KNOWLEDGE ORGANIZATION
RESEARCH OBSERVATORY

LOD = **L**inked **O**pen **D**ata
KOS = **K**nowledge **O**rganization **S**ystems
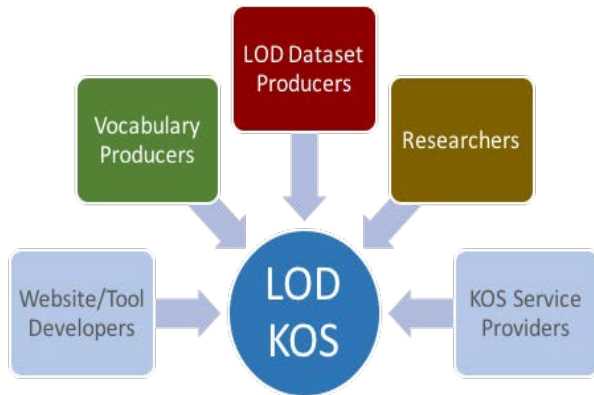
LOD Dataset Producers
Vocabulary Producers
Researchers
Website/Tool Developers
LOD KOS
KOS Service Providers

FAIR + FIT
Functional Metrics for
LOD KOS Products

Marcia Lei Zeng
College of Communication and Information (CCI)
Kent State University, USA

KO Research Observatory, 2021-10-27, ISKO-UK

Metrics development for LOD KOS

Outline

1. FAIR

- a LOD KOS **as an open dataset**

    Findable
    Accessible
    Interoperable
    Reusable

2. FIT

- a LOD KOS **as a value vocabulary**

    Functional
    Impactful
    Transformable

Zeng, M. L. & Clunis, J. (2020). FAIR+ FIT: guiding principles and functional metrics for linked open data (LOD) KOS products. *Journal of Data and Information Science*, 5(1), 93-118. https://doi.org/10.2478/jdis-2020-0008
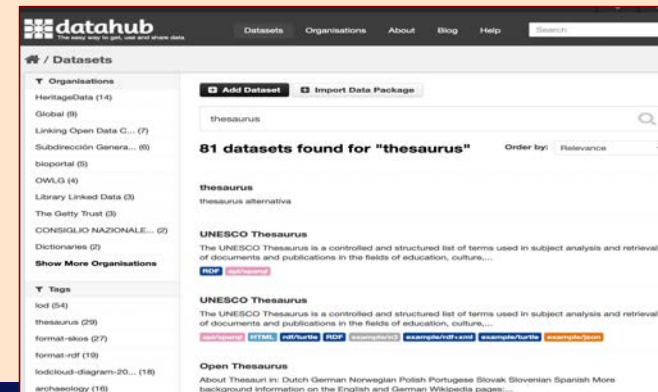
2

The metrics are developed based on:
- Data collected 2015-16, 2017-18, & 2019-20
    - Datahub and other vocab services
- Case studies on LOD KOS in the SW, 2015-2020

(Refer to the full paper: Zeng & Clunis, 2020)

LOD KOS found in the Datahub:
 -(2016): 1251 found
 -(2017): 1662 found   https://datahub.io/
 -(2021): 1176 found   & https://old.datahub.io/

with tags: "thesaurus", "classification", "taxonomy", "ontology", "terminology"
(Note: Some are tagged with multiple categories. Some have multiple editions.*)*
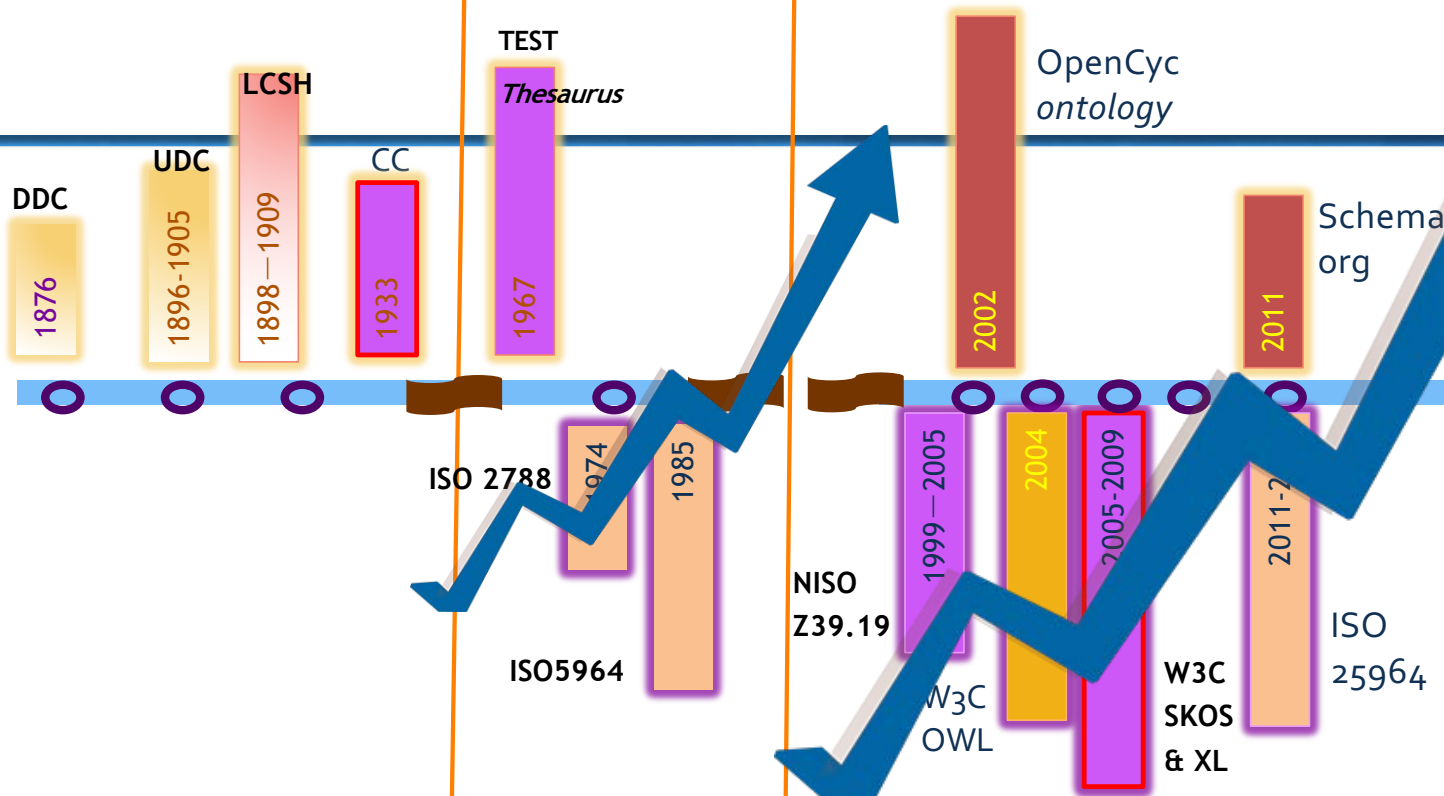
FAIR
– a **LOD KOS** as an open dataset

Findable
Accessible
Interoperable
Reusable

Growth of the KOS for information and knowledge organization tasks

KOS

Standards

Creator

|		
Classification, Subject headings	+ Thesaurus	++ Ontology, knowledge base
• LIS professionals	• LIS professionals • I&A database producers	• LIS professionals • I&A database producers • Anyone

Growth of the KOS for the information and knowledge organization tasks

Conventional KOS have always been quick adapters of
of new technologies in their publishing venues and applications

The release of a LOD KOS product represents a turning point for the producer or provider of a vocabulary.

RDF datasets

HTML Websites

XML files

CD-ROMs

I&A databases

printed
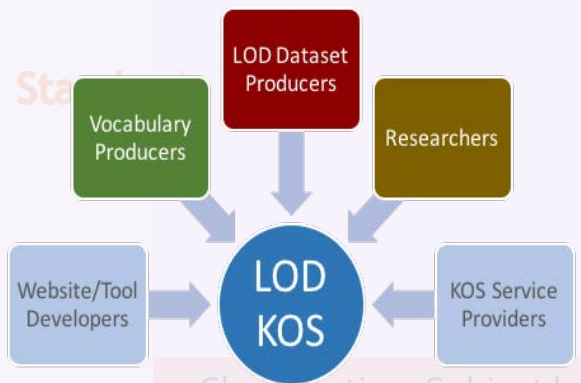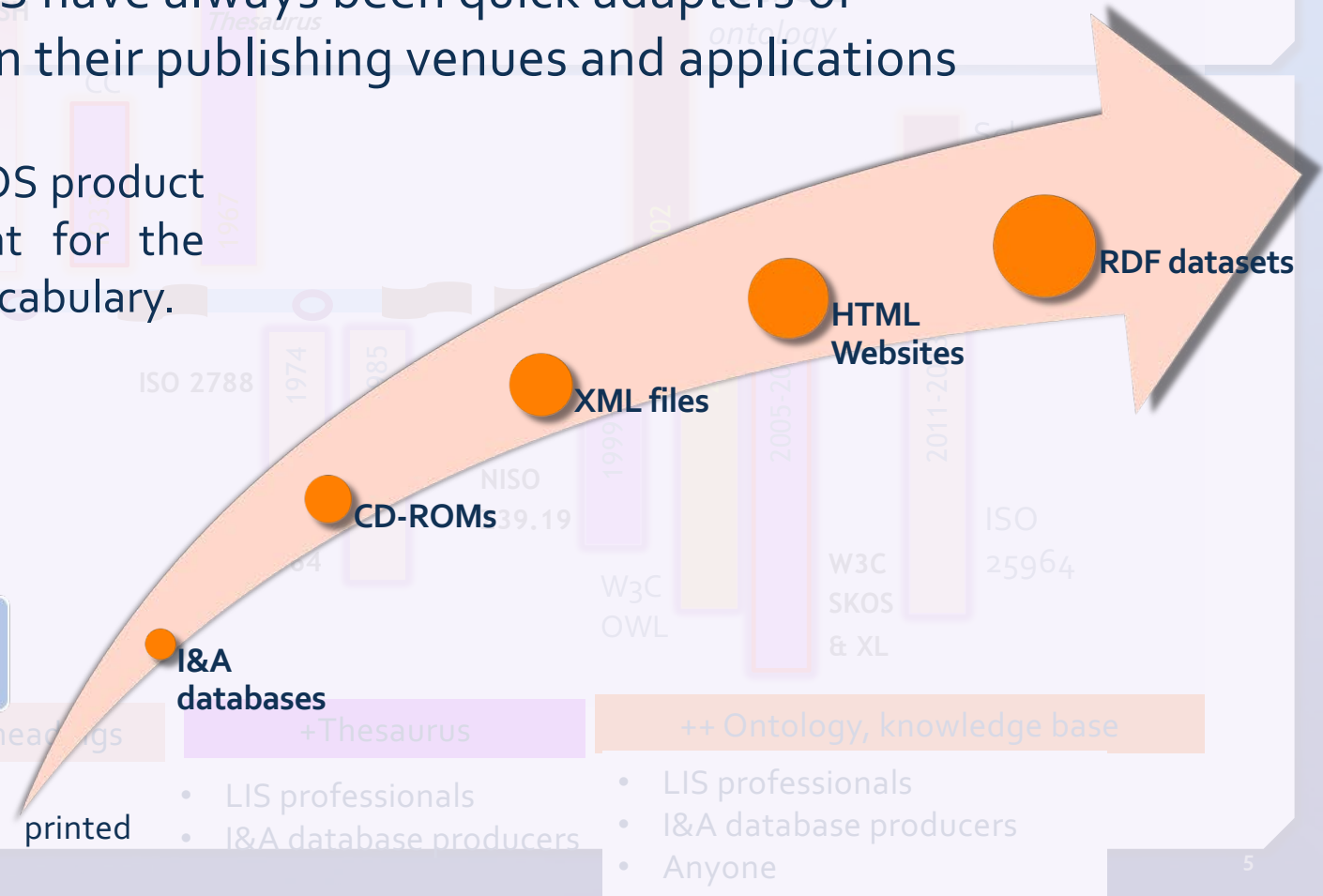
LOD KOS' FAIR: **F**indable

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.
FINDABLE

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

https://www.go-fair.org/fair-principles/

Examples from the datahub:
– Various levels of F[indable]

vs.

❖ Our additional recommendation:
⇒ Enrich metadata about KOS as much as possible to enable data discovery processes

LOD KOS' FAIR: **A**ccessible

Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.
ACCESSIBLE

Accessible

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A2. Metadata are accessible, even when the data are no longer available

https://www.go-fair.org/fair-principles/

❖ Our additional recommendation:

⇒ Provide multiple pathways for accessing KOS datasets

UNESCO Thesaurus
The UNESCO Thesaurus is a controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture...

Open Thesaurus

Data and Resources

SPARQL Endpoint
No description for this resource

Home page for browsing and searching
No description for this resource

Complete dataset in Turtle
No description for this resource

Complete dataset in RDF/XML
No description for this resource

Dataset description in VoID
No description for this resource

Example in HTML
No description for this resource

Example in Notation-3
No description for this resource

Example in RDF/XML
No description for this resource

Example in Turtle
No description for this resource

Unnamed resource
No description for this resource

Example in JSON
No description for this resource

Example in JSON-LD
No description for this resource

vs.

Data and Resources

Linked Data entry point (RDFa)
Linked Data entry point (RDFa)

Download (RDF/XML)
Download (RDF/XML)

7

LOD KOS' FAIR: Interoperable

Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
INTEROPERABLE

Interoperable

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (Meta)data use vocabularies that follow FAIR principles
I3. (Meta)data include qualified references to other (meta)data

https://www.go-fair.org/fair-principles/

❖ Our additional recommendation:
⇒ Utilize the *KOS Types Vocabulary** to standardize the way vocabulary types are categorized

*https://nkos.slis.kent.edu/nkos-type.html

BARTOC.org
Vocabularies 3391 vocabularies
Registries
Software
About
Contact

BARTOC unites information about vocabularies and knowledge organization systems. more...

Search | Filter

KOS Type
Languages — Gazetteer / Glossary / Categorization schema / Classification schema / List / Name authority list / Ontology / Subject heading scheme / Semantic network / Synonym ring / Taxonomy / Terminology / Thesaurus / Dictionary
License
Subject
Registry
Sorting

Examples from the datahub:

Search **Type of KOS/DATASET**	# found (initial)	# found (verified)
Authority Files	164	18
List	825	71
Terminology	39	35
Thesaurus	80	91
Taxonomy	37	22
Classification	478	43
Ontology	531	266
Totals	1623 (+531 ontologies)	280 (+ 266 ontologies)

Note: Some are tagged with multiple categories. Some have multiple editions.

LOD KOS' FAIR: **R**eusable

Data and collections have a clear usage licenses and provide accurate information on provenance.
REUSABLE

Reusable
R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
  R1.1. (Meta)data are released with a clear and accessible data usage license
  R1.2. (Meta)data are associated with detailed provenance
  R1.3. (Meta)data meet domain-relevant community standards

https://www.go-fair.org/fair-principles/

❖Our additional recommendation:
  ⇒ Adequately supply license and provenance metadata to enable datasets' reusability

Examples from the datahub:

provenance-metadata   publications   published-by-producer   rdf   thesaurus

vocab-mappings

vs.

lodcloud-diagram-20...   lodcloud-diagram-20...   no-license-metadata   no-proprietary-vocab

no-provenance-metadata   no-vocab-mappings   publications   published-by-producer

F I T
– a **LOD KOS** as a value vocabulary
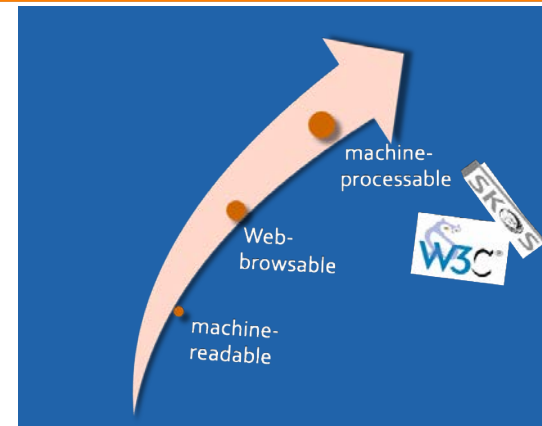
Functional
Impactful
Transformable

Metrics for LOD KOS

-- as a value vocabulary

Functional

[The vocabulary is...]
Made available in ways that enhance its inherent purpose

Metrics:

F1. The vocabulary is delivered in consumable formats

F2. The provided SPARQL endpoints are operational

F3. Dataset properties and structures are informed effectively

F4. Services are user-friendly, making vocabulary contents reachable

11

F1. The vocabulary is delivered in *consumable* formats

➢ **Available in various data serialization formats**

AGROVOC 2021-10

The October 2021 version of the AGROVOC thesaurus has been released. It now contains 39 500 concepts and 890 000 terms, in up to 41 languages.

NT RDF NQ ttl

consumable

Findings based on data from datahub:
Available serialization formats of KOS datasets
(sorted based on data collected 2019)

Functional

Format	2016	2017	2019
JSON	54	42	74
HTML	47	37	71
XML	55	42	69
TSV	44	30	63
RDF+XML	40	30	61
DEFAULT/AUTO	37	27	51
TURTLE	30	26	39
CSV	34	20	39
N-TRIPLES	26	18	36
JAVASCRIPT	23	11	31
SPREADSHEET	22	3	30
PLAIN/TEXT	20	21	28
QUERY STRUCTURE	15	15	23
SERIALIZED PHP	15	15	22
JSON-LD		3	1

Semantic Web standards enable → **LOD KOS**

- SKOS
- OWL
- RDFS
- SPARQL
- RDF

- Many KOS schemes have been turned into
  - OWL ontologies, or
  - SKOS-ified datasets
- Such datasets are usually available
  - as data dumps, or
  - through SPARQL endpoints.

➢ **Accessible through SPARQL**

with **SPARQL Endpoints**
*In addition to **data dumps** …*
- Data collected from datahub

2016			2017			2019		
Search Type of KOS/DATASET	# found	# with SPARQL endpoints	Search Type of KOS/DATASET	# found	# with SPARQL endpoints	Search Type of KOS/DATASET	# found	# with SPARQL endpoints
Thesaurus	67	39	Thesaurus	79	40	Thesaurus	80	41
Classification	458	29	Classification	476	31	Classification	478	31
Taxonomy	26	8	Taxonomy	35	8	Taxonomy	37	10
Terminology	35	7	Terminology	39	8	Terminology	39	8
List	665	52	List	821	58	List	825	59
Total	1251	135	Total	1450	145	Total	1459	149

❖ Our recommendation for F1:
⇒ a KOS vocabulary should be delivered in consumable formats: available in various data serialization formats and accessible through a SPARQL endpoint

13

F2. SPARQL endpoints are *operational*

Findings (2019): Near 80% are operational

Reality check

Number of Non-Functioning SPARQL endpoints

more than 20% are no longer working

after removing duplicates and not including ontologies:

Ensures sustainability!

❖Our recommendation for F2:
⇒ institutions should commit to ensuring the sustainability of access to their KOS dataset deliverables by providing a persistently available SPARQL endpoint

F3. Dataset properties and structures are *informed* effectively

UNESCO vocabularies - SPARQL service

http://vocabularies.unesco.org/sparql-form/

In order to master a query, one must understand:

- the syntax,
- forms,
- operators,
- result set modifiers,
- variables, and
- functions of the SPARQL query language.

+

- <u>properties</u> used by various involved RDF vocabularies

(cont.) **F3. Dataset properties and structures are *informed* effectively**

Reality check: For the KOSs which provided SPARQL Endpoints, are they **similar in structure** and do they **use a similar set of properties**?
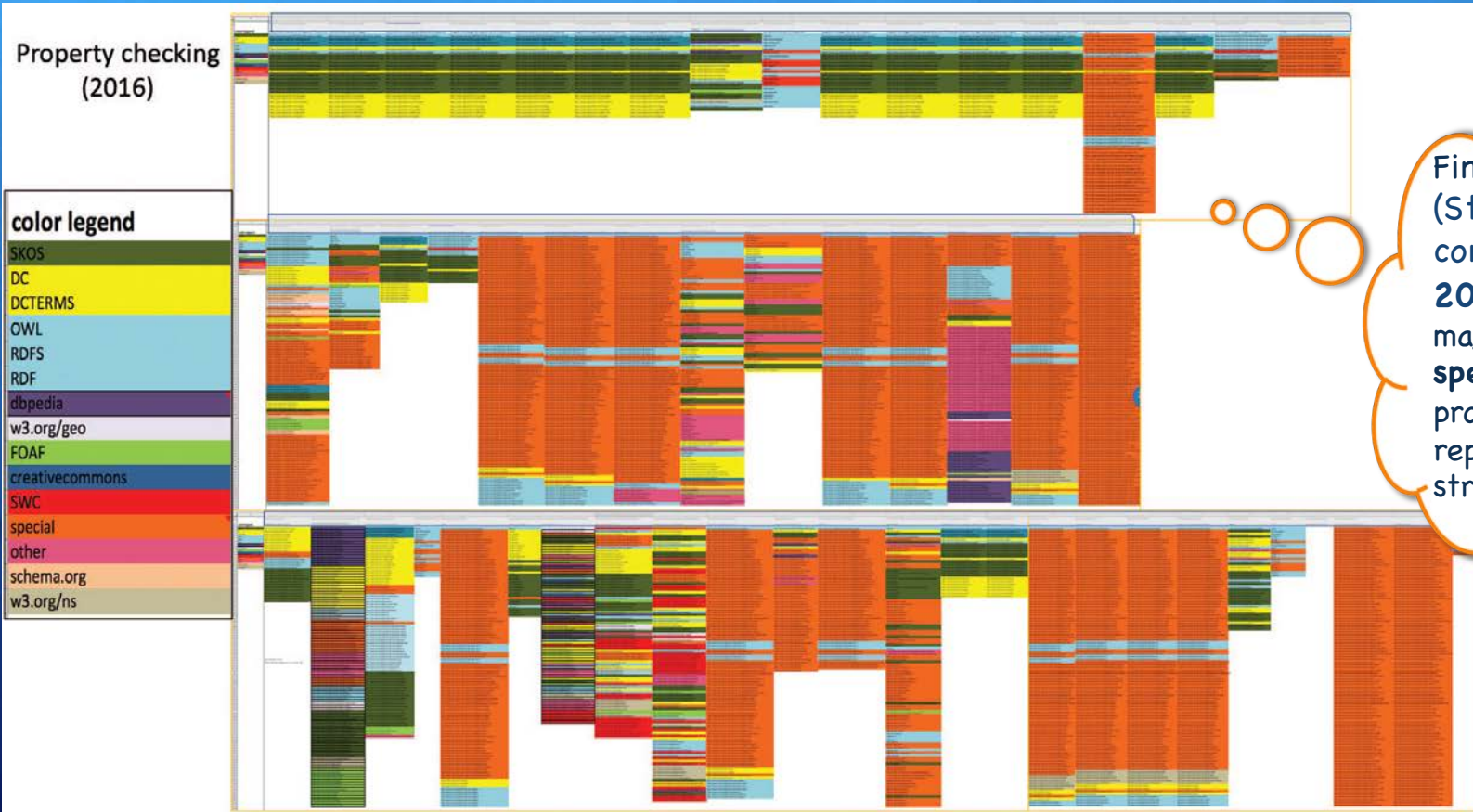
Finding (Study conducted in **2016**): The majority have **specialized** properties to represent their structure.

Property checking (2016)

color legend

SKOS
DC
DCTERMS
OWL
RDFS
RDF
dbpedia
w3.org/geo
FOAF
creativecommons
SWC
special
other
schema.org
w3.org/ns

Reality check: For the KOSs which provided SPARQL Endpoints, are they similar in structure and do they use a similar set of properties?

[Study conducted in **2019**.]

Property checking
(2019)

color legend
- DublinCore
- DBPedia
- SKOS
- OWL
- W3.org
- FOAF
- RDFS
- RDF
- schema.org
- creative commo
- special
- other
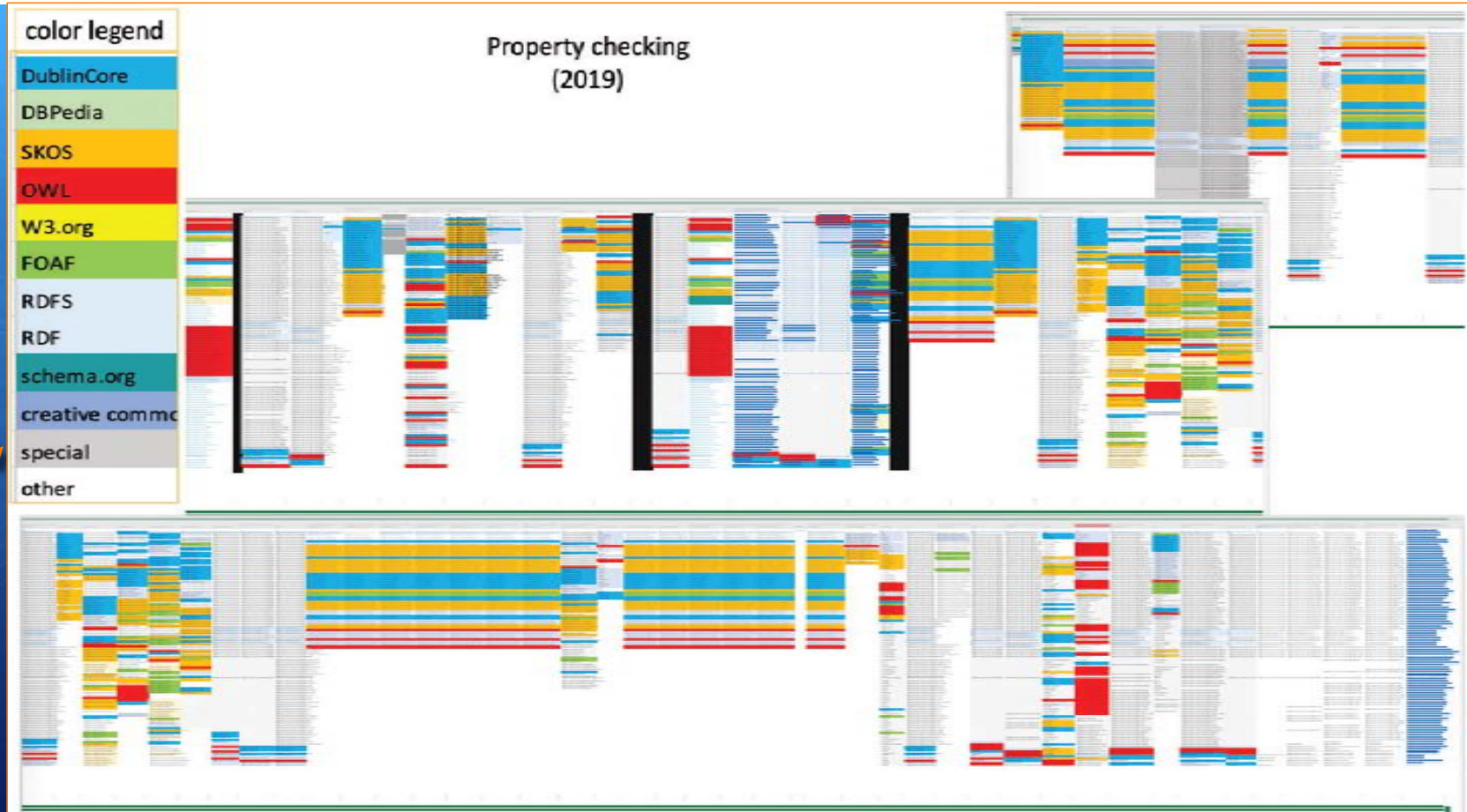
(cont.) **F3. Dataset properties and structures are *informed* effectively**

- **Contains multiple refined query examples to *inform* knowledge of <u>dataset properties</u> and <u>structures</u>**
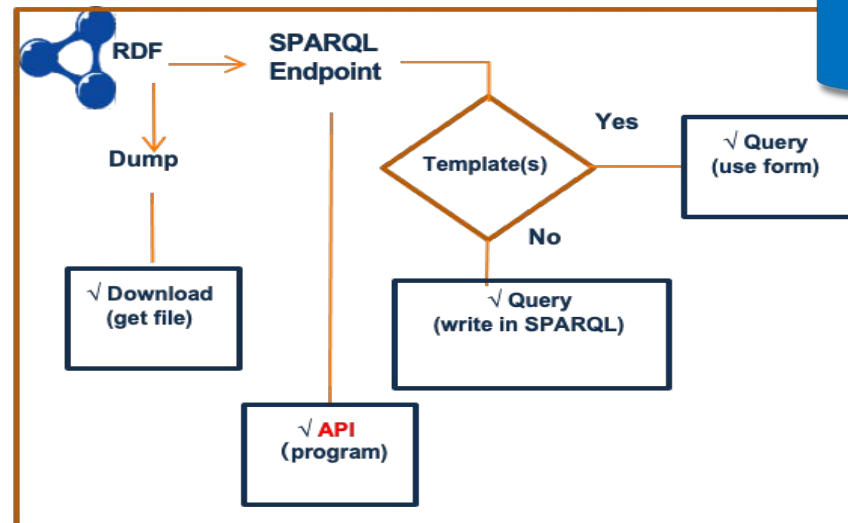
❖ Our recommendation for F3:

⇒ Dataset properties and structure information should be more effectively and readily available

⇒ A SPARQL service should at least contain refined query examples to reveal the internal structures of the datasets

F4. Services are user-friendly, making vocabulary contents reachable

Findings (2019) of 149 endpoints:
- 66 offered a default query.
- 26 offered example queries of any kind.
- 9 provided more than 3 example queries.

❖ We highly recommend KOS producers adopt and adhere to best practices like this to enhance usability

Query examples available by year

Year	# of datasets	Endpoint operational		# providing default query	# providing example queries	# providing more than 3 example queries
2019	1459	149		66 (41%)	26 (<20%)	9 (a few)
2017	1450	145		33	21	10
2016	1251	135		-	16	6

(cont.) **F4. Services are user-friendly, making vocabulary contents reachable**

❖ Our recommendation for F4:
⇒ Datasets with SPARQL endpoints should provide query examples or forms and templates to enable the easy creation of queries allowing users to interact with the data.

Summary

Functional

[The vocabulary is…]
Made available in ways that enhance its inherent purpose

When a KOS is <u>F</u>unctional, it could further its <u>I</u>mpacts and potential <u>T</u>ransformable usages

21

Metrics:

F1. The vocabulary is delivered in consumable formats

F2. Provided SPARQL endpoints are operational

F3. Dataset properties and structures are informed effectively

F4. Services are user-friendly, making vocabulary contents reachable

Metrics development for LOD KOS

(cont.) F I T
– a LOD KOS as a value vocabulary

Functional
Impactful
Transformable

Metrics for LOD KOS

-- as a value vocabulary

Impactful

Maximizes the impact of a LOD KOS vocab

Metrics:

I1. Exposed through terminology services

I2. Used by data providers

    a) as a primary value Vocab

    b) in semantic enrichment

I3. Mapped with other KOS vocabs

I4. Showed/discussed at professional conferences and publications

I1. Exposed through terminology services

a. **Registry of KOS**

- ✓ **BARTOC** (Basel Register of Thesauri, Ontologies & Classifications)
  - • 3390+ https://bartoc.org/
- ✓ **Taxonomy warehouse** http://www.taxonomywarehouse.com/default.aspx
- ✓ **Taxobank**: 2000 vocabularies http://www.taxobank.org/

b. **Registry of LOD vocabularies ("property vocabularies" & "value vocabularies")**

- ➢ **LOV** (Linked Open Vocabularies) http://lov.okfn.org/dataset/lov
  - • 760+ registered, some are value vocabularies
- ➢ **BioPortal** https://bioportal.bioontology.org/
  - • 900+ ontologies, some are classifications and thesauri

c. **Registry of LOD products, including KOS**

- ❖ **DataHub** https://datahub.io/ & https://old.datahub.io/

I2. Used by data providers
a) as a primary value vocabulary

Impactful

Projects using MESH ⊕

ARRS GoldMiner

Biomedical Semantic QA

Cell line ontology

DisGeNET-RDF

eagle-i

Epidemic Marketplace

Kino

Lexigram

Neuronal Morphologies and Species Metadata Classification System

Plant Ontology

PubChem

Retrospective Analytics System

Semantic Indexing of French Biomedical Data Resources

Socrates MD

The Ontological Discovery Environment

http://bioportal.bioontology.org/ontologies/MESH

Impactful

b) used in semantic enrichment

Europeana Dereferenceable vocabularies

The Getty - Union List of Artist Names (ULAN)	edm:Agent
Virtual International Authority File (VIAF)	edm:Agent
Wikidata	edm:Agent
Gemeinsame Normdatei (GND)	edm:Agent, edm:Place, skos:Concept
Getty Thesaurus of Geographic Names (TGN)	edm:Place
Geonames	edm:Place
The Getty - Art & Architecture Thesaurus (AAT)	skos:Concept
IconClass	skos:Concept
Israel Museum Jerusalem Concepts	skos:Concept
Library of Congress Subject Headings (LCSH)	skos:Concept
data.europeana.eu WWI Concepts from Library of Congress Subject Headings (LCSH)	skos:Concept
Europeana Sounds Genres	skos:Concept
UDC	skos:Concept
UNESCO Thesaurus	skos:Concept
YSO - General Finnish ontology	skos:Concept
Fashion Thesaurus	skos:Concept
MIMO Concepts	skos:Concept

agents
places
concepts

[Europeana enriches <u>xxx</u> by aligning to (xxx)]
agent names →
places →
concepts →
time period → (Semium Time).

✓ Relate objects to concepts, agents, places, etc., using the properties in the Europeana Data Model (EDM) (e.g., *dc:subject, dc:creator*).

***Update* from Europeana**
enriched in 2021 (by May):

Type	June 19	Nov 20	May 21
Place	14,019,045	15,206,861	16,101,291
Concept	15,734,342	15,218,899	15,765,470
Agent	1,309,614	1,429,242	1,311,327
Time Period	17,406,100	18,597,882	16,452,469

Europeana semantic enrichment - - >
Europeana Semantic Enrichment Framework

- Source: Europeana Semantic Enrichment Framework *Documentation*
Version: 17th November 2016 (updated 2017, 2018, 2020, 2021).
Available from https://pro.europeana.eu/page/**europeana-semantic-enrichment**
--> several vocabularies (Compiled by MZ 2020-11-18, validated 2021-10-21)

I3. Mapped with other KOSs

Alignments require **interoperability** in **syntax** & **structure**

Impactful

STW Thesaurus for Economics

STW Mappings

Here you find mappings to other thesauri and vocabularies, which can also be downloaded.

- Integrated Authority File (GND)
- Wikidata
- DBpedia
- EuroVoc
- Thesaurus Social Sciences (TheSoz)
- AGROVOC
- WKD German labor law thesaurus
- JEL classification
- SDMX subject-matter domains classification

https://zbw.eu/stw/version/latest/mapping/about.en.html

Table 1 below provides some figures about the vocabularies to which AGROVOC is aligned (December 2020):

AGROVOC

Vocabulary	Alignments
CAAS	20704
NALT	14922
DBPEDIA	11059
DNB	6211
BNCF	1967
EARTh	1389
EUROVOC	1381
GEMET	1235
STW	1135
LCSH	1090
TheSoz	847
RAMEAU	671
WORLDBANK	601
GBIF	234
GeoNames	206
UNBIS	186
CRAI	69
CALATHE	
ICD	
CABI	
WIKIDATA	
UKAT	
UNESCO	
INRA	
SDG	

http://aims.fao.org/standards/agrovoc/linked-data

📦 Asset

EuroVoc
Version: 20210604–0 LAT
URI: http://publications.europa.eu/resource/dataset
Type of dataset: Thesaurus

About | Downloads | Documentation | Alignments

- 📄 EuroVoc Alignment ThesSoz
- 📄 EuroVoc Alignment ESCO
- 📄 EuroVoc Alignment Agrovoc
- 📄 EuroVoc Alignment UMTHES
- 📄 EuroVoc Alignment Unbis
- 📄 EuroVoc Alignment LCSH
- 📄 EuroVoc Alignment Eige
- 📄 EuroVoc Alignment Rameau
- 📄 EuroVoc Alignment Unesco
- 📄 EuroVoc Alignment Eclas
- 📄 EuroVoc Alignment gnd
- 📄 EuroVoc Alignment Inspire
- 📄 EuroVoc Alignment ZBW
- 📄 EuroVoc Alignment Gemet
- 📄 EuroVoc Alignment mesh
- 📄 EuroVoc Alignment WikiData
- 📄 EuroVoc Alignment Country

https://publications.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc

I4. Showed/discussed at professional conferences and publications

-- as a KOS vocabulary

- NKOS workshops

- LODLAM Summit

- ISKO and ISKO-chapter events

- DCMI conferences

- Books and journal articles

- … …

Summary

Impactful

Maximizes the impact of a LOD KOS vocab

Impactful

Metrics:

I1. Exposed through terminology services

I2. Used by data providers

    a) as a primary value vocab

    b) in semantic enrichment

I3. Mapped with other KOS vocabs

I4. Showed/discussed at professional conferences and publications

2. F I T
– a **LOD KOS** as a value vocabulary

Functional
Impactful
Transformable

Metrics for LOD KOS

-- as a value vocabulary

Transformable

Extends the functionality and impact through innovative adaptations

Metrics:

T1. Allows special KOS products to be derived from the original data

T2. The user is given autonomy to determine what structure and information is desired and can be reproduced from the vocabulary

T3. Enables extensibility to fit diverse needs

T4. Supports innovative and transformative uses beyond normal "value vocabularies"

T1. Allows special KOS products to be *derived* from the original data

Transformable

Top - -down

About 100 micro-thesauri can be obtained

http://vocabularies.unesco.org/sparql-form/

UNESCO vocabularies - SPARQL service

Default graph (IRI)

Query

```
1  PREFIX skos: <http://www.w3.org/2004/02/skos/core#> PREFIX isothes: <http://purl.org
   /iso25964/skos-thes#> SELECT ?domainNotation (STR(?domainLabel) AS ?domainEnglish)
   ?mtNotation (STR(?englishLabel) AS ?english) WHERE {
2    ?collection a <http://vocabularies.unesco.org/ontology#MicroThesaurus> .
3    ?collection isothes:superGroup ?domain .
4    ?collection skos:notation ?mtNotation .
```

Query exemples

derivable

- Explore a sample of the data
- List all concepts of a micro-thesaurus in french
- List all concepts of a domain
- List all concepts
- List all micro-thesauri
- List all the translations english-french
- List the translations english-russian
- List concepts created after a given date
- Get the list of countries
- Select all the properties of a concept
- Make a search on all the concept labels
- Get all concepts in english and french, with synonyms, notes, broaders, narrowers and related (with IDs)
- Get the hierarchical table of all the concepts (with IDs)
- Get the hierarchical table of all the concepts (with labels)

Result format: HTML

- HTML
- Text
- XML
- JSON
- NTriples
- RDF/XML
- CSV
- TSV

HTML

Run Query   Reset

"6"	"Politics, law and economics"	"6.05"	"Legal systems"
"6"	"Politics, law and economics"	"6.10"	"Human rights"
"6"	"Politics, law and economics"	"6.25"	"Economics"
"6"	"Politics, law and economics"	"6.30"	"Economic and social development"
"6"	"Politics, law and economics"	"6.35"	"Agriculture"
"6"	"Politics, law and economics"	"6.40"	"Industry"
"6"	"Politics, law and economics"	"6.45"	"Civil, military and mining engineering"
"6"	"Politics, law and economics"	"6.50"	"Manufacturing and transport engineering"
"6"	"Politics, law and economics"	"6.55"	"Materials and products"
"6"	"Politics, law and economics"	"6.60"	"Equipment and facilities"
"6"	"Politics, law and economics"	"6.65"	"Services"
"6"	"Politics, law and economics"	"6.70"	"Finance and trade"
"6"	"Politics, law and economics"	"6.75"	"Organization and management"
"6"	"Politics, law and economics"	"6.80"	"Personnel management"
"6"	"Politics, law and economics"	"6.85"	"Labour"

http://vocabularies.unesco.org/sparql-form/
Image captured 2019-08-21

Transformable

Query: http://vocab.getty.edu/sparql

```
1 select * {?x gvp:broaderExtended aat:300194567; skos:inScheme aat: ;
2 gvp:prefLabelGVP/xl:literalForm ?l}
```

bottom - - up

reproducible

☑ Include inferred

☐ Expand results over equivalent URIs

Submit

2.2 Descendants of a Given Parent

Let's look for AAT descendants of 300194567 "drinking vessels". This finds "rhyta" and other interesting records, including "Fichtelgebirgehumpen":

```
select * {?x gvp:broaderExtended aat:300194567; skos:inScheme aat: ; gvp:prefLabelGVP/xl:literalForm ?l}
```

SPARQL

2.3 Subjects by Contributor Id

You can easily find subjects contributed by a particular Contributor if you know the id. E.g. the Getty Conservation Institute (GCI) in AAT is aat_contrib:10000088. Let's find their contributions to aat:300033618 paintings (visual works):

```
select * {
  ?x a gvp:Subject; dct:contributor aat_contrib:10000088;
    gvp:broaderExtended aat:300033618;
    gvp:prefLabelGVP/xl:literalForm ?l}
```

T2. The user is given *autonomy* to determine what structure and information is desired and can be reproduced

Transformable

Endless reuse potentials.

AAT descendants of 300194567 "drinking vessels"

AAT descendants of 300212133 "costume by function"

T3. *Extended* to fit diverse needs

➢ **Extended to fit diverse needs**

- Culture

- Language

- Domain

- Structure

➢ **Virtual harmonization through linking**

- E.g., *Faceted Application of Subject Terminology* (FAST)
  → VIAF URI (schema:sameAs)
  & Wikidata URI (foaf:focus)

T4. Supports *innovative* **transformative uses beyond normal "value vocabularies"**

➢ LOD KOS can be used for
  ➢ obtaining special graphs or datasets for very complicated questions
  ➢ revealing unknown relationships

beyond being a "vocabulary"

Could a LOD KOS dataset be considered
  ➢ a knowledge base?
  ➢ the foundation of a network analysis?
  ➢ the building blocks of a framework for research in humanities and science?

Query examples can lead users to explore he rich contents of the datasets

Transformable

Examples

1. Select all taxa from the UniProt taxonomy: (show)
2. Select all bacterial taxa, and their scientific name, from the UniProt taxonomy: (show)
3. Select all E-Coli K12 (including strains) UniProt entries and their amino acid sequence: (show)
4. Select the UniProt entry with the mnemonic 'A4_HUMAN': (show)
5. Select a mapping of UniProt to PDB entries using the UniProt cross-references to the PDB database: (show)
6. Select all cross-references to external databases of the category '3D structure databases' of UniProt entries that are classified with the keyword '3Fe-4S': (show)
7. Select all UniProt entries, and their recommended protein name, that have a preferred gene name that contains the text 'DNA': (show)
8. Select the preferred gene name and disease annotation of all human UniProt entries that are known to be involved in a disease: (show)

innovative

9. Select all human UniProt entries with a sequence variant that leads to a 'loss of function': (show)
10. Select all human UniProt entries with a sequence variant that leads to a tyrosine to phenylalanine substitution: (show)
11. Select all UniProt entries with annotated transmembrane regions and the regions' begin and end coordinates on the canonical sequence: (show)
12. Select all UniProt entries that were integrated on the 30th of November 2010: (show)
13. Was any UniProt entry integrated on the 9th of January 2013? (show)
14. Construct new triples of the type 'HumanProtein' from all human UniProt entries: (show)
15. Select all triples that relate to the EMBL CDS entry AA089367.1: (show)
16. Select all triples that relate to the taxon that describes *Homo sapiens*: (show)
17. Select the average number of cross-references to the PDB database of UniProt entries that have at least one cross-reference to the PDB database: (show)
18. Select the number of UniProt entries for each of the EC (Enzyme Commission) second level categories: (show)

UniProt SPARQL Downloads

http://sparql.uniprot.org/

Your SPARQL query

Add common prefixes

```
1  PREFIX up:<http://purl.uniprot.org/core/>
2  PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
3  PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
4  PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
5  SELECT ?name ?text
6  WHERE
7  {
8      ?protein a up:Protein .
9      ?protein up:organism taxon:9606 .
10     ?protein up:encodedBy ?gene .
11     ?gene skos:prefLabel ?name .
12     ?protein up:annotation ?annotation .
13     ?annotation a up:Disease_Annotation .
14     ?annotation rdfs:comment ?text
15 }
16
```
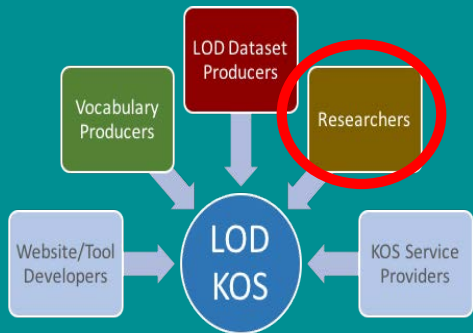
UniProt
Universal Protein Resource
- with many entries being derived from genome sequencing projects

(cont.) T4. Supports *innovative* transformative uses

LOD KOS products can be transformed **beyond** being just "value vocabularies"

They can become knowledge bases and provide semantic-rich discoveries

Getty Vocabularies: LOD   SF

Transformable

4.20   Places Nearby Each Other
5   ULAN-Specific Queries
5.1   Agents by Type
5.2   Associative Relations of Agent
5.3   Female Artists
5.4   Female Artists as a Hobby
5.5   Native American Painters
5.6   Names of Native American Painters
5.7   Architects Born in the 14th or 15th Century
5.8   Indian and Pakistani Architectural Groups
5.9   Non-Italians Who Worked in Italy
5.10   Artists Associated to a Given Patron or His Family
5.11   German, Dutch, Flemish printmakers, listed with their teachers
5.12   Artists Whose Identity May be Associated or Confused With Another
5.13   Ordered Hierarchy of Given Subject
5.14   Ancient Artists or Groups by Nationality
5.15   Art Repositories in the USA by State
5.16   Popes and Their Reigns
5.17   Pope Reign Durations
5.18   Life Events

Union List of Artist Names (ULAN) – Specific Queries

http://vocab.getty.edu/queries#ULAN-Specific_Queries

ocab.getty.edu/queries

Name authorities offer foundational structured data for network analyses

innovative

38

Summary

LOD KOS

Transformable

Extends the functionality and impact through innovative adaptations

Metrics:

T1. Allows special KOS products to be derived from the original data

T2. The user is given autonomy to determine what structure and information is desired and can be reproduced from the vocabulary

T3. Enables extensibility to fit diverse needs

T4. Supports innovative and transformative uses beyond normal "value vocabularies"

Conclusion   **F I T**   Metrics for LOD KOS (as value vocabularies)

➢ Functional

[The vocabulary is…]

Made available in ways that enhance its inherent purpose

Metrics:

F1. The vocabulary is delivered in consumable formats

F2. Provided SPARQL endpoints are operational

F3. Dataset properties and structures are informed effectively

F4. Services are user-friendly, making vocabulary contents reachable

➢ Impactful

[The vocabulary…]

Maximizes the impact of a LOD KOS vocab

Metrics:

I1. Exposed through terminology services

I2. Used by data providers

    a) as a primary value Vocab

    b) in semantic enrichment

I3. Mapped with other KOS vocabs

I4. Showed/discussed at professional conferences and publications

➢ Transformable

[The vocabulary…]

Extends the functionality and impact through innovative adaptations

Metrics:

T1. Allows special KOS products to be derived from the original data

T2. The user is given autonomy to determine what structure and information is desired and can be reproduced from the vocabulary

T3. Enables extensibility to fit diverse needs

T4. Supports innovative and transformative uses beyond normal "value vocabularies"

FAIR + F I T   Metrics for LOD KOS

FAIR as a dataset

- ➢ **F**indable

- ➢ **A**ccessible

- ➢ **I**nteroperable

- ➢ **R**eusable

FIT as a value vocabulary

➢ **Functional**
- • Consumable
- • Operational
- • Use-friendly, Reachable
- • Informative

[The vocabulary …]
Is made available in ways that enhance its inherent purpose

➢ **Impactful**
- • Exposed
- • Used
- • Mapped
- • Showed /Discussed

[The vocabulary…]
Maximizes the impact of a LOD KOS vocab

➢ **Transformable**
- • Derivable
- • Autonomous
- • Extendable
- • Innovative

[The vocabulary…]
Extends the functionality and impact through innovative adaptations

41

Full paper: **FAIR + FIT : Guiding Principles and Functional Metrics for Linked Open Data (LOD) KOS Products**

Marcia Lei Zeng & Julaine Clunis
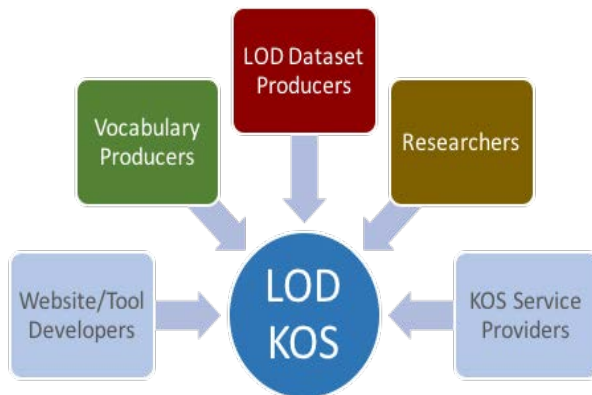*Journal of Data and Information Science* (JDIS) 5(1), 2020.
https://doi.org/10.2478/jdis-2020-0008

Thank you!

FAIR + FIT
Functional Metrics for
LOD KOS Products

Marcia Lei Zeng
College of Communication and Information (CCI)
Kent State University, USA

References

➢ European Commission Expert Group on FAIR Data. (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data.* Publications Office of the European Union. https://op.europa.eu/s/nF4t

➢ FORCE11. [2014]. *Guiding principles for findable, accessible, interoperable and re-usable data publishing* version b1.0. FORCE11. https://www.force11.org/fairprinciples

➢ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 160018. doi:10.1038/sdata.2016.18

➢ Zeng, M. L. & Clunis, J. (2020). FAIR+ FIT: guiding principles and functional metrics for linked open data (LOD) KOS products. *Journal of Data and Information Science*, 5(1), 93-118. https://doi.org/10.2478/jdis-2020-0008

➢ Zeng, M. L. & Mayr. P. (2018). Knowledge Organization Systems (KOS) in the Semantic Web. *International Journal on Digital Libraries, 20*(3), 209-230. https://doi.org/10.1007/s00799-018-0241-2