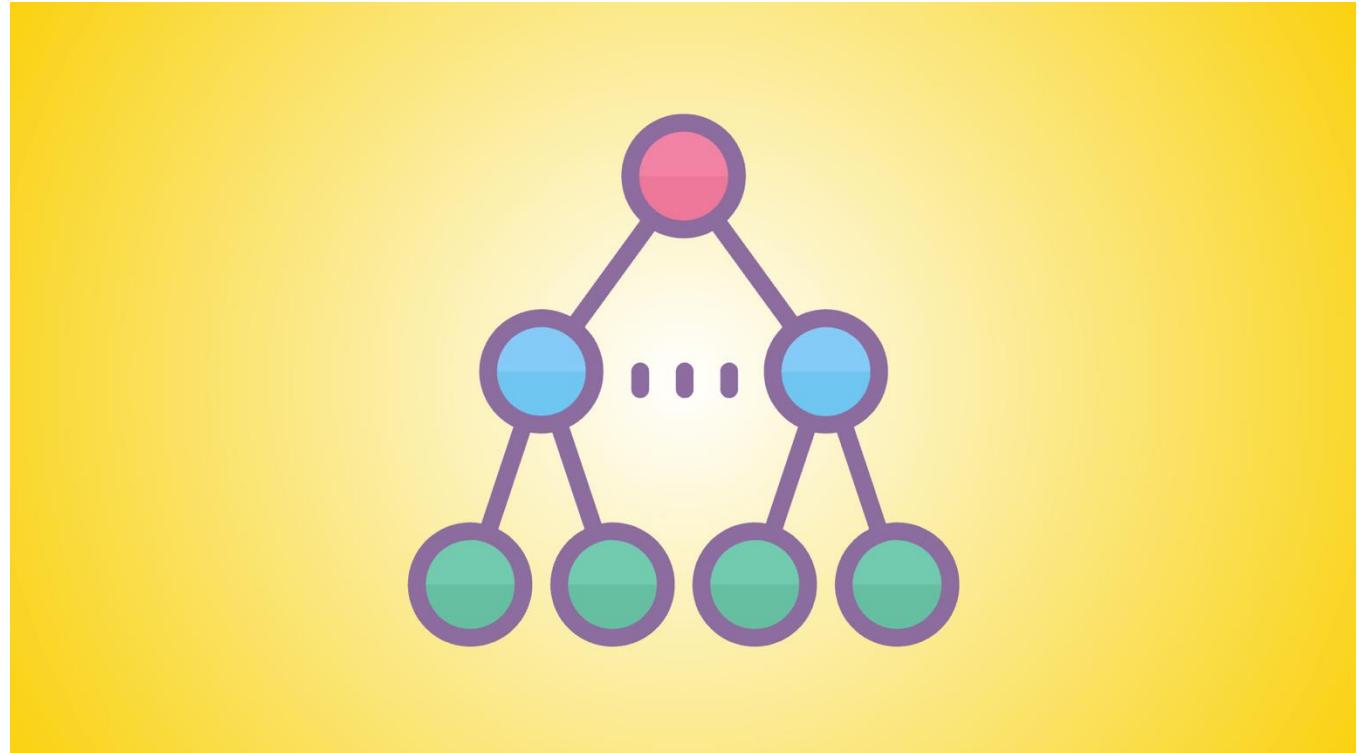


# Taxonomy and Thesaurus Development Using SKOS

---



graftilo

# 01 Introduction

# Course Audience

Modern knowledge graph-based methods for developing data taxonomies and augmented data catalogs



Metadata managers

Data governors

Data stewards

Data Management  
Professionals



Business data  
champions

Data consultants

Course  
Audience



ata

ultants



- Information architects
- Metadata architects
- Software developers



Build taxonomies & thesauri  
to drive content navigation  
structures, recommender  
systems, faceted search, data  
classification schemes, etc.

# ence



- Library science professionals
- Knowledge managers
- Information managers



Practical outlook on working with  
Semantic Web standards for  
developing taxonomies & thesauri

Develop information  
schemes & learn about  
taxonomy standards



Content managers

Document managers

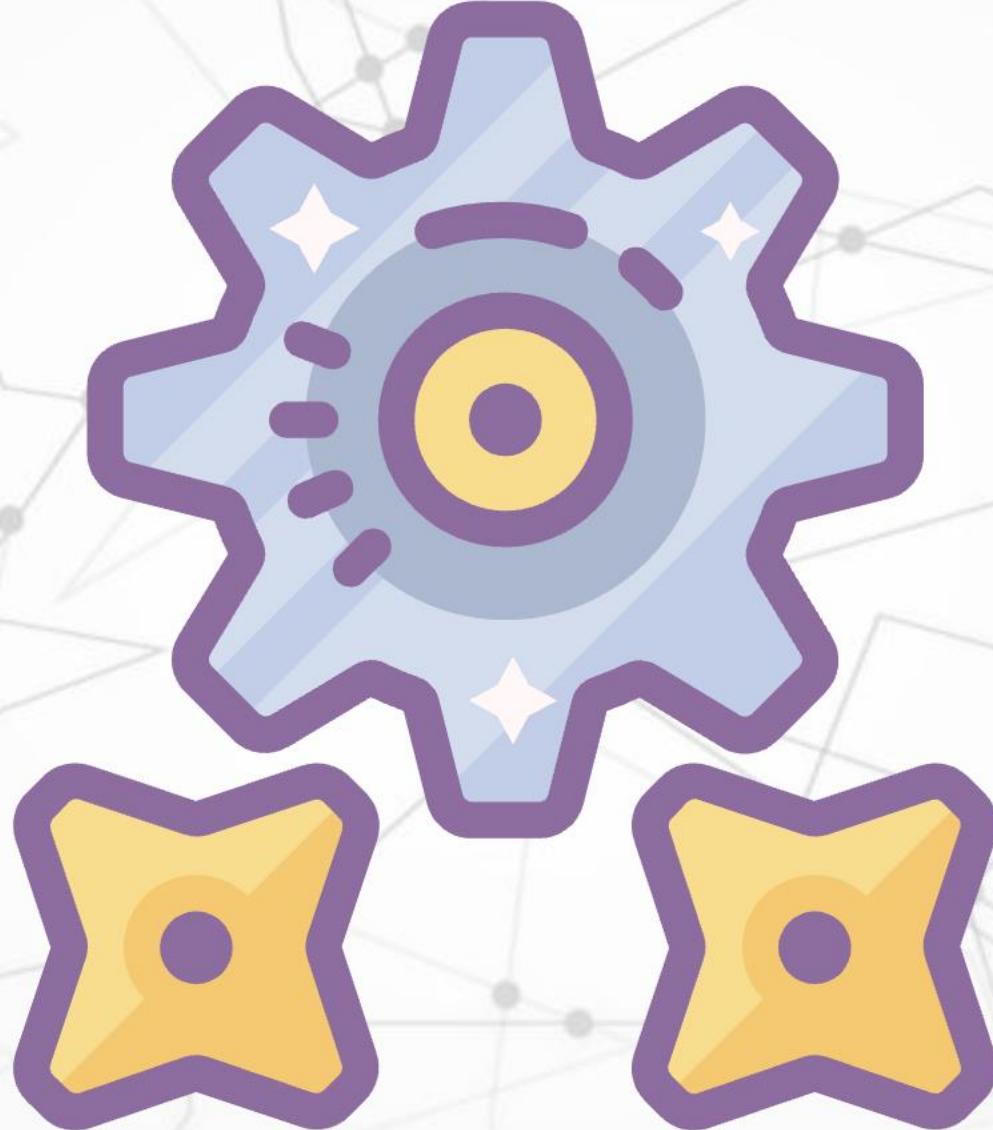


# Course Audience



## Prerequisites

- Basic RDF & SPARQL
- Basic ontology development using Web Ontology Language (OWL)



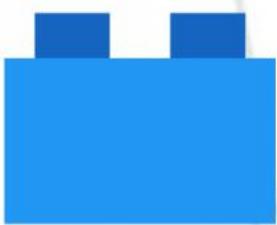
Theory and outlook on  
taxonomies & thesauri

Build taxonomies &  
thesauri based on SKOS

# Learning Outcomes



Articulate the purpose and applications of taxonomies & thesauri and recognise different types of taxonomies



Understand the building blocks provided  
in SKOS Core



Utilise SKOS Core to create taxonomies & thesauri and perform semantic curation of concepts and terms



Become familiar with using the SPARQL query language to interrogate SKOS taxonomies & thesauri



This course is not about:

- End-to-end application development
- Automating data transformation into SKOS taxonomies
- SKOS-XL
- Computational linguistics

## 02 The purpose of taxonomies and thesauri

"The spacious is trunk"

 Syntactically incorrect

"The trunk is spacious"

 Syntactically correct



"The car trunk is spacious"

✓ Syntactically correct

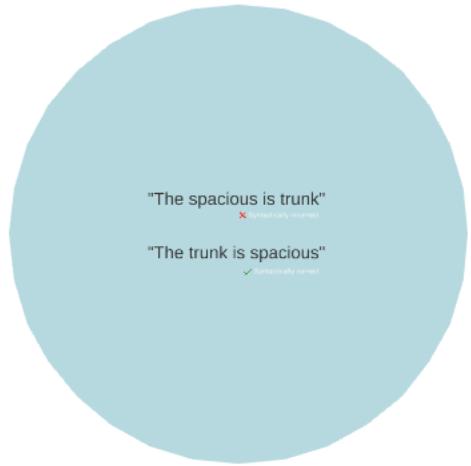


"The tree trunk is spacious"

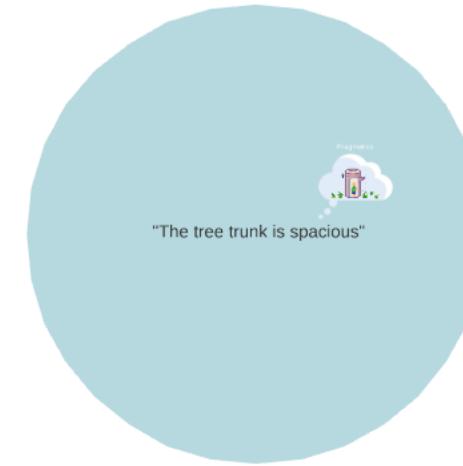
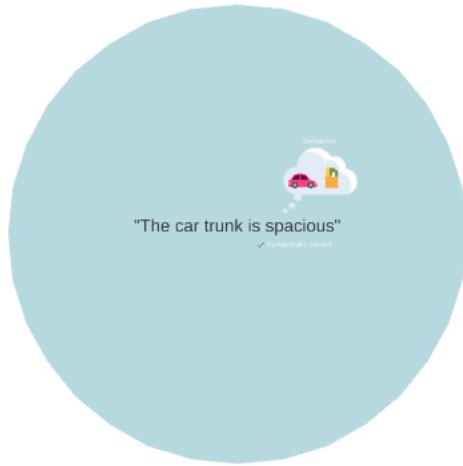
# Pragmatics



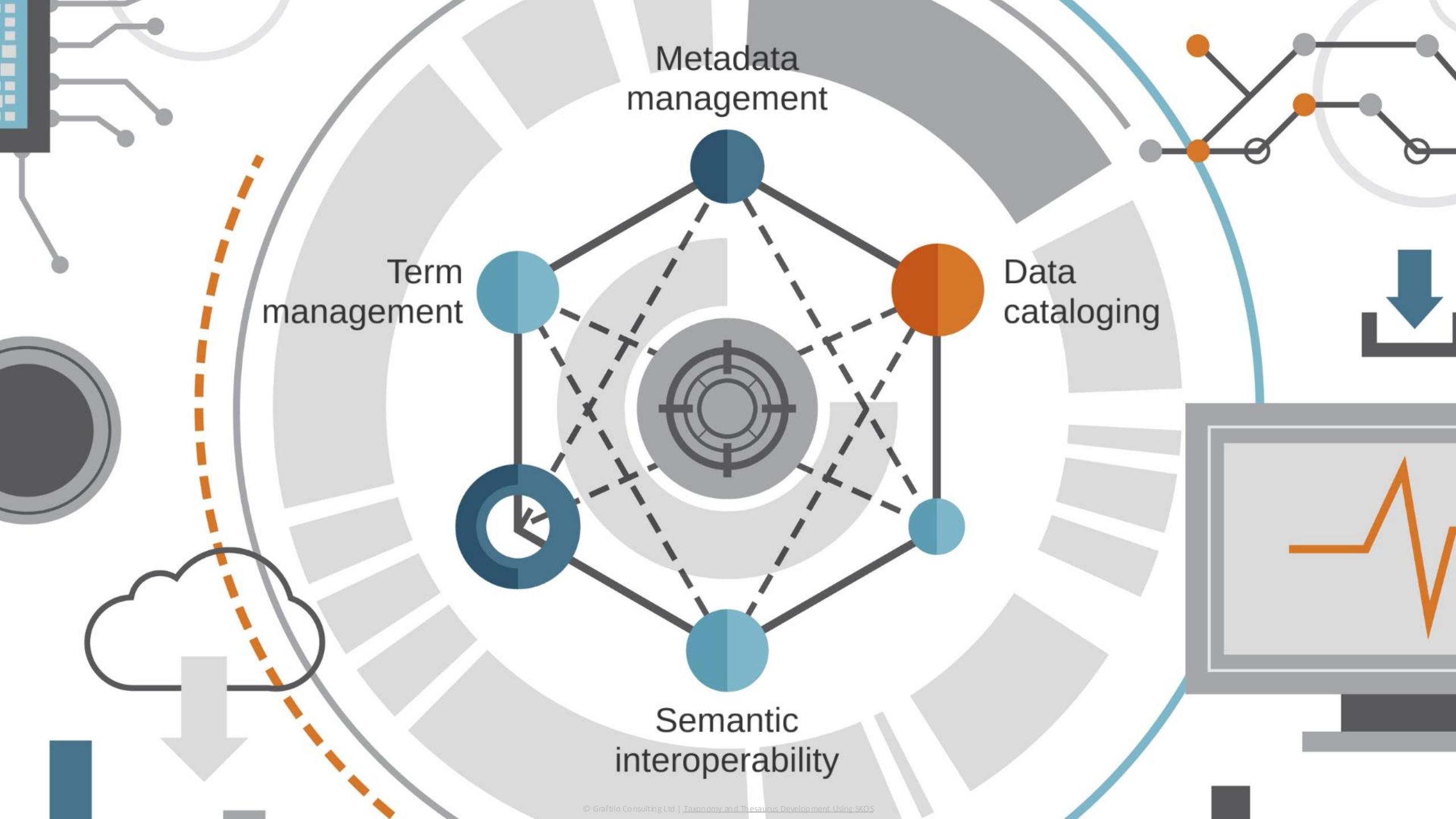
# Syntax



# Pragmatics



# Semantics



Hypernym

Hyponym

Vehicle storage compartment

Trunk



Glove box

Passenger compartment storage

type of

type of

type of

# Hypernym

# Hyponym

Vehicle

Car



Van



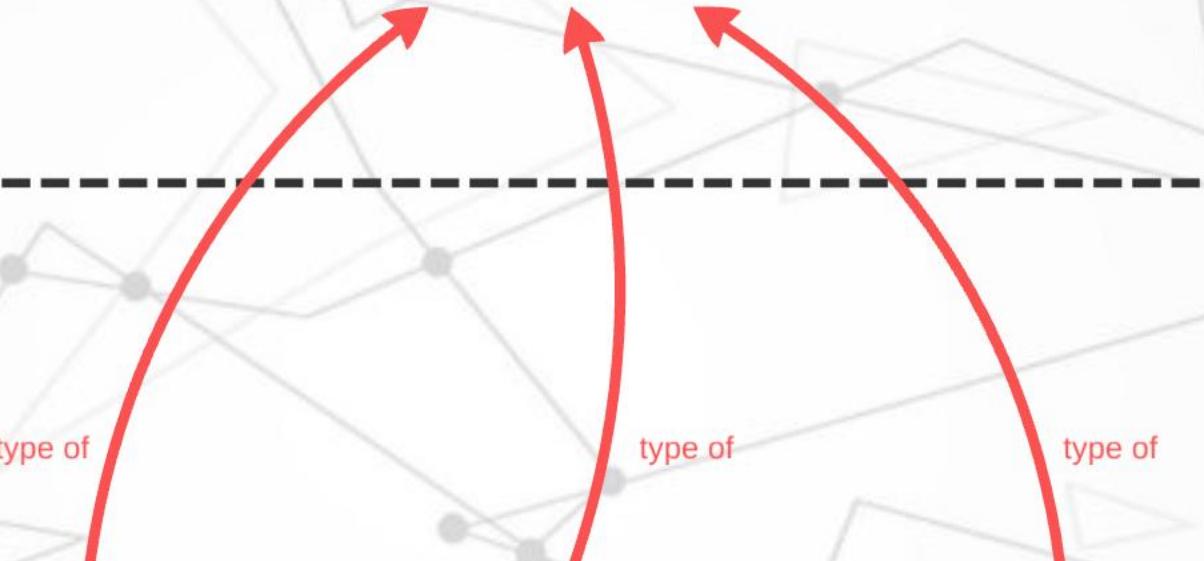
Lorry



type of

type of

type of



Hyponym

Automobile      Car

synonym of

type of



type of

# Car



Holonym

type of

# Van



type of

# Lorry



type of

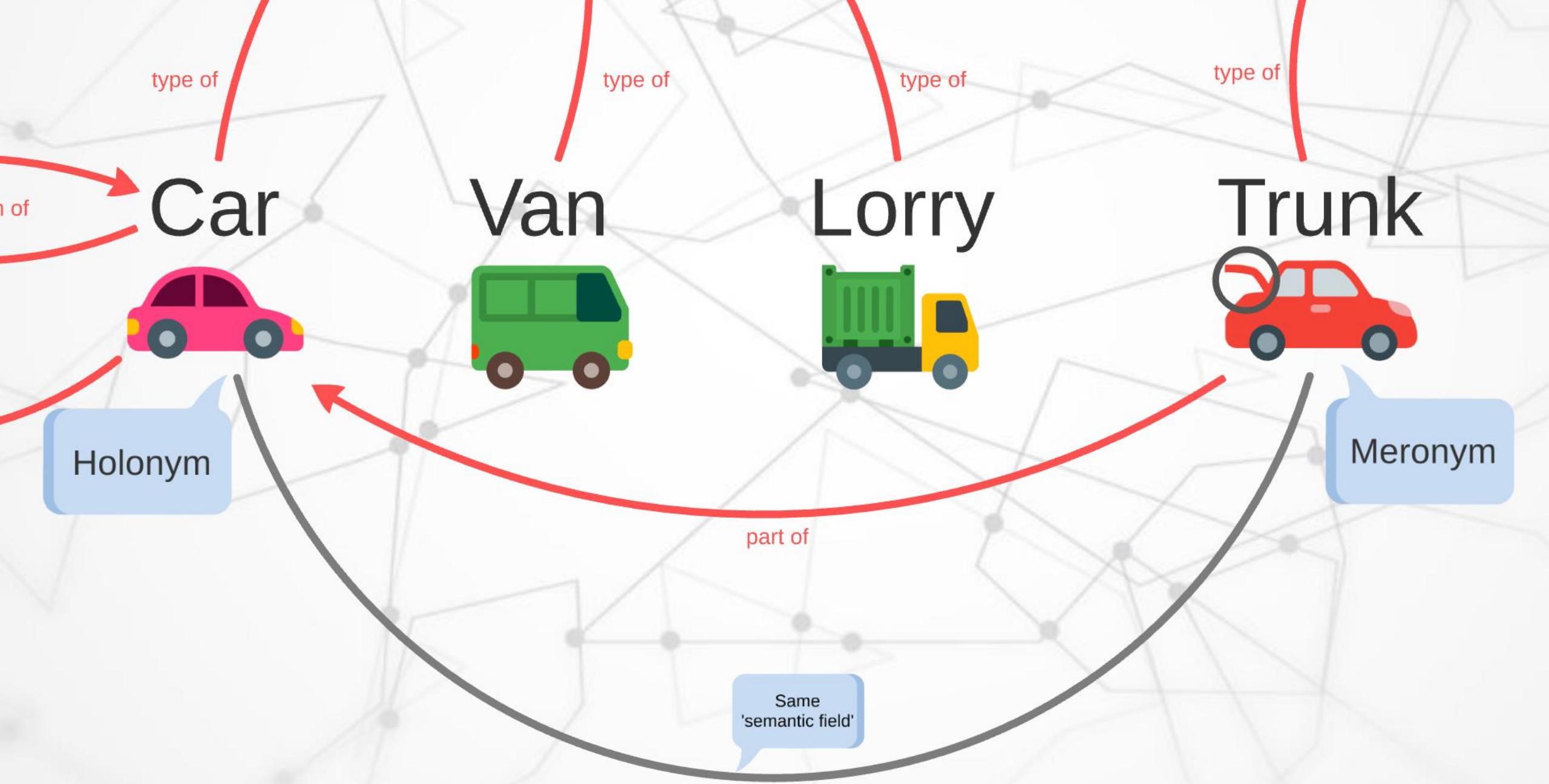
# Trunk



Meronym

part of

Same  
'semantic field'



Terms

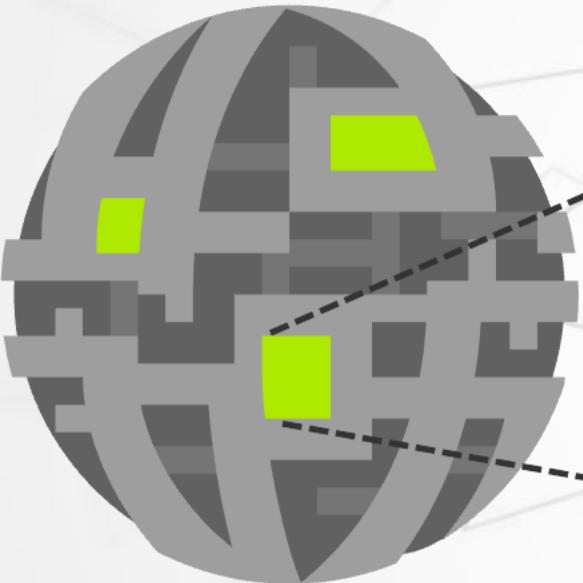
Concepts

Trunk

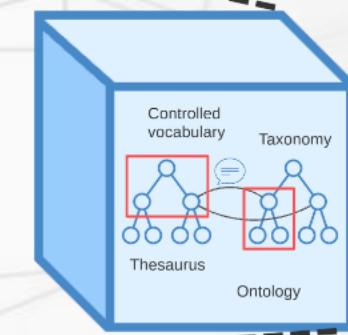




Worldview

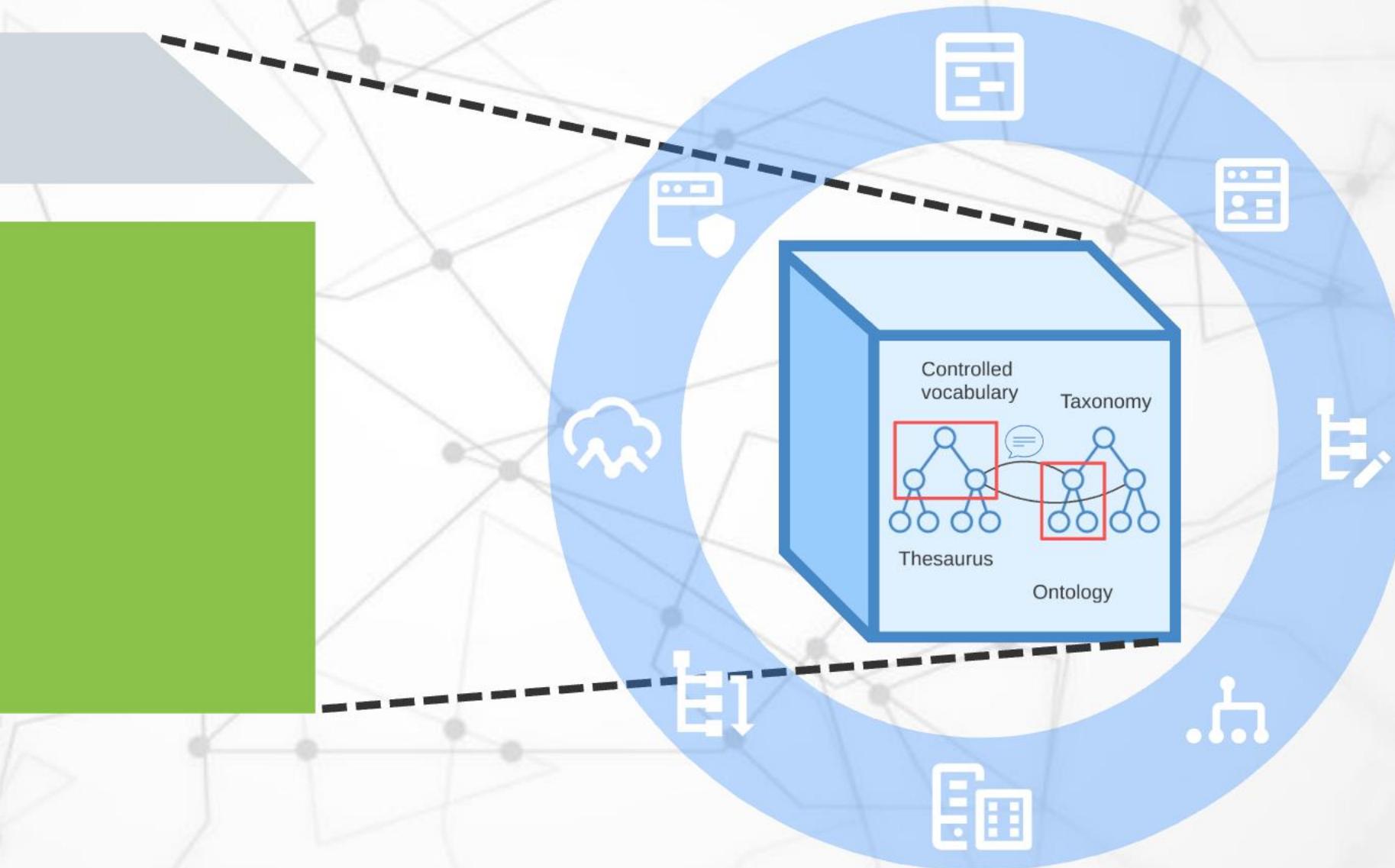


Domain of interest



f interest

## Information structures





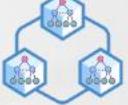
**Controlled  
Vocabulary**



**Taxonomy**



**Thesaurus**



**Ontology**

# Controlled vocabulary



Method of organising data & information



Collection of terms



Simple & lightweight term list to capture domain information



Starting point for the governance of terms

# Taxonomy



Classification scheme for organising information

- Parent-child
- Type-subtype
- Instance-of
- ...

- Broader-narrower

# Taxonomy



Classification scheme for organising information



More robust and more explicit than controlled vocabularies



Different types of taxonomies

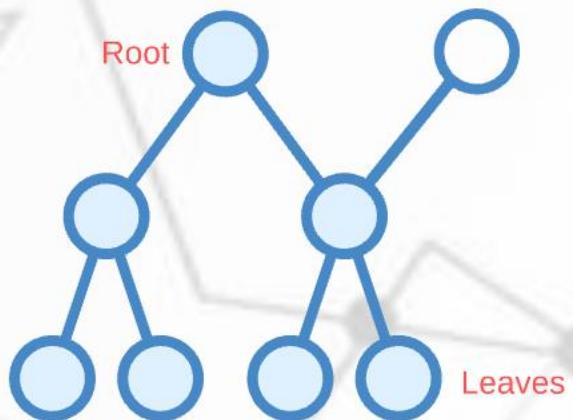
## Flat taxonomy



- Nodes live at the top-most level
  - No further categorisation
  - Content tagging can evolve into a folksonomy



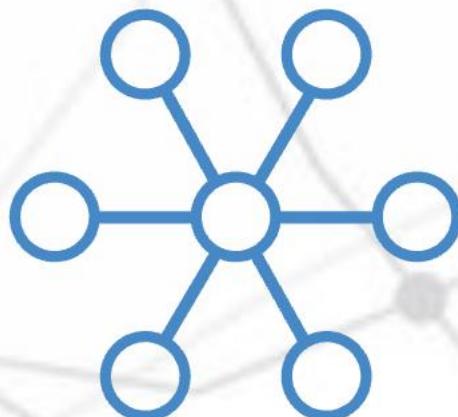
## Hierarchical taxonomy



- Tree structure
  - Manage levels of granularity
  - In polyhierarchical taxonomies, nodes can have multiple parents



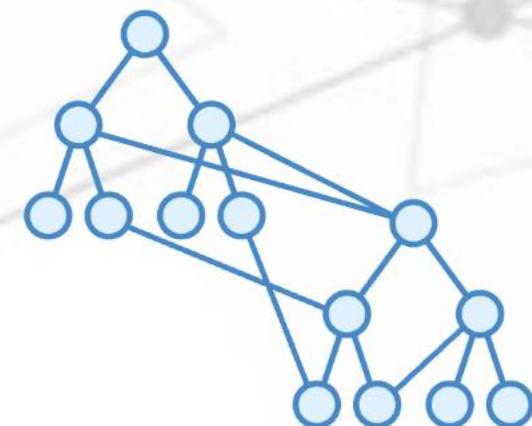
## Facet taxonomy



- Nodes can participate in multiple branches
  - Flexible ways of classifying and ordering



## Network taxonomy



- Leverages hierarchical and custom-defined relations to organise nodes
  - Relations carry meaning

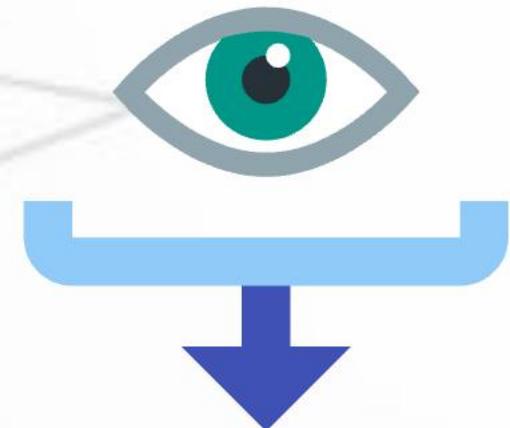
# organise nodes

- Relations carry meaning

Most popular



Most viewed



Recommended



# Taxonomy



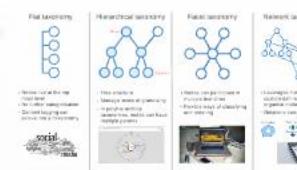
Classification scheme for organising information



More robust and more explicit than controlled vocabularies



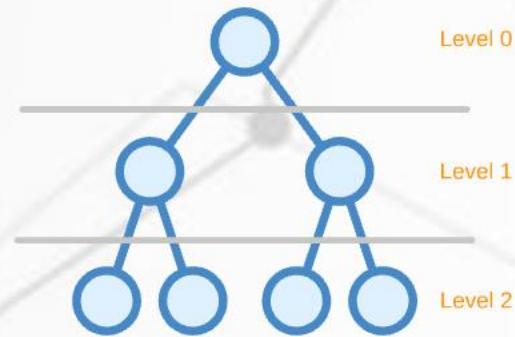
Different types of taxonomies



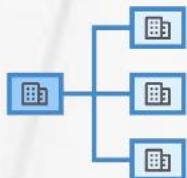
Greater flexibility in managing the governance of concepts & terms

# Balanced & ragged taxonomies

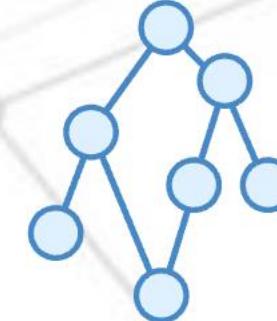
## Balanced hierarchy



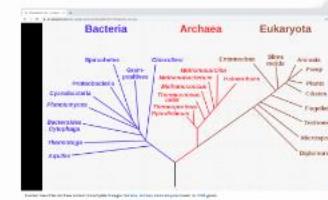
- Parent node sits at exactly one level above child node
  - Uniformity over flexibility
  - Can lead to unnecessary data and inaccurate domain representation



## Ragged hierarchy

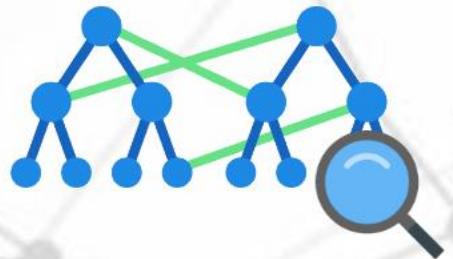


- Parent and child nodes can live at any level
  - Polyhierarchical taxonomy
  - Flexible structures that are brain-friendly

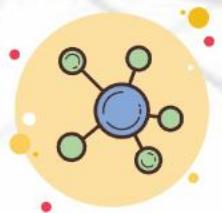


concrete	concentration	corporation	Crown corporation	corporate attorney
desacrom	desacrum	dateuse	<b>empire</b>	entomology
<b>establishment</b>	estate agent	financial technology	fintech	firm
franchise	franchises	frontline	gross	group-wide
headcount	holding company	<b>house</b>	Inc.	incorporated
<u>independent</u>	<u>industrial</u>	industry	internally	joint venture
joint-stock company	label	limited	limited company	line
<u>listed company</u>	LLC	Ltd	management consultancy	marriage bureau
microbusiness	MLM	monolithic	<b>multiple</b>	office

# Thesaurus

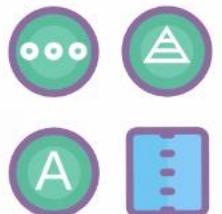


Augmented taxonomy aimed at promoting semantics



Rich representation and organisation of **knowledge**

Organisation System (KOS)

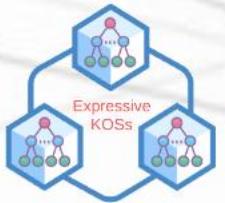


Concepts, labels, concept schemes & semantic mappings



Great flexibility in managing the governance of concepts and terms

# Ontology



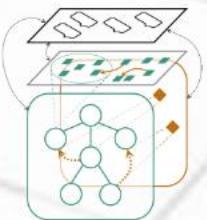
Formal specification of a domain of interest



Basis for sharing meaning



Reasoning and inference mechanisms

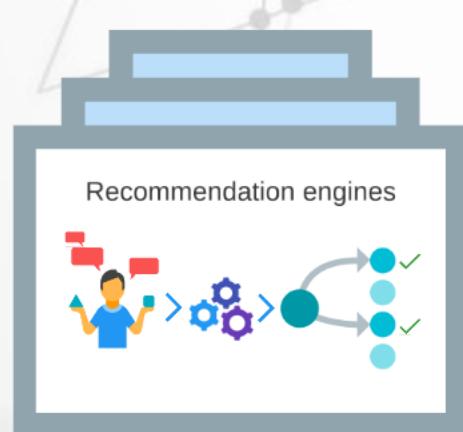
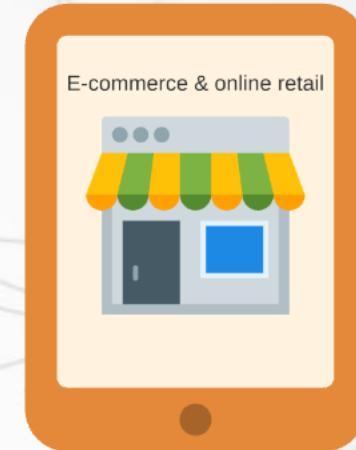
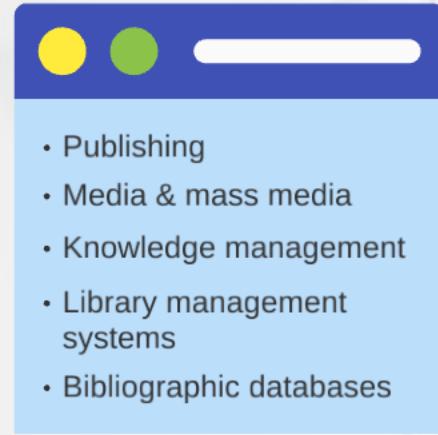


Classes, properties (relations), rules, individuals (instances)

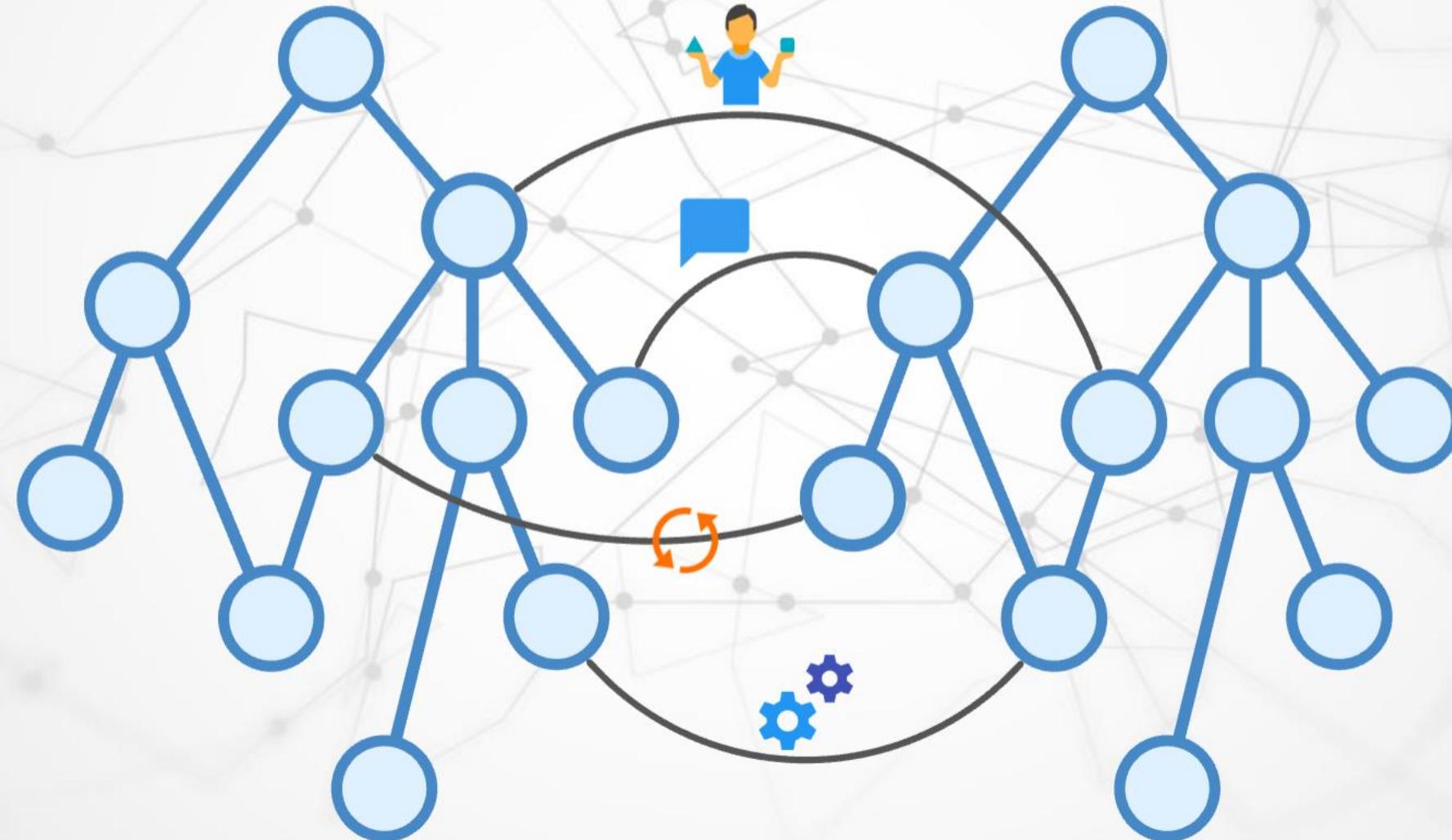


Superior governance of concepts and terms

# Examples of taxonomy and thesaurus applications



# Considerations for taxonomy & thesaurus development



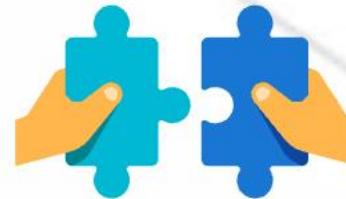
Adopt standard  
modelling methods



Portability



Interoperability

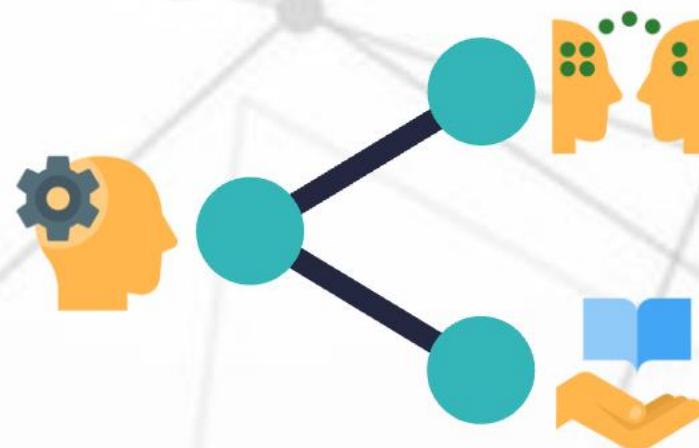


Rich set of constructs to accurately  
model domain knowledge

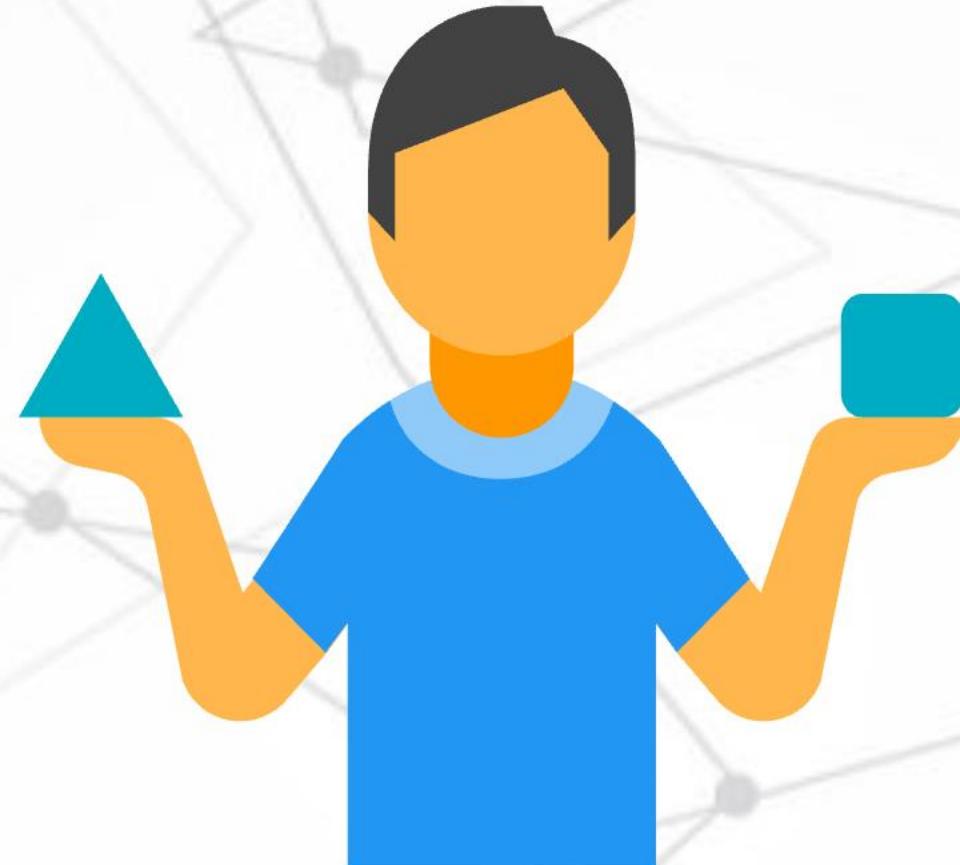


# Expressiveness

Approaches that promote shared  
meaning instead of enforcing  
single common vocabularies



# Simple Knowledge Organization System (SKOS)



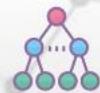
# 03 A quick tour of SKOS



Rich taxonomies & thesauri:

**Simple Knowledge Organization System (SKOS)**

- Portability
- Interoperability
- Extensibility
- Standardisation



Rich taxonomies & thesauri:

**Simple Knowledge Organization System (SKOS)**



Ontologies: **Web Ontology Language (OWL)**



Simple taxonomies: **RDF Schema (RDFS)**



Graph model: **Resource Description Framework (RDF)**

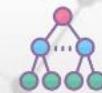


Syntax: **XML, Turtle, etc.**

Lightweight OWL-based ontology for  
developing controlled vocabularies,  
taxonomies and thesauri



End user  
applications



Rich taxonomies & thesauri:

**Simple Knowledge Organization System (SKOS)**



Ontologies: **Web Ontology Language (OWL)**



Simple taxonomies: **RDF Schema (RDFS)**



Graph model: **Resource Description Framework (RDF)**



Syntax: **XML, Turtle, etc.**

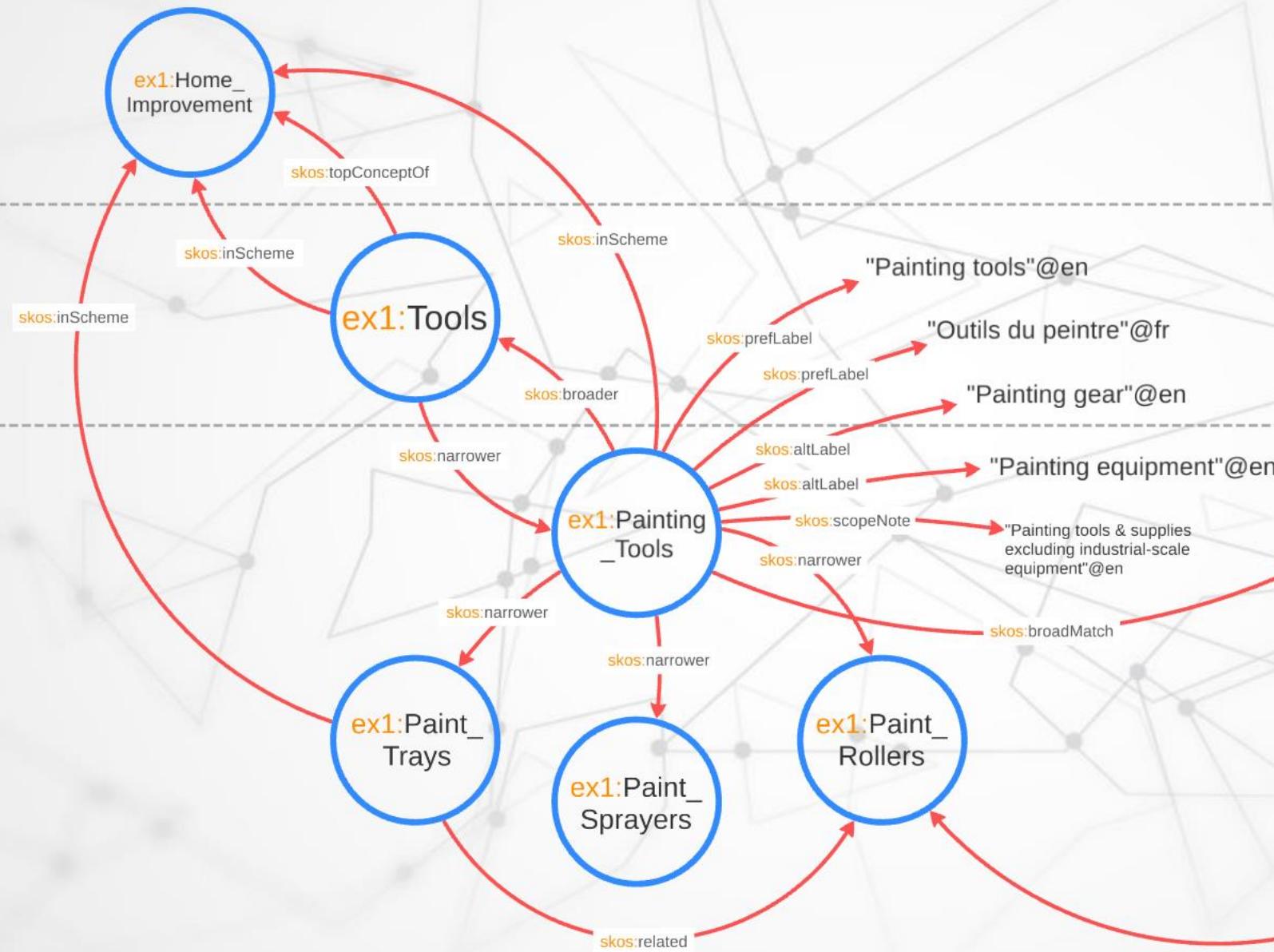
## Domain A

PREFIX ex1: <<http://example1.com/>>

Concept Scheme

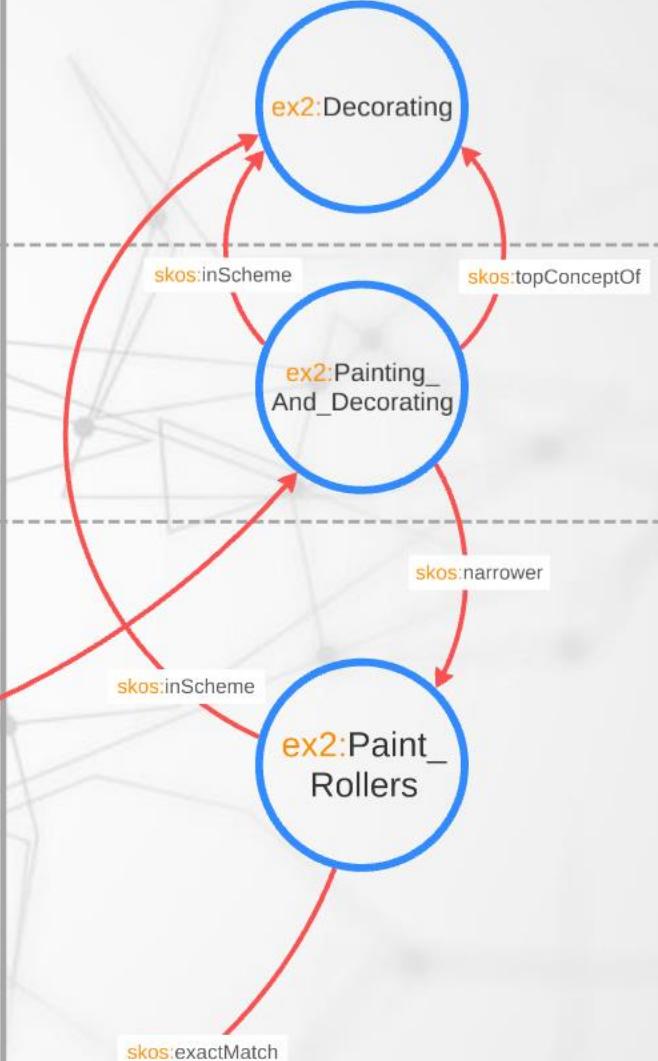
Top Concepts

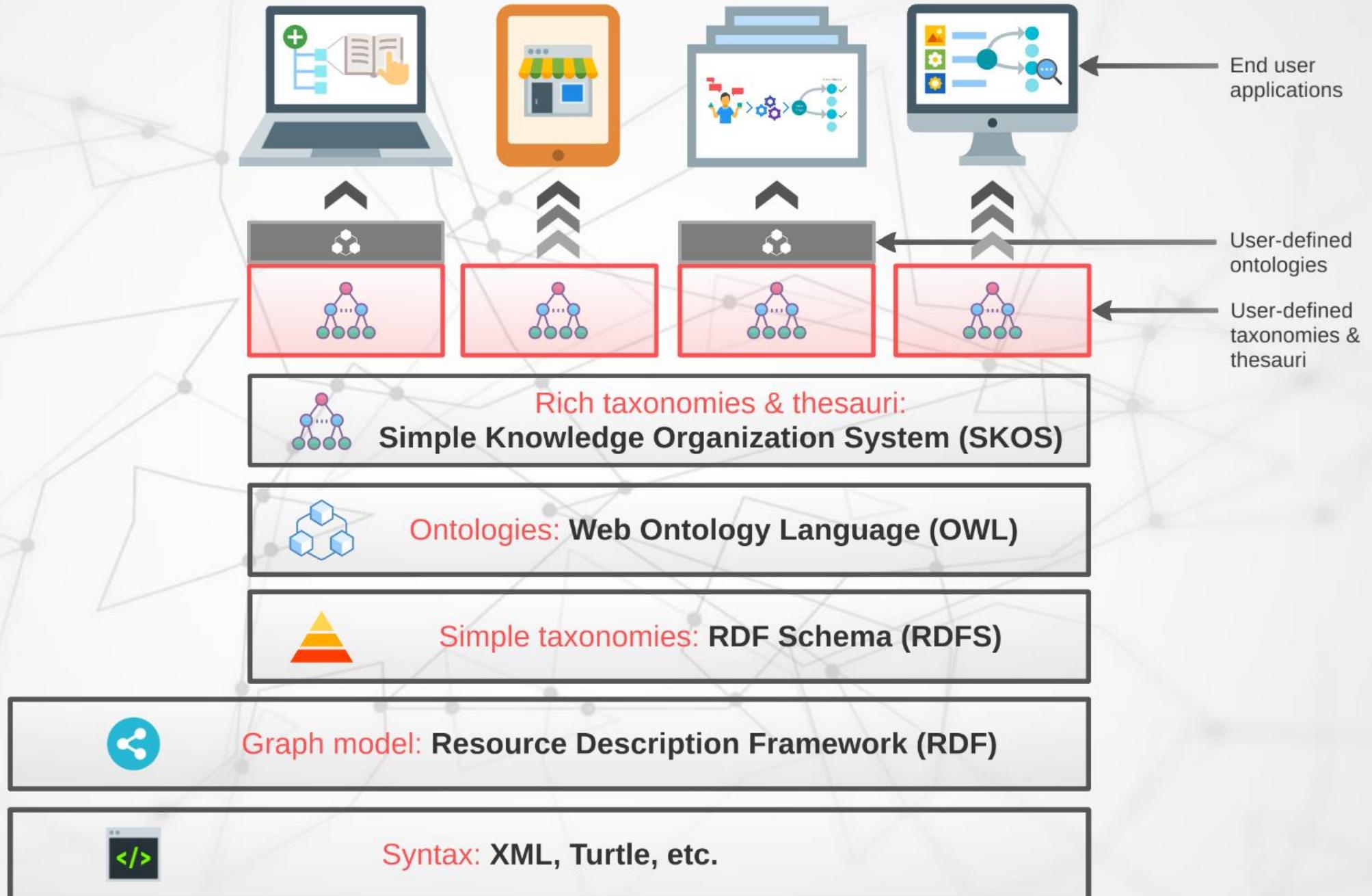
Concepts

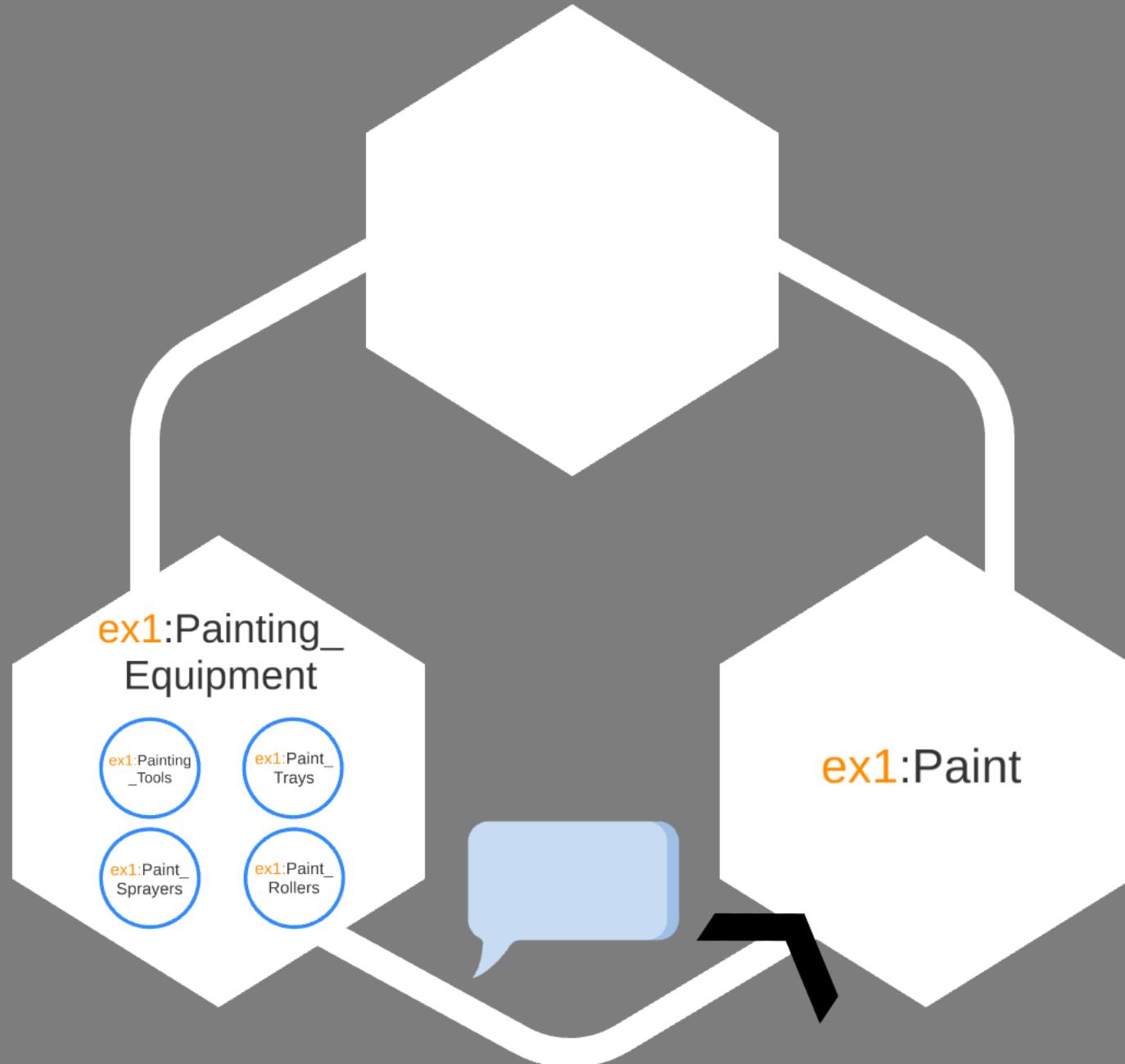


## Domain B

PREFIX ex2: <<http://example2.com/>>

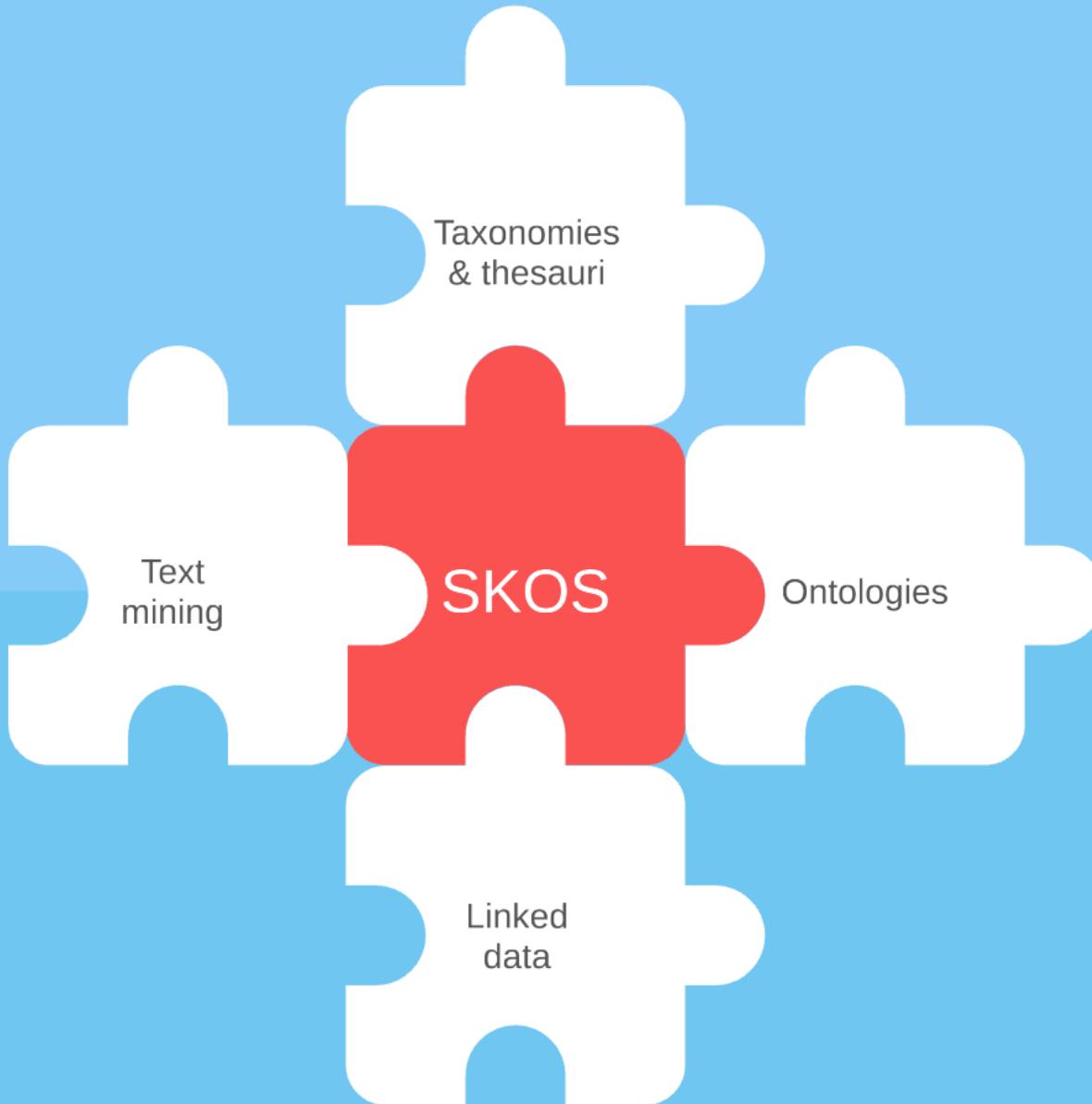


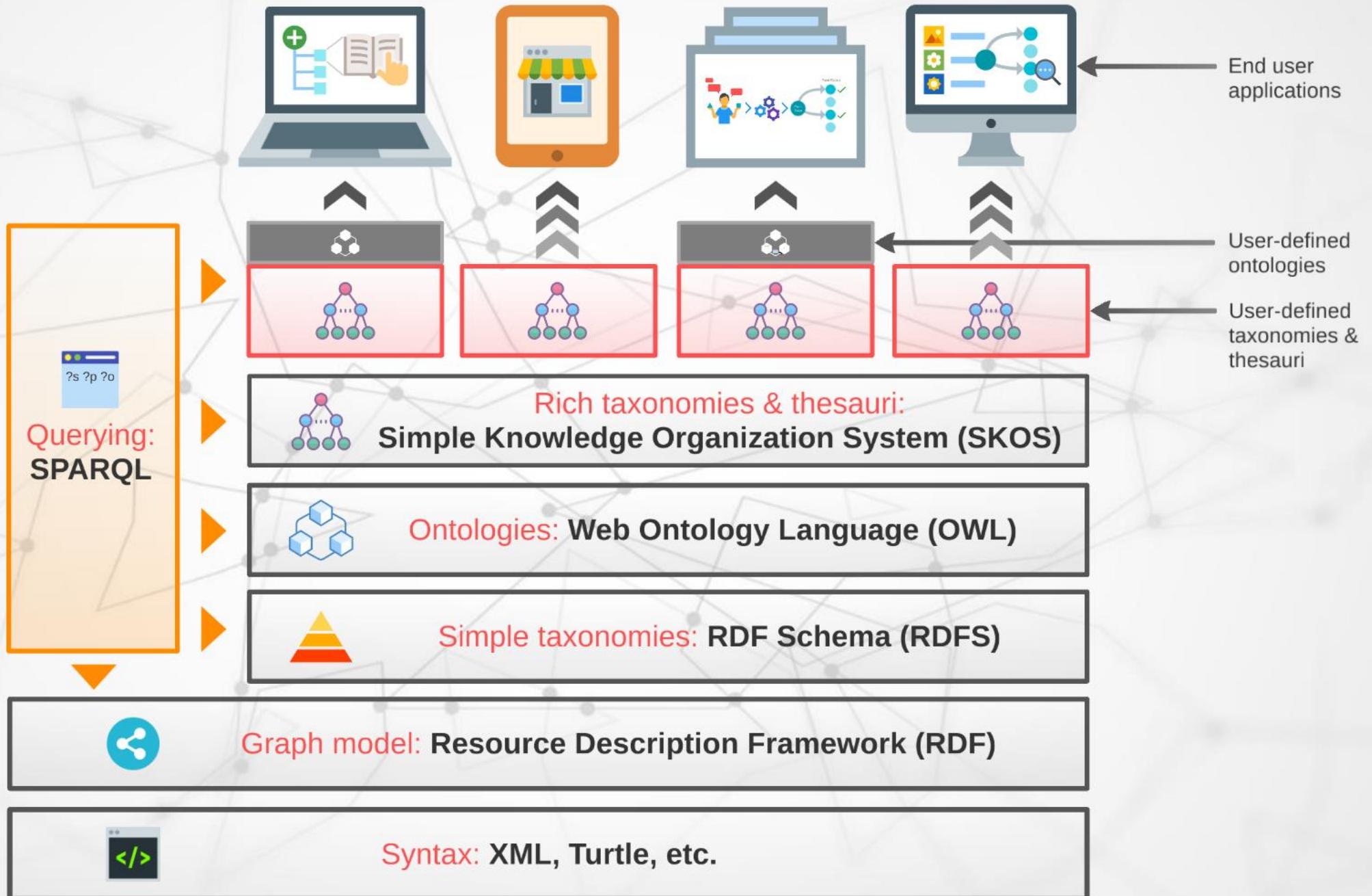




# Necessary condition:

- Every painting equipment uses some paint







?s ?p ?o

# Business questions

- What are the top five concepts in a taxonomy with the highest density of incoming and outgoing related concepts via skos:related predicate?

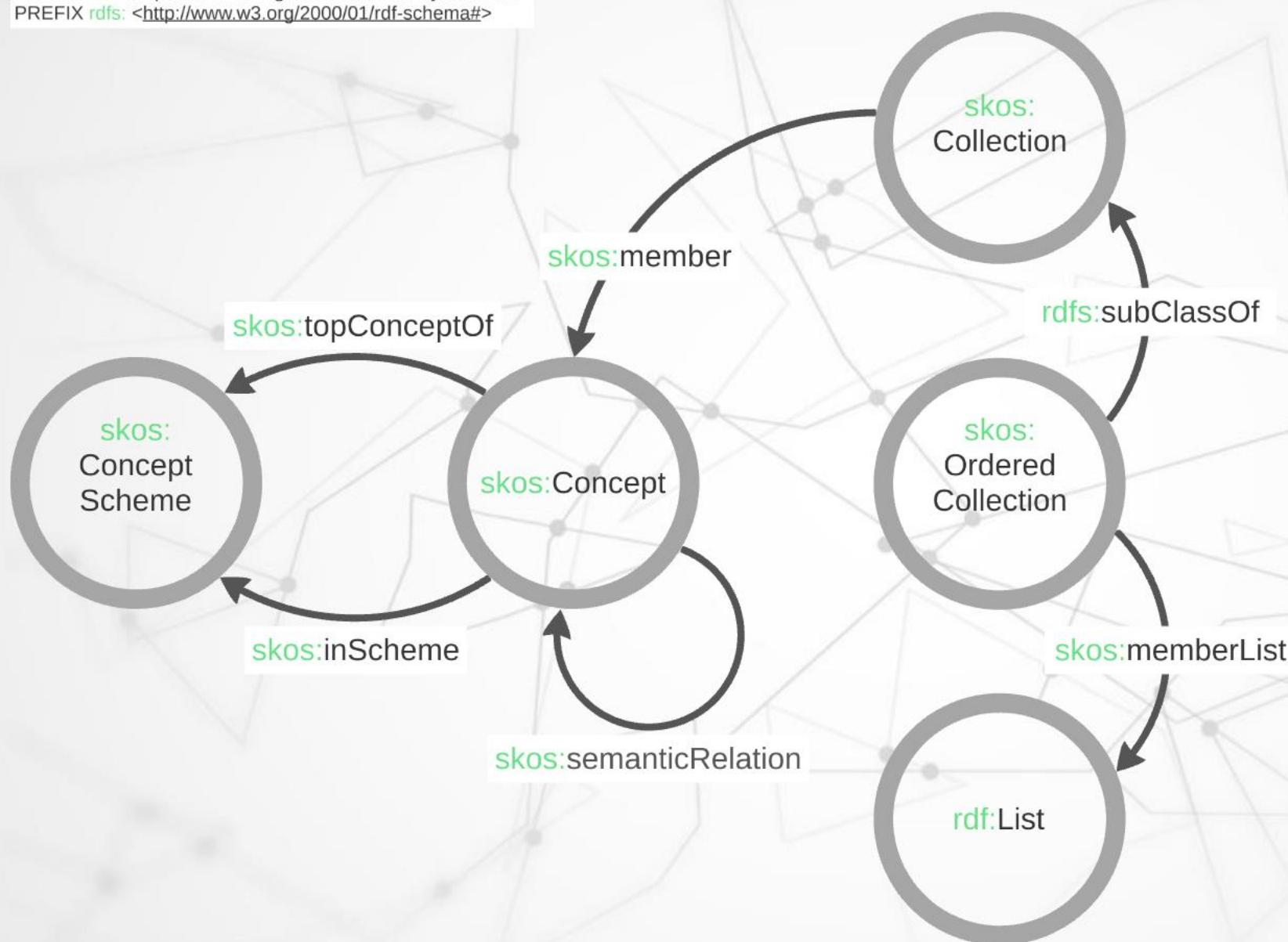
## SKOS graph statistics

- Count of concept schemes
- Count of concepts
- Relation statistics, etc.

# SKOS graph integrity

- Are there any concepts for which we defined more than one preferred label in the same language?

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#>>



### Labels

- Preferred label
- Alternative label
- Hidden label



### Documentation

- Note
- Change note
- Definition
- Editorial note
- Example
- History note
- Scope note



### Notation

- Notation



### Semantic

- Broader
- Broader transitive
- Narrower
- Narrower transitive
- Related



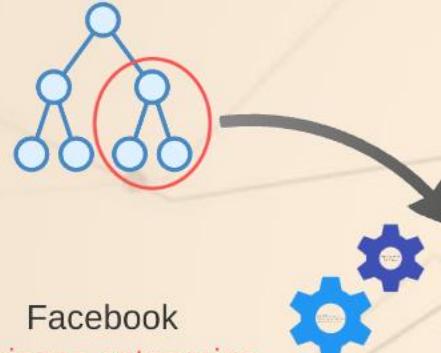
### Mapping

- Broad match
- Close match
- Exact match
- Narrower match
- Related match

# 04 SKOS taxonomy and thesaurus development



Facebook  
Product categories

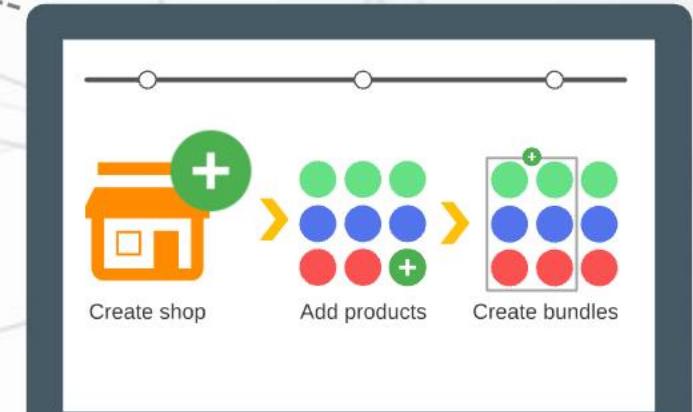


Facebook  
Business categories

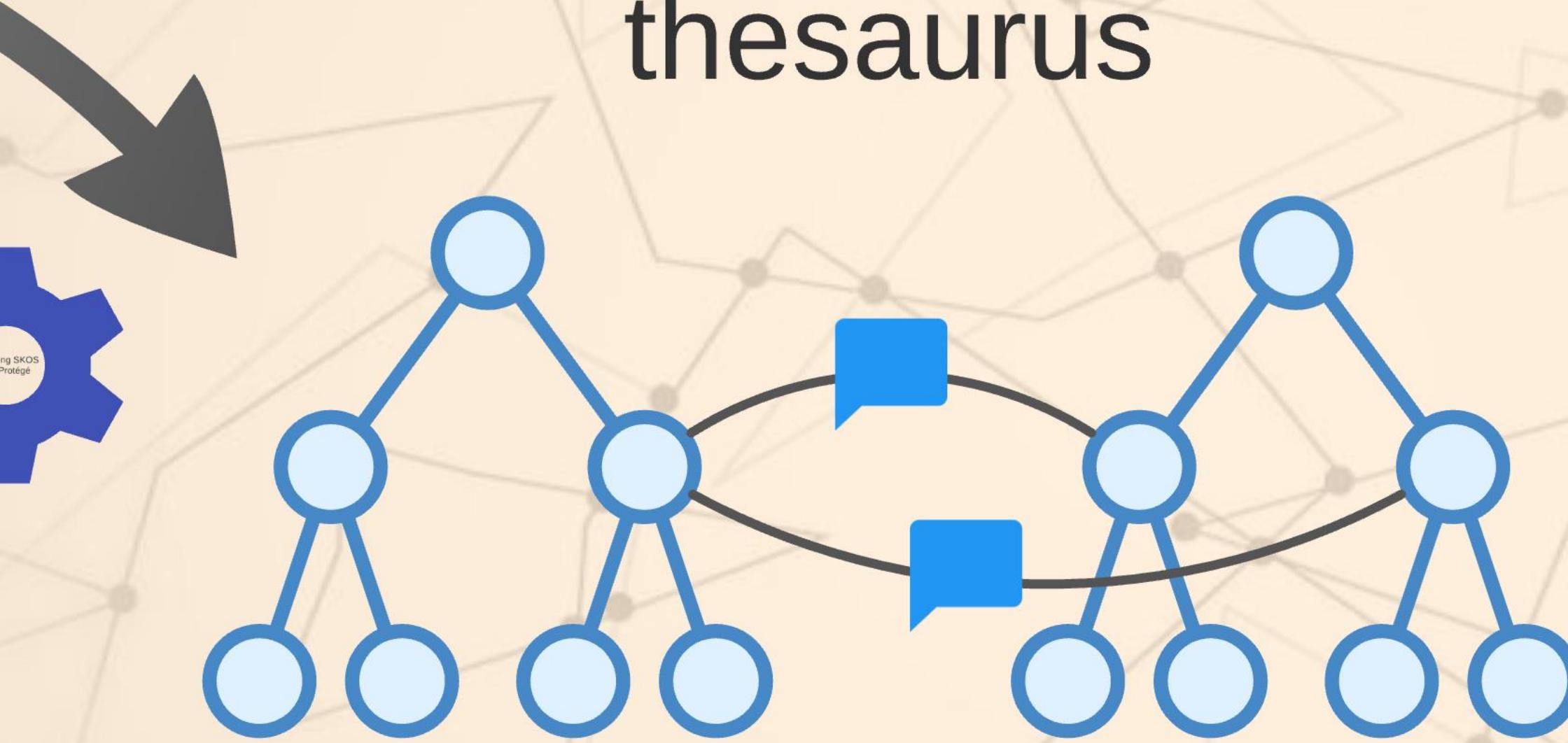


Example SKOS  
thesaurus

B2C and B2B  
e-commerce platform



# Example SKOS thesaurus



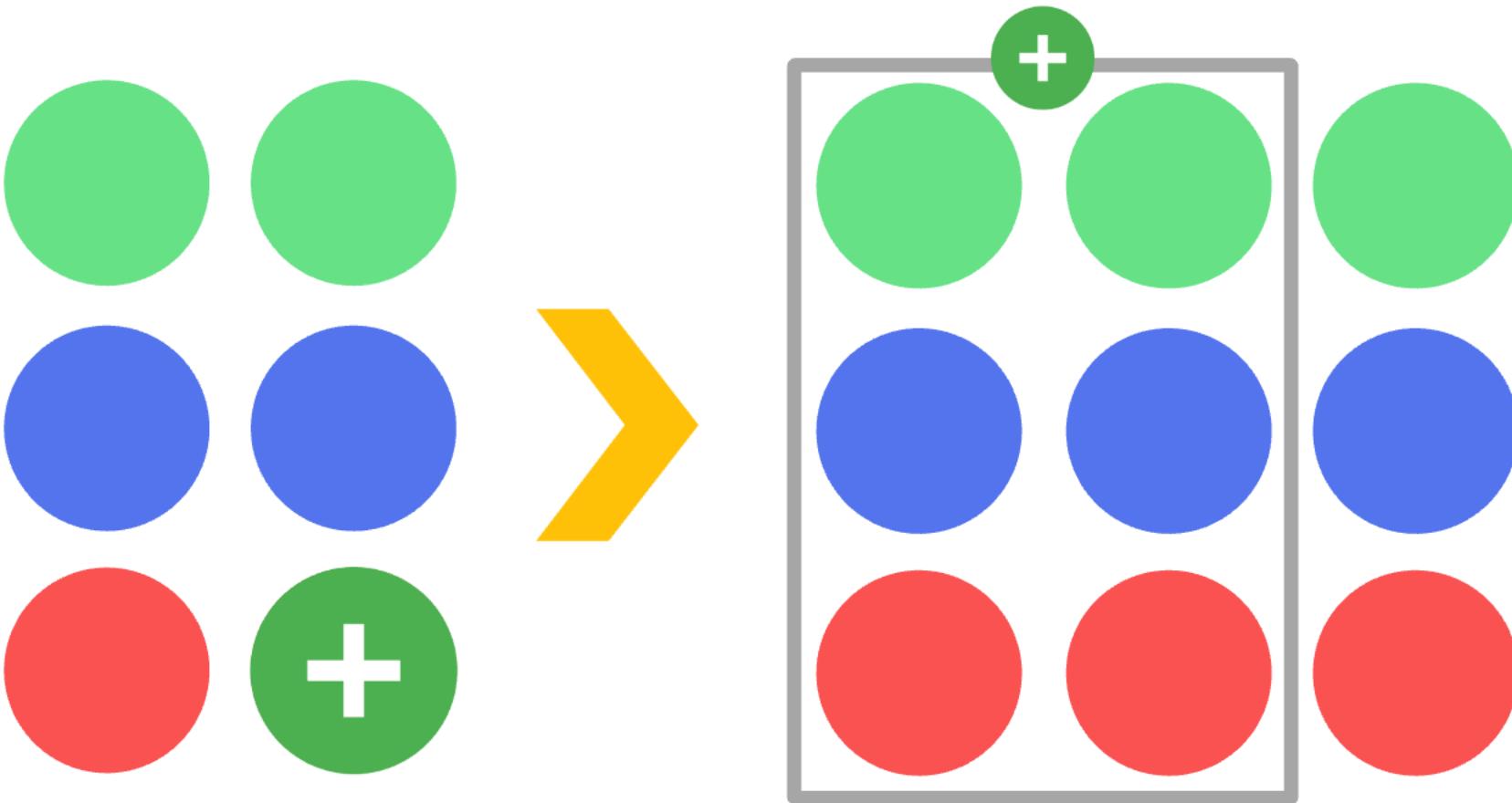


'skos:related' for  
generic concept  
relations



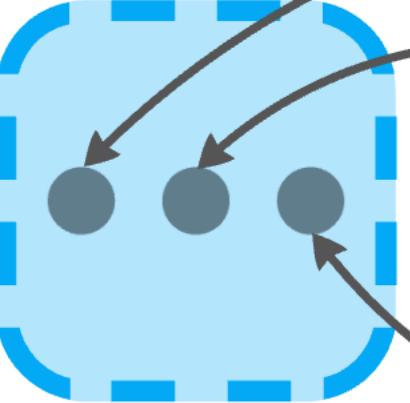
- Product A
  - Product B
  - Product C
  - Product D
  - Product E
  - Product F
- 
- The diagram illustrates a classification structure. On the left, a vertical list of six products is shown. To the right of this list, three red curly braces group the products into three distinct categories. Each category is accompanied by a small icon: a fork and knife for Flatware, a wine glass for Drinkware, and a dinner plate for Dinnerware. The products are grouped as follows: Products A and B are in the Flatware category; Products C and D are in the Drinkware category; and Products E and F are in the Dinnerware category.
- Product A
  - Product B
  - Product C
  - Product D
  - Product E
  - Product F

# 05 Collections



products

Create bundles



Product Bundle 1

- Product A
  - Product B
  - Product C
  - Product D
  - Product E
  - Product F
- Flatware      {  }
- Drinkware      {  }
- Dinnerware      {  }

Create shop

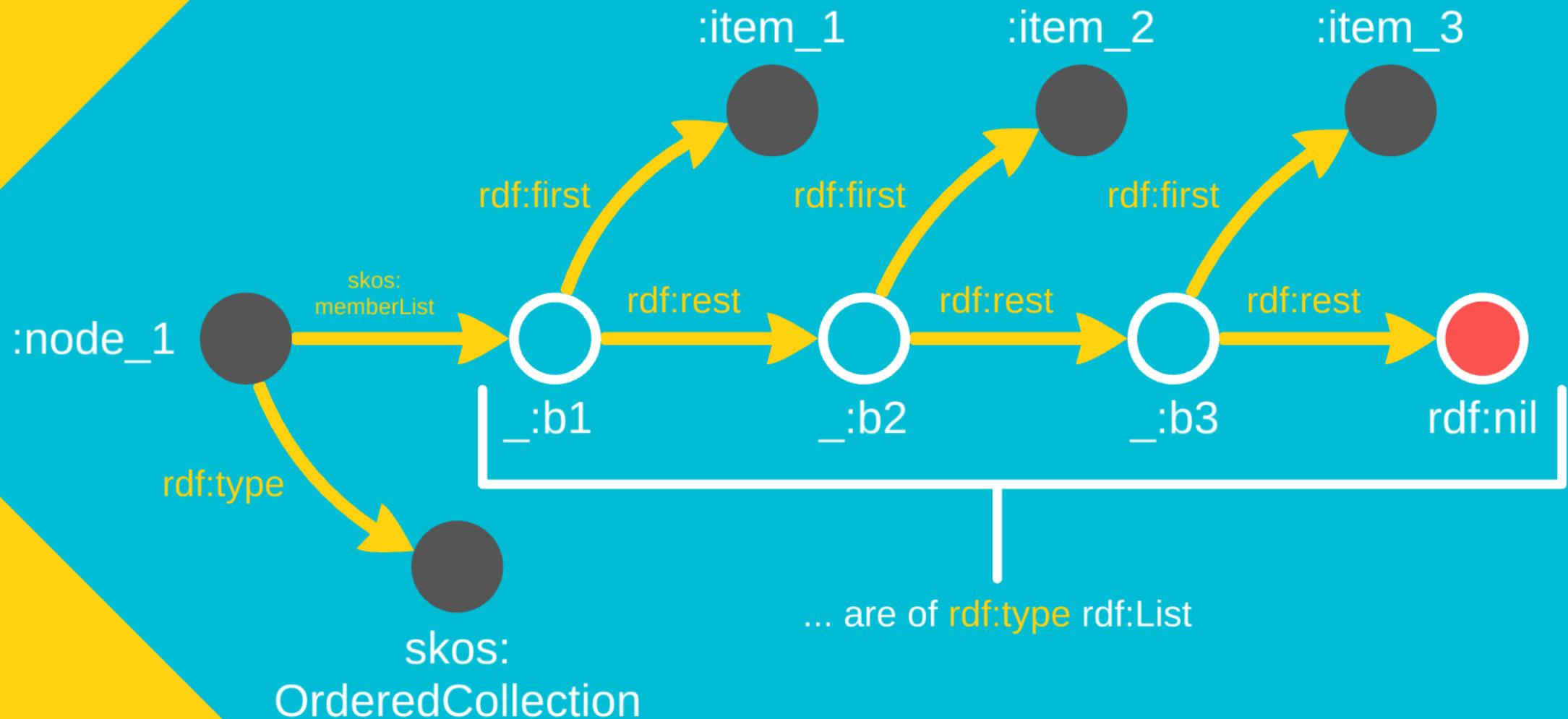
Add products

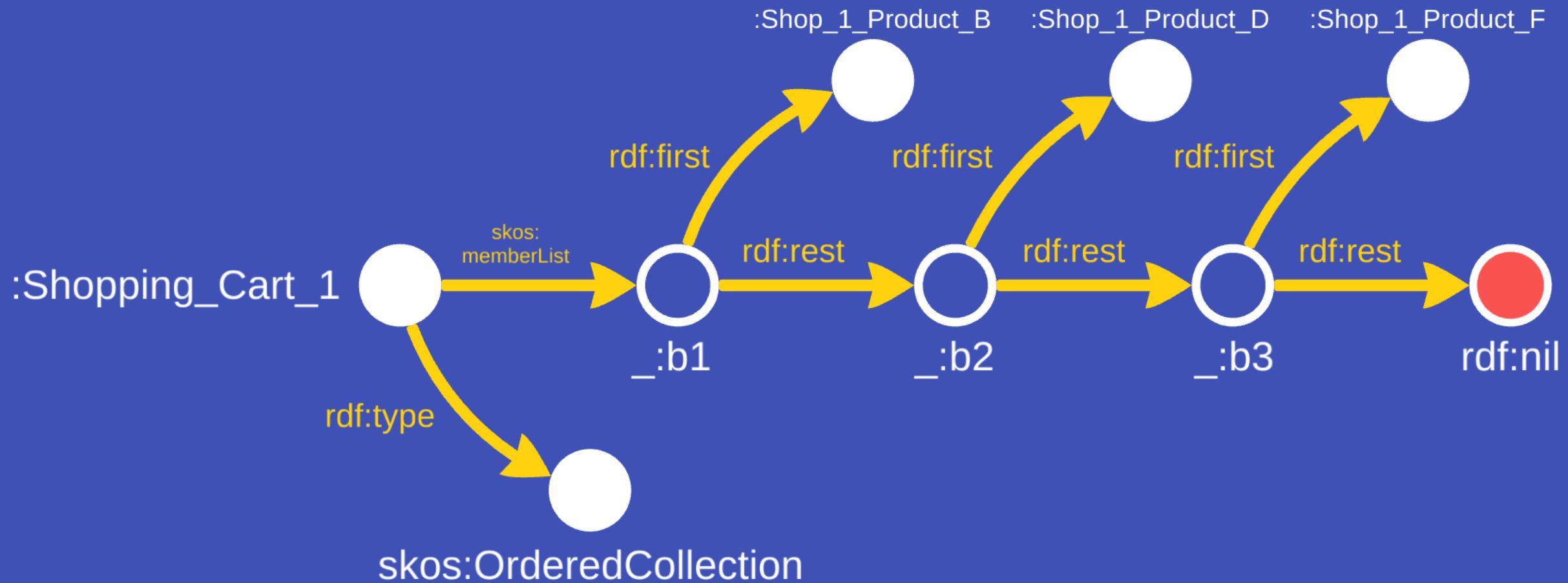
Create bundles



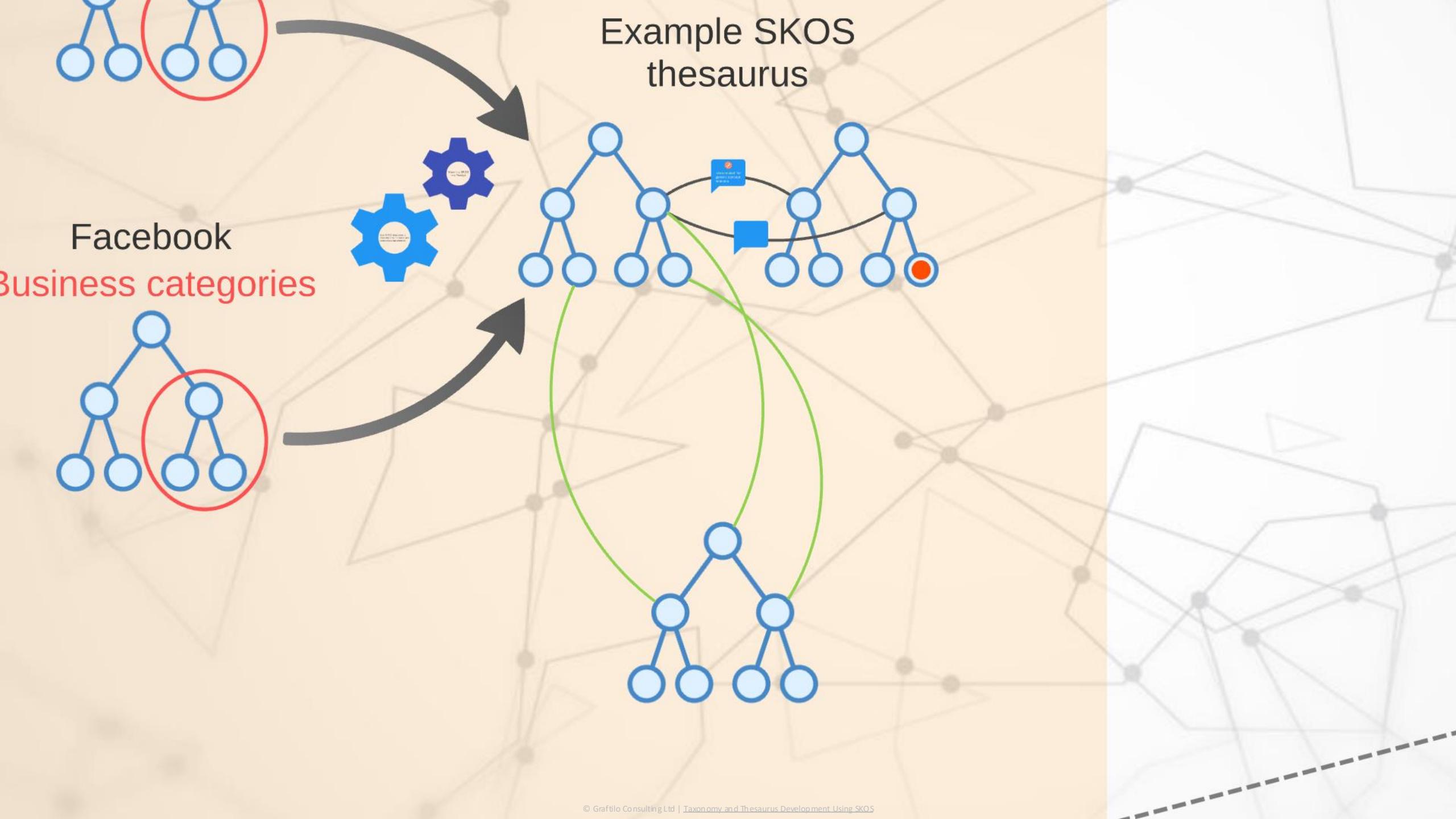
Save shopping cart  
with items in the  
correct order







# 06 Semantic mappings

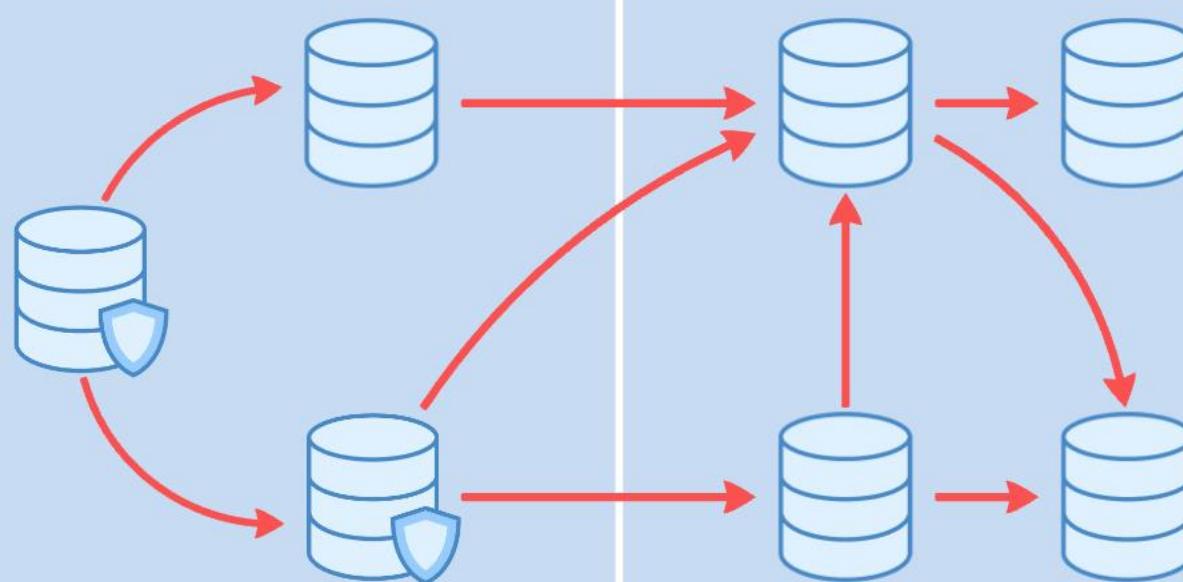




## Banking & financial services

Front-office systems

Back-office systems

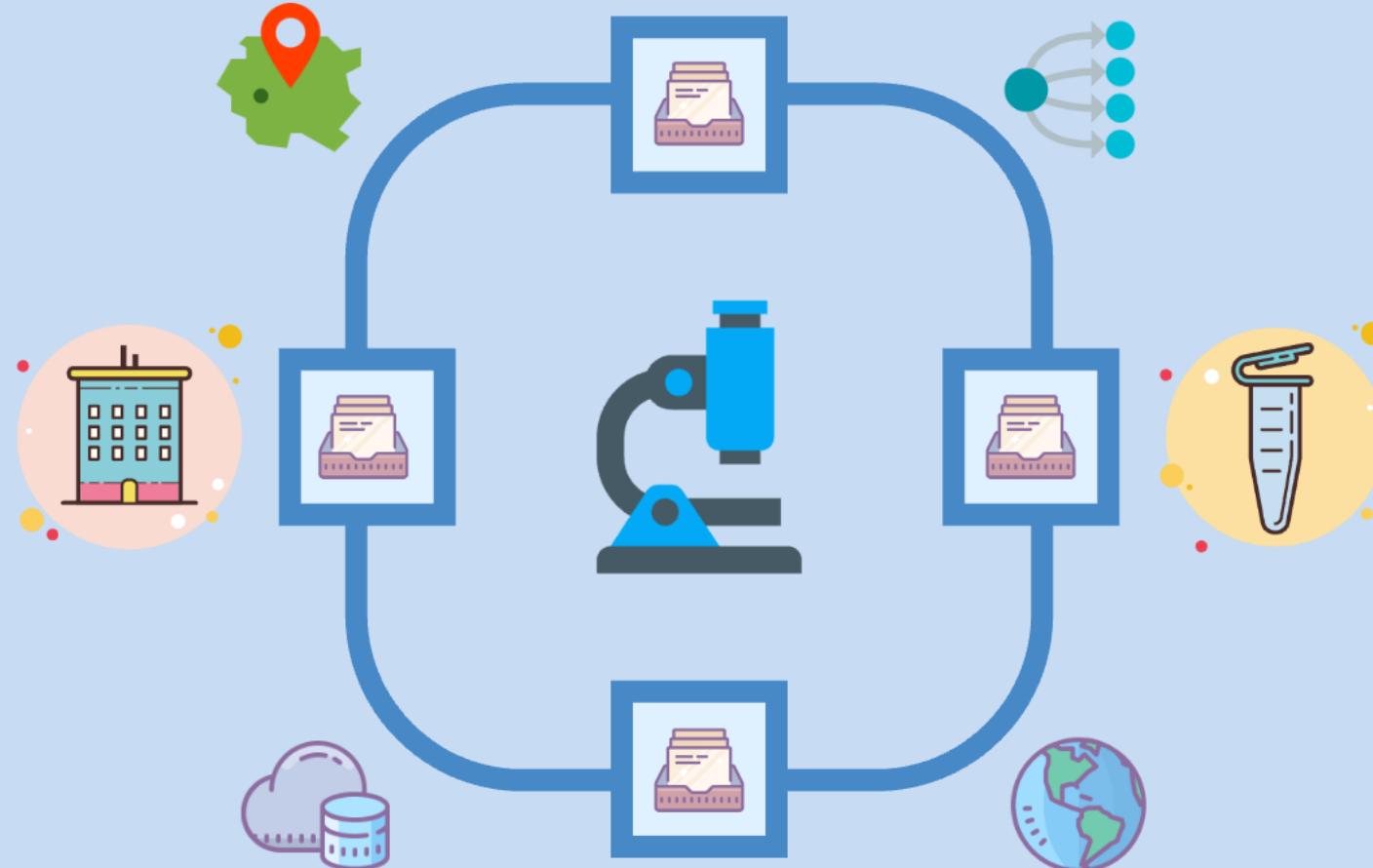


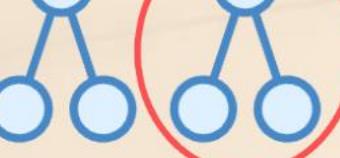
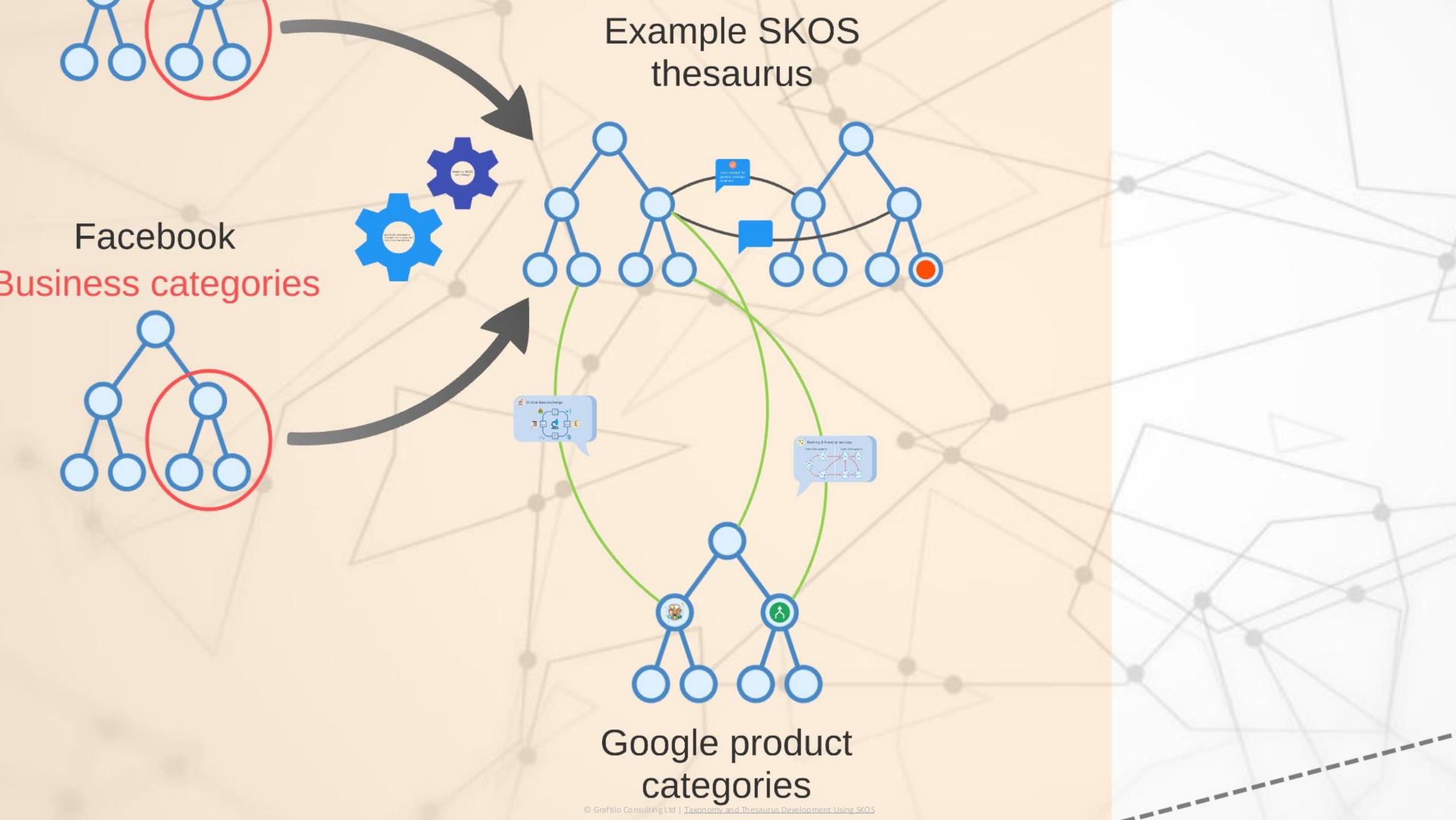




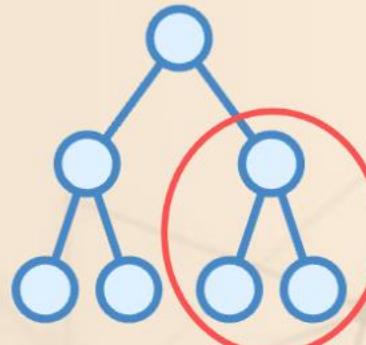


# Clinical data exchange





Facebook  
Business categories



Google product  
categories

# 07 Querying SKOS graphs

# Business questions

- What are the top five concepts in a taxonomy with the highest density of incoming and outgoing related concepts via skos:related predicate?

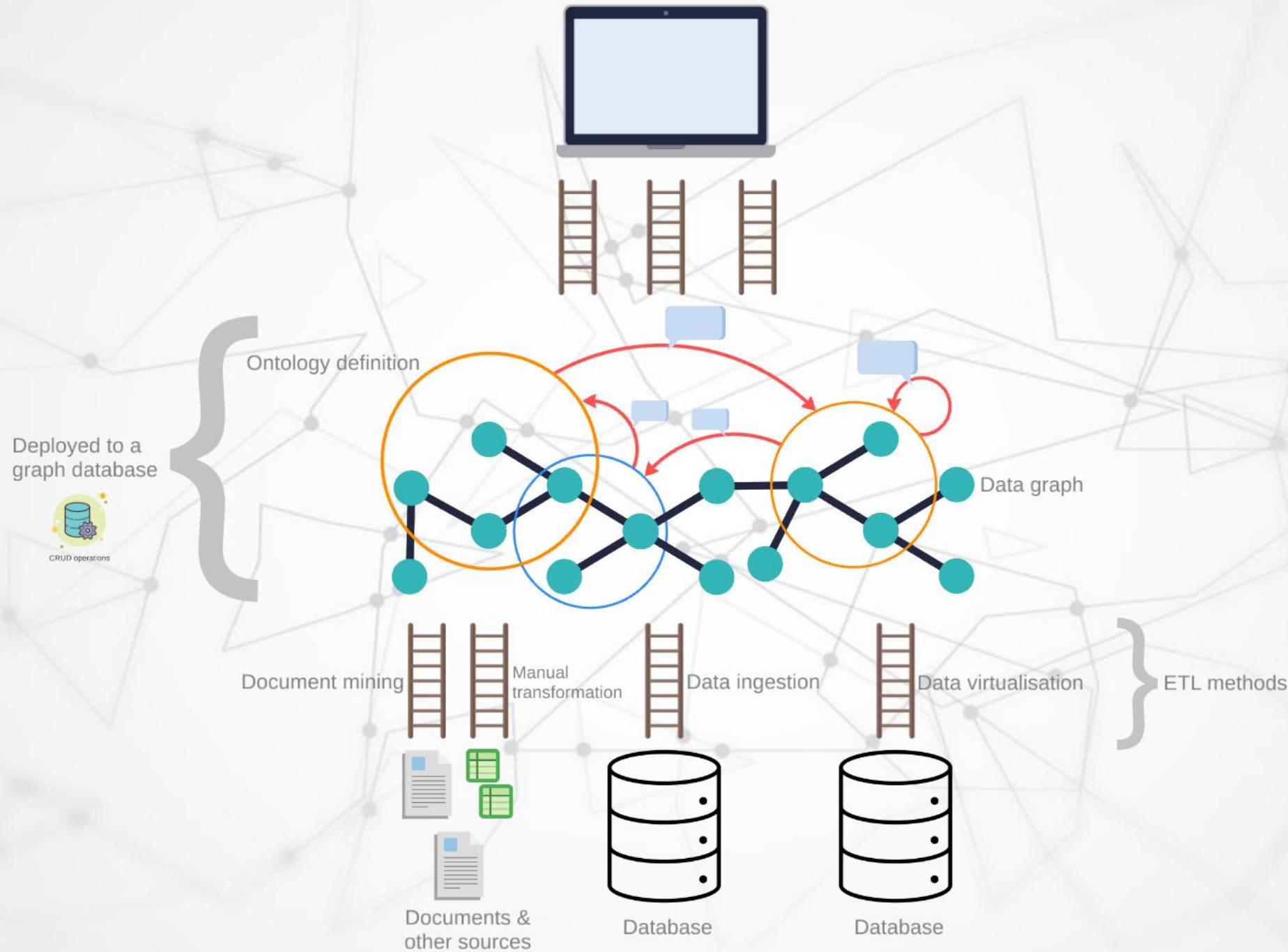
## SKOS graph statistics

- Count of concept schemes
- Count of concepts
- Relation statistics, etc.

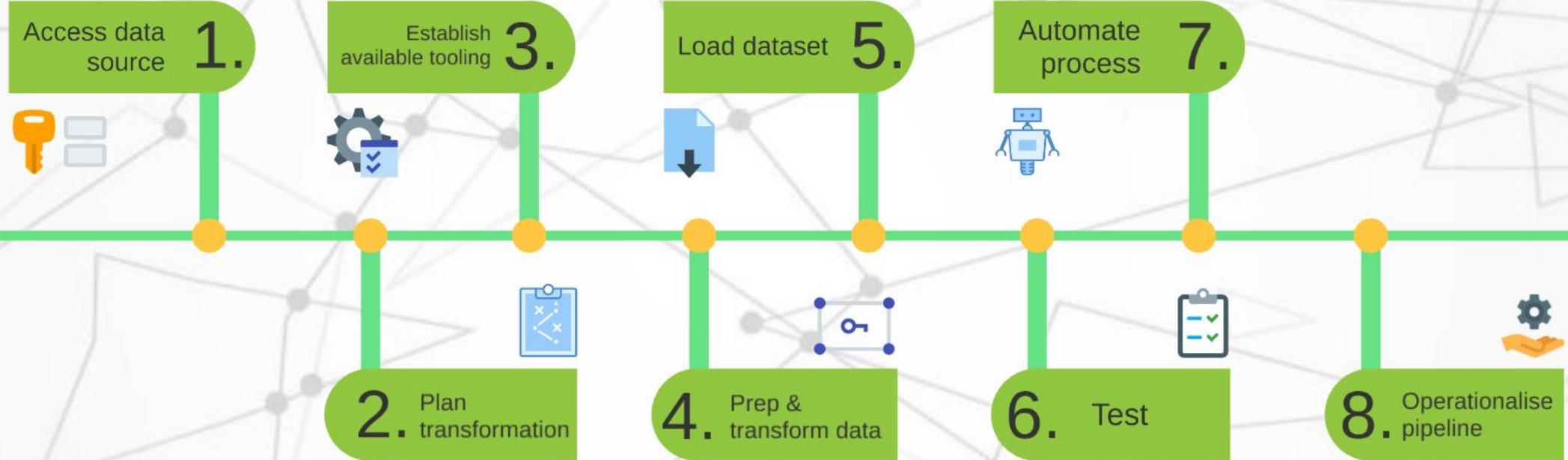
# SKOS graph integrity

- Are there any concepts for which we defined more than one preferred label in the same language?

# 08 End-to-end architecture



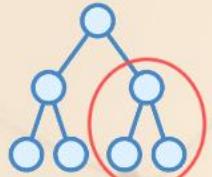




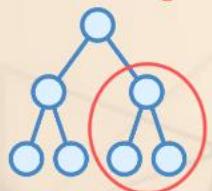




Facebook  
Product categories



Facebook  
Business categories



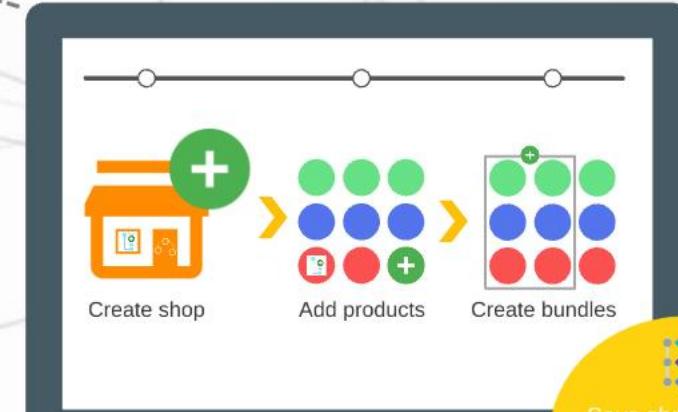
Example SKOS  
thesaurus



Google product  
categories

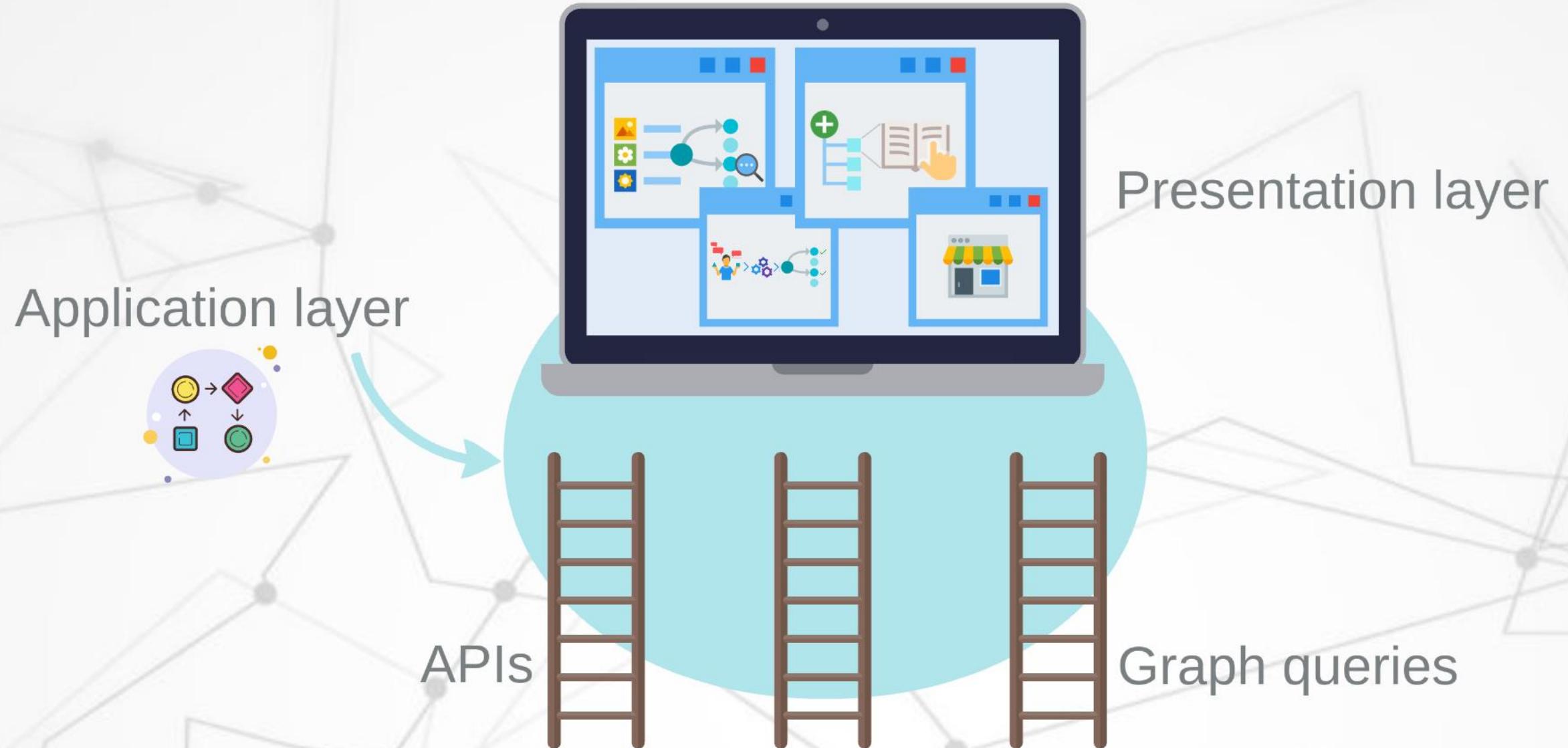


B2C and B2B  
e-commerce platform



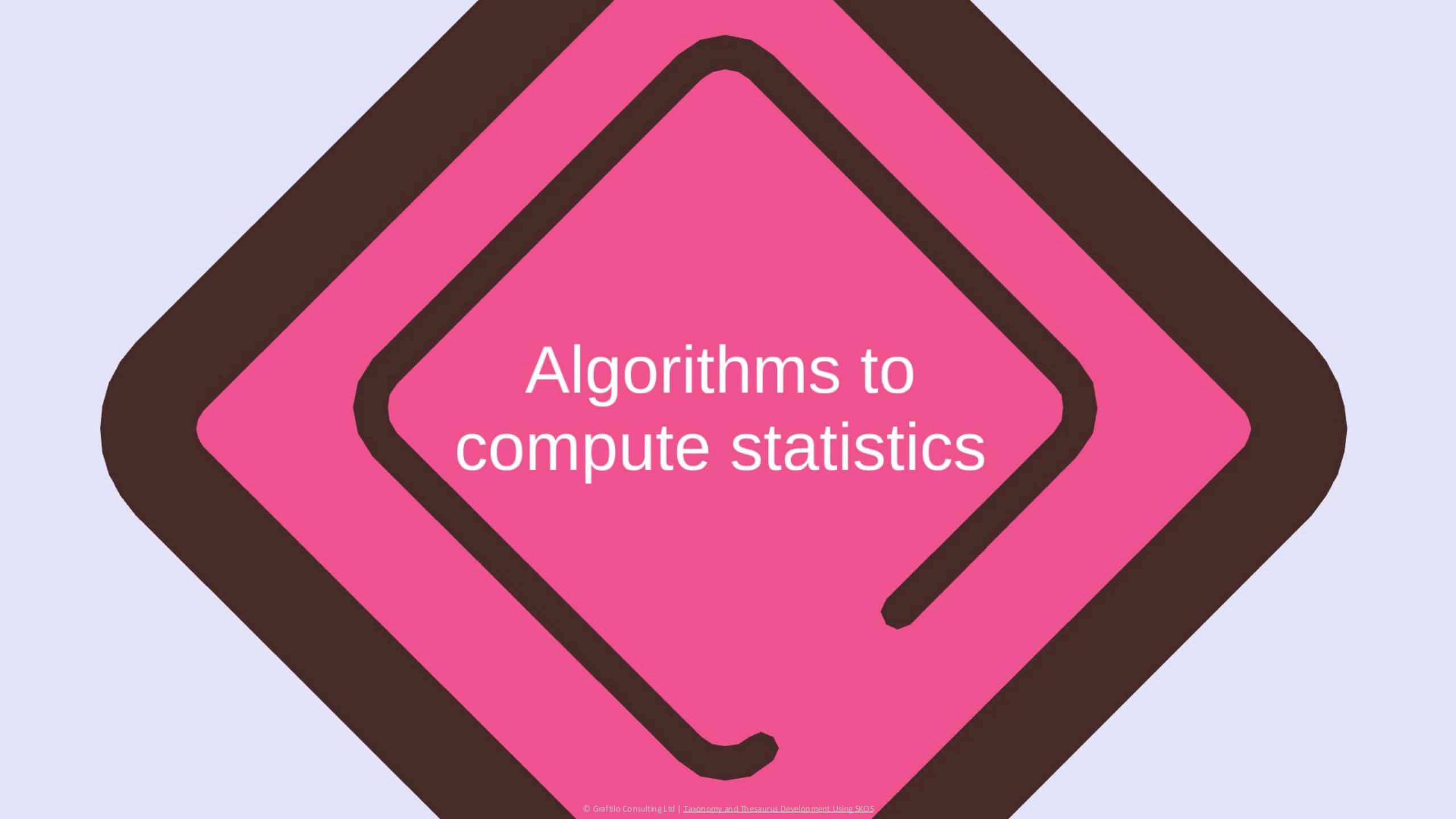
Save shopping cart  
with items in the  
correct order



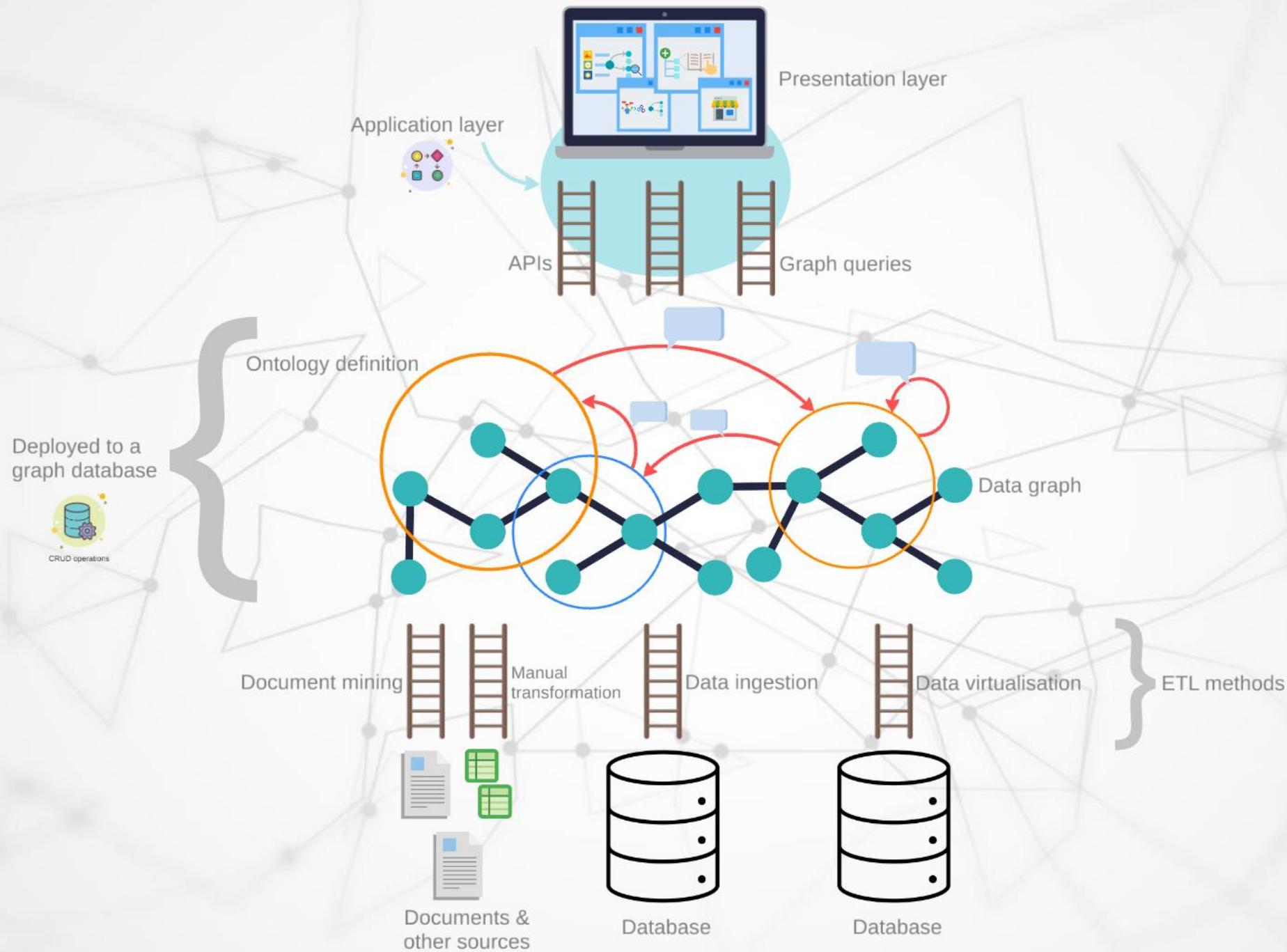




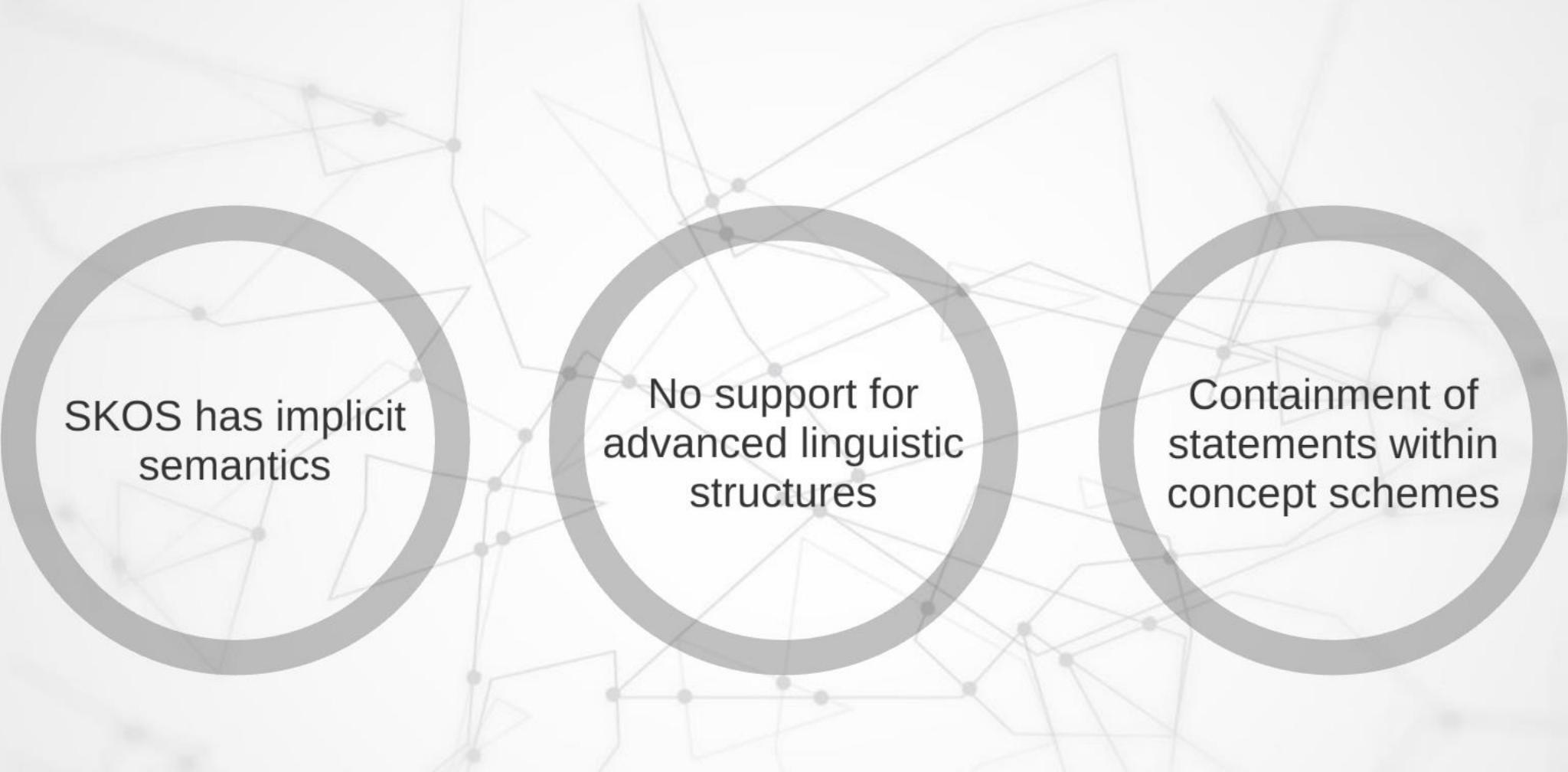
Recommender  
engine



# Algorithms to compute statistics



# 09 Course wrap-up



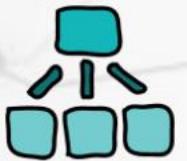
SKOS has implicit semantics

No support for advanced linguistic structures

Containment of statements within concept schemes



SKOS has implicit  
semantics



SKOS has somewhat loose  
structural definition



Design choices are needed



SKOS is a set of guidelines



No support for  
advanced linguistic  
structures



icit



No direct support for  
modelling parts of speech



Extensions needed for  
linguistic support

# Containment of statements within concept schemes

