

**Author(s): Rafael Barros and Bhagyasha Patil Date: November 5th 2024**



**Context and Meaning, Concepts, Structure, Applications, Limitations and an example with SPARQL and Wikidata**

**Wikidata Environment** - Before Wikidata, data across Wikimedia projects such as Wikipedia existed in a *unstructured and decentralized format*; - **Example:** each wikipedia language version maintained its own datasets for items like birthdates, coordinates and so on; - **Result:** it led to redundancy, inconsistency and lack of coherence across languages;

### Multi-Language centralized Data Hub needed

- The **growing use of linked data in web applications and knowledge graphs** revealed the need for a centralized, *language-independent*, machine-readable repository;

# Purposes

- Wikidata's goal is centralizing data for all Wikimedia projects and freely available for all users for reuse;
- **mission:** make the structured knowledge available for both humans and machines;
- **Core focus:** *Open Data* that can be edited by anybody and also reused anywhere;

- defining **items** (unit) like people, places and **property** which describes its attributes and relationships, like 'birth date', 'occupation', 'educated at' etc;
- **Statements and references**: a combo ( item-property-value ) define a factual information;

## Structure and Model

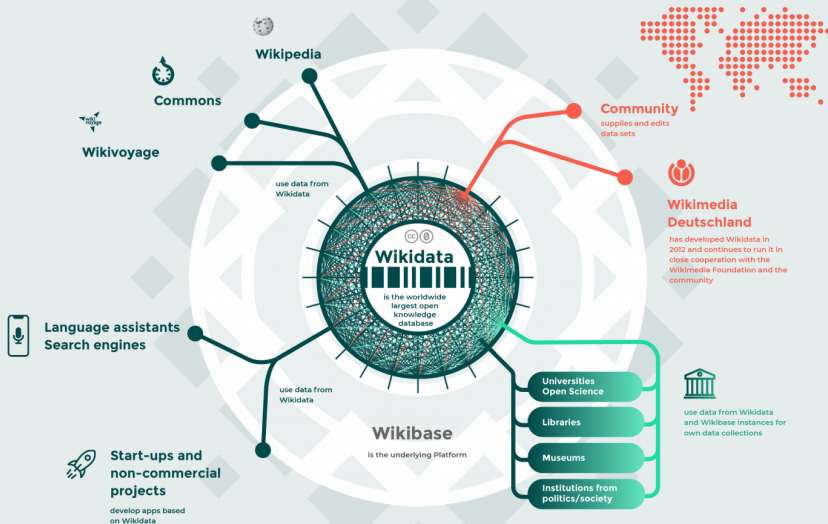
- same principle as with RDF based on **triples format** (subject-predicate-object);
- **Example:**

```
- ASK WHERE {  
  dbr:Cologne dbo:country dbr:Germany .  
}
```



Figure 2: order

# What is Wikidata?





- *a free, collaborative, multilingual knowledge base.*
- part of the **Wikimedia Foundation** (like Wikipedia - an encyclopedia hosted by Wikimedia which provides the infrastructure for thousands of global volunteers to create and edit pages).
- Wikidata stores information as structured data in a database.
- the data stored and shared as LOD is discernable by machines and intelligent devices like Alexa, Siri, and others.
- **Link to NLP (see in applications):** LOD provides structured, semantically rich data, which helps NLP algorithms interpret words and concepts in context;  
**Example:** if a virtual assistant like Siri queries a dataset containing LOD, it can retrieve precise information about entities (like people, places, or objects) and their relationships, improving its ability to understand and respond accurately.
- **Bullet point:** Central source for structured data that can be used across Wikimedia projects and beyond

# Thinking of Contributing?

**How to get started?** - Set up an account - Basic editing tools

**Types of Contributions** - Adding new entities - Editing or updating statements - Translating labels and descriptions

**Community Guidelines** - Notability of entities - Citing reliable sources - Respecting data accuracy and applications

**Why Contribute?** - Improve global knowledge

- **Wikidata** as a training dataset for AI/ML models;
- **NLP**: Entity Recognition and Machine Translation as linked in the picture shown;
- **chatbots** and **voice assistants** like alexa need entity databases ( well-structured by wikidata );
- **Integration with other platforms** - Knowledge Graph, VIAF and Europeana;

- **Healthcare:** medical research databases, linking diseases and treatments;
- **Cultural Heritage:** Museum data, cataloging artifacts;
- **E-Commerce:** Product categorization and recommendation systems;

## Others Real-World User Cases

- Digital Libraries and archives
- Cultural and Educational projects
- Academic and scientific research
- Used to populate infoboxes and update articles across languages

- Data quality and vandalism issues due to open editing data
- Data management and maintenance in terms of complexity
- Multilingual data challenges and need for standardization

# Querying Wikidata with SPARQL

The image shows a Wikidata profile for Douglas Adams (Q42) with various annotations explaining SPARQL query components. The annotations are as follows:

- label**: Points to the name "Douglas Adams".
- item identifier**: Points to the identifier "(Q42)".
- description**: Points to the text "English writer and humorist".
- aliases**: Points to the text "Douglas Noël Adams | Douglas Noel Adams".
- property**: Points to the "educated at" property.
- value**: Points to the value "St John's College".
- qualifiers**: Points to the table of qualifiers for "St John's College".
- rank**: Points to the "rank" column in the table.
- statement group**: Points to the "statement group" column in the table.
- opened references**: Points to the "opened references" section.
- collapsed reference**: Points to the "collapsed reference" section.

The Wikidata profile content is as follows:

**Douglas Adams** (Q42)

English writer and humorist

Douglas Noël Adams | Douglas Noel Adams

► In more languages

**Statements**

rank	statement group	property	value	qualifiers								
		educated at	St John's College	<table border="1"><tr><td>end time</td><td>1974</td></tr><tr><td>academic major</td><td>English literature</td></tr><tr><td>academic degree</td><td>Bachelor of Arts</td></tr><tr><td>start time</td><td>1971</td></tr></table>	end time	1974	academic major	English literature	academic degree	Bachelor of Arts	start time	1971
end time	1974											
academic major	English literature											
academic degree	Bachelor of Arts											
start time	1971											

▼ 2 references

rank	statement group	property	value	qualifiers
		stated in	Encyclopædia Britannica Online	
		reference URL	http://www.nndb.com/people/731/000023662/	
		original language of work	English	
		retrieved	7 December 2013	
		publisher	NNDB	
		title	Douglas Adams (English)	

+ add reference

Brentwood School

end time	1970
start time	1959

► 0 references

+ add (statement)

The following highlighted SPARQL query retrieves all people who studied at Trinity College:

**?person wdt:P69 wd:Q332342 .**

wdt	wd
<property>	<entity>
<educated at>	Trinity College



```
SELECT ?person ?personLabel ?birth_placeLabel ?coordinates
```

```
WHERE{
```

```
    ?person wdt:P69 wd:Q332342 . # all those who studied at Trinity College
```

```
    ?person wdt:P106 wd:Q170790 . # ordered by occupation listed
```

```
    ?person wdt:P21 wd:Q6581097 . # male who studied at Trinity College
```

```
    ?person wdt:P19 ?birth_place . # place of birth , make sure the name given
```

```
    ?birth_place wdt:P625 ?coordinates . # with Label after the name ensure the
```

```
    SERVICE wikibase:label {
```

```
        bd:serviceParam wikibase:language "en" . }
```

```
    #service provided by examples on wikidata query service
```

```
}
```

- Wikidata – Wikimedia Germany
- Wikidata Query Service
- Much more than a mere technology: A systematic review of Wikidata in libraries - ScienceDirect

Thank you for attending our presentation!

