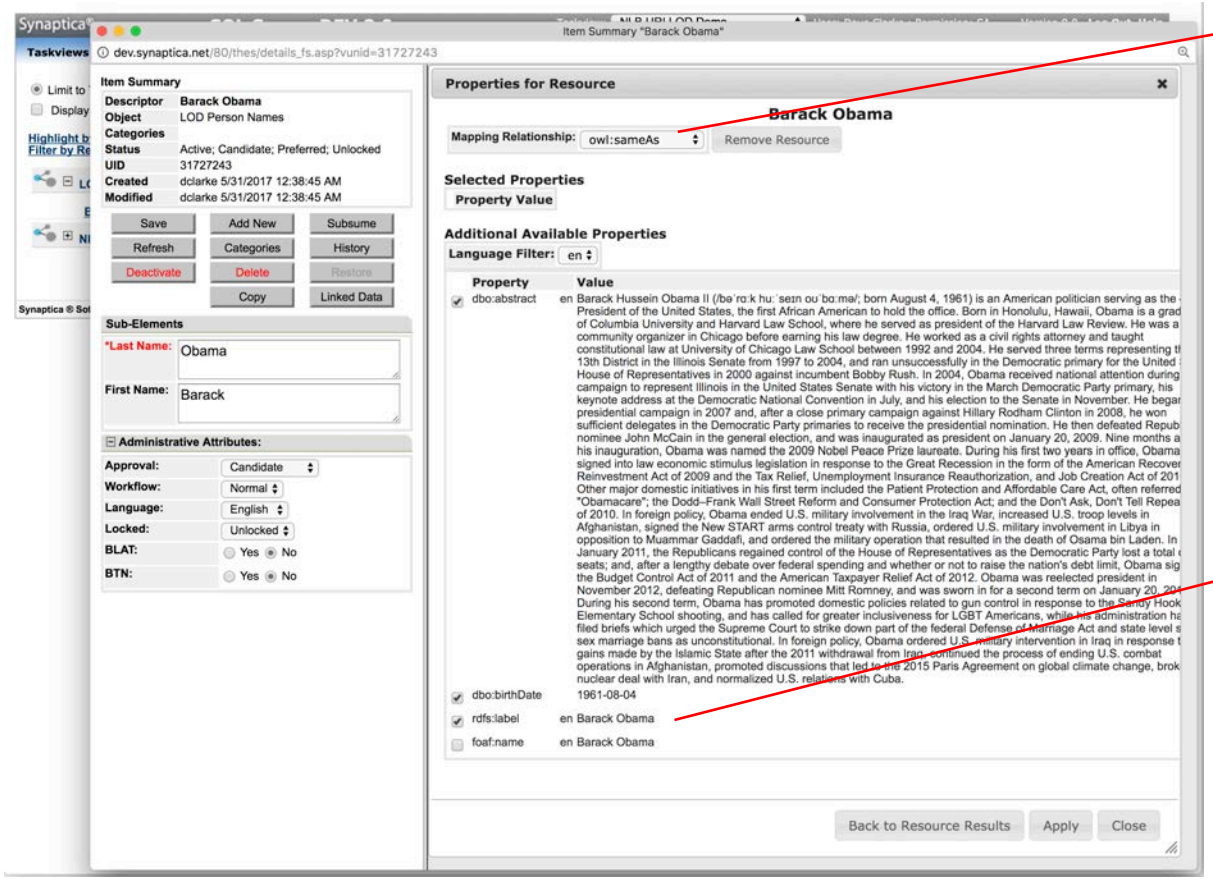
Searching LoD sources to identify a mapping target



After pressing the Linked Data button the system allows the user to choose a pre-configured data source (SPAQRL endpoint) in this case Dbpedia people, and search for matching entities. The keywords of the descriptor and non-preferred terms can automatically be used to initiate a search.

All matching entities in the LoD data source are displayed and can be previewed. Language filters are provided. In this example the second entity will be selected.

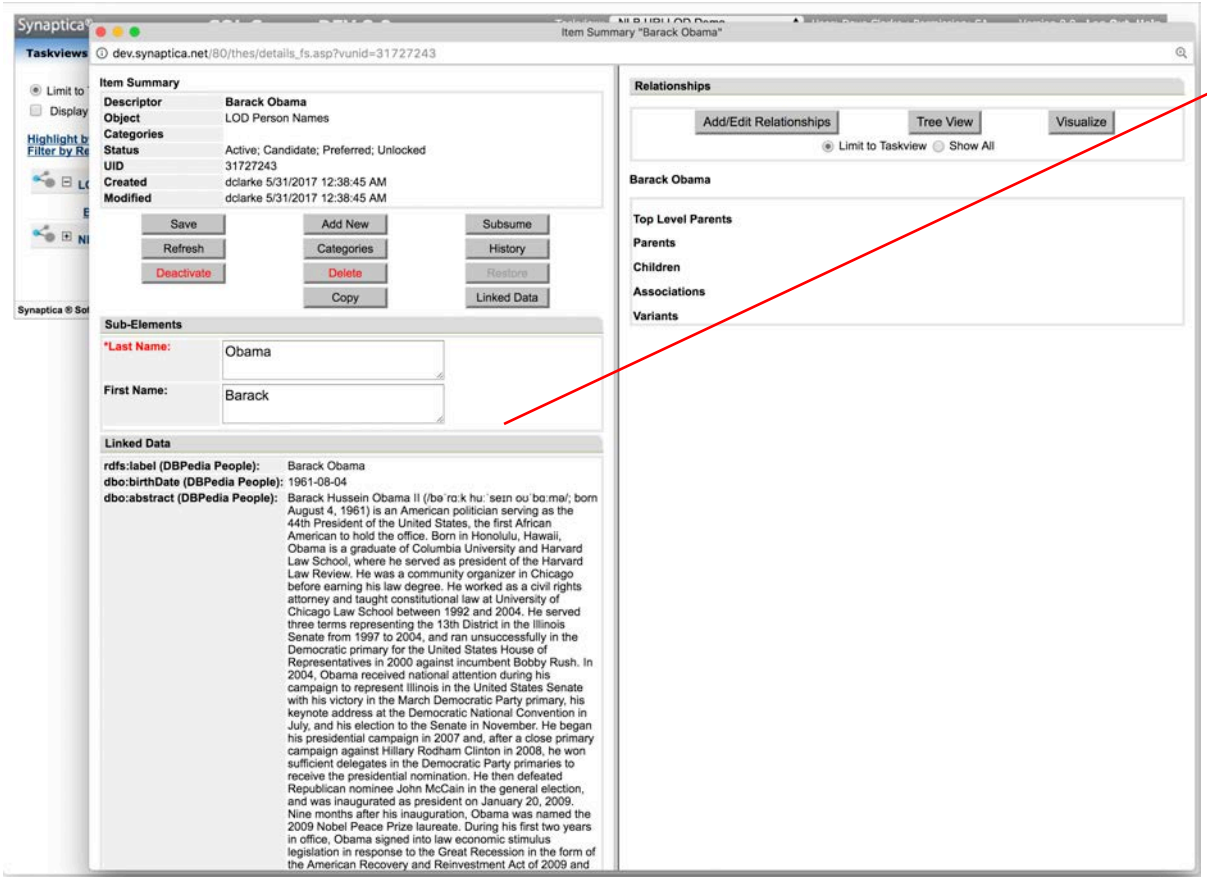
Selecting a mapping relationship and target properties



A mapping relationship is then selected – in this case 'owl:sameAs'.

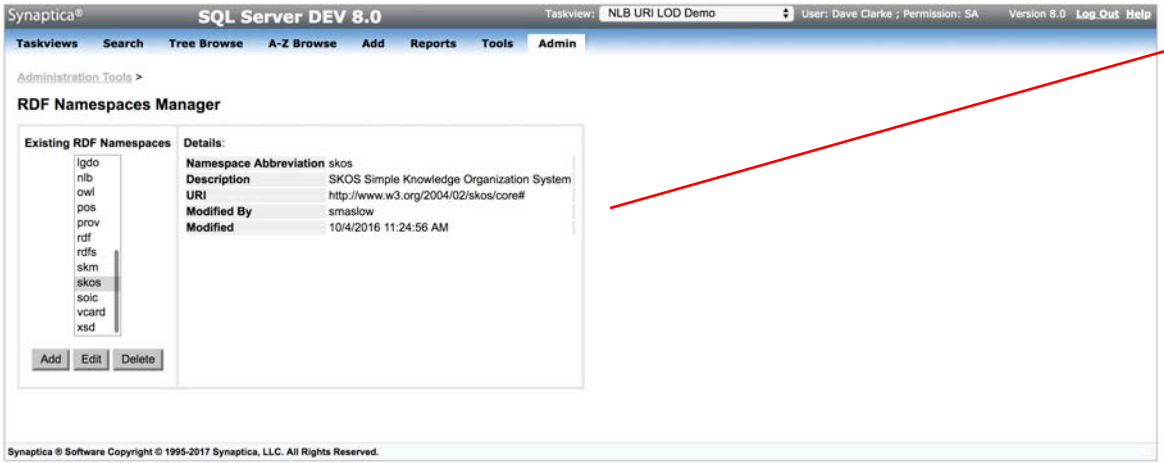
Individual properties belonging to the target entity can then be selected. When the Apply button is pressed two things happen: (i) a link is created between the internal entity and the external entity using the specified mapping relationship; (ii) the selected properties are copied (dereferenced) from the target data source and stored inside Synaptica, where they form an integral part of the internal entity record.

Viewing an internal record with mapped Linked Data



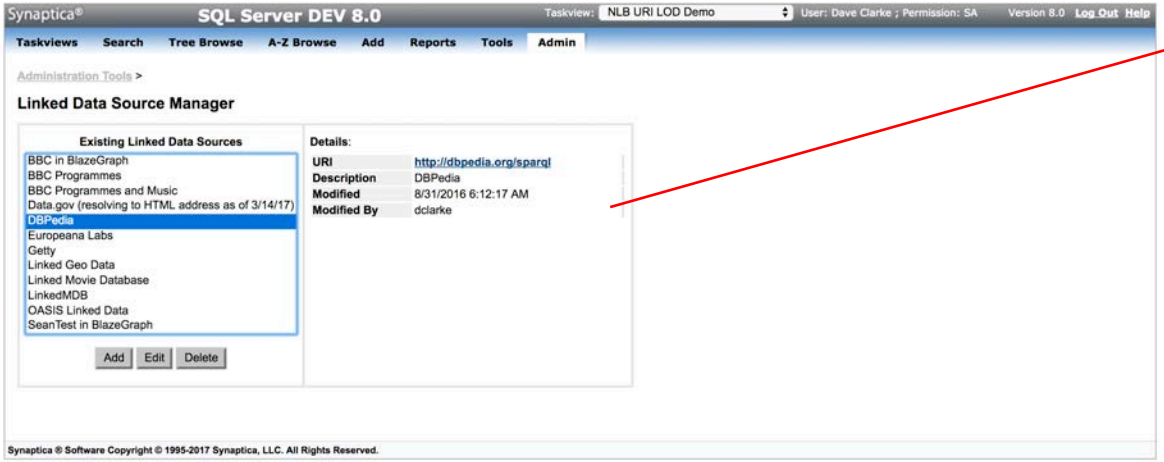
After mapping the term record of an internal entity will display both internal and external data together. In a combined view. Reports and exports can also be generated that deliver a combination of the two data sources.

RDF Namespace Manager



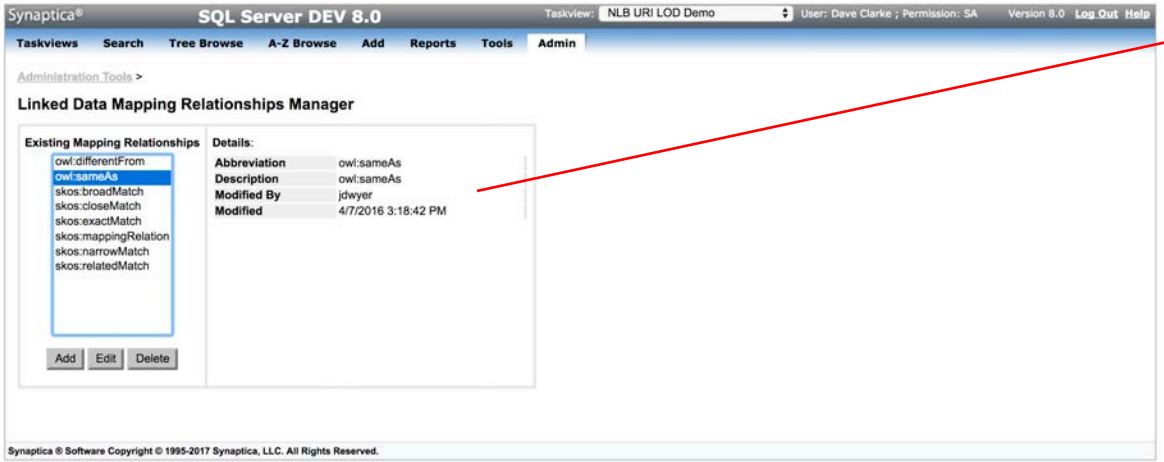
RDF Namespace Manager allows namespaces and prefixes to be created and managed in the system.

Linked Data Source Manager



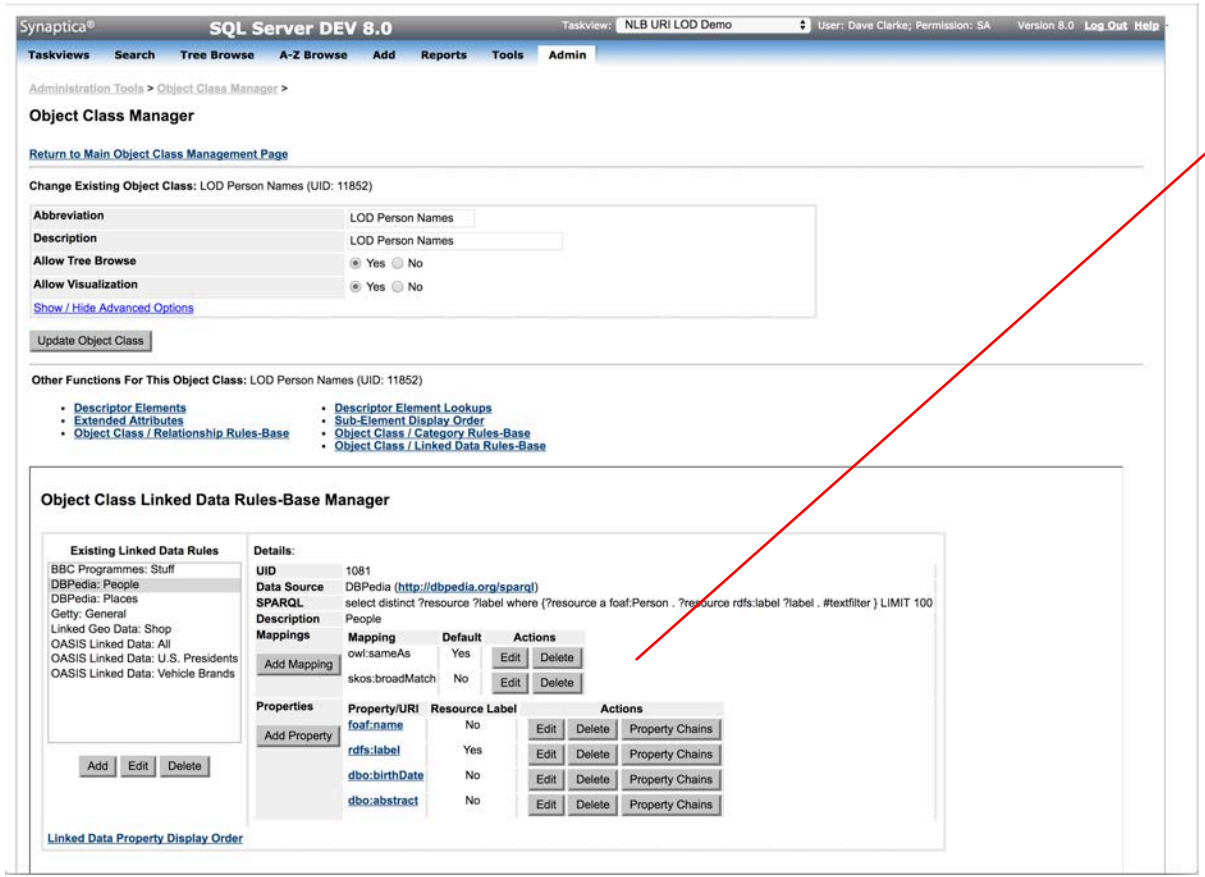
Linked Data Source Manager allows any publicly accessible SPARQL endpoint to be configured.

Linked Data Mapping Relationships Manager



Linked Data Mapping Relationships Manager allows any mapping predicate to be configured.

Concept Scheme to Linked Data Mapping Manager



The Object Class / Linked Data Rules-base link inside the Object Class Manager tool allows an Admin user to configure which Linked Data sources an internal taxonomy can link to; which mapping relationships it can use; and which specific property fields belonging to the external Linked Data source can be selected for dereferencing.