

# Linked data for libraries

## An introduction

**By Ki Tat LAM**

Formerly Associate University Librarian, Hong Kong University of Science and Technology

ORCID: <https://orcid.org/0000-0003-2625-9419>

# Agenda

- What is linked data?
- Why linked data for libraries?
- Examples
- Linked data basics
  - RDF triple, serialization, triplestore, SPARQL and vocabulary

## Semantic web and linked data

**Linked data** was coined in 2006 by **Tim Berners-Lee**, the founder of the World Wide Web.

He advocated the idea of **semantic web**, in which resources (data) are published and linked on the Web according to a set of principles.

### Linked data principles

- Use **URI**s as names for **things**
- Use **HTTP** URIs so that people can **look up** those names
- When someone looks up a URI, **provide useful information, using the (web) standards** (RDF, SPARQL)
- **Include links** to other URIs so that they can **discover** more things

<https://www.w3.org/DesignIssues/LinkedData.html>

## Linked data facilitates semantic inference

**Given** the following knowledge that is structured as linked data (each line is a triple):

```
:oxford :locatedIn :oxfordshire.  
:oxfordshire :locatedIn :england.  
:england :locatedIn :uk.
```

and **a rule**:

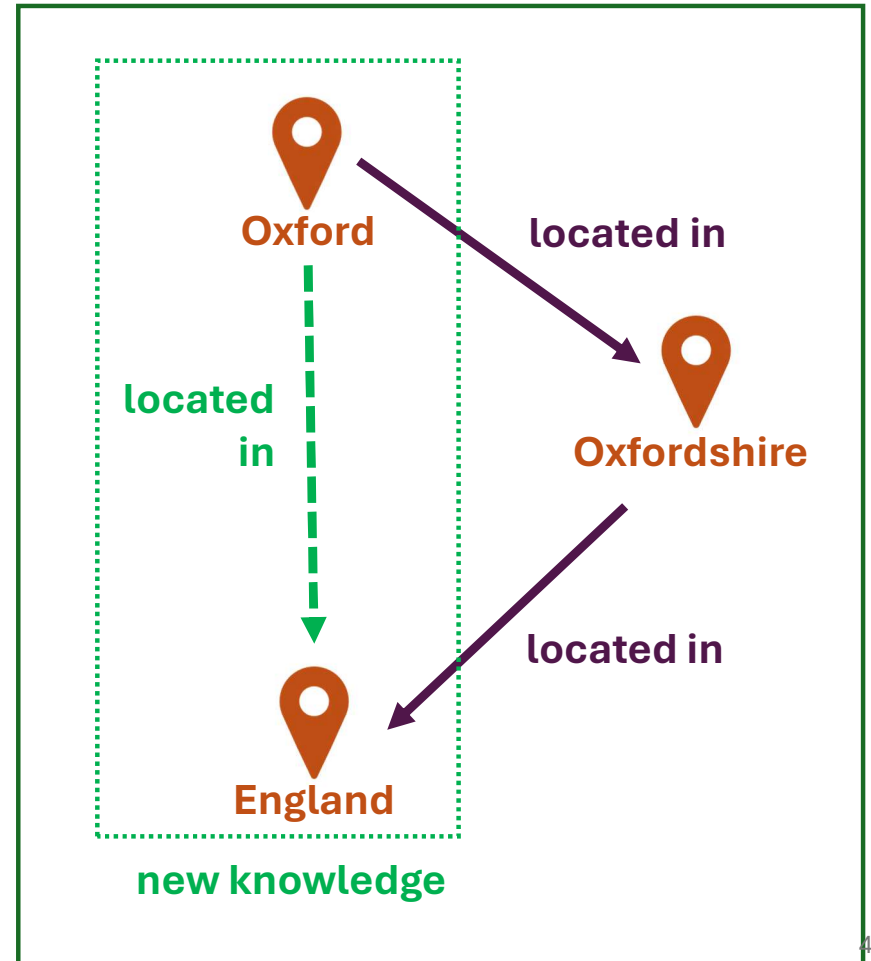
**IF (:a :locatedIn :b) AND (:b locatedIn :c)  
THEN (:a :locatedIn :c)**

then machines can **infer** and generate the following **new knowledge**:

```
:oxford :locatedIn :england.  
:oxford :locatedIn :uk.  
:oxfordshire :locatedIn :uk.
```

Example from: <https://docs.oxfordsemantic.tech/reasoning.html>

## Reasoning by inference



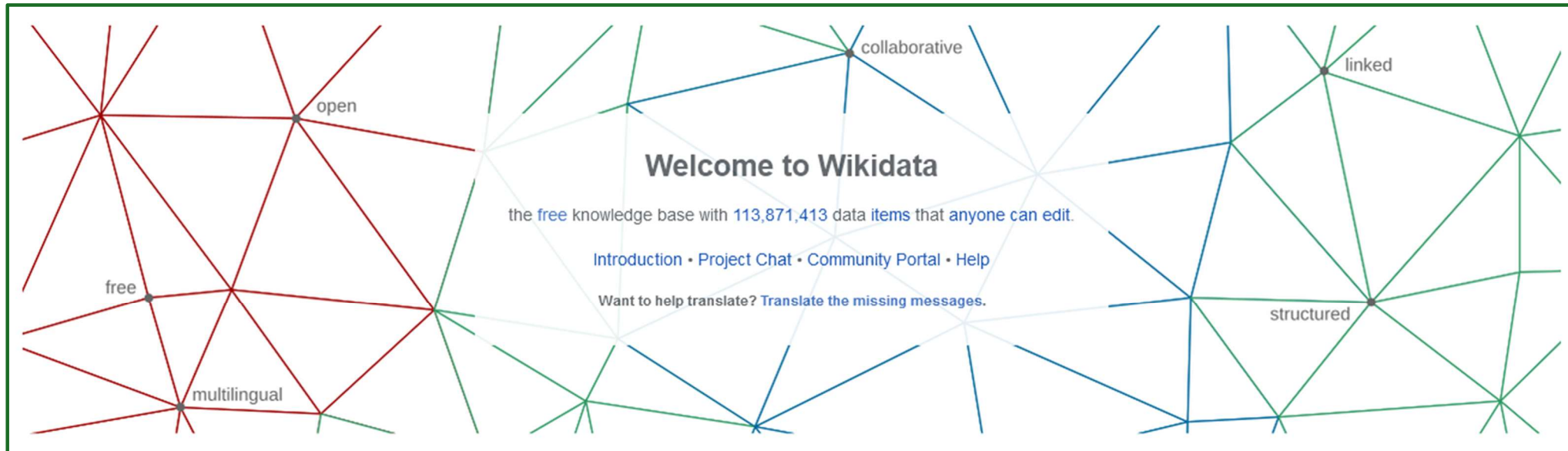
## Wikidata – an example of linked data



<https://www.wikidata.org>

**Wikidata** is a huge dataset published as linked data on the Web. It is also known as **Linked Open Data** as it is openly accessible by anyone.

**An open, free, multilingual, structured, linked and collaborative knowledge base**

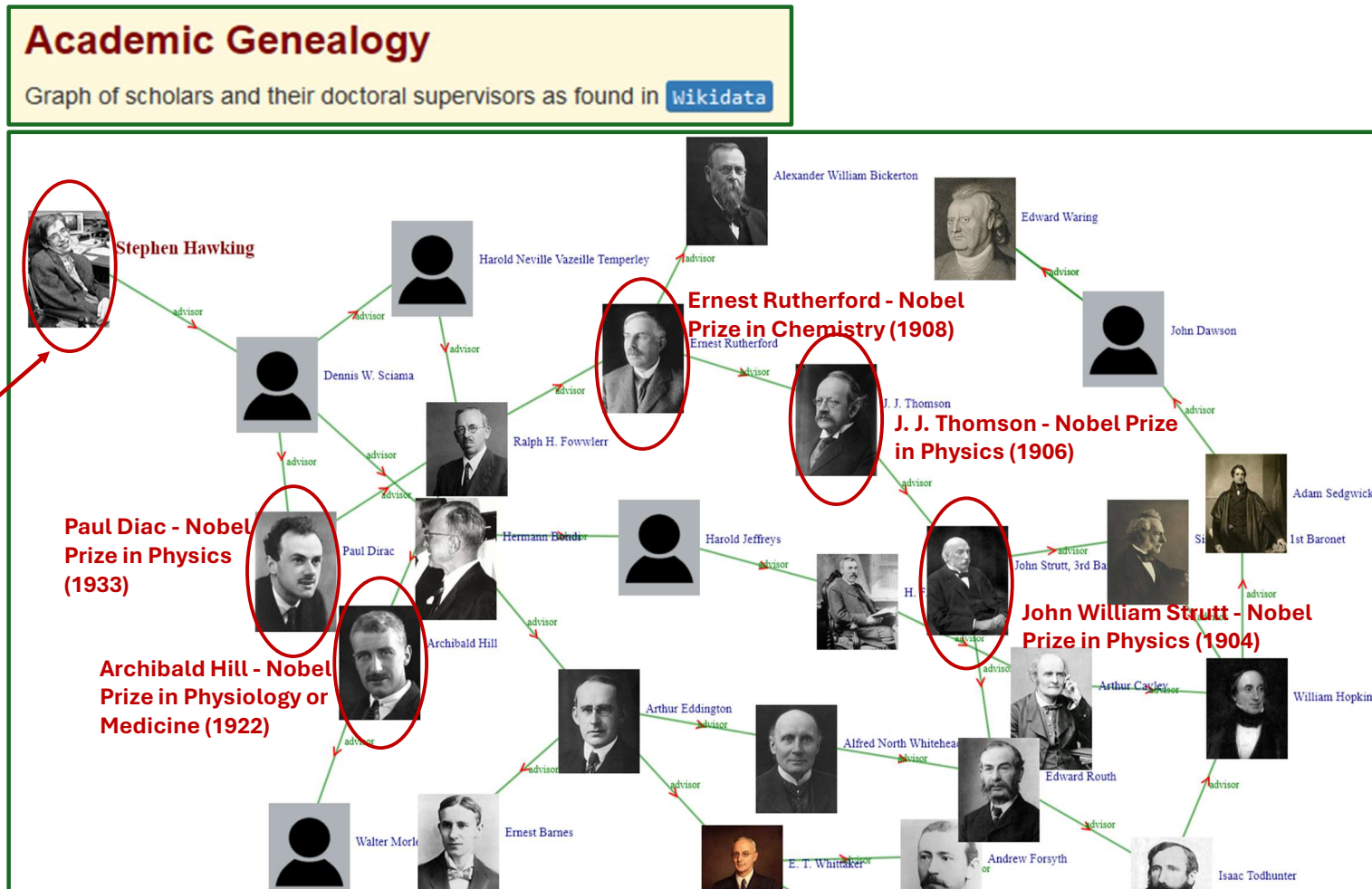


**Example:** Item of **Stephen Hawking**: <http://www.wikidata.org/entity/Q17714>

# Aggregating and blending linked data

Data from linked data sources, such as Wikidata, can be **aggregated** to give **new insights**.


**Academic genealogy of Stephen Hawking**



<https://www.sls.org.hk:8443/ktdemo/ag/ag.php?qnum=q17714>

## Aggregating and blending linked data [cont.]

**Scholia** – aggregates scholarly information in Wikidata



Scholia is a service that creates visual scholarly profiles for [topics](#), [people](#), [organizations](#), [species](#), [chemicals](#), etc using bibliographic and other information in Wikidata. [More info...](#)

Scholia relies on Wikidata, and Wikidata contains only a limited albeit growing subset of the corpus of scholarly literature, its authors and citations. Read more about the limitations in the [FAQ](#) or check the [statistics](#).






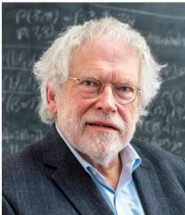


### Search

<https://scholia.toolforge.org>

### Nobel Prize in Physics

<https://scholia.toolforge.org/award/Q38104>

Images of recipients

 <small><a href="#">commons:John Hopfield 2016.jpg</a> 2024 <a href="#">John Hopfield</a></small>	 <small><a href="#">commons:Geoffrey Hinton at Collision ...</a> 2024 <a href="#">Geoffrey Hinton</a></small>	 <small><a href="#">commons:US Embassy Sweden 20...</a> 2023 <a href="#">Pierre Agostini</a></small>	 <small><a href="#">commons:Anne L'Huillier 01.JPG</a> 2023 <a href="#">Anne L'Huillier</a></small>
 <small><a href="#">commons:Ferenc Krausz (cropp...</a> 2023 <a href="#">Ferenc Krausz</a></small>	 <small><a href="#">commons:A. Zeilinger (cropped).jpg</a> 2022 <a href="#">Anton Zeilinger</a></small>	 <small><a href="#">commons:John Francis Clauser (...</a> 2022 <a href="#">John Clauser</a></small>	 <small><a href="#">commons:Alain Aspect.jpg</a> 2022 <a href="#">Alain Aspect</a></small>

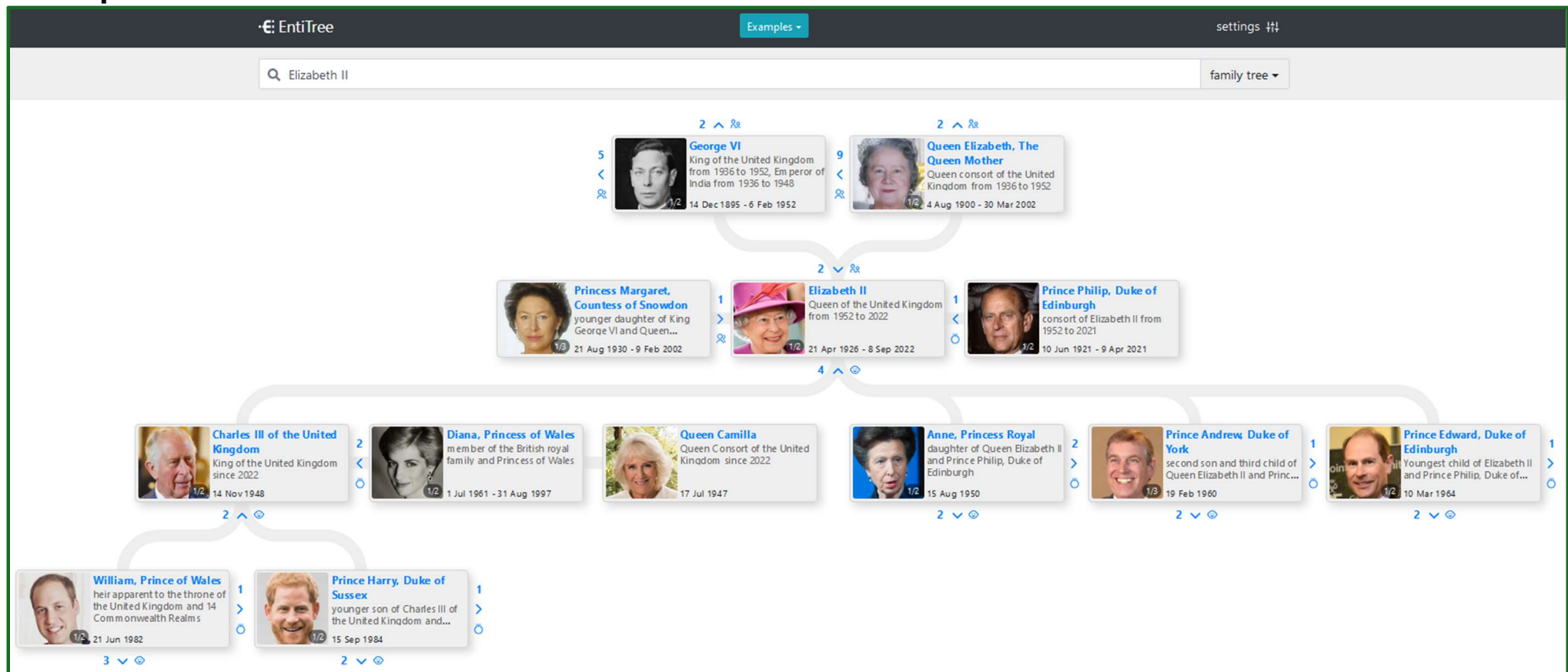
Wikidata Query Serviceaward: images-of-recipients.sparql

# Aggregating and blending linked data [cont.]

**EntiTree** – an application that blends Wikidata content into its websites

## Example: Elizabeth II

[https://www.entitree.com/en/family\\_tree/Elizabeth\\_II](https://www.entitree.com/en/family_tree/Elizabeth_II)



# Google Knowledge Graph

Google search  
result page –  
searching “winter  
solstice 2024”

Google maintains a  
database of linked  
(structured) data known as  
**Knowledge Graph**.

It serves content of  
knowledge panel and other  
rich results that are found in  
a Google Search Result  
Page.

The screenshot shows a Google search result for "winter solstice 2024". The search bar at the top contains the query. Below the search bar, the main result is for "Winter solstice / Date (2024)" with the date and time "Sat, Dec 21, 2024, 5:19 PM" in Hong Kong Standard Time. To the right of this is a "Winter solstice (December 21)" knowledge panel. The panel includes a globe image and text explaining the solstice, its date, and its significance. Below the main result, there is a "People also search for" section with links to "Summer solstice", "Qiyun", "Lidong", "March Equinox", "Chunfer", "Autumn Equinox", and "Xiaohan". At the bottom, there is a "People also ask" section with questions like "What is the shortest day in 2024?", "What is the darkest day of 2024?", "Is the winter solstice always on the 21st?", and "What is the winter solstice in Australia?". Red circles and arrows highlight specific features: one circle around the date and time, another around the "People also search for" section, and a large circle around the "Winter solstice" knowledge panel. Arrows point from the text "rich result" and "knowledge panel (rich result)" to the knowledge panel. Another arrow points from the text "Google search result page – searching “winter solstice 2024”" to the search bar.

Google

winter solstice 2024

All Images News Videos Shopping Web Maps More Tools

Winter solstice / Date (2024)

Sat, Dec 21, 2024, 5:19 PM

Hong Kong Standard Time

Winter solstice (December 21)

The winter solstice, also called the hibernal solstice, occurs when either of Earth's poles reaches its maximum tilt away from the Sun. This happens twice yearly, once in each hemisphere. [Wikipedia](#)

**Date:** Saturday, December 21, 2024

**Also called:** Midwinter; the Shortest Day; the Longest Night

**Celebrations:** Feasting

**Observed by:** Various cultures

**Related to:** Winter festivals

**Significance:** Beginning of lengthening days and shortening nights

More about Winter solstice →

Feedback

People also search for

Summer solstice Qiyun Lidong March Equinox Chunfer Autumn Equinox Xiaohan

People also ask

What is the shortest day in 2024?

What is the darkest day of 2024?

Is the winter solstice always on the 21st?

What is the winter solstice in Australia?

Feedback

rich result

knowledge panel (rich result)

## Linked data for Search Engine Optimization (SEO)

<https://repository.hkust.edu.hk/ir/Record/1783.1-101451>

```
<!DOCTYPE html>
<html>
<head>...</head>
.
.
.
<script
type="application/ld+json">{"@context":{"schema":"http://schema.org/","spdtype":"ht
file":"https://repository.ust.hk/profile/","spdfunding":"https://repository.ust.
erpa.ac.uk/id/publication/","@graph":[{"@type":"schema:CreativeWork","@id":"spdtype
://repository.ust.hk/ir/Record/1783.1-101451","schema:name":"Preliminary Identifi
the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological
Studies","schema:author":[{"@type":"schema:Person","@id":"spdtype:X19424","schema:r
Faraz","schema:affiliation":{"@type":"schema:Organization","@id":"spdtype:101451#Affil
HKUST"}},{"@type":"schema:Person","@id":"spdtype:16129","schema:name":"Quadeer, Ahr
Abdul","schema:affiliation":{"@type":"schema:Organization","@id":"spdtype:101451#Affil
HKUST"}},{"@type":"schema:Person","@id":"spdtype:11106","schema:name":"Mckay, Matth
R","schema:affiliation":{"@type":"schema:Organization","@id":"spdtype:101451#Affiliat
HKUST"}]}],"schema:datePublished":"2020","schema:sameAs":["https://www.scopus.com/re
-85081251137","https://gateway.isiknowledge.com/gateway/Gateway.cgi?GWVersion=2&S
Type=FullRecord&DestApp=WOS&KeyUT=000525486800026","https://doi.org/10.3390/v12030
2020 has seen the emergence of COVID-19 outbreak caused by a novel
coronavirus...","schema:inLanguage":"en","schema:about":["Coronavirus","2019-nCoV","20
coronavirus","SARS-CoV-2","COVID-19","SARS-CoV","MERS-CoV","T cell epitopes","B cell
epitopes","vaccine"],"schema:identifier":["rgc-funded","oa-article","oa-version-Publis
:article","schema:isPartOf":{"@type":"schema:PublicationIssue","@id":"spdtype:101451#
(3), 25 February 2020, article number 254","schema:issueNumber":"v. 12 (3), article nu
February
2020","schema:isPartOf":{"@type":"schema:Periodical","@id":"spdtype:101451#Periodicals
:1999-4915","schema:offers":{"@type":"schema:Offer","@id":"romeopub:17528"}]},"schema
12/3/254/pdf","schema:funder":[{"@type":"schema:Grant","@id":"spdtype:grf","sche
ema:Grant","@id":"spdtype:hkpfs"}]}]}</script>
.
.
.
</body>
</html>
```

## Embed linked data in web page

preliminary identification potential vaccination ta X

All Images News Videos Shopping Web Books : More Tools

Scholarly articles for **preliminary identification potential vaccination targets**

**Preliminary identification of potential vaccine targets** for ... - Ahmed - Cited by 1449

**Identification of vaccine targets** in pathogens and ... - Rawal - Cited by 83

In silico **identification of vaccine targets** for 2019-nCoV - Lee - Cited by 98

National Institutes of Health (NIH) (.gov)  
<https://www.ncbi.nlm.nih.gov/articles/PMC7150947> :

**Preliminary Identification of Potential Vaccine Targets** for ...  
 by SF Ahmed · 2020 · Cited by 1448 — Our findings provide a screened set of epitopes that can help guide experimental efforts towards the development of **vaccines** against SARS-...

National Institutes of Health (NIH) (.gov)  
<https://pubmed.ncbi.nlm.nih.gov/...>

**Preliminary Identification of Potential Vaccine Targets** for ...  
 by SF Ahmed · 2020 · Cited by 1448 — In this study, we sought to gain insights for **vaccine** design against SARS-CoV-2 by considering the high genetic similarity between SARS-CoV-2...

HKUST SPD  
<https://repository.hkust.edu.hk/Record> :

**Preliminary Identification of Potential Vaccine Targets** for ...  
 by SF Ahmed · 2020 · Cited by 1448 — **Preliminary Identification of Potential Vaccine Targets** for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological...

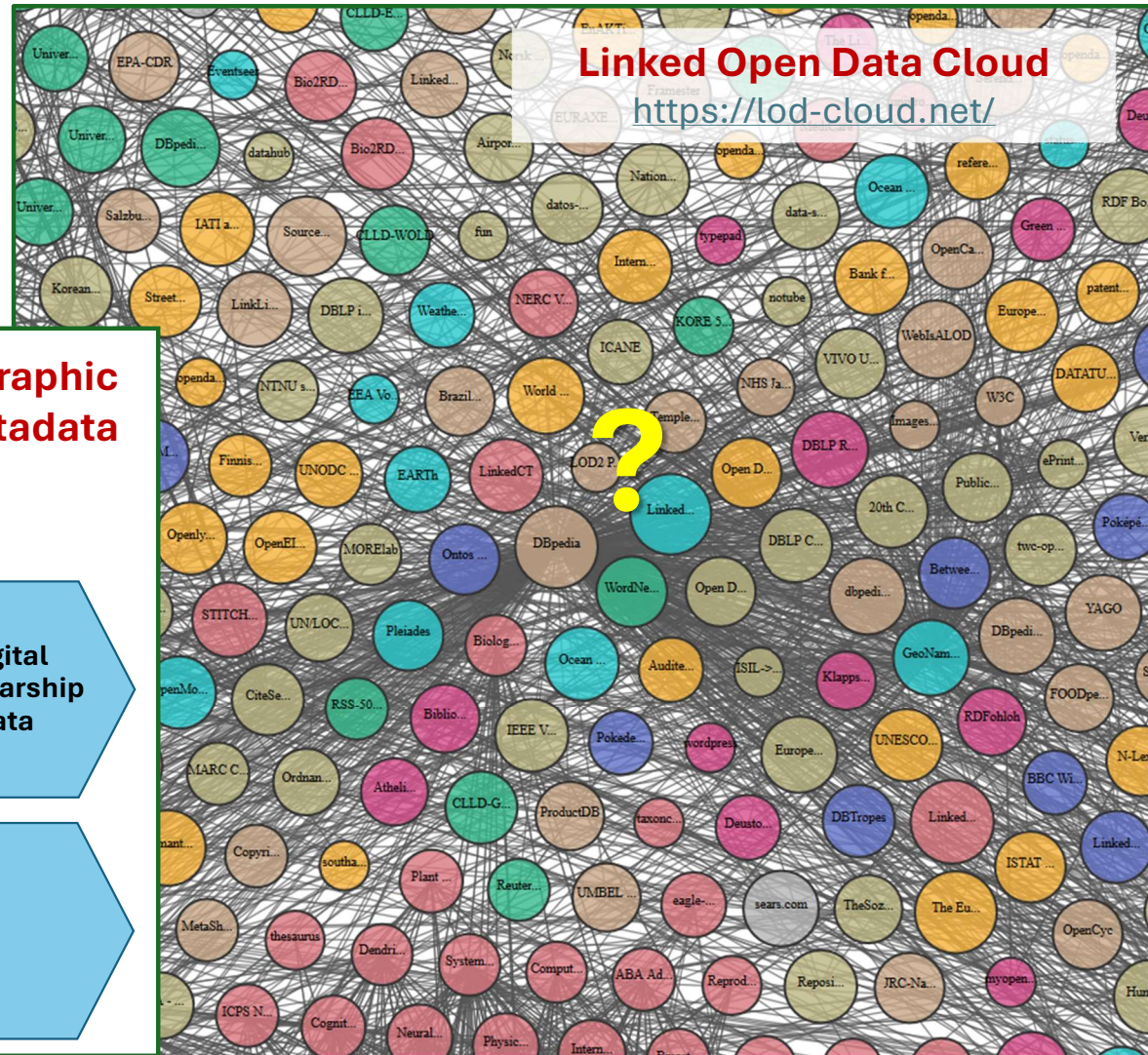
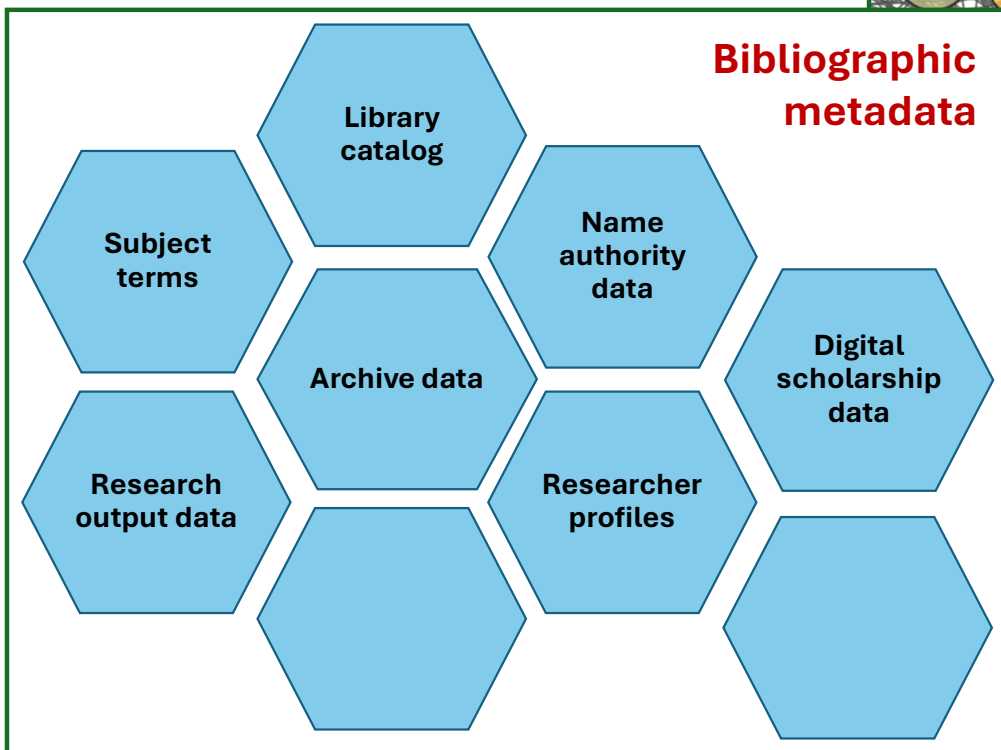
ResearchGate  
<https://www.researchgate.net/.../COVID-19> :

**(PDF) Preliminary identification of potential vaccine targets ...**  
 25 Jun 2024 — Conclusion The present study posits three **potential** epitopes of S protein of SARS-CoV-2 predicted by immunoinformatic methods based on their ...

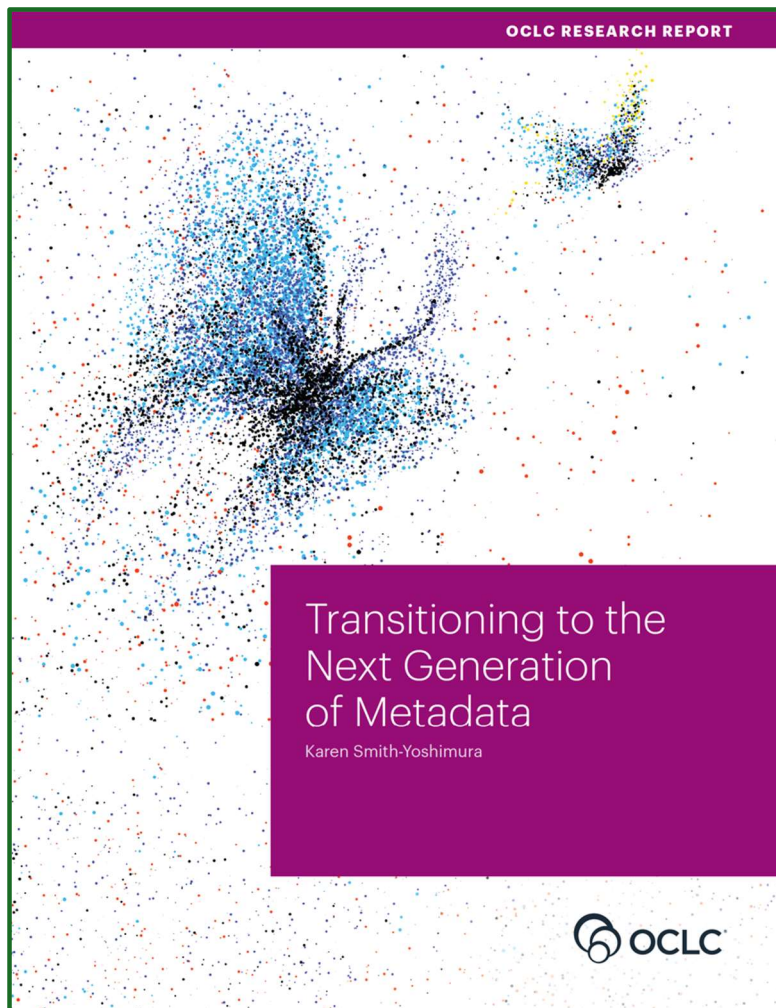
## Improve ranking and increase visibility

# Bibliographic presence in linked open data

Any datasets in the library domain are present in the LOD cloud?



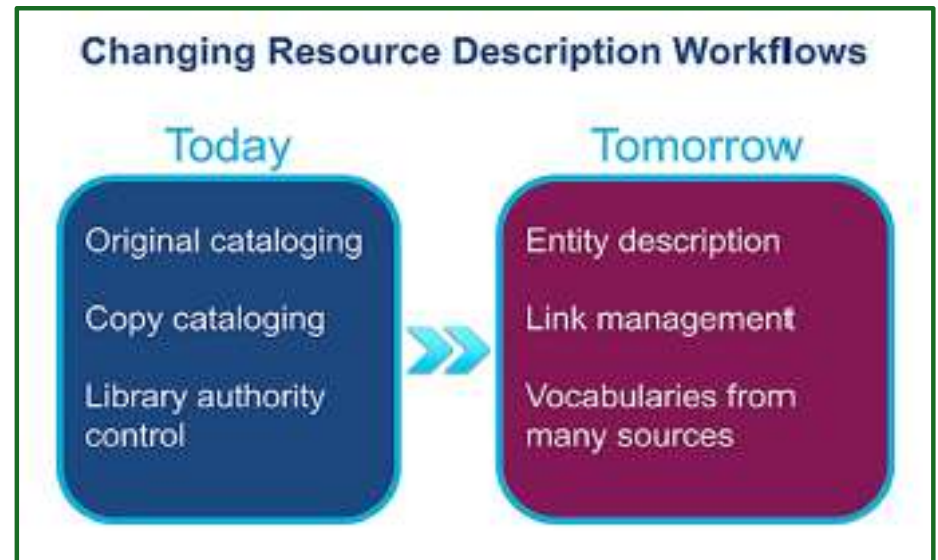
# Next generation of bibliographic metadata



## OCLC Research Report

Smith-Yoshimura, Karen. 2020. *Transitioning to the Next Generation of Metadata*. Dublin, OH: OCLC Research.  
<https://doi.org/10.25333/rqgd-b343>.

## Call for transition to linked data and identifiers



Source: p.4 of the Report

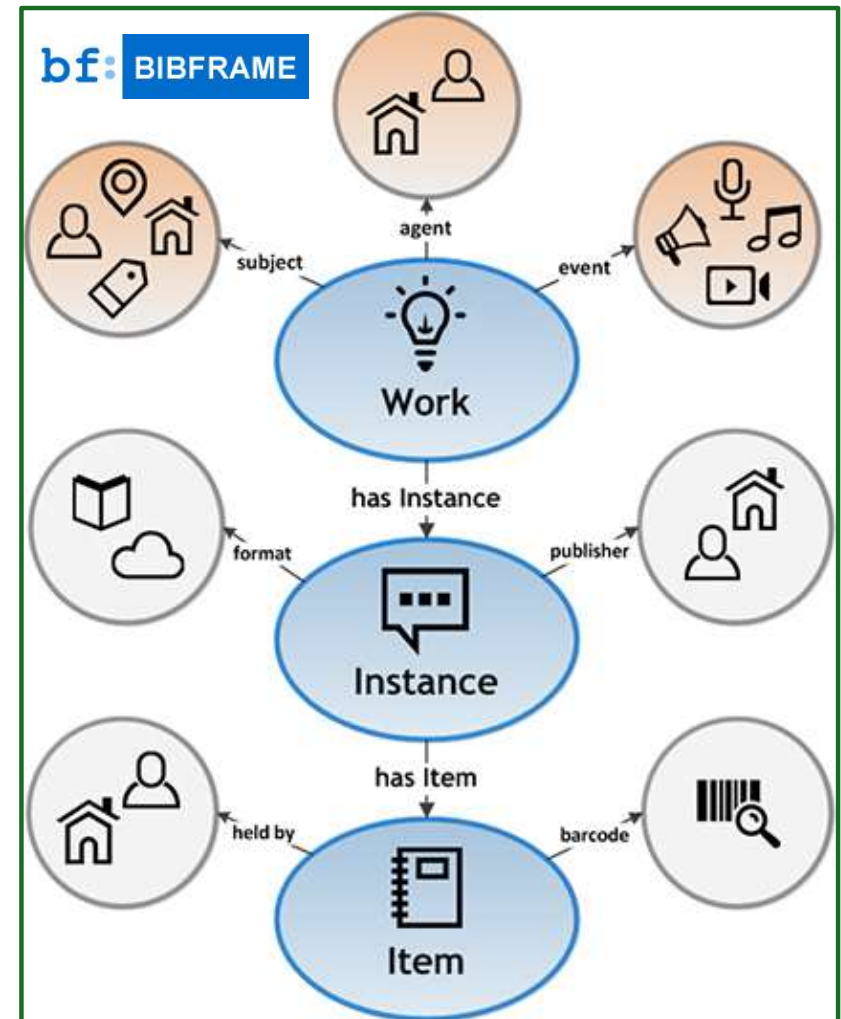
## MARC is linked data incapable!

```
LDR 00000nam a2200000 a 4500
001 ABcS
008 200424s2009 ch a 000 0 chi d
020 |a9789862131206
090 |a859.6 2290
100 0 |a幾米|eartist|0http://id.loc.gov/authorities/names/
nr97011804
245 10 |a星空 =|bThe starry starry night /|c幾米
246 31 |aStarry starry night
250 |a初版
260 |a台北市 :|b大塊文化出版股份有限公司,|c2009
300 |a1 v. (unpaged) :|bchiefly col. ill. ;|c27 cm
650 0 |aCaricatures and cartoons||0http://id.loc.gov/
authorities/subjects/sh85020312
650 0 |aChinese wit and humor, Pictorial|0http://id.loc.gov/
authorities/subjects/sh85024383
586 |a第二十一屆中學生好書龍虎榜. 最受中學生歡迎十本好書
830 0 |a幾米作品 ;|v32
856 42 |ahttps://www.jimmyspa.com/tw/Books/FullLengthStories/
StarryStarryNight/|z內文試閱 (幾米網站)
```

- “**MA**chine **R**eadable **C**atalog” – defined in 1960s, originally meant for printing catalog cards by machines
- **String**-based; not structured; does not facilitate **semantic inference**
- No build-in **linking capability** of things in between records; nor within a record
- MARC data is not understandable in digital space outside of the library domain; and it is **totally invisible** in linked open data

# BIBFRAME ontology

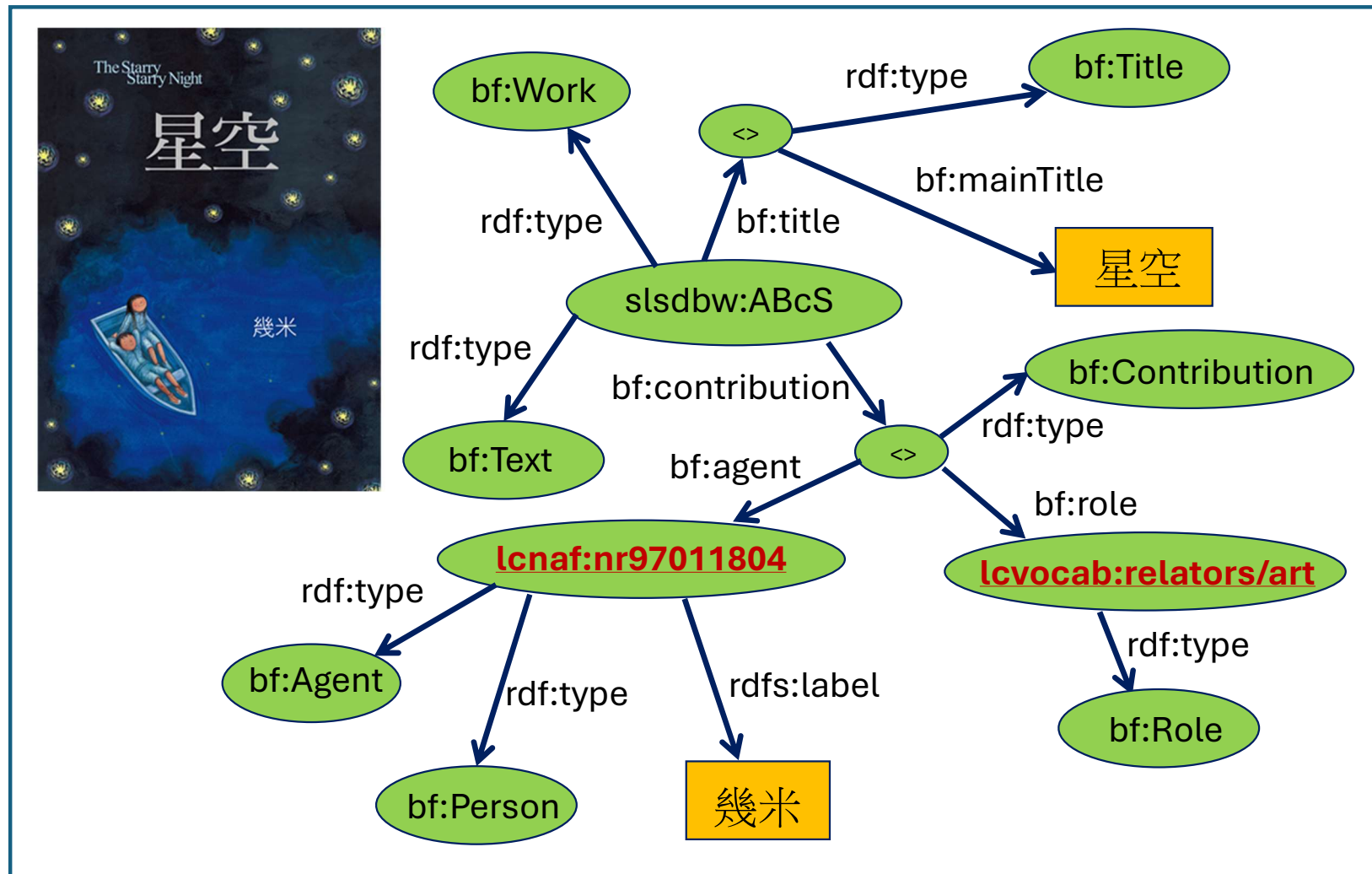
- **BIBFRAME** (Bibliographic Framework) is developed by Library of Congress, with the intention to **replace MARC**.
- First draft was released in 2012. Momentum of adoption began catching up only after **BIBFRAME 2.0** was released in 2016.
- Library of Congress has been cataloguing its own collections with BIBFRAME. Some American and European libraries are either testing it or in the early stage of implementation.
- Find out more from recent presentations:
  - BIBFRAME Workshop in Europe 2024 ([https://www.bfwe.eu/helsinki\\_2024](https://www.bfwe.eu/helsinki_2024))
  - BIBFRAME July 2024 Update Forum (<https://www.loc.gov/bibframe/news/bibframe-update-jul2024.html>)



<https://id.loc.gov/ontologies/bibframe.html>

# Bibliographic metadata in BIBFRAME linked data

<https://www.sls.org.hk:8443/opac/lod/ABcS>



# SLS BIBFRAME Editor


## Design principles

- Light-weight, simple, single-page.
- Meant for library staff who may not have taken proper cataloging training.

<https://www.sls.org.hk:8443/opac/edit/ABcS>

The screenshot displays the SLS Demo BIBFRAME Editor interface. At the top, the 'SLS Demo' logo is on the left, and navigation links 'My Account', 'Logout', and '中文' are on the right. A 'BIBFRAME Editor' button is prominently displayed. Below this, there are links for 'Export: RDF/XML', 'JSON-LD', 'MARC/XML' and 'View: Graph', 'Catalog', 'MARC record'. The main header shows the instance ID: 'Instance [slsdb:ABcS--20220428--20240829]'. The record is organized into sections: 'Title' (Main title: '星空', Subtitle: 'The starry starry night', Type: 'Title'), 'Statement of responsibility' (幾米), 'Edition' (初版), and 'Provision activity' (Type: 'Publication', Place: '台北市', Agent: '大塊文化出版股份有限公司', Date: '2009', Place code: 'Taiwan', Date 1/2 (MARC): '2009'). Each field has a small icon for adding, deleting, or duplicating the entry.

# Enrichment – blending knowledge cards in library catalog



**星空 : The starry starry night**

幾米  
初版. 台北市 : 大塊文化出版股份有限公司, 2009  
1 v. (unpaged) ; 27 cm  
**Book**


**Links**  
▶ [內文試閱 \(幾米網站\)](#)

**Item availability**


Location	Call number	Barcode	Status
Chinese ERS	859.6 2290	A1706299	Available

[Redirect search to: Google | HKPL | HKALL](#)


**Related items**




急不了先生 : Mr Not Urgent




幾米科學童話: 7. the 山 of music



老少女心事






我住在台灣了!: 港人居台灣. 教書辛酸史. 5. 露思兔兔






課不停學

**More about...**





**Cartoon**   
卡通   
Illustration telling a comic or satirical story in a single image 一種圖畫類型






**Caricature**   
諷刺畫   
Rendered image showing the features of its subject in a simplified or exaggerated way



**Knowledge cards** about **subjects** and **names** are generated by blending content from Wikidata


**More about...** 





**Cartoon**   
卡通   
Illustration telling a comic or satirical story in a single image 一種圖畫類型



**Jimmy Liao**   
幾米   
Taiwanese picture book writer and illustrator 台灣作家和藝術家



**Caricature**   
諷刺畫   
Rendered image showing the features of its subject in a simplified or exaggerated way

<https://www.sls.org.hk:8443/opac/bib/ABcS>

## Library dedicated chatbot

*Coming soon?*

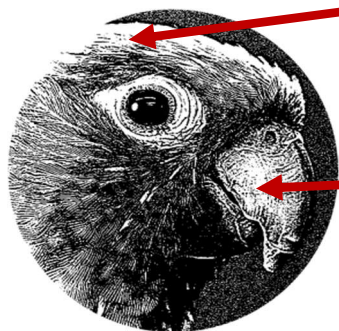
Library dedicated **chatbot** deploying **linked data** and **large language model** to serve library users' information needs

Hi chatbot,

Can you **find** me some books **about** Alan Turing that I can **borrow** from my **nearby** public library for reading in this **weekend**?

Hi chatbot,

I am reading this article (...) on quantum computing. Please **find** me ten other **similar** scholarly articles that I have **permission** to read.

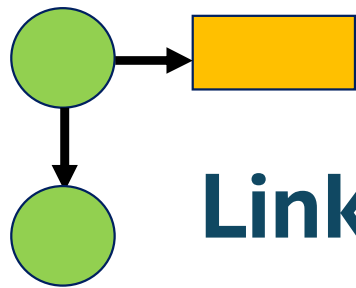


### **AI Reasoning**

**Linked data** facilitates information retrieval through inference

### **Generative AI**

**LLM (Large Language Model)** facilitates natural language communication between library users and chatbot



**Linked data explained**

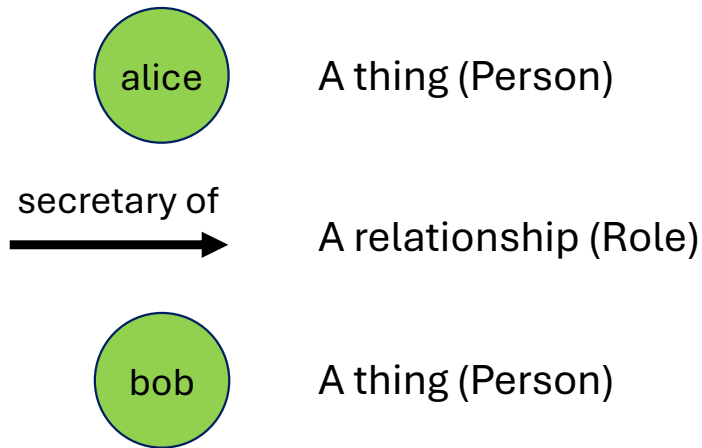
# RDF triple

## A statement in English text (literal):

**“Alice is a secretary of Bob”**

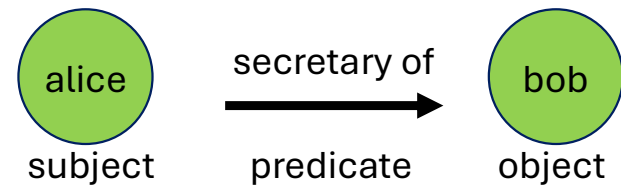
It is difficult for computers to understand the semantics conveyed in an unstructured text string

## This statement has three parts – two things and a relationship connecting them



Things – can be anything, such as people, places, concepts, events, etc. The relationship connecting things is also a thing!




## Using triple to express a statement structurally



Triple contains subject, predicate and object. It provides the semantics for computers.

## RDF triple [cont.]

Subject, predicate and object of a triple are identified by URIs

	<u>URI</u>		<u>URI shortened by using prefix</u>
<u>Subject</u>		<http://example.com/person/alice>	exp:alice
<u>Predicate</u>	secretary of 	<http://example.com/vocab/secretaryOf>	ex:secretaryOf
<u>Object</u>		<http://example.com/person/bob>	exp:bob

Prefix	Namespace
exp:	<http://example.com/person/>
ex:	<http://example.com/vocab/>

Triple defined in this way is known as RDF triple. Computer can look up the URI to obtain more description about the thing.

## RDF triple [cont.]

### Triple in URIs

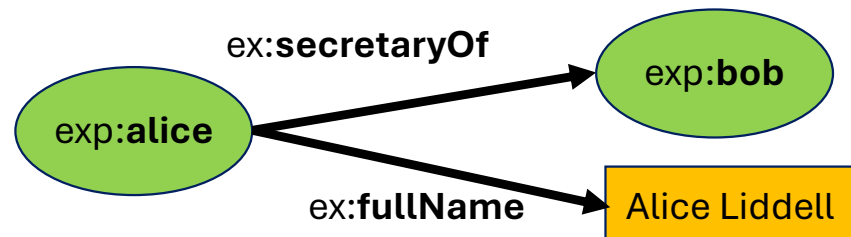
<u>Subject</u>	<u>Predicate</u>	<u>Object</u>
exp:alice	ex:secretaryOf	exp:bob

### Object can be an URI or a literal

<u>Subject</u>	<u>Predicate</u>	<u>Object</u>
exp:alice	ex:fullName	"Alice Liddell"

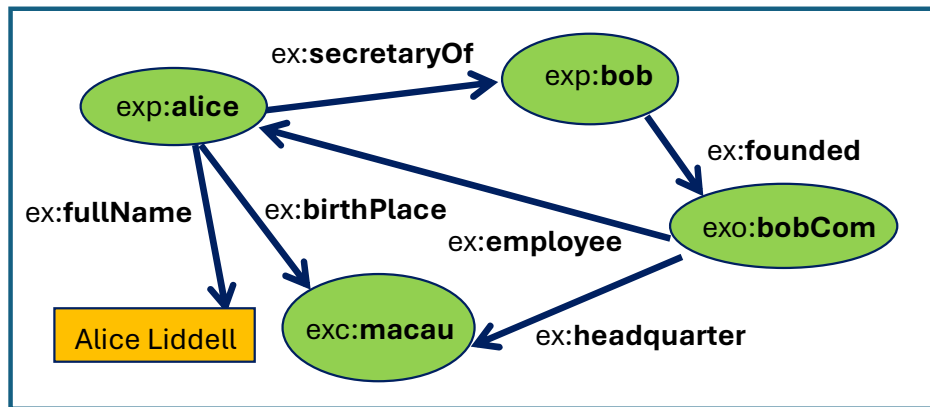
This triple tells the computer that Alice's full name is "Alice Liddell", which is a **literal** (text in English)

### A graph showing these two triples



# Semantic Web

Things in URIs and literals are linked within the same dataset, thus forming linked data.



Things can also be linked across multiple datasets forming “web of data”.

URIs are “understandable” by computers, resulting “semantic web”

The term “semantic web” was coined by Tim Berners-Lee, in 1999. Berners-Lee is the inventor of World Wide Web.

## Principles of publishing structured data on the Web - Tim Berners-Lee, 2006.

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs. so that they can discover more things.

<https://www.w3.org/DesignIssues/LinkedData>

# Serialization

File formats for transferring triples between computers:

## N-Triples

```
<http://example.com/person/alice> <http://example.com/vocab/secretaryOf> <http://example.com/person/bob> .  
<http://example.com/person/alice> <http://example.com/vocab/fullName> "Alice Liddell" .
```

## Turtle

```
@prefix ex: <http://example.com/vocab/> .  
@prefix exp: <http://example.com/person/> .  
exp:alice  
  ex:secretaryOf exp:bob ;  
  ex:fullName "Alice Liddell" .
```

## RDF XML

```
<?xml version="1.0" encoding="utf-8" ?>  
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ex="http://example.com/vocab/">  
  <rdf:Description rdf:about="http://example.com/person/alice">  
    <ex:secretaryOf rdf:resource="http://example.com/person/bob"/>  
    <ex:fullName>Alice Liddell</ex:fullName>  
  </rdf:Description>  
</rdf:RDF>
```

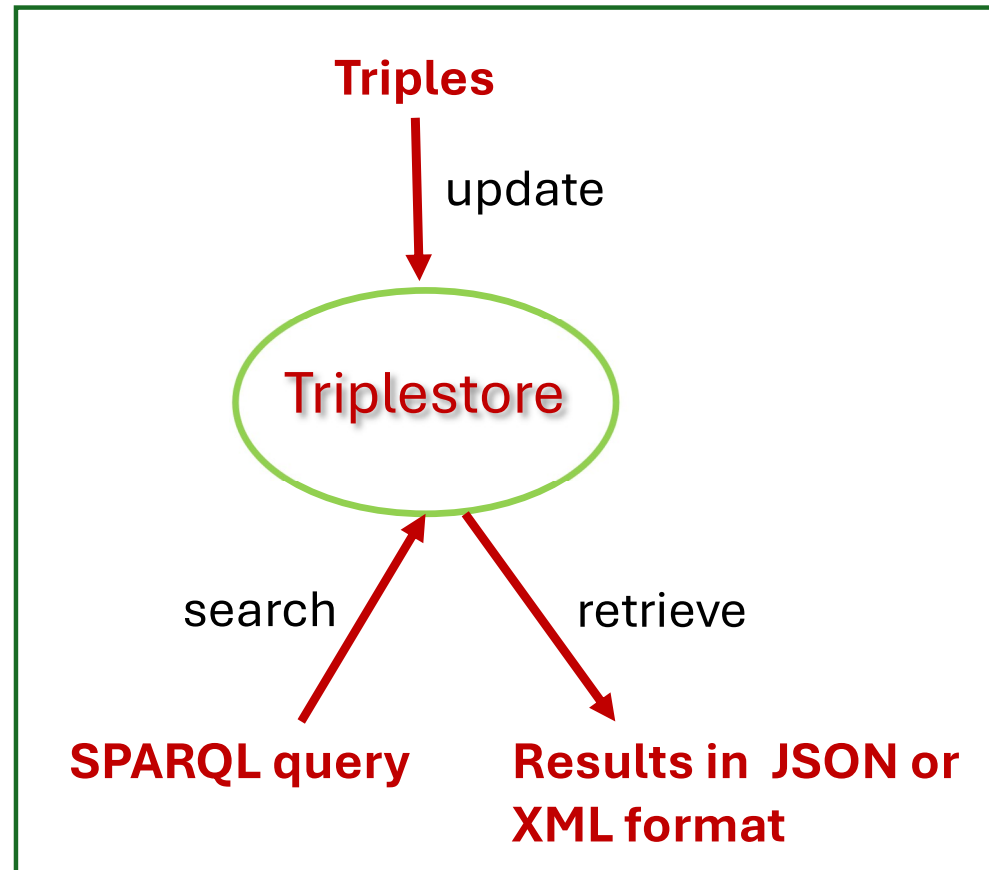
## Serialization [cont.]

### JSON-LD

```
{
  "@context": {
    "ex": "http://example.com/vocab/",
    "exp": "http://example.com/person/"
  },
  "@graph": [
    {
      "@id": "exp:alice",
      "ex:secretaryOf": {
        "@id": "exp:bob"
      },
      "ex:fullName": {
        "@value": "Alice Liddell"
      }
    }
  ]
}
```

# Triplestore

- Just as tabular data is stored in a relational database, **triples** are stored in a **triplestore**.
- SQL is the query language for relational database, while **SPARQL** is used to query triples in a **triplestore**.



## Triplestore [cont.]

### Linked data query service

provides a front-end interface to search triplestore by SPARQL

SPARQL query editor

Query result in table

### Example: SLS linked query data

The screenshot shows the 'Linked data query service' interface. At the top, there's a 'Query' section with a SPARQL query editor. Below it, there's a 'Table' tab showing the query results. The query is a PREFIX query for 'slsdb' with a WHERE clause using a GRAPH pattern. The results table has three columns: 's', 'p', and 'o'. The first row shows 'slsdb:ACQs' linked to 'bf:responsibilityStatement' with the value '泰絲·格里森著；尤傳莉譯.'.

```
1 PREFIX slsdb: <https://www.sls.org.hk:8443/bf/instance/>
2 SELECT ?s ?p ?o
3 WHERE {
4   GRAPH slsdb:ACQs {
5     ?s ?p ?o.
6   }
7 }
8 LIMIT 2000
```

163 results in 0.202 seconds

	s	p	o
27	slsdb:ACQs	bf:responsibilityStatement	泰絲·格里森著；尤傳莉譯.
70	slsdb:ACQs#Bnode-72	bflc:simpleAgent	春天出版國際
129	slsdb:ACQs#Agent999-11	rdfs:label	尤傳莉
80	slsdb:ACQs#Bnode-71	rdfs:label	喀邁拉空間：Gravity
81	slsdb:ACQs#Bnode-71	bf:mainTitle	喀邁拉空間

<https://www.sls.org.hk:8443/opac/lod/ABcS>

# Vocabulary

**Vocabulary** (or **ontology**, when used in more formal and complex situation) **involves naming, grouping and relating things within the domain area concerned.**

## Commonly used vocabularies of linked data in the **bibliographic description** domain:

- **DCMI Metadata Terms**

- Dublin Core Metadata Initiative
- <https://www.dublincore.org/specifications/dublin-core/dcmi-terms>

- **Schema.org vocabulary**

- General purpose, for any things
- Maintained by web community
- <https://schema.org/docs/schemas.html>

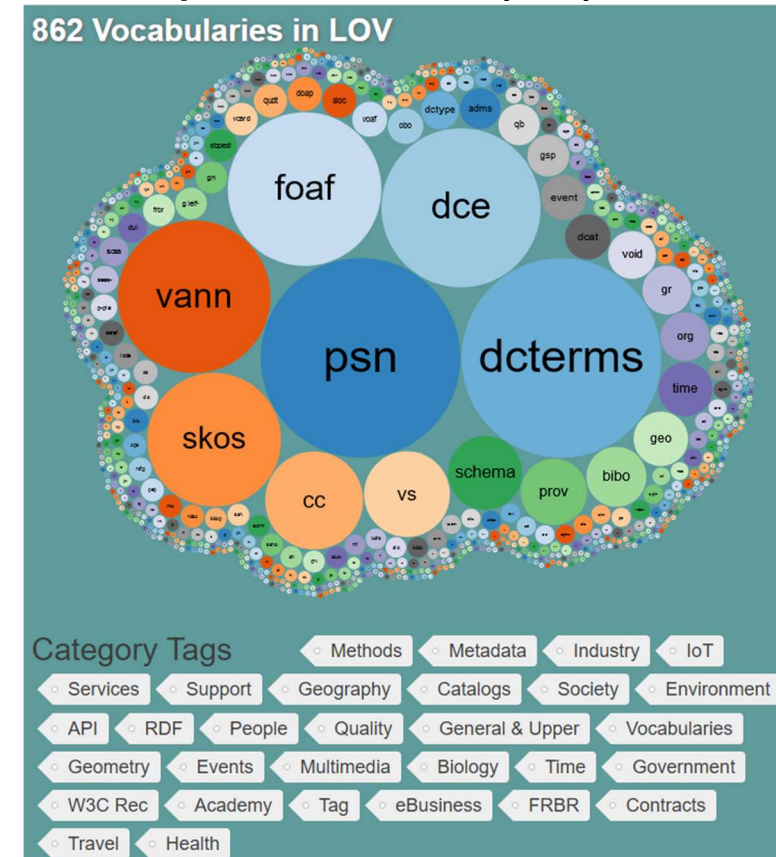
- **BIBFRAME 2.0 vocabulary**

- Library of congress
- <https://id.loc.gov/ontologies/bibframe.html>

- **RDA (Resource Description and Access) vocabularies**

- RDA Steering Committee
- <http://www.rdaregistry.info>

## Linked Open Vocabularies (LOV)



<https://lov.linkeddata.es/dataset/lov>

# Vocabulary – an example

## Classes

Class	URI	Description
Person	<http://example.com/vocab/Person>	A person

## Properties

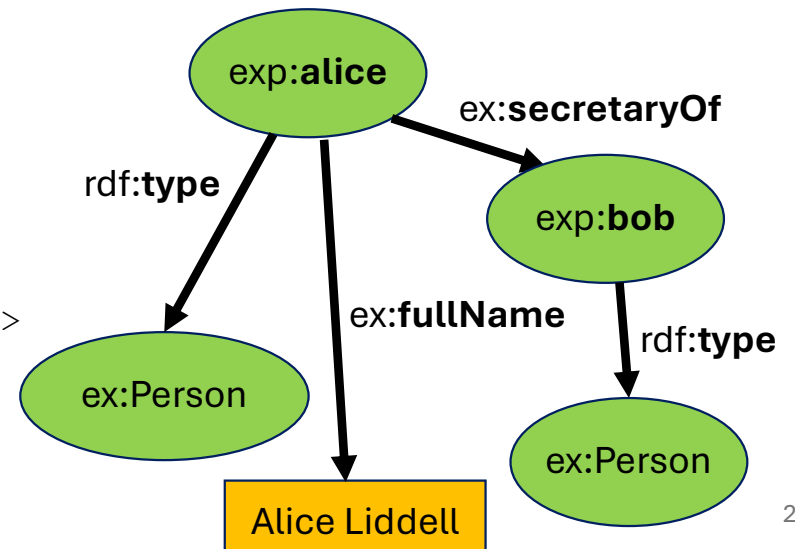
Property	URI	Type of value	Description
fullName	<http://example.com/vocab/fullName>	Text	Full name of the person
secretaryOf	<http://example.com/vocab/secretaryOf>	Person	A secretary of another person

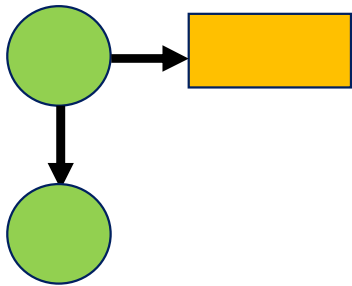
## Example data in English

Alice, whose full name is Alice Liddell, is a secretary of Bob.

## Described in linked data

```
PREFIX exp: <http://example.com/person/>
PREFIX ex: <http://example.com/vocab/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
exp:bob rdf:type ex:Person .
exp:alice rdf:type ex:Person .
exp:alice ex:fullName "Alice Liddell" .
exp:alice ex:secretaryOf exp:bob .
```





**Questions?**