



# Needing Permanent Identifiers: Wikidata solutions

*David Lindemann*

Connecting Women Writers with Digital Tools  
Women Writers in History DARIAH Working Group Annual Meeting

Faculty of Arts, Universidad Complutense de Madrid

Thu, Jan 15, 2026

# Overview

- ▼ Intro
- ▼ The Semantic Web in three slides
- ▼ Five Star Open Data
- ▼ Linked entities describing persons
- ▼ What can the Wikibase ecosystem do for you?
- ▼ Wikidata
- ▼ Wikibase



<https://1.eus/womenwriters>



# About me



- Political Science > Translation and Interpretation > Basque Philology > Lexicography
- Ayudante Doctor at EHU, active in [DARIAH-WG DHwiki](#), [CLARIAH-EUS](#)
- Role in local research group: Consultancy in DH methods
- Wikimedia Basque Country > [Workshop Wikidata \(Basque\)](#) > [Workshop Wikibase \(English\)](#)
- Involved in some [projects involving Wikibase](#) / in a [new Horizon project](#) about entity linking
- Online profiles listed at: <https://www.wikidata.org/wiki/Q57694630>

# Open Science: FAIR Principles



**F**indable

Data and materials enriched with metadata assigned with a unique identifier



**A**ccessible

Data and metadata stored in a trusted repository with an open and free protocol. Accessible by machines and humans



**I**nteroperable

Using vocabularies and public domain ontologies the metadata can be referenced and linked



**R**eusable

Additional documentation and protocols describing the acquisition of the data, licensed with a detailed provenance



**WIKIBASE**



**WIKIBASE**



**WIKIBASE**

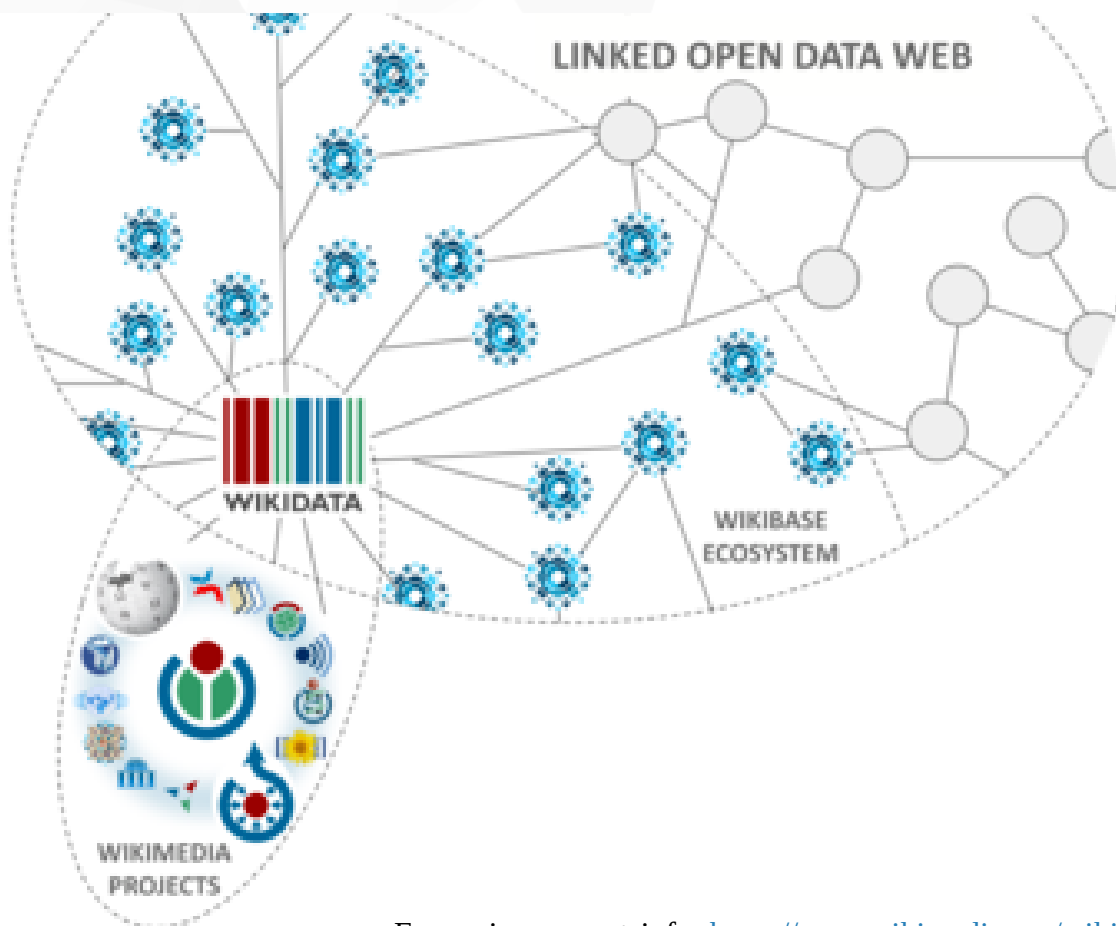


**WIKIBASE**

Fuente de la imagen

# Spoiler: Wikibase is the best

- ▼ "An empty Wikidata" for everybody, for own content
- ▼ Federation with Wikidata ("Wikibase Ecosystem")



A federated query:  
Wikibase-Wikidata

Fuente imagen y + info: [https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021/Joint\\_Vision](https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021/Joint_Vision)

# Intro: FAIR Linked Open Data - some theory

# FAIR: Linked Open Data

- For fulfilling the FAIR criteria, Linked Open Data is probably the best option
- Linked Open Data: a paradigmatic shift
  - from *web* to *semantic web*
  - form *hypertext* to *semantic triple*
- We will see in the following:
  - Semantic triples
  - Entities instead of literal values, concepts instead of terms
  - Examples: persons as linked entities

## La Mujer ilustrada

Eladia Bautista y Patier  
La Mujer



### Resumen

Poema de Eladia Bautista y Patier. Compuesto por veintinueve estrofas de cuatro versos endecasílabos con rima consonante. Riman el primer verso con el tercero y el segundo con el cuarto.

### Violencias de/contra la mujer

**1. Metaforización (mujer como planta o flor que necesita de cuidados [de instrucción]); estereotipos patriarcales de género (se presenta a la mujer como alguien nacida para complacer, pero que tiende al mal y al vicio y que, sin guía ni educación, se desvía del buen camino)**

Cuando la bella planta que ha nacido

Para ser con esmero cultivada,

Por inercia culpable, en el olvido

Se la deja crecer abandonada;

[...]

<https://www.ub.edu/genvipref/revistas/la-mujer>

- Hipertextuality in a catalogue

# Tim Berners-Lee and the Web 1.0: From text to hypertext

## Tim Berners-Lee

142 languages

Article Talk

Read View source View history Tools

From Wikipedia, the free encyclopedia

**Sir Timothy John Berners-Lee** (born 8 June 1955),<sup>[1]</sup> also known as **TimBL**, is an English **computer scientist** best known as the inventor of the **World Wide Web**, **HTML**, the **URL** system, and **HTTP**. He is a professorial research fellow at the **University of Oxford**<sup>[2]</sup> and a professor **emeritus** at the **Massachusetts Institute of Technology** (MIT).

Berners-Lee proposed and implemented the **Transfer Protocol**.<sup>[10][11]</sup> He devised and helped foster the development of the **Web Foundation** and the **Academy of Sciences**.<sup>[12][13]</sup>

Berners-Lee was previously a senior researcher and holder of the **3Com** founder's chair at the **MIT Computer Science and Artificial Intelligence Laboratory** (CSAIL).<sup>[14]</sup> He is a director of the **Web Science Research Initiative** (WSRI)<sup>[15]</sup> and a member of the advisory board of the **MIT Center for Collective Intelligence**.<sup>[16][17]</sup> In 2011, he was named a member of the board of trustees of the **Ford Foundation**.<sup>[18]</sup> He is a founder and president of the **Open Data Institute** and is an advisor at social network **MeWe**.<sup>[19]</sup> In 2004, Berners-Lee was knighted by Queen **Elizabeth II** for his

Sir

**Tim Berners-Lee**

OM KBE FRS RDI FRSA DFBCE FREng



Berners-Lee in 2024

<b>Born</b>	Timothy John Berners-Lee 8 June 1955 (age 70) London, England
<b>Other names</b>	TimBL • TBL
<b>Education</b>	Queen's College, Oxford (BA)
<b>Known for</b>	Invention of the <b>World Wide Web</b>

# From hypertext to the *Semantic Web*

## ▼ Hyptertext – *Link* from subject to object

### ▼ Berners-Lee was born in [London](#).

```
<html>
```

```
  Berners-Lee was born in <a href="https://en.wikipedia.org/wiki/London">London</a>.
```

```
</html>
```

### ▼ Semantic value: Subject (word “London”) > Object (URL of Wikipedia article)

## ▼ RDF: Semantic Triple – Subject, predicate, object

### ▼ Berners-Lee was born in London.

BERNERS-LEE > PLACE OF BIRTH > LONDON

SUBJECT > <http://www.wikidata.org/entity/Q80>

PREDICATE > <http://www.wikidata.org/entity/P19>

OBJECT > <http://www.wikidata.org/entity/Q84>

## ▼ Uniform Resource Identifier (URI)

### ▼ unique and unambiguous, for subject, predicate and object

### ▼ semantic triple: an assertion with a meaning, transparent for human and machine

# Two more things about RDF

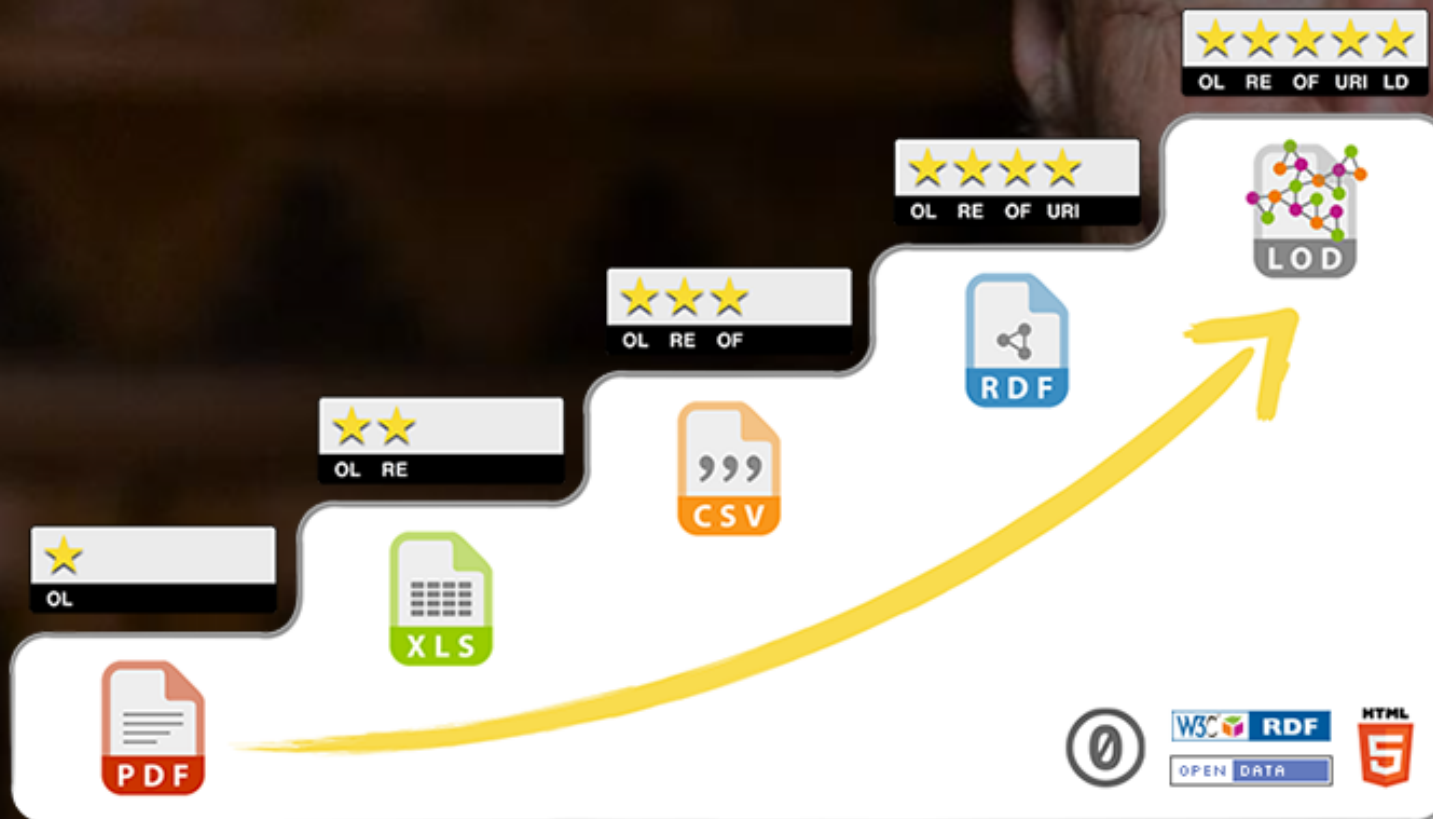
- ▼ The object of a triple may be...
  - ▼ a URI representing an entity
  - ▼ a *data value*, with a *data type*, for example:
    - ▼ a *string*
    - ▼ a *date*
    - ▼ a web page URL
- ▼ Recommended vocabularies / ontologies
  - ▼ There are a lot of them, the [W3C](#) recommends
  - ▼ see <https://lov.linkeddata.es>
  - ▼ for example, regarding bibliographical data
    - ▼ Dublin Core <https://www.dublincore.org/schemas/rdfs/>
    - ▼ BIBO <http://purl.org/ontology/bibo/>
    - ▼ LRM, by IFLA [https://en.wikipedia.org/wiki/IFLA\\_Library\\_Reference\\_Model](https://en.wikipedia.org/wiki/IFLA_Library_Reference_Model)
    - ▼ BIBFRAME, by LOC <https://www.loc.gov/bibframe/>

# Berners-Lee's Five Star Open Data

# 5 ★ OPEN DATA

<https://5stardata.info>

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data. Here, we give examples for each step of the stars and explain costs and benefits that come along with it.



# 1 star: a PDF document



## BIBLIOGRAFÍA

OBSERVACIONES: Las reseñas se consignan a continuación del libro (o artículo) reseñado. Cuando éste se ha descrito en una entrega anterior de la Bibliografía, se pone entre paréntesis el número de ficha correspondiente. También se hacen referencias a ediciones anteriores de una obra, a tomos anteriores de una obra en curso de publicación, etc. En estos casos, la cifra que precede al número de referencia indica el volumen de la NRFH en que figura la ficha en cuestión. Así, el "núm. 39-39695" corresponde a la ficha 39695, la cual está en el volumen 39 de la NRFH. Cuando no hay cifra inicial, ha de entenderse que el número remite a una ficha incluida en el presente fascículo.

### LINGÜÍSTICA HISPÁNICA

#### OBRAS GENERALES

- 45-1363. ALARCOS LLORACH, EMILIO – "Apostillas leves a las ponencias". – EAC, 61 (1994), 81-86.
- 45-1364. ARIZA VIGUERA, MANUEL – "África y América: dos ejes de la investigación del profesor Germán de Granda". – ALM, 34 (1996), 273-290.
- 45-1365. BENIERS, ELIZABETH – "Bibliografía de trabajos descriptivos del español de México". – ALM, 34 (1996), 293-349.
- 45-1366. BRUMME, JENNY – "El español moderno y el siglo XIX, en especial, como objeto de estudio en la historia de la lengua. (Balance, lagunas y tareas)". – *Estudis de lingüística i filologia oferts a Antoni M. Badia i Margarit* (Barcelona, 1996-97), t. 1, pp. 131-140.
- 45-1367. BRUNO, LIDIA, y MIRTA SERAFINI – "Lenguaje, idiosincrasia y educación entre los mapuches". – S&S, 1995, núm. 4, 223-235.
- 45-1368. CABRÉ, M. TERESA – "Sobre manlleus i diccionaris". – *Estudis de lingüística*

*ca i filologia oferts a Antoni M. Badia i Margarit* (Barcelona, 1996-97), t. 3, pp. 39-61.

- 45-1369. CORTÉS RODRÍGUEZ, LUIS – *Tendencias actuales en el estudio del español hablado*. – Universidad, Almería, 1994. 153 pp. (*Monografías Humanidades*, 2). II *Verba*, 22 (1995), 677-680 (Casanovas).
- 45-1370. GÓMEZ TORREGO, L. – *Manual del español correcto*. 4ª ed. corr. y aum. – Arco Libros, Madrid, 1994. 2 ts. II EAC, 60 (1993), 109 (Albalá).
- 45-1371. *Historia de la alfabetización y la educación para adultos en México*. – SEPEI Colegio de México, México, 1994. xxxi + 772 pp. II *Ciencia y Desarrollo*, México, 1996, núm. 130, 90-92 (Barriga Villanueva).
- 45-1372. LEBSANFT, FRANZ – *Spanische Sprachkultur*. – Max Niemeyer, Tübingen, 1997. xv + 322 pp. (BZRP, 282).
- 45-1373. LÓPEZ MORALES, H. – *Métodos de investigación lingüística* (N. 44-22). II RFE, 75 (1995), 379-381 (Hernández).
- 45-1374. MACRÌ, ORESTE – *Studi Ispanici*. T. 2: *Critici*. Ed. de Laura Dolfi. – Liguori Editore, Napoli, 1996. 463 pp.

NRFH, XLV (1997), núm. 2, 541-608

**arranparrila**. iz. (1912; *arranbarrilla* \*1883, 1894; *arramarrilla* \*1905). Paraleloki ezarritako burdin hariz osatutako tresna giderduna, janariak sutan prestatzeko erabiltzen dena. Ik. **grise-la matxarro** Zartagia eta arranparrilak. Okela arranparrilan errea.

**arrantxo**. iz. (\*XVIII, 1930). *Heg.* (Beh.). [esp]. Soldaduen jana-ria, jatena. *Arrantxo ederra egin dut hementxe jarrita, diru gabe nintzena zeuden igarrita.*

**1 arrantza**. iz. (1627; *arrainza* 1571). Arrainak harrapatzeko jarduna eta lanbidea. Ik. **arrantzu**. *Arrantzara joan. Ur handitako, itsas zabaleko, itsas bazterreko arrantza. Urrutietako arrantza eta merkataritza. Balearen edo bakailaoaren arrantza. Arrantzatik bizi direnak. Arrantzan ikasteko. Arrantzan egin. Ik. arrantzatu Arrantzan ari direnak. Hamabi marinel, arrantzan usaturik. Bere lagunekin itsas arrantzan aurkitzen zen batez. Arrantza ontzia. Arrantza handia: itsas zabaleko arrantza. Arrantza txikia: itsasbazterreko arrantza.*

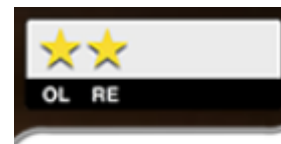
**2 arrantza**. iz. (\*1746, 1800). **1.** Astoaren orroa. *Astoaren arrantza min gabea* (esr. zah.). **2.** Negar zaratatsua. *Seme bat hil eta amak aterako ditu buruko ileak, deiadarraz eta arrantzaz.* — ARRANTZA EGIN. (1842). *Astoa arrantza egiten. II Kantatu beharrean egiten arrantza.*

**arrantzaka**. nola *adlag.* (1842). Arrantza eginez; oihuka. *Astoa arrantzaka hasten denean ez belarri eman. Tramana eta Brix, bata txilioka eta arrantzaka bestea.*

**arrantzale**. iz. (1571; *arrantzale* \*XVII ea., ~1660). Arrainak harrapatzen dituen, berezk., lanbidetzat arrantza duen pertsona. *Arrantzaleak sareak garbitzen ari ziren. Arrantzale Anaidia. II Giza arrantzale egingo zaituztet.*

- If the PDF has a text layer (in addition to the image layer), full text search is possible
- Hypertextual elements (cross-references, http links)

# 2 stars: an excel table (or a MARC catalogue entry)

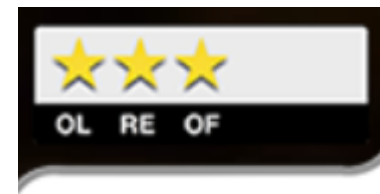


	A	B	C
1	<i>Author</i>	<i>Year</i>	<i>Title</i>
2	Michelena Elissalt, Luis	1953	Apellidos vascos
3	Michelena, Luis	1964	Textos arcaicos vascos
4	Michelena Elissalt, Luis	1970	Estudio sobre las fuentes del Diccionario de Azkue
5	Mitxelena, Koldo; Sarasola, Ibon	1986	Diccionario general vasco - Orotariko euskal hiztegia

040			a KM  b spa
080			a 929.5.09:809.169
080			a 809.169-313.2
100	1		a Mitxelena, Koldo,  d (1915-1987)
245	1	0	a Apellidos vascos /  c Luis Michelena.
260			a San Sebastián :  b Biblioteca Vascongada de los Amigos del País,  c 1953.
300			a 158 p. ;  c 23 cm
490	1		a Monografías vascongadas ;  v 11

- Structured data (and the structure can be explained to a machine)
- The format has limits
  - Proprietary format?
  - A machine can't interpret everything (here: colour)

# 3 stars: a CSV table



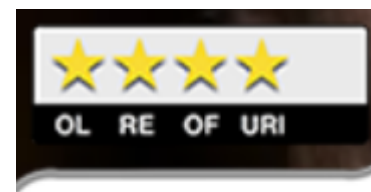
	A	B	C	D
1	<i>Author</i>	<i>Year</i>	<i>Title</i>	<i>status</i>
2	Michelena Elissalt, Luis	1953	Apellidos vascos	green
3	Michelena, Luis	1964	Textos arcaicos vascos	green
4	Michelena Elissalt, Luis	1970	Estudio sobre las fuentes del Diccionario de Azkue	red
5	Mitxelena, Koldo; Sarasola, Ibon	1986	Diccionario general vasco - Orotariko euskal hiztegia	green
1	"Author","Year","Title","status"			
2	"Michelena Elissalt, Luis",1953,"Apellidos vascos","green"			
3	"Michelena, Luis",1964,"Textos arcaicos vascos","green"			
4	"Michelena Elissalt, Luis",1970,"Estudio sobre las fuentes del Diccionario de Azkue","red"			
5	"Mitxelena, Koldo; Sarasola, Ibon",1986,"Diccionario general vasco - Orotariko euskal hiztegia","green"			

▼ Universal format (does not depend on certain apps)

▼ All content can be interpreted by a machine

MARC to CSV: <https://pypi.org/project/marctable/>

## 4 stars: RDF



	A	B	C
1	<i>Author</i>	<i>Year</i>	<i>Title</i>
2	Michelena Elissalt, Luis	1953	Apellidos vascos
3	Michelena, Luis	1964	Textos arcaicos vascos
4	Michelena Elissalt, Luis	1970	Estudio sobre las fuentes del Diccionario de Azkue
5	Mitxelena, Koldo; Sarasola, Ibon	1986	Diccionario general vasco - Orotariko euskal hiztegia

**PREFIX** example: <http://example.org/entity/>

example:work1 example:author "Michelena Elissalt, Luis"

example:work1 example:publicationYear "1953"

example:work1 example:title "Apellidos vascos"

example:work5 example:author "Mitxelena, Koldo"

example:work5 example:author "Sarasola, Ibon"

example:work5 example:publicationYear "1986"

example:work5 example:title "Diccionario general vasco"@es

example:work5 example:title "Orotariko euskal hiztegia"@eu

▼ Semantic triples:  
subject – predicate – object

▼ Subject and predicate are  
expressed using URI

# 5 stars: LOD (Linked Open Data)



```
example:work1 example:author "Michelena Elissalt, Luis"  
example:work1 example:publicationYear "1953"  
example:work1 example:title "Apellidos vascos"  
  
example:work5 example:author "Mitxelena, Koldo"  
example:work5 example:author "Sarasola, Ibon"  
example:work5 example:publicationYear "1986"  
example:work5 example:title "Diccionario general vasco"@es  
example:work5 example:title "Orotariko euskal hiztegia"@eu
```

▼ Semantic triples using  
**standard vocabularies**

<http://purl.org/dc/elements/1.1/creator>

**PREFIX** dc: <<http://purl.org/dc/elements/1.1/>>

```
example:work1 dc:creator "Michelena Elissalt, Luis"  
example:work1 dc:date "1953"  
example:work1 dc:title "Apellidos vascos"  
  
example:work5 dc:creator "Mitxelena, Koldo"  
example:work5 dc:creator "Sarasola, Ibon"  
example:work5 dc:date "1986"  
example:work5 dc:title "Diccionario general vasco"@es  
example:work5 dc:title "Orotariko euskal hiztegia"@eu
```

# How many stars for a SQL database with web interface?

The screenshot displays the RECIRC website, which focuses on the reception and circulation of early modern women's writing from 1550-1700. The navigation bar includes links for Home, Search the Database, Case Studies, About, Contact, Team Members, Publications, Blog, and Conference. The main content area shows the profile of Ana de Jesús, with tabs for People, Works, Time & Place, and a button for Citing this Data. The Biographical Details section contains a table with the following information:

Maiden Name	de Lobera y Torres
Alias	• Ann Lobera
Name in Religion	Ana de Jesús
Date of Birth	1545
Date of Death	1621
Religious Order	Carmelite
Religious House	Avila, Carmelites: Convento de San José


Summary statistics for Ana de Jesús:


- Is in the Top 50% of women received, in our dataset
- Connected to 4 receptions, in our dataset (Explore them)
- Is linked to 1 person in our dataset (View them)

A bar chart titled "Receptions Connected with Ana de Jesús" shows the distribution of roles within linked receptions:

Role within Linked Receptions	No of Linked Receptions
Named Receiver	0
Author	4
Owner/Compiler/Scribe	0

# How many stars for a SQL database with web interface?

Radboud Universiteit  SHEWROTE ▼ [Persons](#) [Collectives](#) [Publishers](#) [Works](#) [Editions](#) [Circulation](#) [Receptions](#) [Log in](#)

 Please note that the data in this database is not exhaustive, and coverage per region and period is partial, as the tool is intended primarily as a research instrument fed by the data and research questions of individual contributors. Should you wish to contribute to the database, please contact [alicia.montoya@ru.nl](mailto:alicia.montoya@ru.nl) for inlog codes.

## Persons

⌵ Reception count (persons and works) ▼

≡ Apply filters

**Name**

**Gender**



☐ **Year of birth**



☐ **Year of death**



**Country/place of birth**

← **1** 2 3 4 5 6 7 8 9 10 11 12 ... →

11328 persons found.

[Sand, George \(pseudonym\)](#) ♀ Nohant-Vic 1804 — Paris 1876  
Full name: Amantine Aurore Lucile Dupin [VIAF](#)   
Alternative names: J. Sand — Aurore Dupin — Aurore Dudevant  


[Marlitt, E. \(pseudonym\)](#) ♀ Arnstadt 1825 — Arnstadt 1887  
Full name: Friederike Henriette Christiane [VIAF](#)   
Alternative names: Eugenie John  
Marlitt, E.  


[Genlis, Stéphanie Félicité de](#) ♀ Champcéri 1746 — Paris 1830  
Full name: Stéphanie Félicité du Crest de Saint-Aubin [VIAF](#)   
Alternative names: Caroline-Stéphanie-Félicité du Crest de Saint Aubin — Madame de Sillery — Madame de Brulart — Marquise de Sillery — Clémence de G.  


Reception count (persons and works):  
1546

Reception count (persons and works):  
593

Reception count (persons and works):  
410

# How many stars for a SQL database with web interface?

```
{
  "person_id": "51acfe1c-176b-4b2f-9da9-4acd8dc8f638",
  "person_name": "Sta\u00ebl, Germaine de",
  "viaf_id": "http://viaf.org/viaf/89204033/"
},
{
  "person_id": "fe3fa376-209c-42c2-b4b9-a3c6fd376c53",
  "person_name": "Stowe, Harriet Elisabeth Beecher",
  "viaf_id": "http://viaf.org/viaf/32003151/"
},
{
  "person_id": "3f603ee3-7c11-4e04-8e00-92ab623b9924",
  "person_name": "Eliot, George (pseudonym)",
  "viaf_id": "http://viaf.org/viaf/89000553/"
},
```

- ▼ Dataset scraped from SHEWROTE web: 3 stars
- ▼ Then I realized: *person\_id* is the same as the (old?) WomenWriters ID on Wikidata
  - ▼ ...that is potentially five stars!

# Linked entities: entities describing persons

# Semantics in a journal web portal. For humans only?

Aldizkari Zientifikoen Ataria

Anuario del Seminario de Filología Vasca "Julio de Urquijo"

ERREGISTRATU HASI SAIOA UNEKOA ONLINE FIRST ARTXIBOAK GUR

ALDIZKARIAREN EDUKIA

Hizkuntza

- > English
- > Español
- > Euskara

Make a Submission

Make a Submission

Informazioa

- > Irakurleentzat

CAMPIONEN "GRAMÁTICA" ETA BESTE

PDF

Argitaratua 1992-01-15

DOI <https://doi.org/10.1387/asju.9357>

Koldo Mitxelena

Laburpena

Nola aipatu

Mitxelena, Koldo. 1992. «Campionen &quot;gramática&quot; Eta Beste». *Anuario Del Seminario De Filología Vasca "Julio De Urquijo"*, urtarrilak, /10.1387/asju.9357.

Abstract 50 | PDF Downloads 88

Entry in an instance of *Open Journal Systems*

# Publication metadata in semantic triples

Aldizkari Zientifikoen Ataria

Anuario del Seminario de Filología Vasca "Julio de Urquijo"

ERREGISTRATU

HASI SAIOA

UNEKOA

ONLINE FIRST

ARTXIBOAK

GUR

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Creator.PersonalName" content="Koldo Mitxelena"/>
<meta name="DC.Date.created" scheme="ISO8601" content="1992-01-15"/>
<meta name="DC.Date.dateSubmitted" scheme="ISO8601" content="2013-06-03"/>
<meta name="DC.Date.issued" scheme="ISO8601" content="1992-01-15"/>
<meta name="DC.Date.modified" scheme="ISO8601" content="2013-06-03"/>
<meta name="DC.Description" xml:lang="eu" content="-"/>
<meta name="DC.Format" scheme="IMT" content="application/pdf"/>
<meta name="DC.Identifier" content="9357"/>
<meta name="DC.Identifier.pageNumber" content="495-501"/>
<meta name="DC.Identifier.DOI" content="10.1387/asju.9357"/>
<meta name="DC.Identifier.URI" content="https://ojs.ehu.eus/index.php/ASJU/article/view/9357"/>
<meta name="DC.Language" scheme="ISO639-1" content="eu"/>
<meta name="DC.Rights" content="Copyright (c) 2015 Anuario del Seminario de Filología Vasca &quot;Julio de Urquijo&quot;"/>
<meta name="DC.Rights" content=""/>
<meta name="DC.Source" content="Anuario del Seminario de Filología Vasca &quot;Julio de Urquijo&quot;"/>
<meta name="DC.Source.ISSN" content="2444-2992"/>
<meta name="DC.Source.URI" content="https://ojs.ehu.eus/index.php/ASJU"/>
<meta name="DC.Title" content="Campionen &quot;gramática&quot; eta beste"/>
<meta name="DC.Type" content="Text.Serial.Journal"/>
<meta name="DC.Type.articleType" content="Artikuluak"/>
<meta name="gs_meta_revision" content="1.1"/>
<meta name="citation_journal_title" content="Anuario del Seminario de Filología Vasca &quot;Julio de Urquijo&quot;"/>
<meta name="citation_journal_abbrev" content="ASJU"/>
<meta name="citation_issn" content="2444-2992"/>
<meta name="citation_author" content="Koldo Mitxelena"/>
<meta name="citation_title" content="Campionen &quot;gramática&quot; eta beste"/>
<meta name="citation_language" content="eu"/>
<meta name="citation_date" content="1992/01/15"/>
<meta name="citation_firstpage" content="495"/>
<meta name="citation_lastpage" content="501"/>
<meta name="citation_doi" content="10.1387/asju.9357"/>
```

# Metadata with complete semantics: linked entities

- Semantic triples:  
subject – predicate – object
- The three of them are entities

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Creator.PersonalName" content="Koldo Mitxelena"/>
<meta name="DC.Date.created" scheme="ISO8601" content="1992-01-15"/>
<meta name="DC.Date.dateSubmitted" scheme="ISO8601" content="2013-06-03"/>
<meta name="DC.Date.issued" scheme="ISO8601" content="1992-01-15"/>
<meta name="DC.Date.modified" scheme="ISO8601" content="2013-06-03"/>
<meta name="DC.Description" xml:lang="eu" content="-"/>
<meta name="DC.Format" scheme="IMT" content="application/pdf"/>
<meta name="DC.Identifier" content="9357"/>
<meta name="DC.Identifier.pageNumber" content="495-501"/>
<meta name="DC.Identifier.DOI" content="10.1387/asju.9357"/>
<meta name="DC.Identifier.URI" content="https://ojs.ehu.eus/index.php/ASJU/article/view/9357"/>
<meta name="DC.Language" scheme="ISO639-1" content="eu"/>
<meta name="DC.Rights" content="Copyright (c) 2015 Anuario del Seminario de Filología Vasca &quot;Julio de Urquijo&quot;"/>
<meta name="DC.Rights" content=""/>
<meta name="DC.Source" content="Anuario del Seminario de Filología Vasca &quot;Julio de Urquijo&quot;"/>
<meta name="DC.Source.ISSN" content="2444-2992"/>
<meta name="DC.Source.URI" content="https://ojs.ehu.eus/index.php/ASJU"/>
<meta name="DC.Title" content="Campioen &quot;gramática&quot; eta beste"/>
<meta name="DC.Type" content="Text.Serial.Journal"/>
<meta name="DC.Type.articleType" content="Artikuluak"/>
<meta name="gs_meta_revision" content="1.1"/>
<meta name="citation_journal_title" content="Anuario del Seminario de Filología Vasca &quot;Julio de Urquijo&quot;"/>
<meta name="citation_journal_abbrev" content="ASJU"/>
<meta name="citation_issn" content="2444-2992"/>
<meta name="citation_author" content="Koldo Mitxelena"/>
<meta name="citation_title" content="Campioen &quot;gramática&quot; eta beste"/>
<meta name="citation_language" content="eu"/>
<meta name="citation_date" content="1992/01/15"/>
<meta name="citation_firstpage" content="495"/>
<meta name="citation_lastpage" content="501"/>
<meta name="citation_doi" content="10.1387/asju.9357"/>
```



Item Discussion Read View history ☆ More ▾

Koldo Mitxelena (Q2457804)

Basque writer

Koldo Mitxelena Elissac Luis Michelena

▼ In more languages

Language	Label	Description
English	Koldo Mitxelena	Basque writer

"Koldo Mitxelena" in Wikidata`

"Koldo Mitxelena" graph viz

# What's the point with names: They are ambiguous...

- ▼ A person may appear with different names
  - ▼ spelling variants
  - ▼ different names that apply to different parts of a person's life
  - ▼ pseudonyms
- ▼ Different persons may have the same name
- ▼ ...how was this solved before Linked Data?
  - ▼ with identifiers (e.g. a passport number)
    - ▼ but this is only used by the State, not e.g. in library records
  - ▼ in so-called *authority files* (as until today in many libraries)
    - ▼ a commission decides on a (hopefully) unique combination of a name and birth & death dates

# Legacy identifiers: a *MARC record* about a person (Basque Govt. authority file)

```
<record>
  <leader>000000nz  22000000n  4500</leader>
  <controlfield tag="001" >00028512</controlfield>
  <datafield tag="100" ind1="1" ind2=" " >
    <subfield code="a" >Mitxelena, Koldo (</subfield>
    <subfield code="d" >1915-1987)</subfield>
  </datafield>
  <datafield tag="370" ind1=" " ind2=" " >
    <subfield code="a" >Erreterria, Gipuzkoa </subfield>
    <subfield code="b" >Erreterria, Gipuzkoa</subfield>
  </datafield>
  <datafield tag="400" ind1="1" ind2=" " >
    <subfield code="a" >Michelena, Luis (</subfield>
    <subfield code="d" >1915-1987)</subfield>
  </datafield>
  <datafield tag="400" ind1="1" ind2=" " >
    <subfield code="a" >Michelena Elissalt, Luis (</subfield>
    <subfield code="d" >1915-1987)</subfield>
  </datafield>
  <datafield tag="400" ind1="1" ind2=" " >
    <subfield code="a" >Mitxelena Elissalt, Koldo (</subfield>
    <subfield code="d" >1915-1987)</subfield>
  </datafield>
</record>
```

- ▼ In MARC, field 100 is the name variant used for linking
- ▼ **example:** in field 100 of a book record, same name literal as in field 100 of the author record
- ▼ The numeric ID in the controlfield is internal to the software system

# MARC record about a book (Fundación Sancho el Sabio)

```
<record>
<leader>01043nam a2200253 i 4500</leader>
<controlfield tag="001">20722</controlfield>
<datafield tag="100" ind1="1" ind2=" " >
  <subfield code="a">Mitxelena, Koldo</subfield>
  <subfield code="d">1915-1987</subfield>
  <subfield code="9">4554</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Basque language</subfield>
  <subfield code="b">: Euskara</subfield>
  <subfield code="c">/ [Luis Michelena]</subfield>
</datafield>
<datafield tag="260" ind1=" " ind2=" " >
  <subfield code="a">[Bilbao]:</subfield>
  <subfield code="b">Euskaltzaindia </subfield>
  <subfield code="a">[Vitoria-Gasteiz]:</subfield>
  <subfield code="b">Basque Government,</subfield>
  <subfield code="c">198-?</subfield>
</datafield>
</record>
```

▼ <https://catalogo.sanchoelsabio.eus/Record/20722>

▼ Contains ID

▼ <https://catalogo.sanchoelsabio.eus/Authority/4554>

▼ The numeric ID can be used for external links

▼ cf. <https://www.wikidata.org/wiki/Q2457804#P13804>

▼ A MARC field 100 is not a permanent ID (Pellizzari in Lindemann et al. 2026)

# MARC record of a book (BNE)

```
<record>
  <leader>01629nam a2200349 i 4500</leader>
  <datafield ind1="1" ind2=" " tag="100">
    <subfield code="a">Michelena, Luis</subfield>
    <subfield code="d">1915-1987</subfield>
  </datafield>
  <datafield ind1="1" ind2="4" tag="245">
    <subfield code="a">The Basque language</subfield>
    <subfield code="h">[Texto impreso]</subfield>
    <subfield code="c">by Koldo Mitxelena ; translated by
Gloria Castresana Waid</subfield>
  </datafield>
</record>
```

Encabezamiento autorizado	Michelena, Luis. 1915-1987
Encabezamientos alternativos	Michelena, L. -- 1915-1987 Michelena Elissalt, Luis -- 1915-1987 Mitxelena, K. -- 1915-1987 Mitxelena, Koldo -- 1915-1987 Mitxelena, Luis -- 1915-1987 Mitxelena Elissalt, Koldo -- 1915-1987 Mitxelena Elissalt, Luis -- 1915-1987
Datos encontrados en la fuente	Diccionario general vasco, 1987 -- port. (Luis Michelena) Comparación y reconstrucción en lingüística, 1981 -- port. K. Mitxelena: gizona eta hizkuntza, 2007 Koldo Mitxelena, 2001 -- port. (Koldo Mitxelena) solapa (R Zenbait hitzaldi, 1990 -- port. (Koldo Mitxelena)
Otro identificador estándar	<a href="http://viaf.org/viaf/49241578/">http://viaf.org/viaf/49241578/</a> <a href="http://www.larramendi.es/aut/POLI20090015302">http://www.larramendi.es/aut/POLI20090015302</a> <a href="https://isni.org/isni/0000000121318905">https://isni.org/isni/0000000121318905</a> <a href="http://www.wikidata.org/entity/Q2457804">http://www.wikidata.org/entity/Q2457804</a>



cf. <https://www.wikidata.org/wiki/Q2457804#P950>

▼ [https://catalogo.bne.es/permalink/34BNE\\_INST/f0g01i/alma991008382699708606](https://catalogo.bne.es/permalink/34BNE_INST/f0g01i/alma991008382699708606)

▼ Author entry (old catalogue)  
[https://bne.primo.exlibrisgroup.com/permalink/34BNE\\_INST/379ebu/981060866676808606](https://bne.primo.exlibrisgroup.com/permalink/34BNE_INST/379ebu/981060866676808606)

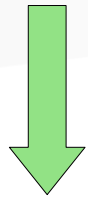
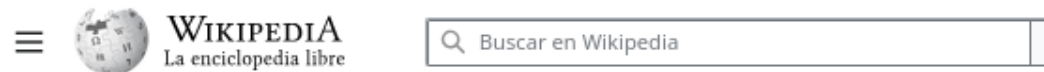
▼ Author URI (LOD)  
<https://datos.bne.es/resource/XX1053652>

▼ Access to the catalogue through ID  
[https://bne.primo.exlibrisgroup.com/discovery/authsearch?authorityQuery=XX1053652&searchField=keyword&vid=34BNE\\_INST:CATALOGO&offset=0](https://bne.primo.exlibrisgroup.com/discovery/authsearch?authorityQuery=XX1053652&searchField=keyword&vid=34BNE_INST:CATALOGO&offset=0)

# Conclusion (looking e.g. at Basque authors): Wikidata is *de facto* the central pillar

## ▼ Koldo Mitxelena (Q2457804)

- ▼ 41 identifiers:  
catalogues at global, state, regional, local level
- ▼ Wikipedia articles in 13 languages
- ▼ Structured data,  
e.g. for an *infobox*



▼ <https://autoritat.eak.eus/eal00089>

Koldo Mitxelena	
	
Información personal	
Apodo	Adiskide berri <span></span>
Nacimiento	20 de agosto de 1915 <span></span> Errenteria (España) <span></span>
Fallecimiento	11 de octubre de 1987 <span></span> (72 años) San Sebastián (España) <span></span>
Nacionalidad	Española <span></span>
Lengua materna	Euskera <span></span> Castellano <span></span>
Educación	
Educado en	Universidad de Salamanca <span></span> Universidad de Madrid (Doctorado en Filosofía y Letras) <span></span>
Supervisor doctoral	José Vallejo Sánchez <span></span>
Información profesional	
Ocupación	Lingüista, escritor, filólogo, profesor universitario y sociolingüista <span></span>
Empleador	Universidad de Salamanca <span></span> La Sorbona <span></span> Universidad del País Vasco <span></span>
Estudiantes doctorales	Pello Salaburu, Lourdes Oñederra, Miren Azkarate Villar, Ibon Sarasola y José María Sánchez Carrión <span></span>
Rama militar	Euzko Gudarostea <span></span>
Conflictos	Guerra civil española <span></span>
Partido político	Partido Nacionalista Vasco <span></span> Eusko Alkartasuna <span></span>
Miembro de	Real Academia de la Lengua Vasca <span></span>

## ≡ Koldo Mitxelena

[Artículo](#) [Discusión](#)

**Koldobika «Koldo» Mitxelena Elissalt** (**Rentería**, 20 de agosto de 1915-**San Sebastián**, 11 de octubre de 1987), también citado como **Luis Michelena Elissalt**, fue un escritor, filólogo y lingüista **español**. Considerado uno de los mejores conocedores de la **lengua vasca**, fue uno de los promotores de su estandarización.<sup>1</sup>

Fue docente en la **Universidad de Salamanca**, en la **Universidad del País Vasco**, en la **Universidad Sorbona de París**. Miembro de la **Real Academia de la Lengua Vasca**, en 1984 fue galardonado con la **Gran Cruz de la Orden Civil de Alfonso el Sabio**.

# The same exercise with Women Writers data

## ■ Maria de Toledo

■ <https://basededatos.visionarias.es/santa/maria-de-toledo>

■ [https://catalogodesantasvivas.visionarias.es/index.php/Mar%C3%ADa\\_de\\_Toledo](https://catalogodesantasvivas.visionarias.es/index.php/Mar%C3%ADa_de_Toledo)



**BASE DE DATOS**

**María de Toledo**

**Información**

Nacimiento: 1437

Fallecimiento: 1507

Orden: Franciscanas

Convento: Convento de Santa Isabel de los Reyes



**CATALOGO DE SANTAS VIVAS**

Página principal  
Cambios recientes  
Página aleatoria  
Ayuda

Herramientas

Lo que enlaza aquí  
Cambios relacionados  
Páginas especiales  
Versión para imprimir  
Enlace permanente  
Información de la página

Página **Discusión**

**María de Toledo**

Nombre	María de Toledo
Orden	<a href="#">Franciscanas</a>
Títulos	Beata, monja, pre
Fecha de nacimiento	1437
Fecha de fallecimiento	1507
Lugar de nacimiento	<a href="#">Toledo</a>
Lugar de fallecimiento	<a href="#">Toledo</a>

**Contenido** [ocultar]

1 Vida manuscrita (1)

1.1 Fuente

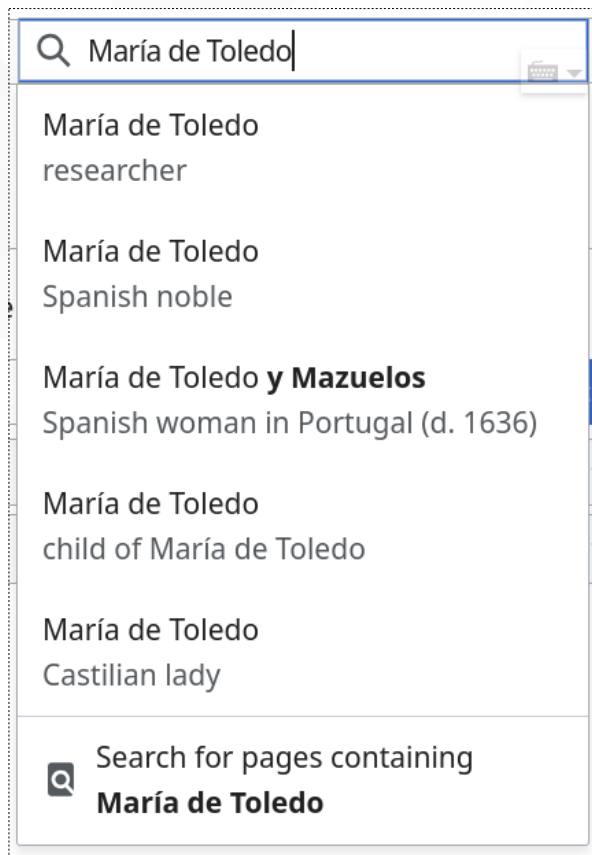
1.2 Criterios de edición

1.3 Vida de María de Toledo

1.3.1 Notas

# Same exercise with Women Writers data

- Wikidata **knows** quite a lot of different “María de Toledo”...



Search results for "María de Toledo":

- María de Toledo researcher
- María de Toledo Spanish noble
- María de Toledo **y Mazuelos** Spanish woman in Portugal (d. 1636)
- María de Toledo child of María de Toledo
- María de Toledo Castilian lady

Search for pages containing **María de Toledo**

- This person in Wikidata is **Q112050023**

- 15 properties
- 1 external ID: Real Acad. de la Historia



Wikidata page for **María Suárez de Toledo** (Q112050023)

Elemento **Discusión**

1437-1507 religiosa terciaria franciscana, clarisa (OSC), fundadora de conventos  
Sor María la Pobre | María la Pobre | **María Suárez**



Real Academia de la Historia | HISTORIA HISPÁNICA

**María Suárez de Toledo**

**Suárez de Toledo, María.** *Sor María la Pobre.* Toledo, 1437 - 3.VII.1507. Religiosa terciaria franciscana, Clarisa (OSC) | Franciscano, na (Orden de Hermanos Menores, OFM) | Fundador, ra de monasterios

**BIOGRAFÍA**

Nació en Toledo, en un linaje muy destacado de la ciudad, emparentado con los miembros de las c... Entre sus ascendientes más directos se encuentran alcaldes mayores de la ciudad, camareros mayor... Toledo —Vasco Fernández o Gutierre Gómez— e incluso el arzobispo de Santiago de Compostela... duda de que el pariente más destacado de la joven noble toledana era el monarca Fernando el Caté...

# Combining MARC and LOD (Czech Academy of Science)

▼ <https://vufind.ucl.cas.cz/Record/002512646#details>

599			a CLB-CPK
600	1	7	a Schulze Wessel, Martin,  d 1962-  7 jo20010090079  2 czenas
610	2	7	a Radio Free Europe  7 kn20010711375  2 czenas
610	2	7	a Československá akademie věd  7 ko20010087906  2 czenas

▼ <https://vufind.ucl.cas.cz/AuthorityRecord/clo000011374#details>

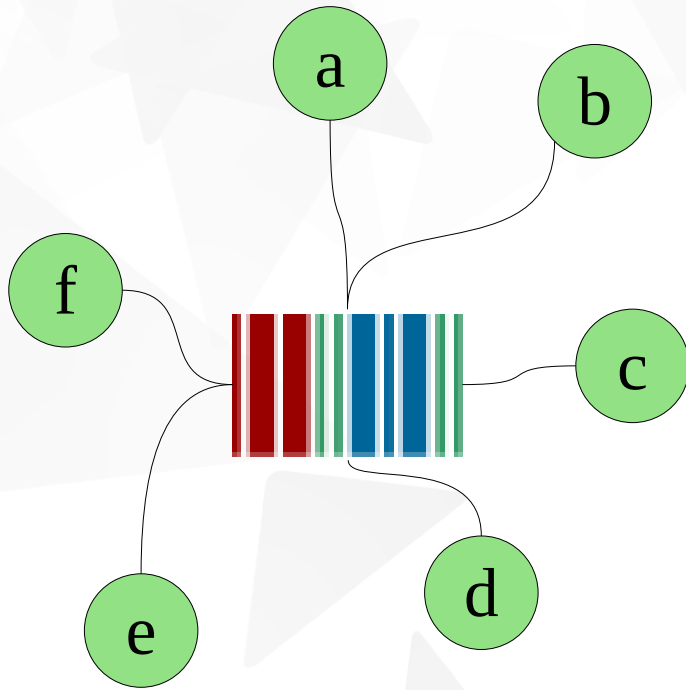
100	1	i skutečné jméno  a Schulze-Wessel, Martin,  d 1962-  7 jo20010090079	Other Databases	Technical Data	Comments
368		d prof.			
370		a Münster in Westfalen (Německo)  7 ge130135			
370		c Německo  7 ge135889			

 [National Authorities Database](#) ↗  
 [Wikidata](#) ↗



▼ véase <https://www.wikidata.org/wiki/Q1904995#P691>

# The idea is to break out of the *silo* using and linking IDs




- ▼ Idea behind the *European Literary Bibliography* (ELB):
  - ▼ In Czech and Polish, who reviews works of Andalusian poet **Lorca**, or of their translations?
    - ▼ **Works about Lorca**, at the moment, searching for the literal "**Lorca**"
    - ▼ When the entity linking step is done, ambiguity is solved, and, moreover: federation with Wikidata!
- ▼ Example BVMC
  - ▼ BVMC federated with Wikidata!
  - ▼ BVMC links works by and about **Lorca**, using the same ID (known to Wikidata)
    - ▼ We can ask **at Wikidata**: Which are the works of and about Lorca in BVMC?
      - ▼ <https://w.wiki/AASW>

# Federated database queries: Wikidata and BVMC

- Works by and about Lorca:  
<https://w.wiki/AASW>
- Expansion of the Wikidata query:  
Apart from Lorca, who else was killed in an extrajudicial execution, and has an entry in BVMC: [https://w.wiki/F\\$Ta](https://w.wiki/F$Ta)
- Federated queries for expositions:  
Lorca in Wikidocumentaries

wd:Q41408



## Federico García Lorca (Q41408)

[Item](#) [Discussion](#) [Star](#)

Spanish poet, dramatist and prose writer (1898–1936)  
García Lorca | García Lorca, Federico | G. F. Lorca | Phenteriko Gkarthia Lorka

**BVMC person ID** 273



[Inicio](#) / [Datos enlazados](#) / García Lorca, Federico, 1898-1936

## García Lorca, Federico, 1898-1936



[Wikimedia Commons](#)

escritor español

Información extraída de [Wikipedia](#) / [\(CC BY-SA 3.0\)](#)

**Identificador:** 273

**Nombre:** García Lorca, Federico

**Fecha de nacimiento:** 1898

**Fecha de fallecimiento:** 1936

**VIAF:** <https://viaf.org/viaf/34456883>

**ISNI:** <https://isni-url.oclc.nl/isni/0000000122781811>

**Variantes de nombre:**

- Federico García Lorca
- Federico García Lorca

**EXPORTAR:**

[RDF](#) [JSON](#)

# Which databases are linked from Wikidata?

- 9,955 (Jan. 26). But which of these belong e.g. to **Cultural Heritage / GLAM**?
- “external id” properties have class declarations (“instance of”). These are the classes:  
<https://w.wiki/FmLx>
- From that list, we have selected the relevant classes, with these results:  
[https://w.wiki/F\\$Ty](https://w.wiki/F$Ty) (9.11.2025: **1749 data bases**)
- BUT, at the moment only **33 of them** seem to have a SPARQL endpoint (and thus can be accessed in federated queries)

performing-arts.ch, Orphanet, Research Vocabularies Australia, OPAC SBN, Gemeinsame Normdatei, dblp computer science bibliography, NFDI4Culture, Linked Open Vocabularies, National Library of Korea Linked Open Data, LDT @ Library, Japan Search, Galician, catalogo.beniculturali.it, Shanghai Library Open Data Platform, BnF authorities, Digital Scriptorium Catalog, Cellosaurus, Union List of Artist Names, Taxonomy of Digital Research Activities in the Humanities, Ontobee, Europeana, Identifiers.org, Nomisma, National Library of Korea, Medical Subject Headings, Hong Kong Chinese Authority Name, Smithsonian American Art Museum, Federico Zeri Foundation, Biblioteca Virtual Miguel de Cervantes, American Art Collaborative, Yale Center for British Art, YCBA LOD














**New:** [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Cultural\\_Heritage\\_Knowledge\\_Graphs](https://www.wikidata.org/wiki/Wikidata:WikiProject_Cultural_Heritage_Knowledge_Graphs)

# Wikidata P2533: WomenWriters ID

- See <https://www.wikidata.org/wiki/Property:P2533>
- See [https://www.wikidata.org/wiki/Property\\_talk:P2533](https://www.wikidata.org/wiki/Property_talk:P2533)
- See [https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/Archive/45#P2533](https://www.wikidata.org/wiki/Wikidata:Property_proposal/Archive/45#P2533)
- The property was created in 2016
- Today, it points to `https://womenwriters.rich.ru.nl/womenwriters/vre/persons/...`
  - that can be changed (and point to SHEWROTE)
- Today, it is used 5,521 times on 5,099 Wikidata person items
- Since 2016, users are asking if Wikidata can be enriched using the linked entries about persons (e.g. birth dates, birth places). Has that been done?
- It does not seem that the womenwriters database has been enriched using Wikidata content (Wikidata knows much more about some persons, e.g. [Selma Lagerlöf](#))

# Items with a WomenWriters ID: what other IDs do they have?

- See <https://qlever.dev/wikidata/ENqIds>

?property	?label	?items
 P2533	WomenWriters ID	5,099
 P214	VIAF cluster ID	4,469
 P213	ISNI	4,116
 P10832	WorldCat Entities ID	3,618
 P227	GND ID	3,568
 P244	Library of Congress authority ID	3,363
 P1006	Nationale Thesaurus voor Auteursnamen ID	2,948
 P269	IdRef ID	2,702
 P13591	Yale LUX ID	2,415
 P268	Bibliothèque nationale de France ID	2,408
 P648	Open Library ID	2,392
 P2163	FAST ID	2,340
 P646	Freebase ID	2,267

## ...and some other queries

- Gender of people with a WomenWriters ID
  - <https://w.wiki/HPpc>
- For how many people with a WomenWriters ID does Wikidata have a birth date? > 5,064
  - <https://w.wiki/HT9n>
- People with a WomenWriters ID that are women, and the number of Wikipedia articles about them
  - <https://qlever.dev/wikidata/qw9ofI>
- How many items are authored by people with a WomenWriters ID?
  - <https://qlever.dev/wikidata/iFyqyh>
- How many items are authored by women with a WomenWriters ID?
  - <https://qlever.dev/wikidata/e50Pz1>

# So many linked IDs... Won't there be many mistakes?

## ▼ Of course, and a lot!

### ▼ errors in a local database

- ▼ hard to correct from the outside (to whom to speak?)

- ▼ can be marked (in Wikidata)

- ▼ example <https://www.wikidata.org/wiki/Q24083#P214>

### ▼ errors in Wikidata

- ▼ easy to correct (everybody can edit)

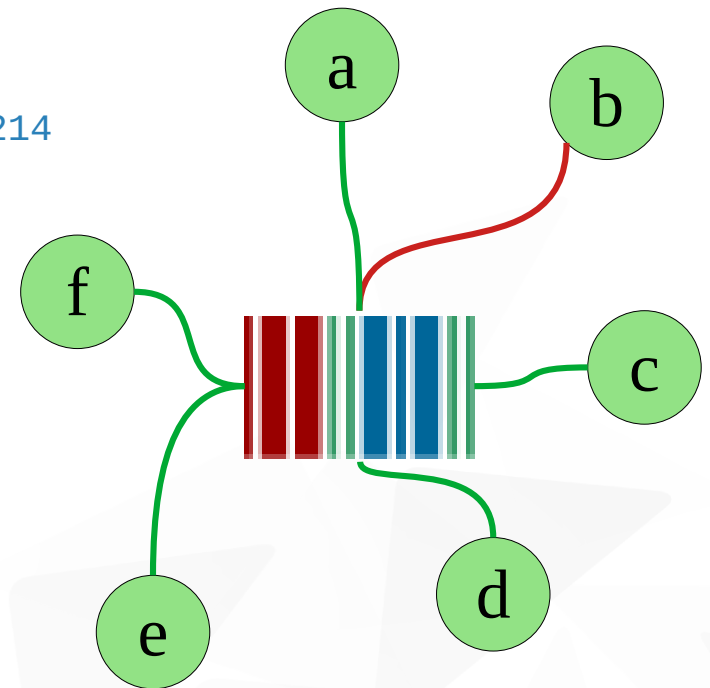
## ▼ Error types

- ▼ duplicates (same entity, 2 or more IDs)

- ▼ conflations (same ID, two or more entities)

- ▼ errors in the data (name form, dates, etc.)

- ▼ gaps in the data (missing birth and death dates, etc.)



*cf.* Pellizzari 2024, 2025

# Problem #1: Duplicates

## Two IDs for the same entity

- ▼ *Luzia* in Wikidata
- ▼ *Luzia* in SHEWROTE

[Luísa de Freitas Lomelino Grande](#) ♀

Portugal 1875 — Portugal 1945

Full name: Luísa Grande

Alternative names: Luzia

[Luzia](#) ♀

Portugal 1875 — Portugal 1945

Full name: Luzia

 **WIKIDATA**

## Luísa Susana Grande de Freitas Lomelino (Q10321814)

Item Discussion

Portuguese writer

Luzia

WomenWriters ID

6dcb1c98-8b8f-4208-98d7-2c74f0d65b4a



▼ 0 references

02b698b7-5a7d-4f3e-bcf4-81ff2c6bf1eb



▼ 0 references

## ▼ WomanWriters ID constraint violations report (Wikidata)

[https://www.wikidata.org/wiki/Wikidata:Database\\_reports/Constraint\\_violations/P2533](https://www.wikidata.org/wiki/Wikidata:Database_reports/Constraint_violations/P2533)

# A word about subject headings

## ▼ Subject headings as used in libraries

- ▼ In MARC, they work like persons: They have their entries in *authority files*
  - ▼ For linking, field 150 is used (the preferred name for a subject)
- ▼ In controlled vocabularies: the *term* is the identifier
  - ▼ problematic example: *agricultura ecológica* / *agricultura biológica*
  - ▼ Basque Govt. authority file: *agricultura biológica* / *nekazaritza biologikoa* only
  - ▼ There are LOD versions, with unambiguous URI
    - ▼ LCSH (example "organic farming"), see <https://www.wikidata.org/wiki/Q165647#P244>
    - ▼ BNE-LOD <https://datos.bne.es/tema/XX528048>

## ▼ Wikidata

- ▼ Also here, Wikidata can be the central pivot, although it is more complicated...
  - ▼ millions of entities, different levels of granularity
  - ▼ solution: for subject headings, use only **those entities that are linked to LCSH**, etc.
    - ▼ More relations in Wikidata than in LCSH (there: only *broader-narrower*)

see Patel-Schneider & Doğan 2025

# Entity linking: to annotate literal values with identifiers

- Any literal addressing an entity can and should be annotated with identifiers
  - In a bibliographical record
    - Luísa Susana Grande de Freitas Lomelino, Luzia* > [Q10321814](#)
    - Praha, Prague, Praga* > [Q1085](#)
    - Agricultura biológica, agricultura ecológica, organic farming, nekazaritza ekologikoa* > [Q165647](#)
    - La Mujer, La Mujer Revista de interés general del bello sexo* > [Q88564895](#)
  - Entities mentioned in title, abstract, full text
  - We see powerful annotation tools emerging (for Wikidata ID annotation)
    - An example without IA for entity linking in full texts: [Dr. Note](#)
    - An example with IA: [Wikidata Embedding Project](#)
      - In the sentence *In the English lesson, they read White Fang by London*, the system proposes Jack London the writer, and not e.g. the city of London. The system does not understand the sentence, but sees the proximity of the entities *White Fang* and *Jack London* in the Wikidata Embedding vector space.
      - Recent more comprehensive tests very encouraging
- [Open Refine](#): the (at the moment) most widely used tool for manual *entity linking*
  - Tools that combine intelligent algorithms with manual validation will appear very soon

# Entity linking: persons

- [Open Refine](#), at the moment the most widely used tool for entity linking
- In your own Wikibase, just by storing the Wikidata ID of an entity, you get everything Wikidata knows about it to your query results (through **federation**)

The screenshot displays the Open Refine interface with a table of authors. The table has two columns: 'fullName\_recon\_Wikidata' and 'fullName\_recon\_Wikibase'. The 'fullName\_recon\_Wikidata' column contains names and Wikidata IDs, while the 'fullName\_recon\_Wikibase' column contains the same names. A modal window is open over the 'Carina Nilstun' row, showing a Wikidata entity card for 'Carina Nilstun (Q63256230)' with a description: 'philologist and lexicographer (\*1978)'; member of Norwegian Academy. Another modal window is open over the 'Isabel Pereira' row, showing a Wikidata entity card for 'Isabel Pereira (Q85647514)' with a description: 'researcher (ORCID 0000-0002-9769-7208)'. The modal window also includes buttons for 'Match this cell', 'Match all identical cells', and 'Cancel'.

fullName_recon_Wikidata	fullName_recon_Wikibase
Anna B. Nikulásdóttir	Anna B. Nikulásdóttir
Carina Nilstun (Q63256230)	Carina Nilstun
Vitor Augusto	Vitor Augusto
Magadán Oliv	Magadán Oliv
Marko Orešković (100)	Marko Orešković
Mehriban Orocova	Mehriban Orocova
Isabel Pereira (Q85647514)	Isabel Pereira
Katya Pertsova	Katya Pertsova
Mike Pham (100)	Mike Pham
David A. Pharies	David A. Pharies
Matthieu Dierens	Matthieu Dierens

[https://eneoli.wikibase.cloud/wiki/Bibliographical\\_Items#Authors\\_with\\_Wikidata\\_ID\\_\(federated\\_query\)](https://eneoli.wikibase.cloud/wiki/Bibliographical_Items#Authors_with_Wikidata_ID_(federated_query))

The diagram illustrates the Wikibase data model components and their relationships, organized into two main sections separated by a dashed line.

**Top Section (External Data Sources):**

- Wikisource:** Contains a **transcribed manuscript** (yellow box) which includes a **paragraph** (blue box). It is linked to a **Wikidata** entity (blue box) via a **referenced WD entity** relationship.
- Wikidata:** Represented by a barcode icon and the text "WIKIDATA". It is linked to a **Token** (yellow circle) via a **referenced WD entity** relationship.
- Bibliographical item (publication metadata):** A blue oval representing a specific Wikibase item.

**Bottom Section (Internal Wikibase Model):**

- Span:** A yellow oval representing a range of text. It has a **has annotation** relationship with a **provenance** (blue oval) and a **has part** relationship with a **Token**.
- Token:** A yellow oval representing a specific instance of text. It has a **has annotation** relationship with a **philological annotation text** (green box) and a **has standard form** relationship with a **Form**.
- Form:** A yellow oval representing a specific form of a word. It has a **form** relationship with a **Lexeme (standard)** and a **Form repres.** (green box).
- Lexeme (standard):** A yellow oval representing a standard lexeme. It has a **has standard lexeme** relationship with a **Sense** and a **POS** (yellow circle). It also has a **form** relationship with a **Form** and a **form** relationship with a **gramm. feat.** (blue circle).
- Sense:** A yellow oval representing a specific sense of a word. It has a **has sense** relationship with a **Token repr.** (green box) and a **sense** relationship with a **Lexeme (standard)**.
- Token repr.:** A green box representing a token representation. It has a **referenced WD entity** relationship with a **Wikidata** entity.
- Lemma:** A green box representing a lemma. It has a **POS** relationship with a **POS** (yellow circle).
- POS:** A yellow circle representing a part of speech.
- Sense gloss:** A green box representing a sense gloss. It has a **sense** relationship with a **Sense**.

**Wikibase Logo:** The Wikibase logo is located in the bottom left corner, consisting of a stylized 'W' made of four colored squares (red, green, blue, yellow) and the text "WIKIBASE".

- see
  - Lindemann 2024 (Wikisource text linking)
  - Lindemann 2025 (Wikibase lexemes intro)
  - Lindemann et al. 2025 (Latin morphology)

# What can Wikibase do for you?

# What can Wikibase do for you?

- You can use your own Wikibase...
  - as a sandbox for experiments
  - for preparing datasets to link to or (when ready) send (in part or as a whole) to Wikidata
  - for implementing your own data model
  - for having your own community working on the data
- Things only Wikibase has (and not other LOD database solutions)
  - total compatibility with Wikidata
  - each entity has its own editable entity page
  - every edit is stored in reversible edit histories, at entity and at user level
  - text pages for textual contents
  - advances user and user permission administration (inherited from MediaWiki)



# What can Wikibase do? Four/Five-Star Open Data!

- Example: Street names in the city of Vitoria-Gasteiz, from [opendata.euskadi.eus](https://opendata.euskadi.eus)
- Street names to Wikidata entities >> enrich the street name dataset with metadata

 Callejero de Vitoria-Gasteiz

Calle	Denominación castellano	Denominación Euskera	Tipo Via
445	ABENDAÑO	ABENDAÑO KALEA	CALLE
6020	ABERASTURI	ABERASTURI	ENTIDAD
6000	ABETXUKO	ABETXUKU	ENTIDAD
830	ABETXUKO CARRETERA	ABETXUKOBIDEA	CARRETERA
4706	ABRAHAM LINCOLN	ABRAHAM LINCOLN KALEA	CALLE
10	ABREVADERO	ASKA KALEA	CALLE
4900	ADANA, PLAZA	ADANA PLAZA	PLAZA
20	ADRIANO VI	ADRIANO VI.AREN KALEA	CALLE
30	ADURTZA	ADURTZA KALEA	CALLE
4697	AERODROMO, PASEO DEL	AERODROMOKO IBILBIDEA	PASEO
50	AGIRRELANDA	AGIRRELANDA KALEA	CALLE
40	AGUIRRE MIRAMON	AGIRRE MIRAMON KALEA	CALLE
55	AIZKORRI	AIZKORRI KALEA	CALLE
4615	AJANALDE	AJANALDE KALEA	CALLE
60	ALAVA	ARABA KALEA	CALLE
4783	ALBA	ALBA KALEA	CALLE
4522	ALBERT EINSTEIN	ALBERT EINSTEIN KALEA	CALLE

Datuak deskargatu



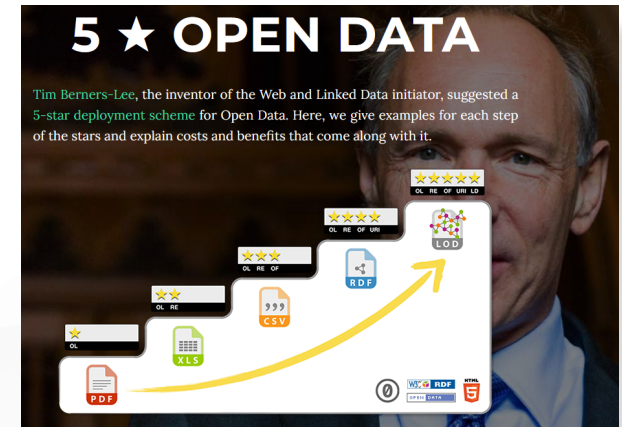
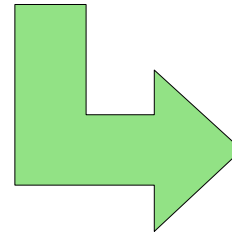
PDF - Kale-izendegi elebiduna pdf



XLS - Kale-izendegi elebiduna xls



CSV - Kale-izendegi elebiduna csv



<https://izendegi.wikibase.cloud>

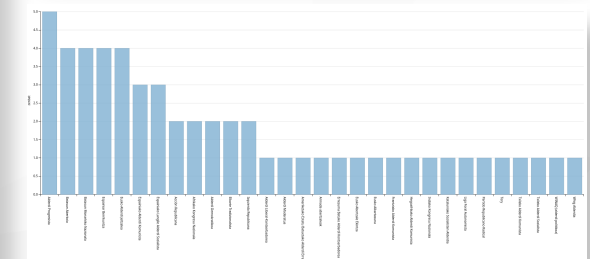
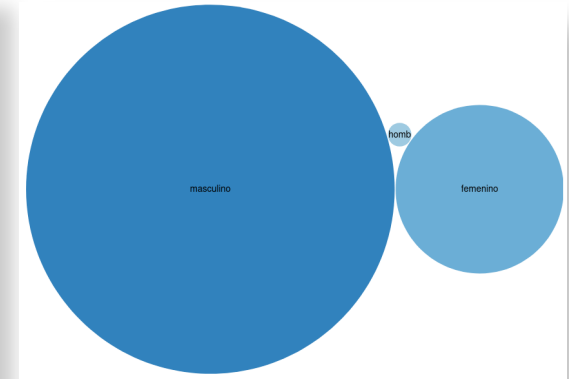
Example: Which are the classes the entities in Vitoria-Gasteiz street names belong to?



<https://izendegi.wikibase.cloud>

# What can Wikibase do? Four/Five-Star Open Data!

- Examples: Birth places, gender, political party of Vitoria-Gasteiz street name entities

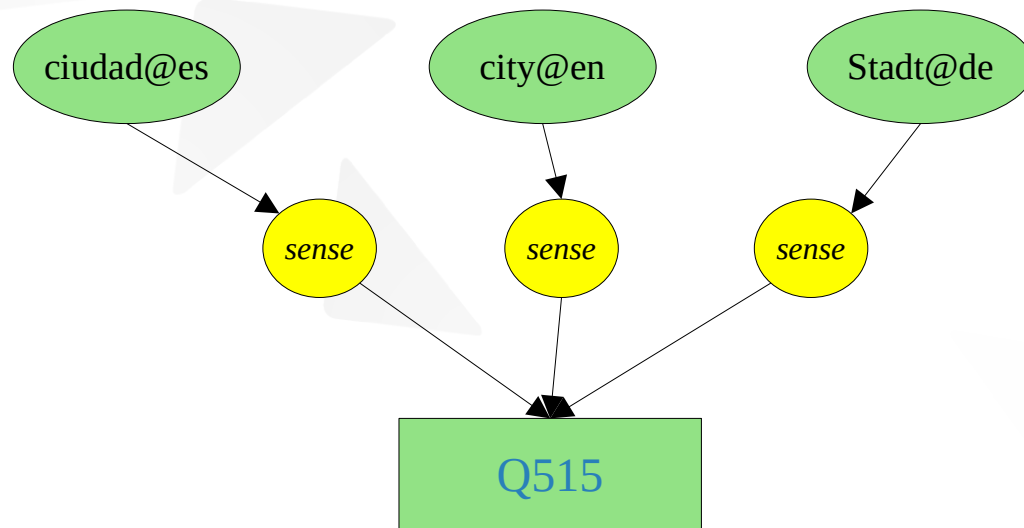


<https://izendegi.wikibase.cloud>

# Wikibase makes the difference: *Items* and *Lexemes*

## Wikibase entity types

- ▼ **items** (Q): entidades ontológicas, objetos del mundo real, conceptos abstractos
  - ▼ *ciudad* (Q515), *Barcelona* (Q1492), *multiculturalismo* (Q190656)
- ▼ **properties** (P): relaciones conceptuales y asociativas
  - ▼ *instance of* (P31), *subclass of* (P279), *lugar de nacimiento* (P19)
- ▼ **lexemes** (L): palabras (diccionario general y especializado)
  - ▼ *ciudad*@es (L313523), *city*@en (L3929), *Stadt*@de (L11539)



# Wikibase makes the difference: items, lexemes, links



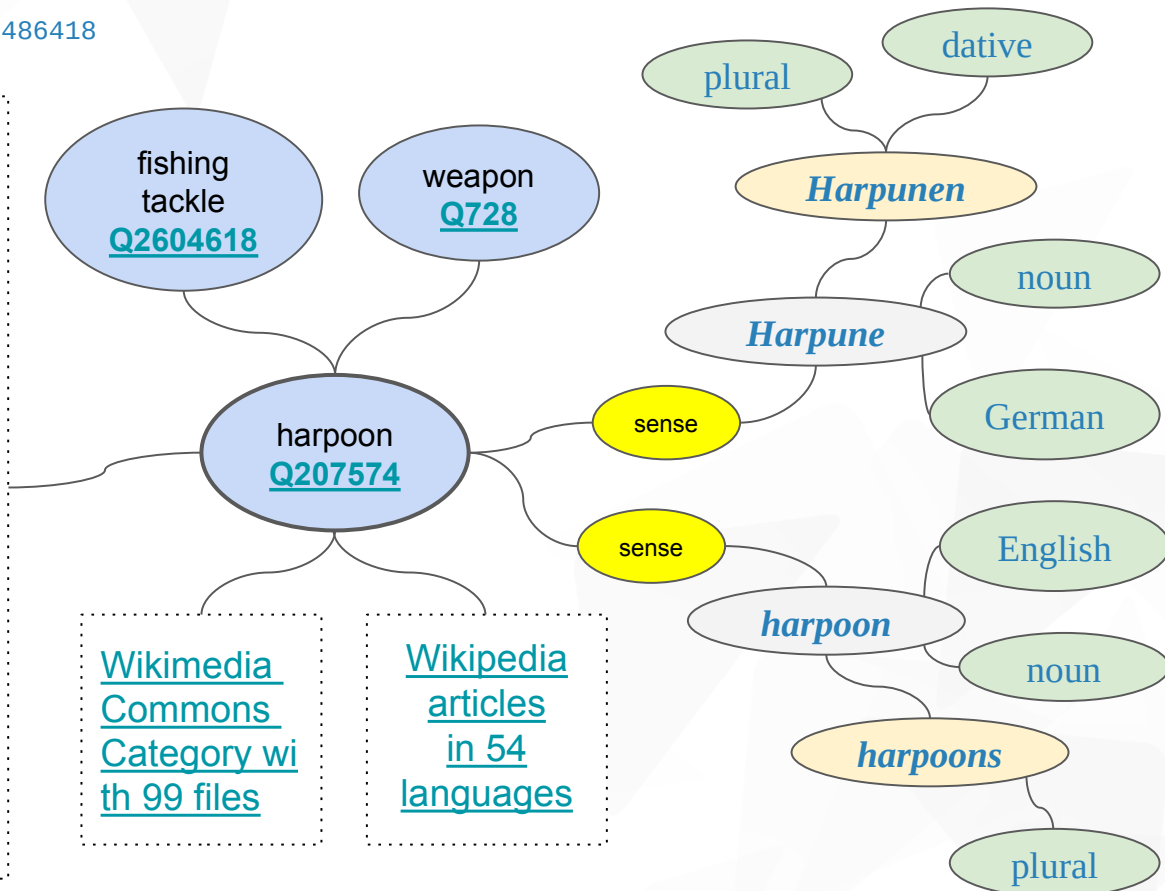
Conjunto de remates  
de arpones  
compuestos.  
Esquimales Chugach,  
Ártico, Estados  
Unidos y Canadá.  
Siglo XVIII.

<https://commons.wikimedia.org/entity/M76486418>

words that denote some fishing tackle <https://w.wiki/ESyr>

24 external ID for harpoon ([list](#)):

[GND ID](#)  
[NDL Authority ID](#)  
[BNCF Thesaurus ID](#)  
[Freebase ID](#)  
[Art & Architecture Thesaurus ID](#)  
[Gran Enciclopèdia Catalana ID \(former scheme\)](#)  
[Encyclopædia Britannica Online ID](#)  
[Great Russian Encyclopedia Online ID \(old version\)](#)  
[JSTOR topic ID \(archived\)](#)  
[PACTOLS thesaurus ID](#)  
[Store norske leksikon \(Great Norwegian Encyclopedia\) ID](#)  
[Microsoft Academic ID](#)  
[Nomenclature for Museum Cataloging](#)  
[Lex ID](#)  
[KBpedia ID](#)  
[TOPCMB ID](#)  
[WordNet 3.1 Synset ID](#)  
[National Historical Museums of Sweden ID](#)  
[Portable Antiquities Scheme object type ID](#)  
[FISH Archaeological Objects Thesaurus ID](#)  
[WikiKids ID](#)  
[Gran Enciclopèdia Catalana ID](#)  
[museum-digital tag ID](#)  
[Vikidia article ID](#)



[Wikimedia Commons](#)  
[Category with 99 files](#)

[Wikipedia articles in 54 languages](#)

# Wikibase, Wikibase Cloud



<https://wikiba.se>



<https://wikibase.cloud>

# A word about “notability” (regarding persons)

## ▼ Notability on Wikipedia

- ▼ What is questioned is the notability of the entity itself
- ▼ Appearance in major reference works is...
  - ▼ a condition?
  - ▼ a good reason?
  - ▼ propagating old gender and other biases

## ▼ Notability on Wikidata

- ▼ What is questioned is not the notability of the entity
  - ▼ Not for persons, but e.g. yes for corpus tokens
- ▼ What is questioned is the granularity of the description
  - ▼ Do we want to have in Wikidata just anything about a person? No.

## ▼ In the case of WomenWriters

- ▼ all writers are notable for Wikidata
  - ▼ and their works and reviews, too
- ▼ nevertheless use an own LOD database?
  - ▼ probably not because of the notability problem

## ▼ Close example:

### Beyond Notability Wikibase

- ▼ Domain: Archaeology
- ▼ Traditionally, only excavation directors are regarded notable, not all participants
- ▼ This Wikibase offers a very fine-grained description of who was involved in what excavation, and in what role
- ▼ Result: A lot more women than in older archaeology databases

# Wikidata and Wikibase: resources

## ▼ DARIAH-EU Working Group DHwiki

### ▼ Our Wikibase

▼ <https://dhwiki.wikibase.cloud>

### ▼ Our Zenodo Community

▼ [https://zenodo.org/communities/dariah\\_dhwiki/](https://zenodo.org/communities/dariah_dhwiki/)

### ▼ Our Zotero group bibliography

▼ [https://www.zotero.org/groups/5639268/dariah\\_dhwiki/library](https://www.zotero.org/groups/5639268/dariah_dhwiki/library)

## ▼ DHwiki Documentation pages

▼ [Overview page](#), with credits to contributors

▼ [Introduction](#)

▼ [Structure and features of Wikibase for DH](#)

▼ [Wikibase and Wikidata, differences and round-tripping](#)

▼ [Wikidata and Wikibase as a hub for identifiers](#)

▼ [Examples of existing Wikibase instances in the field of DH and GLAM](#)

▼ [Pages in Category:Wikibase instances](#)

▼ [Requirements from a DH perspective](#)

▼ [Wikibase Lexical Data](#)

▼ [Resource directory and Bibliography](#)





Eskerrik asko

*[david.lindemann@ehu.eus](mailto:david.lindemann@ehu.eus)*

<https://www.wikidata.org/wiki/Q57694630>