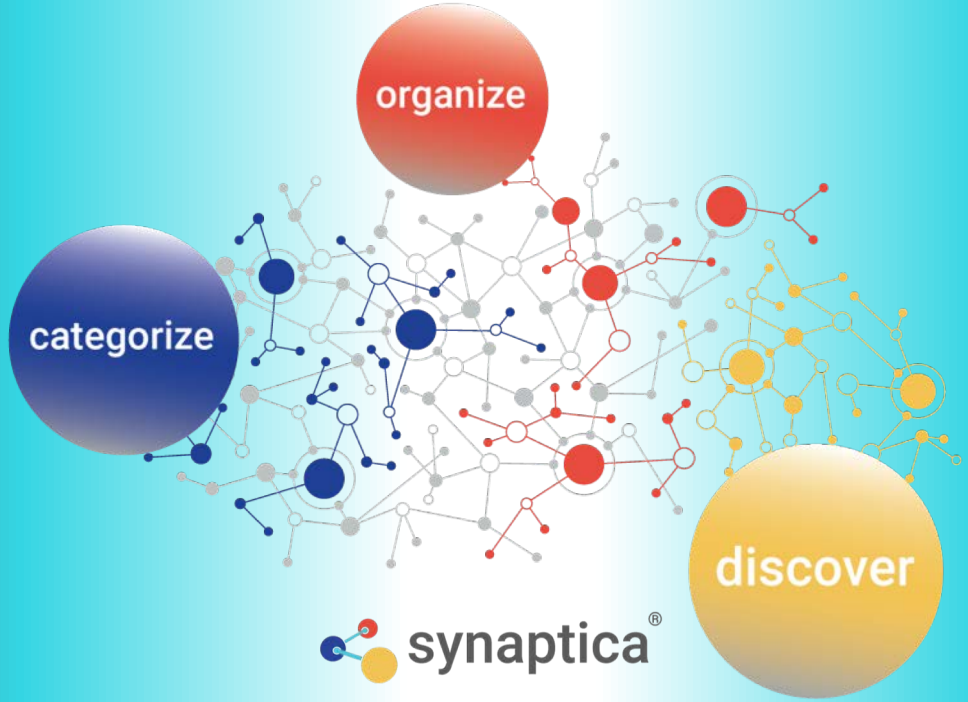
# Using Linked Data as a framework for enterprise KOS


**Dave Clarke**

CEO Synaptica

@Synaptica

@DavidClarkeBlog


**Dublin Core**

Sep. 10-13, 2018

Porto, Portugal

#dcmi18

1. What is LOD-KOS
2. Examples of open data LOD-KOS
3. What is LED
4. How are enterprises using LED and LOD
5. How do enterprises manage confidentiality

International Journal on Digital Libraries
https://doi.org/10.1007/s00799-018-0241-2

CrossMark

## Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review

Marcia Lei Zeng[1] · Philipp Mayr[2]

Received: 26 September 2017 / Revised: 3 May 2018 / Accepted: 3 May 2018


**Abstract**
Since the *Simple Knowledge Organization System* (SKOS) specification and its *SKOS eXtension for Labels* (SKOS-XL) became formal W3C recommendations in 2009, a significant number of conventional Knowledge Organization Systems (KOS) (including thesauri, classification schemes, name authorities, and lists of codes and terms, produced before the arrival of the ontology-wave) have made their journeys to join the Semantic Web mainstream. This paper uses "LOD KOS" as an umbrella term to refer to all of the value vocabularies and lightweight ontologies within the Semantic Web framework. The paper provides an overview of what the LOD KOS movement has brought to various communities and users. These are not limited to the colonies of the value vocabulary constructors and providers, nor the catalogers and indexers who have a long history of applying the vocabularies to their products. The LOD dataset producers and LOD service providers, the information architects and interface designers, and researchers in sciences and humanities, are also direct beneficiaries of LOD KOS. The paper examines a set of the collected cases (experimental or in real applications) and aims to find the usages of LOD KOS in order to share the practices and ideas among communities and users. Through the viewpoints of a number of different user groups, the functions of LOD KOS are examined from multiple dimensions. This paper focuses on the LOD dataset producers, vocabulary producers, and researchers (as end-users of KOS).

**Keywords** Linked Open Data · Knowledge Organization Systems · LOD KOS functions · Personas

### 1 Introduction

Conventional knowledge organization systems (including thesauri, classification schemes, taxonomies, subject heading systems, name authorities, and lists of codes and terms, produced before the arrival of the ontology-wave) have always been quick adapters of new technologies in their publishing venues and applications. They have had timely appearances in the earliest indexing and abstracting (I&A) databases, online information services, CD-ROMs, Adobe PDF files, HTML Web sites, and XML databases since the 1950s. Recently, they have made their journeys to join the Semantic Web mainstream and turned their products into Linked Open Data (LOD) datasets, along with ontologies that have been developed in the twenty-first century.

The *Simple Knowledge Organization System* (SKOS) and *SKOS eXtension for Labels* (SKOS-XL) became formal W3C recommendations in 2009, as a separate, lightweight, intuitive language for developing and sharing new knowledge organization systems. SKOS may be used on its own, or in combination with formal knowledge representation languages such as the Web Ontology Language (OWL) [30]. Eight years later, by the end of 2017, there are over 1000 valid LOD KOS datasets registered in the DataHub,[1] while many LOD KOS services also exist. The KOS products that have adopted the LOD approach using the standardized data model syntax recommended by SKOS and OWL can be found in a variety of domains and formats, from general-purpose to specialized domains, from monolingual to multilingual, from classification systems, thesauri, taxonomies to name-authority files, from extracted portions or a complete version of an original vocabulary to the end-products that are made

✉ Marcia Lei Zeng
mzeng@kent.edu

✉ Philipp Mayr
philipp.mayr-schlegel@gesis.org

1 School of Information, Kent State University, Ohio, USA

2 GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

[1] https://old.datahub.io/dataset.

Published online: 25 May 2018

Springer

https://doi.org/10.1007/s00799-018-0241-2

*LOD-KOS is 'all the value vocabularies and lightweight ontologies within the semantic framework.'*

- **LOD:** Linked Open Data – open data structured as RDF triples (subject-predicate-object) in which entities and predicates are uniquely identified by HTTP URIs
- **KOS**: Knowledge Organisation Systems – conceptual and named entity terminologies organised within a structured scheme using hierarchical, associative and equivalency relationships
- **Semantic framework**: family of W3C specifications for KOS and semantic resource description including RDF, SKOS, SKOS-XL, OWL, Linked Data Platform
- **Value vocabularies**: descriptive terminologies of concepts and named entities, e.g. lexicons, thesauri, taxonomies, classification schemes and name authorities
- **Property vocabularies**: sets of metadata elements or property types including the data properties and object properties (i.e. predicates) of an ontology
- **Lightweight ontologies:** KOS that use selections of ontological classes and properties derived from one or more property vocabulary to fit the needs of a specific community (as contrasted with reference ontologies containing semantic and axiomatic rules)

### 2.2 Features of a LOD KOS vocabulary

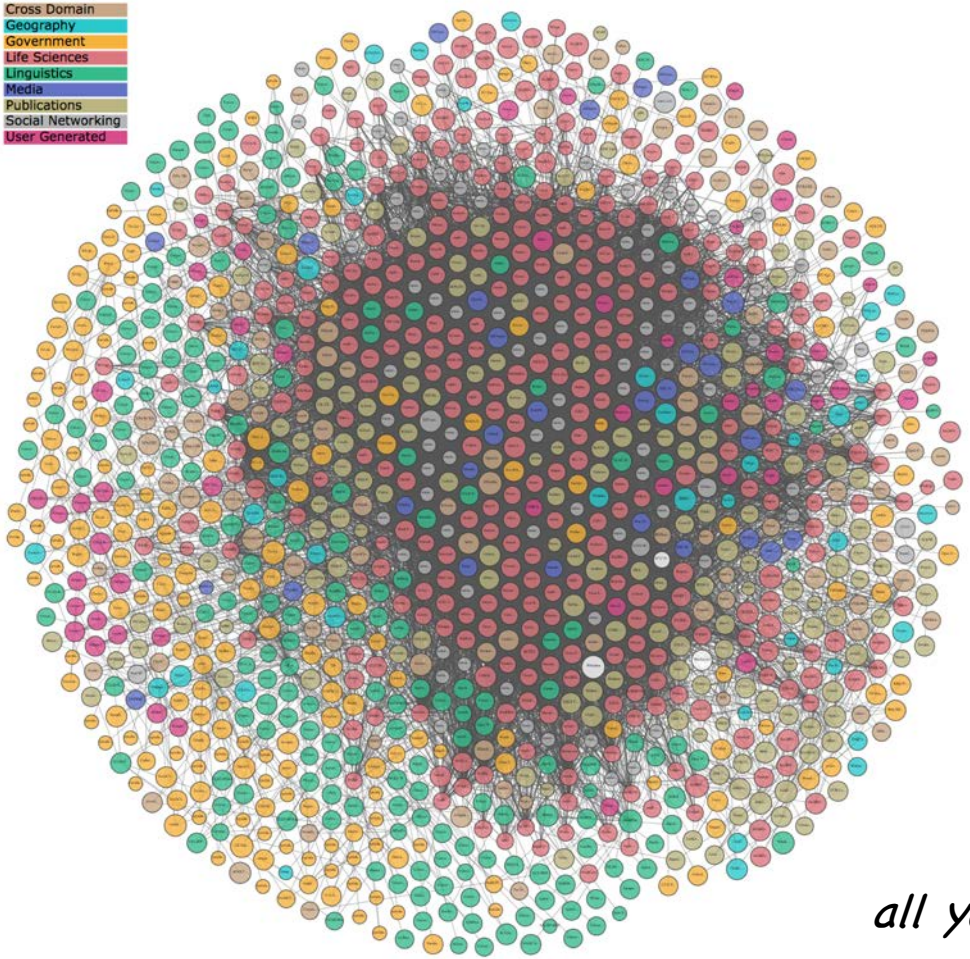
A LOD KOS vocabulary must follow the principles of Linked Data [4] and must be openly available. The SKOS data model views a KOS as a *concept scheme* comprising a set of *concepts* [30], where each *concept* must be named by a URI (Uniform Resource Identifier) or IRI (Internationalized Resource Identifier). Using a unique identifier to represent an entity or resource is one of the basic solutions for providing machine-processable, disambiguated data. Furthermore, HTTP URIs should be used when releasing a dataset as LOD.

Data of a LOD KOS are expressed as RDF triples and may be encoded using any concrete RDF syntax such as RDF/XML, Turtle, TriG, N-Quads, and JSON-LD, allowing the data to be passed between computer applications in an interoperable way, enabling a KOS to be used in distributed, decentralized metadata applications.

A LOD KOS end-product may be available as a RDF data dump or accessed through a SPARQL endpoint. Templates for forming SPARQL queries, visualized relationships, on-the-fly mapping/matching services, and other innovative delivery methods may also enrich the presence of LOD KOS on the Web.

- **Individual Vocabulary Providers**: e.g. EuroVoc multilingual thesaurus of the EU http://eurovoc.europa.eu/
- **Individual Institution Providers**: e.g. Library of Congress Data Services comprising LC Subject Headings, LC Classification, LC Name Authority File http://id.loc.gov/, or Getty Vocabularies LOD comprising the Art & Architecture Thesaurus (AAT), the Getty Thesaurus of Geographic Names (TGN), and the Union List of Artist Names (ULAN) http://vocab.getty.edu/
- **National Portals** and **Domain Portals**: e.g. http://finto.fi/en/ and http://www.bioportal.bioontology.org
- **Registries**: e.g. BARTOC (Basel Register of Thesauri, Ontologies & Classifications) over 2740 KOS, 315 of these are available in RDF format https://bartoc.org/ and LOV (Linked Open Vocabularies) over 600 registered vocabularies; many of the vocabularies are property vocabularies http://lov.okfn.org/dataset/lov

Linked Open Data cloud

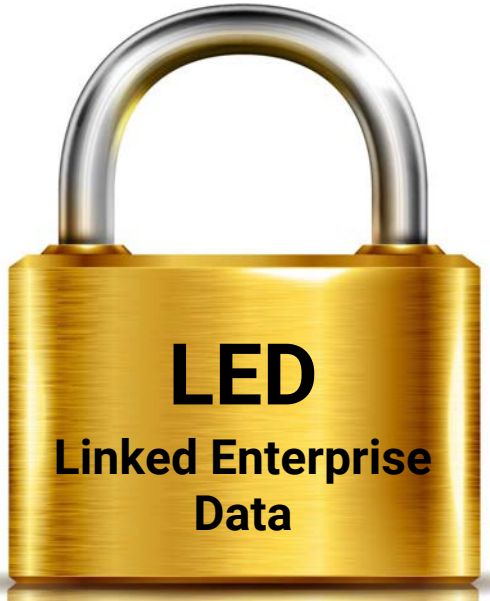



big and growing

*all you need is L♥D*

What about using Linked Data as a framework for enterprise KOS?


shared openly




behind the firewall


## Use of Standards

- 39% use CV standards like ISO 25964
- 21% use SKOS
- 3% use SKOS-XL
- 5% use OWL
- 24% use Linked Data Platform
- 36% use metadata like Dublin Core
- 12% use Schema.Org
- 12% use LOD as reference resources

## Public v. Private

- 52% behind firewall
- 36% partially exposed
- 12% made available as LOD

## Expectations

- 57% Linked Data impacting their business
- 51% Semantic Web impacting their business

Survey open from pinned Tweet at: https://twitter.com/DavidClarkeBlog

Pharmaceuticals

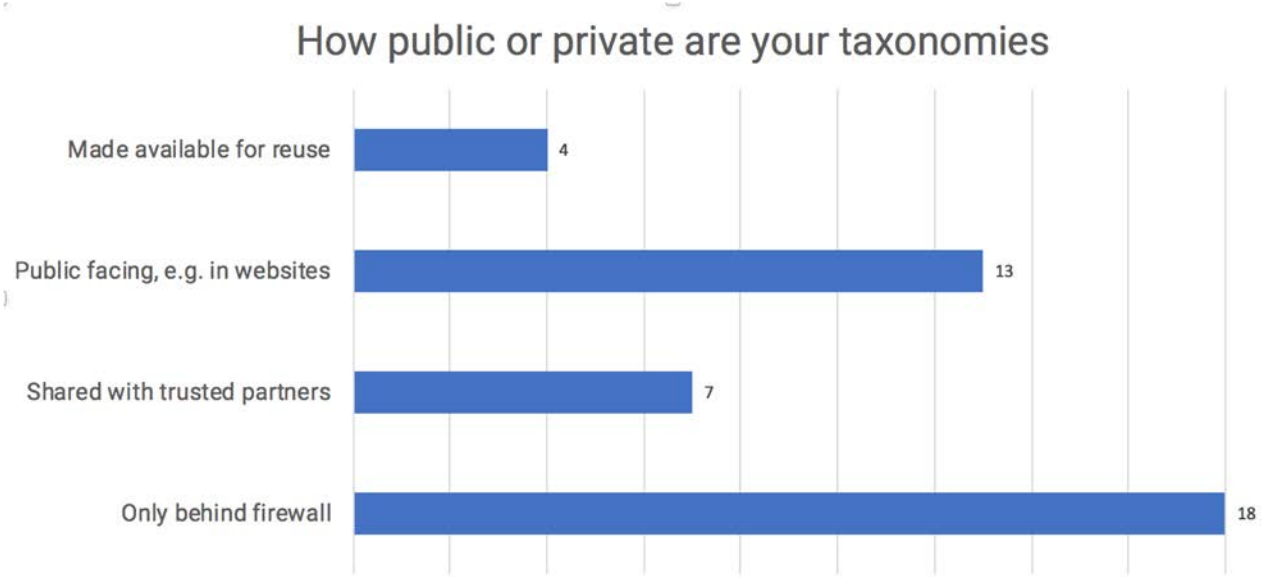
Media & Games

Academic Publishing

+

Consumer Electronics

Healthcare and Life Sciences


How public or private are your taxonomies
Made available for reuse
4
Public facing, e.g. in websites
13
Shared with trusted partners
7
Only behind firewall
18

1. Flexible compartmentalization
2. Lightweight ontologies
3. Mix-and-match property vocabularies
4. Ease-of-use value vocabulary editing
5. Data portability

## Property Vocabularies

## Value Vocabularies

### Predicates:

**links**
connecting a thing to another thing (object properties), or a thing to a string (data properties)

### Classes:

**types**
of individuals that are grouped together because they share a common set of properties

### Schemes:

**sets**
of concepts or named entities that belong together because they describe a domain of knowledge

### Concepts:

**individuals**
(concepts or names) defined by their classes and properties and linked together within a scheme

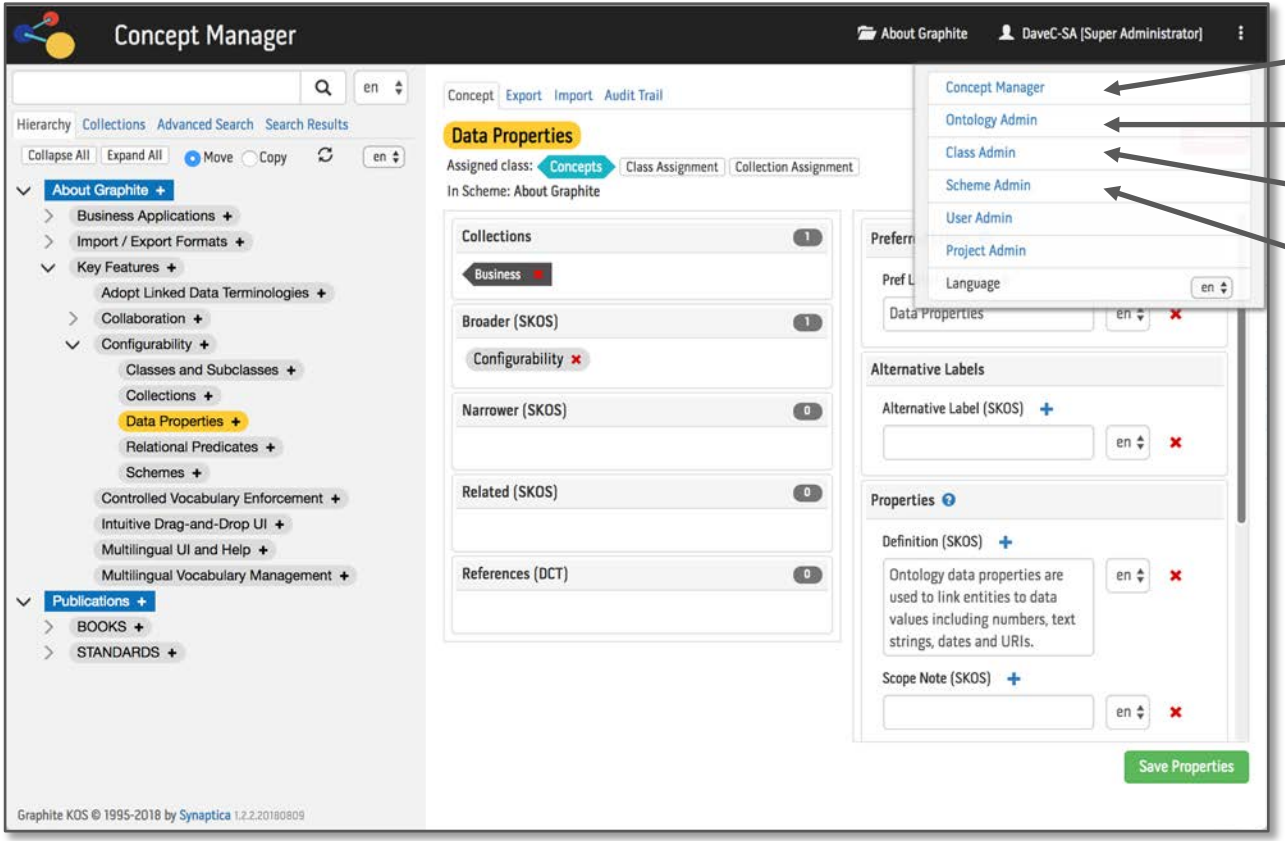
### Ontology

Coherent sets of predicates and classes defining the logical structure of a domain

### Taxonomy

Coherent sets of terminology defining the entities within a domain

## Knowledge Organization Systems




**Concepts**

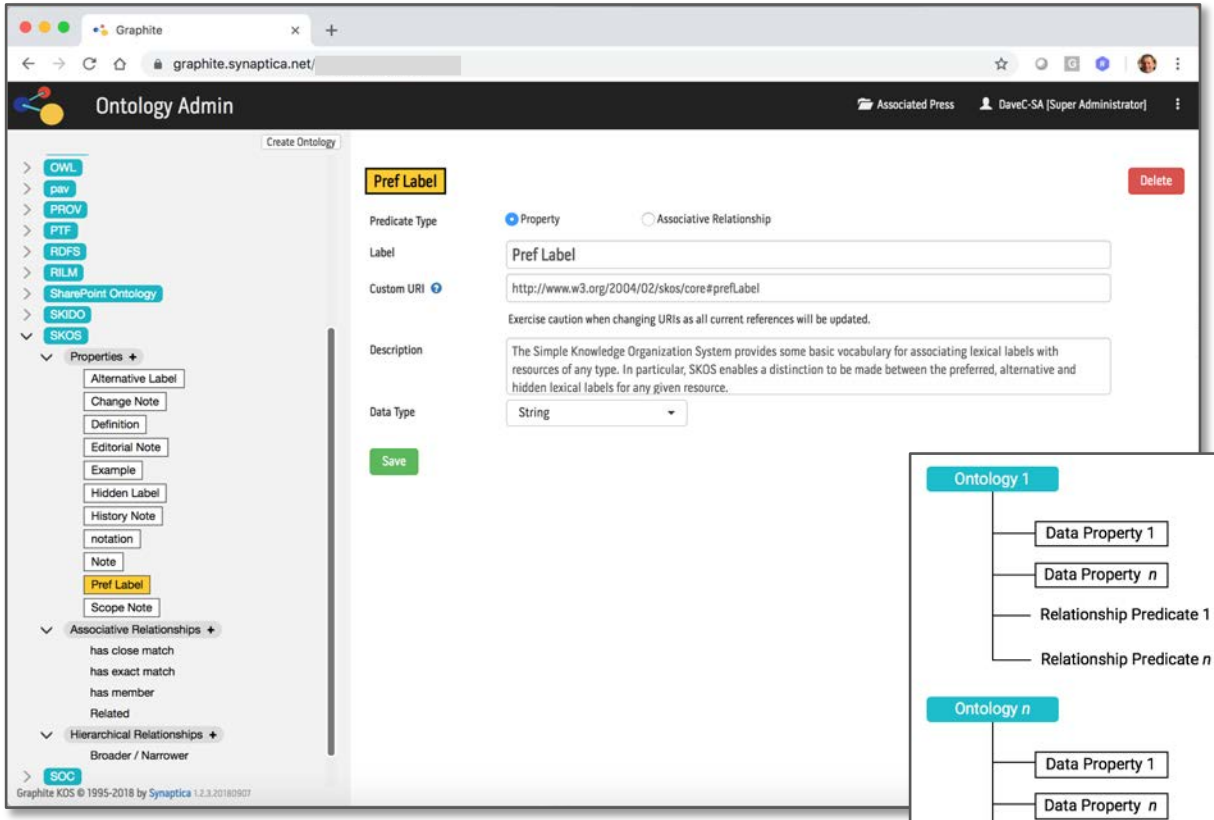
**Predicates**

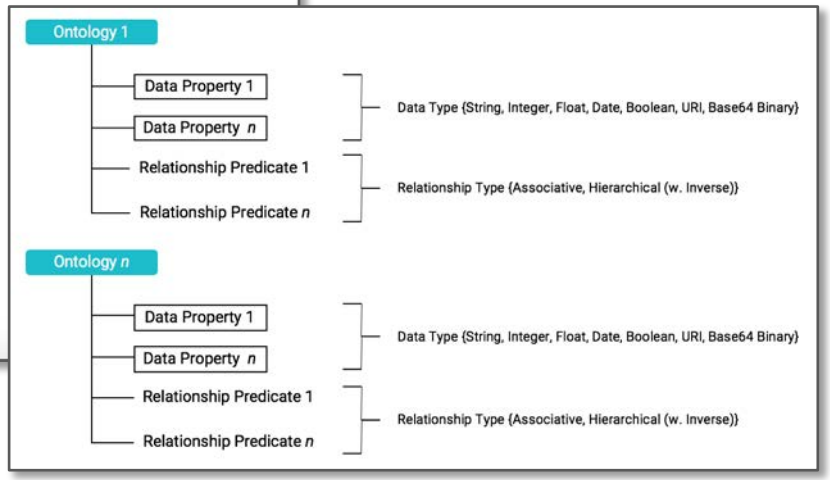
**Classes**

**Schemes**

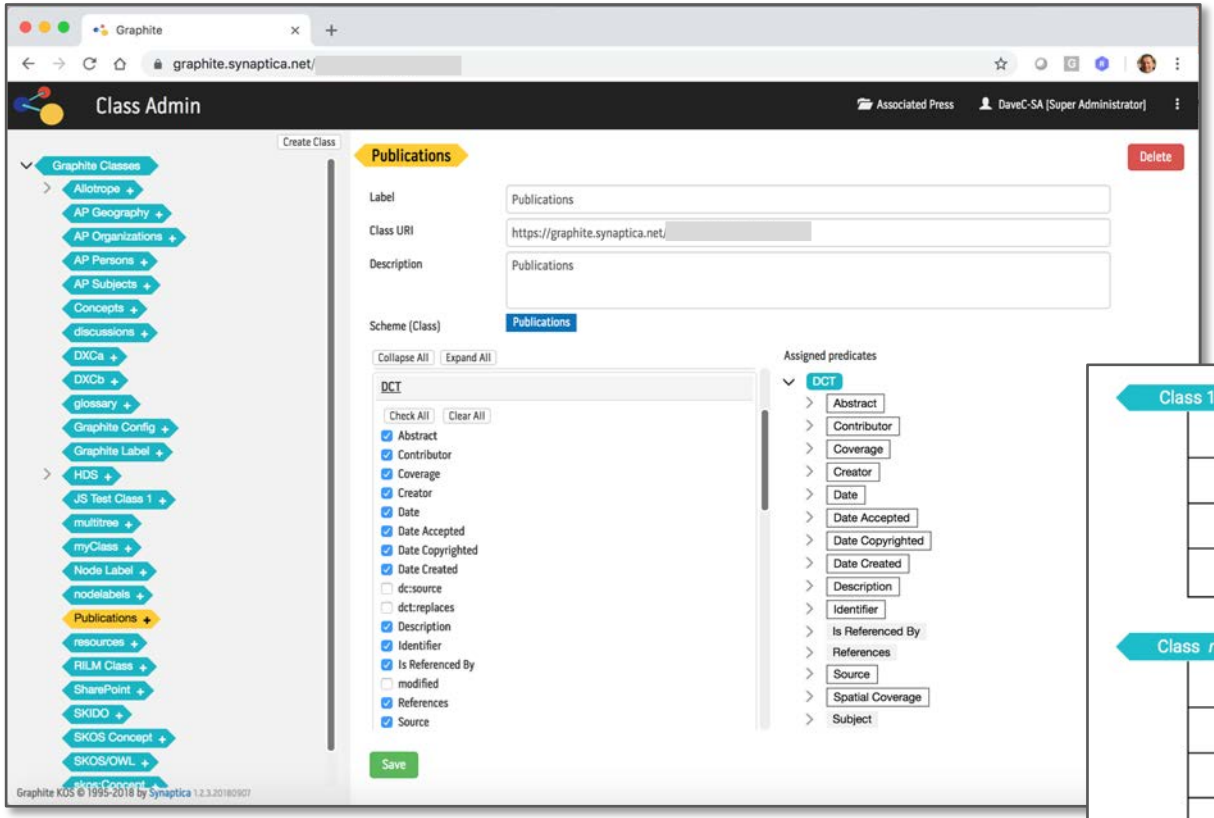
1. Manage a library of proprietary and/or public domain ontology classes and predicates
2. Design KOS schemes using plug-and-play ontology classes and predicates
3. Create and curate descriptive taxonomies that conform with the business rules of each scheme

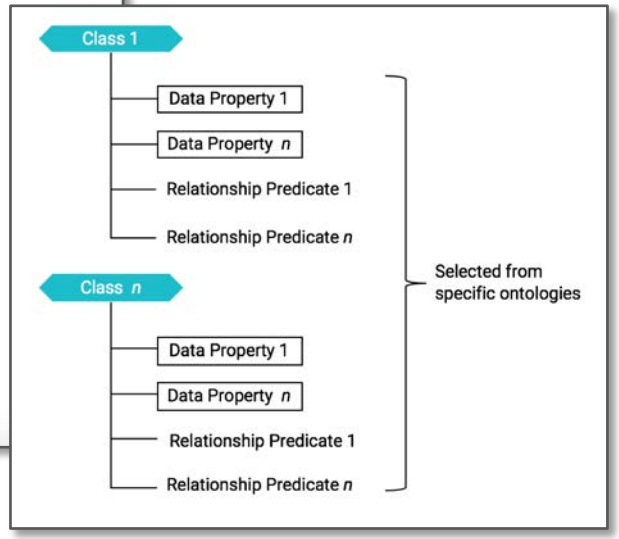
# Predicates


Graphite
graphite.synaptica.net/
Ontology Admin
Associated Press
DaveC-SA [Super Administrator]
Create Ontology
OWL
pav
PROV
PTF
RDFS
RILM
SharePoint Ontology
SKIDO
SKOS
Properties
Alternative Label
Change Note
Definition
Editorial Note
Example
Hidden Label
History Note
notation
Note
Pref Label
Scope Note
Associative Relationships
has close match
has exact match
has member
Related
Hierarchical Relationships
Broader / Narrower
SOC
Graphite KOS © 1995-2018 by Synaptica 1.2.3.20180907
Pref Label
Delete
Predicate Type
Property
Associative Relationship
Label
Pref Label
Custom URI
http://www.w3.org/2004/02/skos/core#prefLabel
Exercise caution when changing URIs as all current references will be updated.
Description
The Simple Knowledge Organization System provides some basic vocabulary for associating lexical labels with resources of any type. In particular, SKOS enables a distinction to be made between the preferred, alternative and hidden lexical labels for any given resource.
Data Type
String
Save


Ontology 1
Data Property 1
Data Property n
Data Type {String, Integer, Float, Date, Boolean, URI, Base64 Binary}
Relationship Predicate 1
Relationship Predicate n
Relationship Type {Associative, Hierarchical (w. Inverse)}
Ontology n
Data Property 1
Data Property n
Data Type {String, Integer, Float, Date, Boolean, URI, Base64 Binary}
Relationship Predicate 1
Relationship Predicate n
Relationship Type {Associative, Hierarchical (w. Inverse)}

# Classes


Graphite
graphite.synaptica.net/
Class Admin
Associated Press
DaveC-SA [Super Administrator]
Create Class
Graphite Classes
Allotrope
AP Geography
AP Organizations
AP Persons
AP Subjects
Concepts
discussions
DXCa
DXCb
glossary
Graphite Config
Graphite Label
HDS
JS Test Class 1
multitree
myClass
Node Label
nodelabels
Publications
resources
RILM Class
SharePoint
SKIDO
SKOS Concept
SKOS/OWL
Publications
Delete
Label
Publications
Class URI
https://graphite.synaptica.net/
Description
Publications
Scheme (Class)
Publications
Collapse All
Expand All
DCT
Check All
Clear All
Abstract
Contributor
Coverage
Creator
Date
Date Accepted
Date Copyrighted
Date Created
dc:source
dct:replaces
Description
Identifier
Is Referenced By
modified
References
Source
Assigned predicates
DCT
Abstract
Contributor
Coverage
Creator
Date
Date Accepted
Date Copyrighted
Date Created
Description
Identifier
Is Referenced By
References
Source
Spatial Coverage
Subject
Save
Graphite KOS © 1995-2018 by Synaptica 1.2.3.20180907


Class 1
Data Property 1
Data Property n
Relationship Predicate 1
Relationship Predicate n
Class n
Data Property 1
Data Property n
Relationship Predicate 1
Relationship Predicate n
Selected from specific ontologies

# Schemes

Graphite
graphite.synaptica.net/

Scheme Admin
Associated Press
DaveC-SA [Super Administrator]

Create Scheme
About Graphite
AP Geography
AP Organizations
AP Persons
AP Subjects
Business Models

**AP Subjects**

Scheme Label: AP Subjects
Scheme URI: https://graphite.synaptica.net/
Description
Classes: Concepts | AP Subjects | Class Assignment
Concepts Base URI
Type of URI: GUID / Graphite ID
Total Concepts: 4189
Enforce uniqueness for Alt Label

Properties | Relationships | Property Paths | Linked Data

Collapse All | Expand All

SKOS

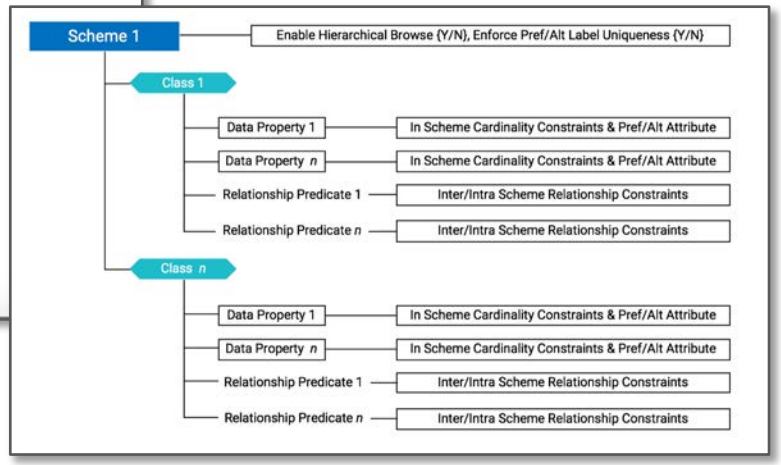
Check All | Clear All

| Properties | | Cardinality | | | | Pref Label | Alt Label |
|---|---|---|---|---|---|---|---|
| Alternative Label | | | | | | | |
| Data Type | string | Min | | Max | | | |
| Change Note | | | | | | | |
| Data Type | string | Min | 0 | Max | 0 | | |
| Definition | | | | | | | |
| Data Type | string | Min | 0 | Max | 0 | | |
| Editorial Note | | | | | | | |
| Data Type | string | Min | | Max | | | |
| Pref Label | | | | | | | |

Assigned Properties | Properties Sorting
Assigned Properties for AP Subjects
- AP Ontology
  - Date Modified
  - DBPediaURI
  - Home Page
  - Is Placeholder
  - Is Reference
- SKOS
  - Change Note
  - Definition

Save | Delete All Concepts | Delete Scheme

Graphite KOS © 1995-2018 by Synaptica 1.2.3.20180907



Enterprises have identifiable needs to filter Linked Enterprise Data for restricted access through both UI and also APIs and Endpoints:

1. By scheme
2. By collection
3. By individual term
4. By individual data property
5. By individual object property (relationship)
6. By approvals, languages, users

By all of the above

Graphite

graphite.synaptica.net/graphite/project

Project Admin

Associated Press | DaveC-SA [Super Administrator]

Create Project

About Graphite
Associated Press
Demonstration Project

Associated Press

Delete

Project Name: Associated Press

Project URI: https://graphite.synaptica.net/taskview/j9fnfw0j2z8z3

Description: Imported from KMS

Public

Select Scheme | Reset All Permissions | Set Individual User Permissions | User Collection and Relationship Exclusion | Set Public Permissions

AhrenL
DaveC-SA
DavidClarke
gene
GeneL
hltay01
JimS
JimS-Editor
Linn
shafizanuar
TayHI

Save

Check to exclude Collections

- All Concepts
- Blue
- Green
- Red

Check to exclude Relationships

- SKOS
  - Broader
  - Narrower
  - Related
- DCT
  - Is Referenced By
  - References
- AP Ontology
  - child
  - parent
  - contains
  - locatedIn
  - formerSignificantOther
  - siblingOf
  - significantOther

synaptica

Graphite

graphite.synaptica.net/graphite/project

Project Admin

Associated Press

DaveC-SA [Super Administrator]

Create Project

About Graphite

Associated Press

Demonstration Project

Associated Press

Delete

Project Name: Associated Press

Project URI: https://graphite.synaptica.net/taskview/j9fnfw0j2z8z3

Description: Imported from KMS

Public

Select Scheme | Reset All Permissions | Set Individual User Permissions | User Collection and Relationship Exclusion | Set Public Permissions

DaveC-Editor — Project Admin

DaveC-SA — Project Admin

*AP Geography* — View | Editor I | Editor II

Create/Delete Concepts | Create/Delete Relationships

| Properties | Hide | View | Edit |
|---|---|---|---|
| Change Note [SKOS] | | ● | |
| Date Modified [AP Ontology] | ● | | |
| DBPediaURI [AP Ontology] | | ● | |
| Is Placeholder [AP Ontology] | | | ● |
| Is Reference [AP Ontology] | | | ● |
| ISO 2 Code [AP Ontology] | | | ● |
| ISO 3 Code [AP Ontology] | | | ● |
| Latitude [AP Ontology] | | | ● |

Save

Graphite KOS © 1995-2018 by Synaptica 1.2.3.20180907

## Ongoing research

1. What are the advantages of using the Linked Data model?
2. What are the challenges of using the Linked Data model?
3. How public facing are enterprise KOS?
4. How do enterprises approach compartmentalization?

# Using Linked Data as a framework for enterprise KOS

# Thank You


**Dave Clarke**

CEO Synaptica

@Synaptica

@DavidClarkeBlog



**Dublin Core**

Sep. 10-13, 2018

Porto, Portugal

@DublinCore