

# Linking Wikidata to the rest of the Semantic Web

Andra Waagmeester<sup>1</sup>, Egon Willighagen<sup>2</sup>, Núria Queralt Rosinach<sup>3</sup>,  
Elvira Mitraka<sup>4</sup>, Sebastian Burgstaller-Muehlbacher<sup>3</sup>, Tim E. Putman<sup>3</sup>,  
Julia Turner<sup>3</sup>, Lynn M Schriml<sup>4</sup>, Paul Pavlidis<sup>5</sup>, Andrew I Su<sup>3</sup>, and  
Benjamin M Good<sup>3</sup>

<sup>1</sup> Micelio, Antwerp 2180, Belgium

`andra@micelio.be`,

<sup>2</sup> Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, the Netherlands

<sup>3</sup> Department of Molecular and Experimental Medicine, the Scripps Research Institute, La Jolla, USA

<sup>4</sup> University of Maryland Baltimore, Baltimore, MD, USA

<sup>5</sup> The University of British Columbia, Vancouver, British Columbia, Canada

**Abstract. Keywords:** federated queries, SPARQL, Wikidata

## Background

Wikidata is the linked database of Wikipedia and other sister projects from the Wikimedia foundation. Similar to Wikipedia, Wikidata is not limited to a select set of knowledge domains, in principle it can capture all knowledge. However, external resources can contain much more granularity on a given topic. Research in the life sciences for example can generate vast amounts of data (e.g. high-throughput screening), so much that storing all in Wikidata will not be possible. More effective would be to consider Wikidata the central hub between more detailed life-science hubs on the semantic web. We present how Wikidata can be linked to and used to link other semantic sources. We compare four federated SPARQL query patterns to create the links.

## Methods

To link Wikidata to the semantic web, we have followed four routes: (1) Using the `FILTER` operator in SPARQL; (2) Using IRIs which are compiled on the fly based on existing -value based- shared identifiers, using SPARQL's `BIND` operator; (3) Storing Wikidata-item identifiers in external data sources as mappings to local identifiers; and (4) Storing remote IRI's as a Wikibase property value of type URL.

## Using the `FILTER` operator in SPARQL queries

If both Wikidata and the external resource supports the same set of identifiers and provided the external SPARQL endpoint supports handling federated

queries, Wikidata and an external source can be linked through a SPARQL query. Since both identifiers will be stored as literals, linking through IRIs is not an option. Using filters in SPARQL is then an option. The FILTER operator from SPARQL limits the results of a graph pattern based on boolean or regular expressions. This only works if the results of both graph patterns are not substantial in size or complexity such that the SPARQL endpoint can handle processing both returns. We demonstrate this using graph patterns from Wikidata and WikiPathways.

### **Using the BIND operator in SPARQL queries**

When, however, the external source contains IRIs for their identifiable concepts, but the equivalent Wikidata property is still available only as a value, a link can be made using the BIND operator, where an IRI is composed based on the property value stored in Wikidata. We demonstrate this using the BIND operator by a federated query which connects Wikidata with Uniprot, allowing merges of protein annotations from both Wikidata and Uniprot.

### **Remote IRI mappings between Wikidata and remote resource**

External resources, like WikiPathways and DisGeNET, actively store mappings between their concepts and the equivalent items in Wikidata. When resources maintain such mappings, a federated query where the remote query pattern and the Wikidata query pattern share the same variables is sufficient to merge results.

### **Local IRI mappings between Wikidata and remote resource**

We have proposed a Wikidata property (P2888) that allows to create mappings between Wikidata and external resources on the semantic web as Wikidata statements. This property has been accepted and we are currently populating items with statements expressing similarity between Wikidata items and others concept descriptions on the semantic web. This Wikidata property is based on `skos:exactMatch`.

### **Conclusion**

Linking Wikidata to other hubs on the semantic web opens up new routes of validation between different resources and toward better integration. We have demonstrated four potential paths to link content from Wikidata with other resources on the semantic web, using federated SPARQL queries.