

Linked data for libraries

An introduction

By LAM Ki Tat

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

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

Email: lam.ki.tat@connect.hku.hk

Abstract

In recent years, **linked data technology** has gained significant momentum in libraries, transforming the way we **discover and interact with information**.

This cutting-edge approach leverages **graph-based** structure of linked data to empower machines with **semantic understanding of bibliographic metadata**.

This talk will provide a brief introduction to the **basics of linked data**, followed by examples of its **practical applications** in library settings.

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 **URIs** 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

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**:

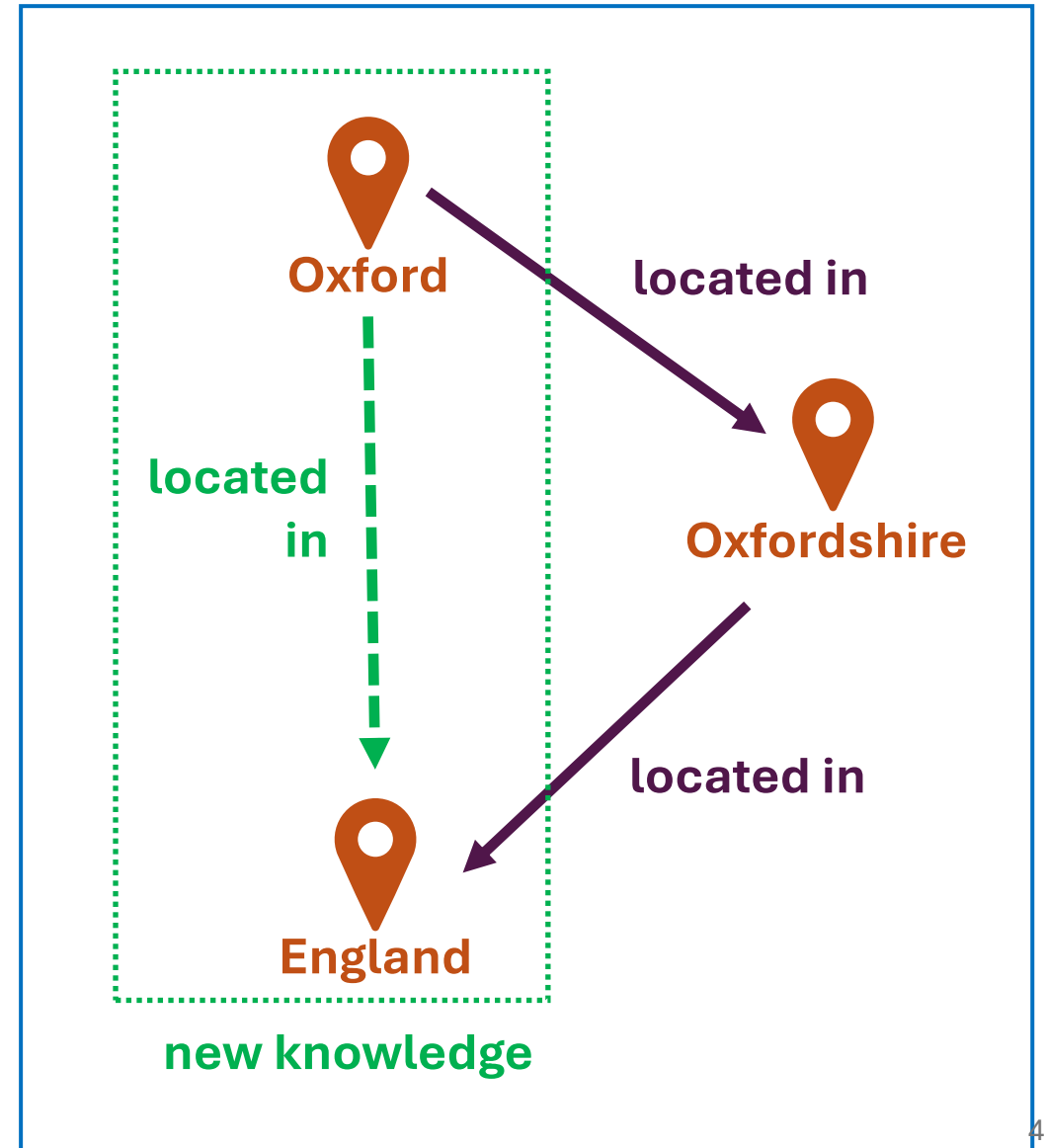
```
IF (:a :locatedIn :b) AND (:b locatedIn :c)  
THEN (:a :locatedIn :c)
```

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

The image shows a screenshot of the Wikidata homepage. The background is a network graph with nodes and edges in red, green, and blue. The text on the page reads: "Welcome to Wikidata", "the free knowledge base with 119,067,903 data items that anyone can edit.", "Introduction • Project Chat • Community Portal • Help", and "Want to help translate? Translate the missing messages." There are several annotations: a red box around the number "119,067,903", a red arrow pointing from a yellow box to this number, and the words "open", "collaborative", "structured", and "multilingual" are placed near different parts of the network graph.

Number of items:
October 2024: 113,871,413
October 2025: 119,067,903
4.6% increase in one year

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

Aggregating, blending and visualizing 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

Nobel Prize in Physics

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

Images of recipients



2025
[commons:John Martinis \(cropp...](#)
[Q:John M. Martinis](#)



2025
[commons:Michel H. Devoret 2017 190x180.jpg](#)
[Q:Michel Devoret](#)



2024
[commons:John J. Hopfield, 2024 N...](#)
[Q:John Hopfield](#)



2024
[commons:Geoffrey E. Hinton, 2...](#)
[Q:Geoffrey Hinton](#)



2023
[commons:US Embassy Sweden 2023 N...](#)
[Q:Pierre Agostini](#)



2023
[commons:Anne LHuillier 01.JPG](#)
[Q:Anne L'Huillier](#)



2023
[commons:Ferenc Krausz \(cropped...](#)
[Q:Ferenc Krausz](#)



2022
[commons:John Francis Clauser \(cro...](#)
[Q:John Clauser](#)

<https://scholia.toolforge.org>

Aggregating, blending and visualizing linked data [cont.]

EntiTree – an application that blends Wikidata content into its websites

Example: Charles III

https://www.entitree.com/en/family_tree/Q43274

The screenshot displays the EntiTree application interface for the family tree of Charles III. The top navigation bar includes the EntiTree logo, an 'Examples' dropdown, and a 'settings' link. A search bar contains 'Charles III' and a 'family tree' dropdown menu. The main content area shows a family tree structure with the following individuals and their details:

- Prince Philip, Duke of Edinburgh**: consort of Elizabeth II from 1952 to 2021; 10 Jun 1921 - 9 Apr 2021.
- Elizabeth II**: Queen of the United Kingdom from 1952 to 2022; 21 Apr 1926 - 8 Sep 2022.
- Prince Andrew, Duke of York**: second son and third child of Queen Elizabeth II and Prince Philip; 19 Feb 1960.
- Prince Edward, Duke of Edinburgh**: British prince; youngest child of Elizabeth II and Prince Philip; 10 Mar 1964.
- Charles III**: King of the United Kingdom and other Commonwealth realms since 2022; 14 Nov 1948.
- Queen Camilla**: Queen Consort of the United Kingdom since 2022; 17 Jul 1947.
- Diana, Princess of Wales**: member of the British royal family and Princess of Wales; 1 Jul 1961 - 31 Aug 1997.
- William, Prince of Wales**: heir apparent to the throne of the United Kingdom and 14 Commonwealth Realms; 21 Jun 1982.
- Prince Harry, Duke of Sussex**: younger son of Charles III of the United Kingdom and Camilla; 15 Sep 1984.

The interface also features a search bar, a 'family tree' dropdown, and a navigation menu with 'Examples' and 'settings'. A sidebar on the left contains navigation icons: a plus sign, a minus sign, a location pin, and a refresh icon.

Google Knowledge Graph

Google search result page –
searching “stephen hawking”

Content served or influenced by
the Google Knowledge Graph

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

It enables Google to
**understand the context of
a query beyond just
keywords** and deliver more
relevant information.

<https://www.google.com/search?q=stephen+hawking>

The image shows a Google search result for "Stephen Hawking". The search bar at the top contains the text "Stephen Hawking". Below the search bar, the "AI Mode" button is circled in red. The main search result is for "Stephen Hawking", an English theoretical physicist and cosmologist. The Knowledge Graph information panel is highlighted with a red oval and contains the following details:

- Overview:** English theoretical physicist and cosmologist
- Images:** A collage of images including a portrait of Stephen Hawking, a video thumbnail of him speaking, and a photo of him in a wheelchair.
- Biographical Information:**
 - Born:** 8 Jan 1942, Oxford, United Kingdom
 - Died:** 14 Mar 2018, Cambridge, United Kingdom
- Related Content:** A YouTube video titled "Stephen Hawking: My life in physics" and a CNN article titled "Stephen Hawking's remarkable life in pictures".

Below the Knowledge Graph panel, there is a Wikipedia snippet for Stephen Hawking, followed by a "People also ask" section with questions like "What illness did Stephen Hawking have?" and "Why is Stephen Hawking so famous?". At the bottom right, there is a "People also search for" section with images of Jane Hawking, Elaine Mason, Lucy Hawking, and Albert Einstein.

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/","spditem":"nt
file":"https://repository.ust.hk/profile/","spdfunding":"https://repository.ust
erpa.ac.uk/id/publication/"},"@graph":[{"@type":"schema:CreativeWork","@id":"spdit
://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":"spdprofile:X19424","schema:r
Faraz","schema:affiliation":{"@type":"schema:Organization","@id":"spditem:101451#Affil
HKUST"}},{"@type":"schema:Person","@id":"spdprofile:16129","schema:name":"Quadeer, Ah
Abdul","schema:affiliation":{"@type":"schema:Organization","@id":"spditem:101451#Affil
HKUST"}},{"@type":"schema:Person","@id":"spdprofile:11106","schema:name":"Mckay, Matth
R","schema:affiliation":{"@type":"schema:Organization","@id":"spditem:101451#Affiliati
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":"spditem:101451#
(3), 25 February 2020, article number 254", "schema:issueNumber":"v. 12 (3), article nu
February
2020", "schema:isPartOf":{"@type":"schema:Periodical","@id":"spditem:101451#Periodical5
:"1999-4915", "schema:offers":{"@type":"schema:Offer","@id":"romeopub:17528"}}}, "schema
12/3/254/
ema:Grant", "
.
.
.
</body>
</html>
```

Embed linked data in web page

Improve ranking and increase visibility

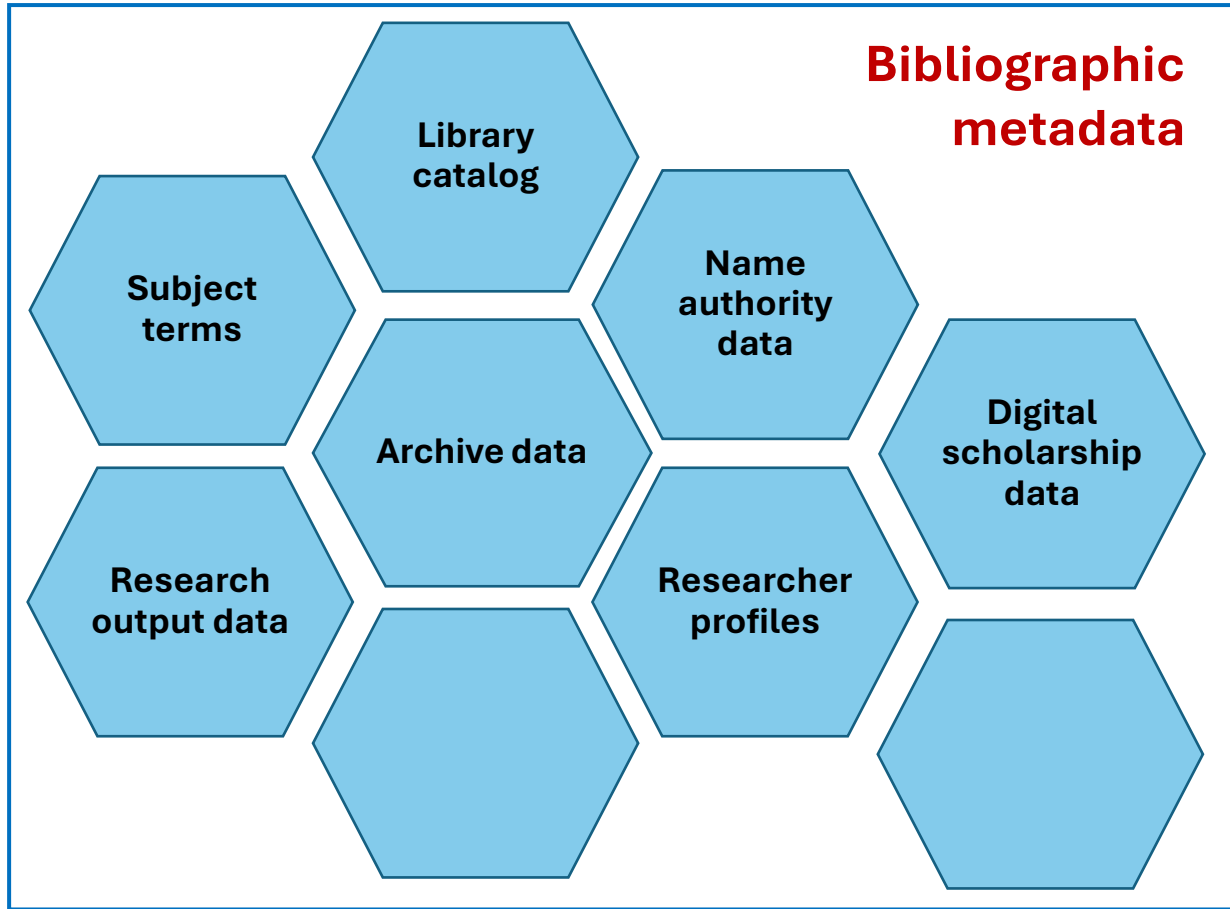
For unknown reasons, the embedded linked data was removed, and the web page is no longer ranked in the Google search results!

The screenshot shows a Google search result page. The search bar contains the text "preliminary identification potential vaccination ta". The search results are displayed in a list format. The top result is a scholarly article titled "Preliminary identification of potential vaccine targets for ... - A". Below it is another result from the National Institutes of Health (NIH) titled "Preliminary Identification of Potential Vaccine Targets for ...". The third result is from HKUST SPD, titled "Preliminary Identification of Potential Vaccine Targets for ...". The fourth result is from ResearchGate, titled "(PDF) Preliminary identification of potential vaccine targets ...". A red oval highlights the HKUST SPD result, and a red arrow points from the text "For unknown reasons, the embedded linked data was removed, and the web page is no longer ranked in the Google search results!" to this result. Another red arrow points from the text "Improve ranking and increase visibility" to the top of the search results.

Google search result page as of October 2024

Bibliographic presence in linked open data

Any datasets in the library domain are available as Linked Open Data (LOD)?



Linked Open Data Cloud

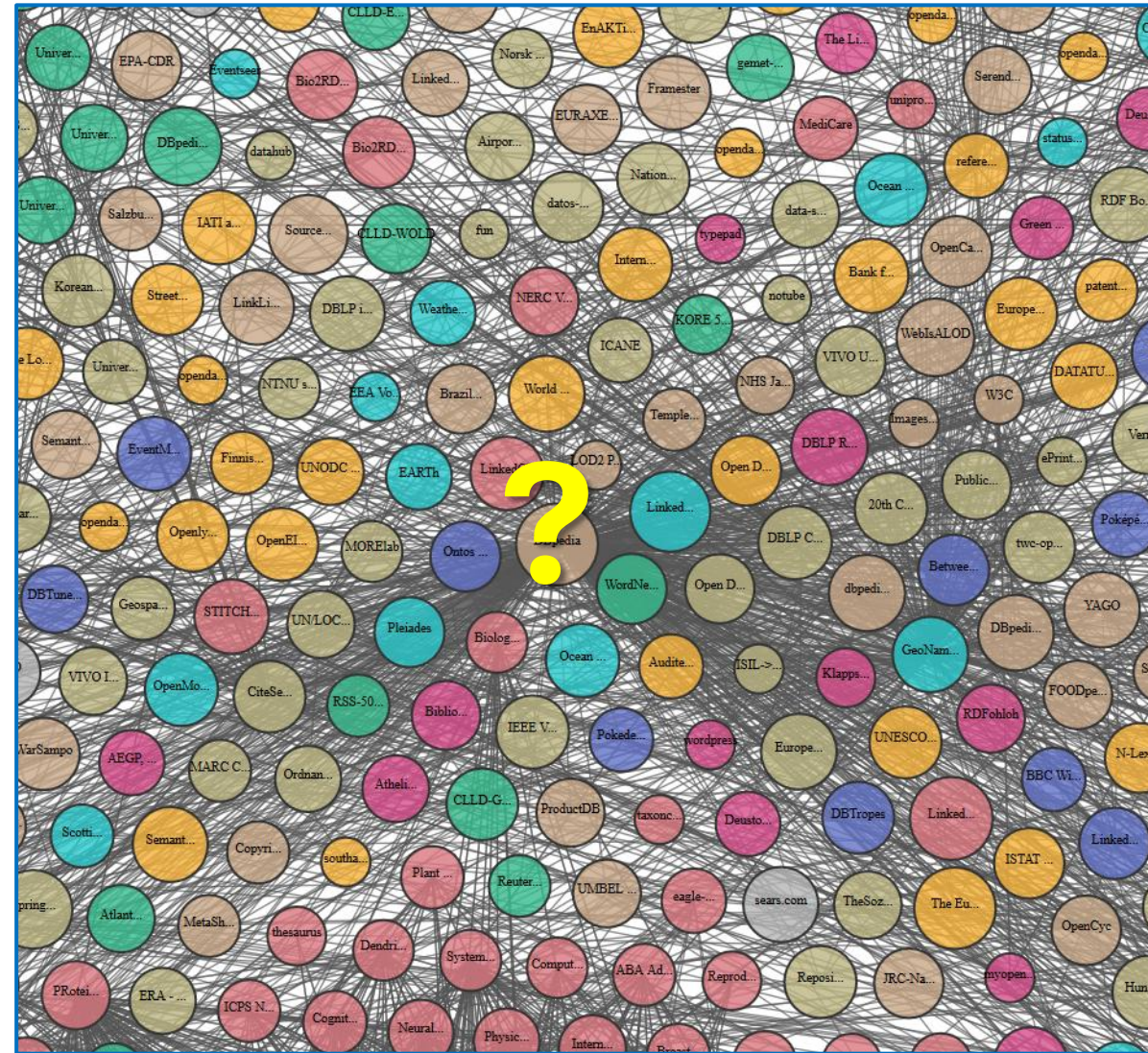
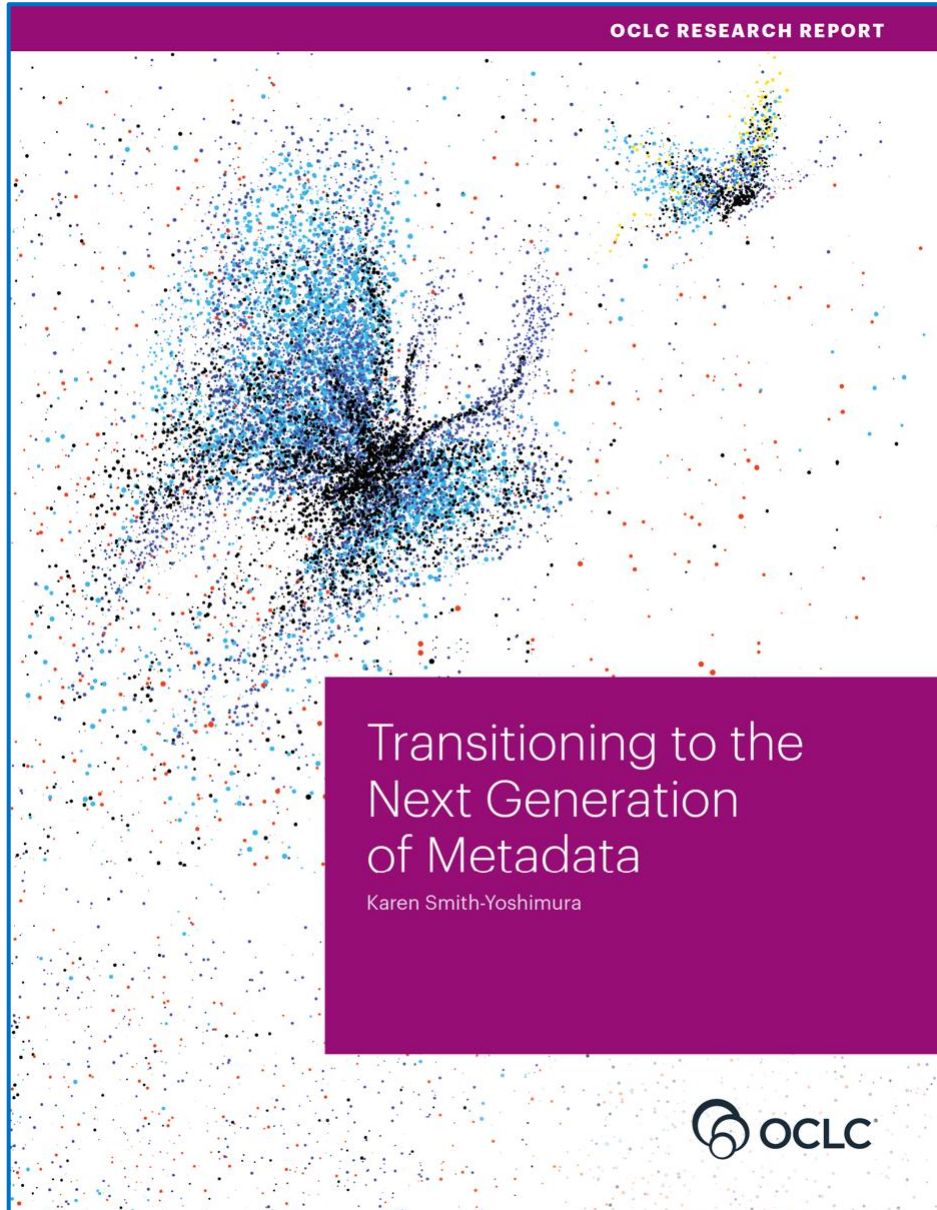


Image from <https://lod-cloud.net> (website no longer available)

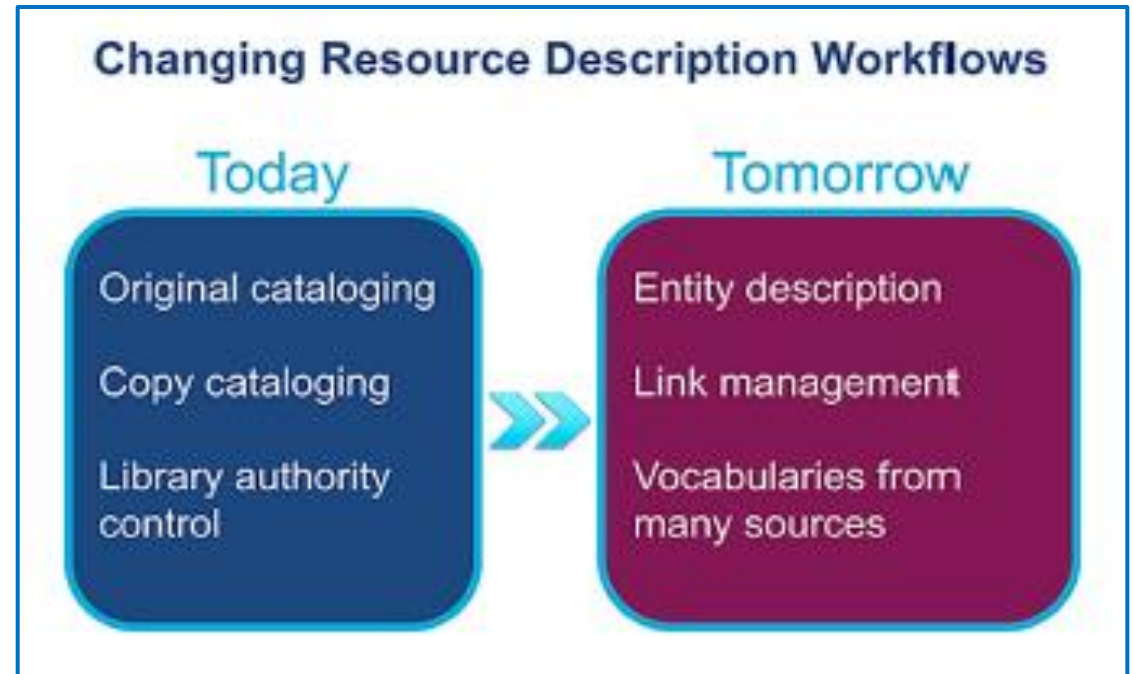
Next generation of bibliographic metadata



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

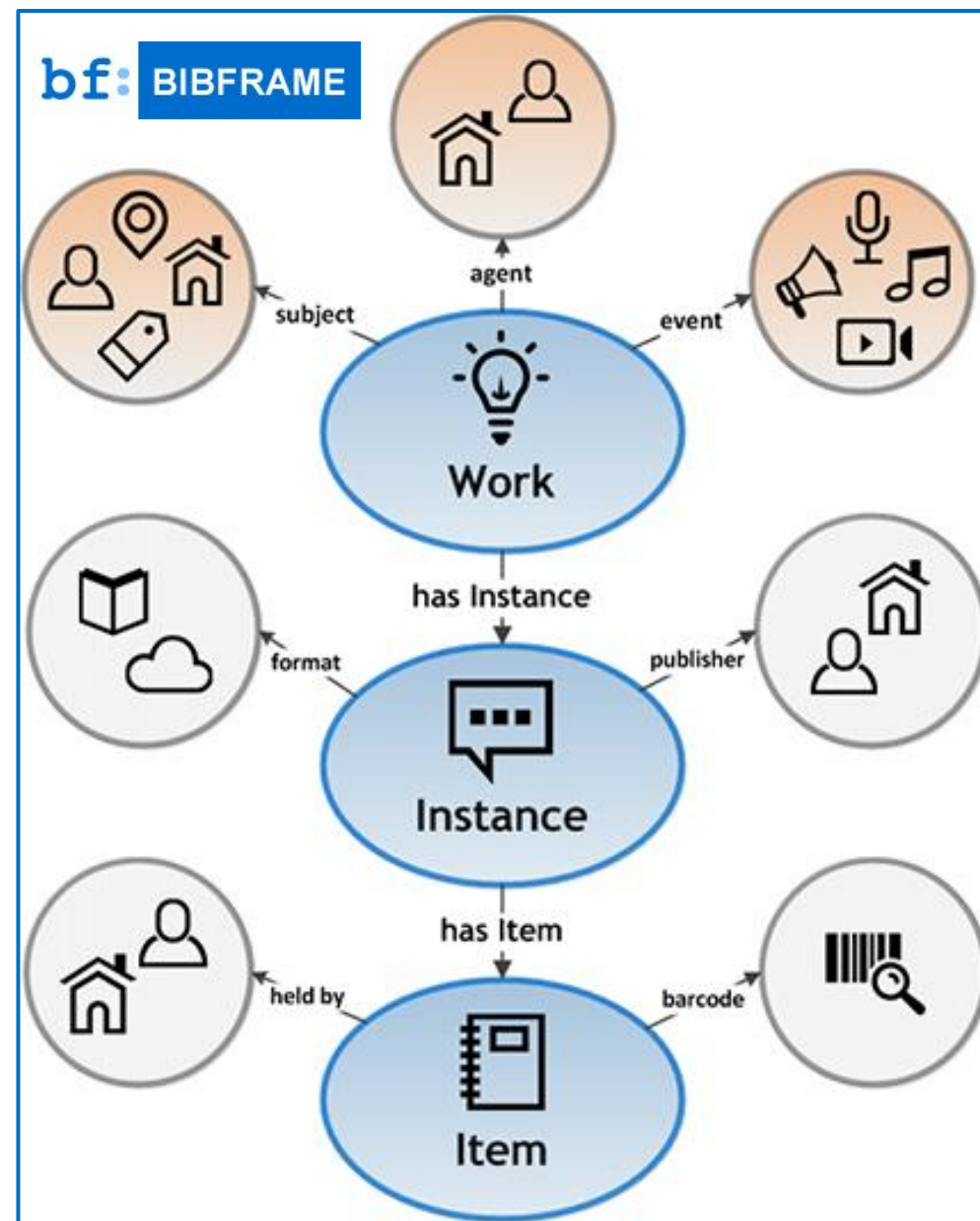
String-based; not structured; does not facilitate **semantic inference**

No build-in **linking capability** of things in between records; nor within a record

MARC data is not readily understandable outside of the library domain; and it is **invisible** in the linked open data

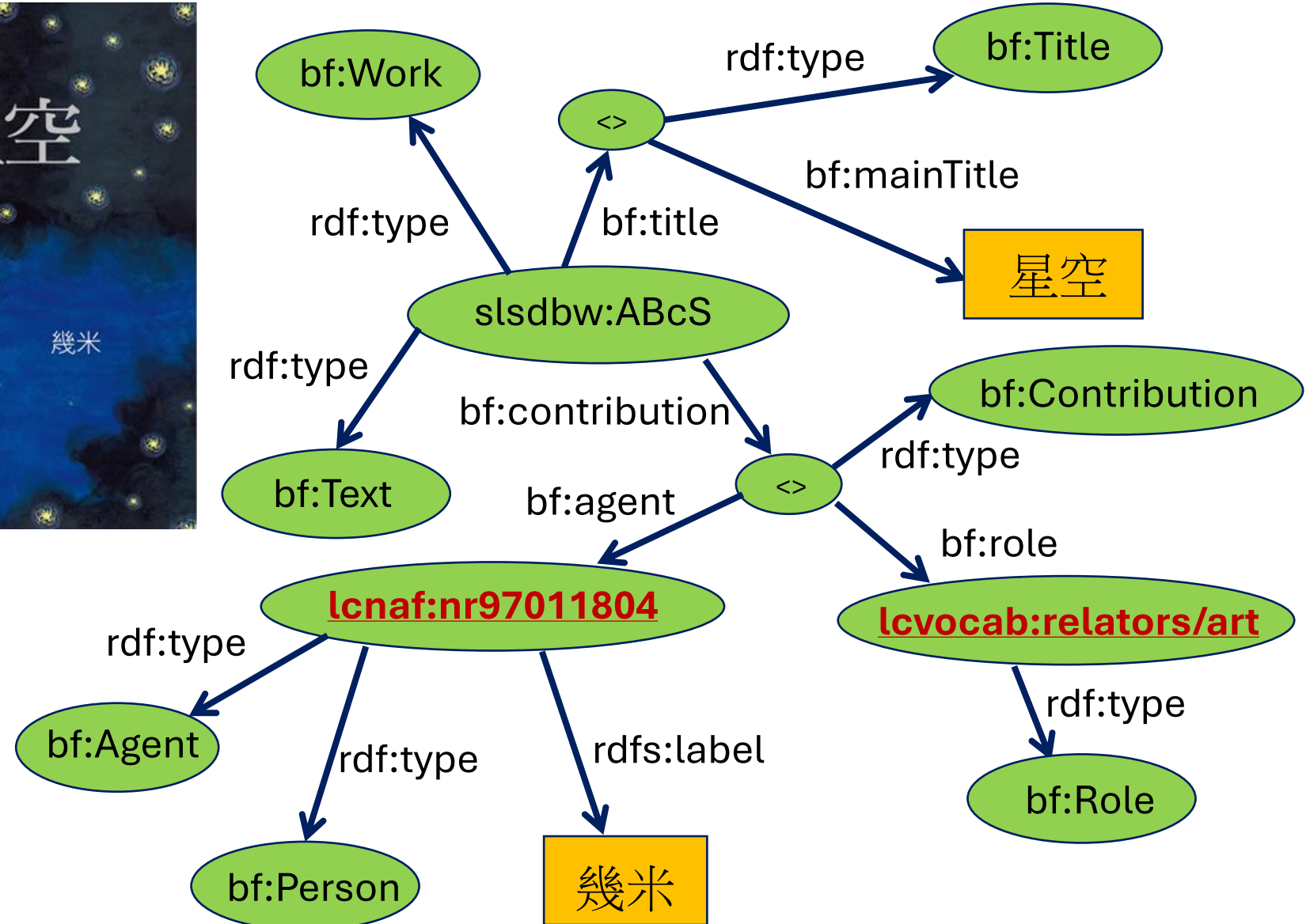
BIBFRAME ontology

- **BIBFRAME** (Bibliographic Framework) is developed by Library of Congress, with the intention to **replace MARC**.
- First draft was release in 2012. Momentum of adoption began catching up only after **BIBFRAME 2.0** was released in 2016.
- Library of Congress is now routinely 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 2025 (https://www.bfwe.eu/naples_2025)
 - BIBFRAME June 2025 Update Forum (<https://loc.gov/bibframe/news/bibframe-update-jun2025.html>)



Bibliographic metadata in BIBFRAME linked data

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



Design principles

- Light-weight, simple, single-page.
- Ideal for library staff who are new to cataloguing or have limited formal cataloging training.

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

The screenshot displays the SLS Demo BIBFRAME Editor interface. At the top, there is a navigation bar with the SLS Demo logo, a user account link, and a language selector. Below this is a header for the 'BIBFRAME Editor' and a list of export and view options. The main content area shows an instance record for 's1sdb:ABcS--20220428--20240829'. The record is organized into sections: 'Title', 'Statement of responsibility', 'Edition', and 'Provision activity'. Each section contains input fields for various attributes, such as 'Main title', 'Subtitle', 'Type', 'Place', 'Agent', 'Date', and 'Place code'. The 'Title' section includes 'Main title' (星空) and 'Subtitle' (The starry starry night). The 'Statement of responsibility' section contains '幾米'. The 'Edition' section contains '初版'. The 'Provision activity' section includes 'Type' (Publication), 'Place' (台北市), 'Agent' (大塊文化出版股份有限公司), 'Date' (2009), and 'Place code' (Taiwan). At the bottom, there is a 'Date 1/2 (MARC)' field with the value '2009'. The interface is clean and user-friendly, with clear labels and intuitive controls.

Enrichment – blending knowledge cards in library catalog

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



Comics 
漫畫 
Creative work in which images and text convey information such as narratives 藝術類型



星空 : The starry starry night
幾米
初版. 台北 : 大塊文化出版股份有限公司, 2009
[138] p. ; 26 cm
Book


Links
▶ [內文試閱 \(幾米網站\)](#)

Item availability


Location	Call number	Barcode	Status
Chinese ERS	859.6 2290	A1706299	Due: 2025-02-19

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


Related items




老少女心事




我住在台灣了!: 港人居台第 三年



教書辛酸史. 5. 露思兔子停課不停學



奶茶通俗學漫畫集: The comix collection of m...

More about... 

Library dedicated chatbot

Coming soon?

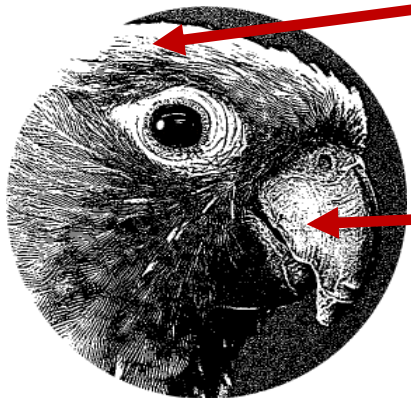
Library dedicated **chatbot** deploying **linked data** and **large language model** to serve library users' information needs

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.

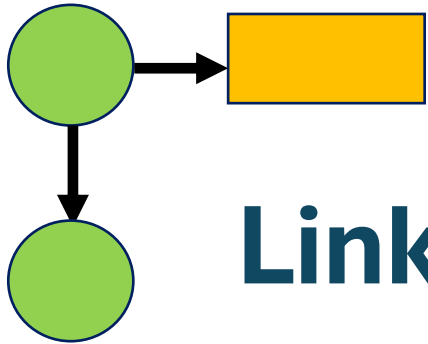


AI Reasoning

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

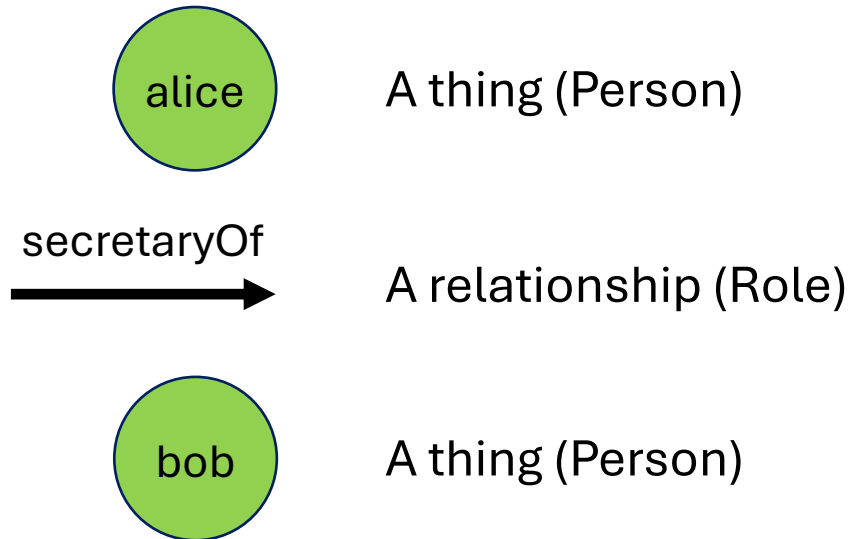
RDF triple

A statement in English text (literal):

“Alice is a secretary of Bob”

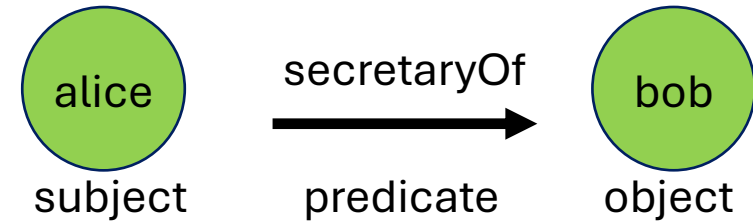
It is difficult for computers to understand the semantics conveyed in an unstructured text string

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!



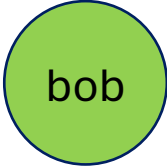
Using triple to express a statement structurally



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>	secretaryOf 	<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

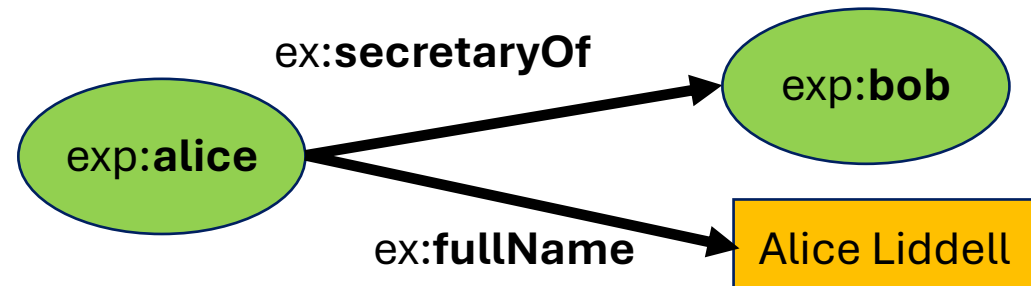
<u>Subject</u>	<u>Predicate</u>	<u>Object</u>
exp:alice	ex:secretaryOf	exp:bob

Object can be an URI or a literal

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

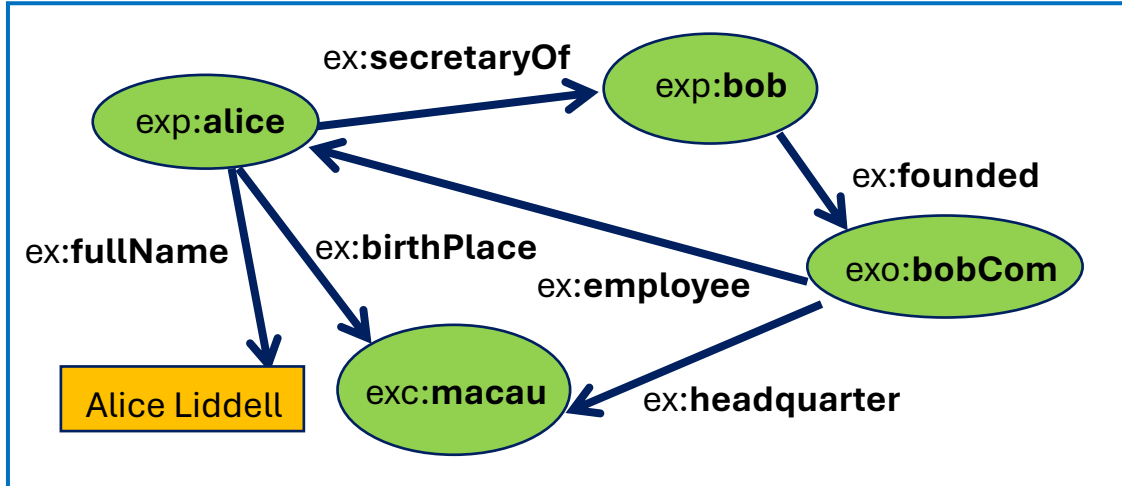
This triple tells the computer that Alice’s full name is “Alice Liddell”, which is a **literal** (text in English)

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>
```

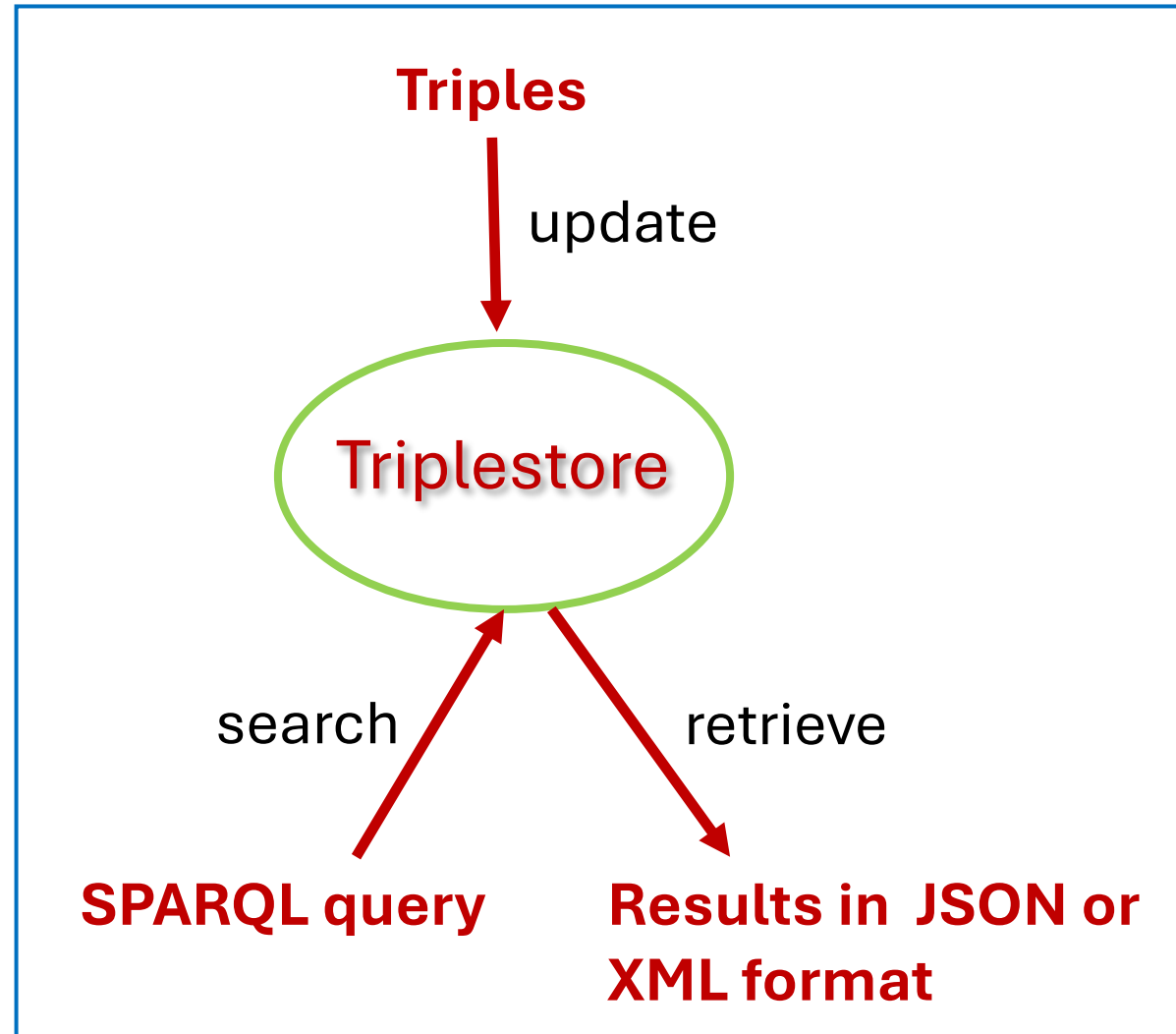
Serialization [cont.]

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

Example: SLS linked query data

SPARQL query editor

Query result in table

The screenshot shows the 'Linked data query service' interface. At the top, there is a 'Query' section with a text editor containing the following SPARQL query:

```
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
```

Below the query editor, there are two tabs: 'Table' (selected) and 'Response'. The 'Table' tab shows the results of the query, indicating '163 results in 0.202 seconds'. The table has three columns: 's', 'p', and 'o'. The first row of results is highlighted in yellow.

	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/ACQs>

Vocabulary – an example

Classes

Class	URI (full form)	URI (prefixed form)	Description
Person	<http://example.com/vocab/Person>	ex:Person	A person

Properties

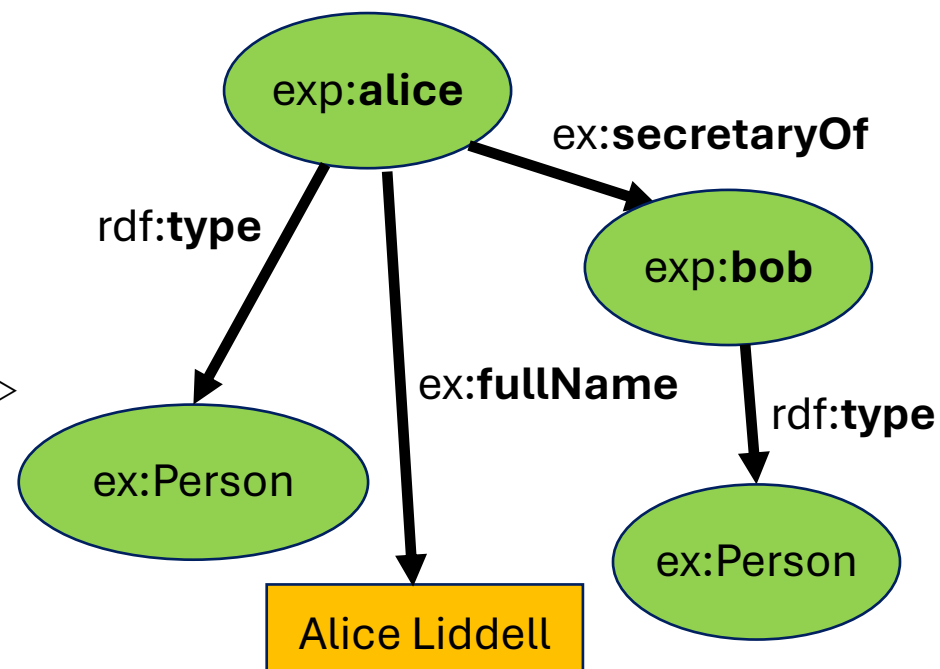
Property	URI (full form)	URI (prefixed form)	Type of value	Description
fullName	<http://example.com/vocab/fullName>	exp:fullName	Text	Full name of the person
secretaryOf	<http://example.com/vocab/secretaryOf>	exp:secretaryOf	Person	A secretary of some person

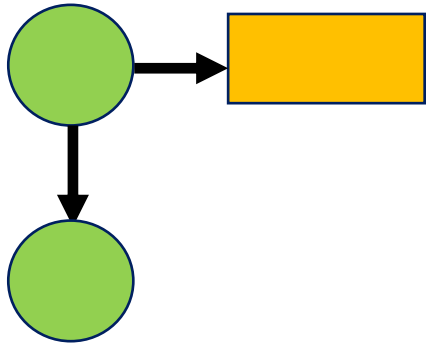
Example data in English

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?