

I prossimi passi di Wikidata

Structured Data e Lexicographical Data

Luca Martinelli, Wikimedia Italia

Como, 17 novembre 2018



Presentazione pubblicata con licenza CC BY-SA 4.0 Internazionale
(consultare la sezione "Credits" per la corretta attribuzione delle immagini usate)

Che cos'è Wikidata?

Il progetto più recente della Wikimedia Foundation



WIKIDATA

per un database secondario, libero,
collaborativo e multilingua

Glossario minimo

- **Wikidata** – il progetto che (più o meno) tutti quanti amiamo alla follia
- **Wikibase** – l'estensione di MediaWiki su cui si basa Wikidata

Glossario minimo

- **Elemento** (Qxxx) – describe un soggetto, un oggetto o un concetto esistente nel Mondo Reale™
- **Proprietà** (Pxxx) – describe a cosa si riferisce un determinato valore

European Union (Q458) **< ETICHETTA**

economic and political union of states mostly located in Europe **< DESCRIZIONE**

EU | E.U. | eu | eu **< ALIAS**

In more languages Configure

Language	Label	Description	Also known as
English	European Union	economic and political union of states mostly located in Europe	EU E.U. eu eu
Italian	Unione europea	entità politica di carattere sovranazionale e intergovernativo	UE
French	Union européenne	union économique et politique d'États européens	UE EU
Sardinian	Unioni europea	No description defined	

All entered languages

Statements

instance of	regional organization	edit
	▼ 0 references	+ add reference
	economic union	edit
	▼ 0 references	+ add reference
	supranational organisation	edit
	▼ 0 references	+ add reference

edit

edit

- Wikipedia (207 entries) **edit**
- ab Европатәи Аидгыла
 - ady Евро Эзкветыныгъэ
 - af Europese Unie
 - als Europäische Union
 - am ኡሮፓ ህብረት
 - ang Europisce Gesamnung
 - an Unión Europea
 - arc ܘܪܘܦܝܬܐ ܥܘܪܘܦܝܬܐ
 - ar الاتحاد الأوروبي
 - arz اتحاد اوروبي
 - ast Xunión Europea
 - as ইউৰোপীয়ান ইউনিয়ন
 - av Европаялъл цолъи
 - azb اوروپا بیرلیگی
 - az Avropa İttifaqı
 - bar Eiropäische Union
 - bat_smg Euruopas Sąjonga
 - ba Европа берләшмәһе
 - bcl Union Europea
 - be_x_old Эўрапейскі Зьвяз
 - be Еўрапейскі Саюз
 - bg Европейски съюз
 - bh यूरोपियन यूनियन
 - bm Eropa Jekulu
 - bn ইউরোপীয় ইউনিয়ন
 - br Unaniezh Europa
 - bs Evropska unija
 - bxr Европын холбоон

LIBERAMENTE MODIFICABILE (cit.)

- Main page
- Community portal
- Project chat
- Create a new item
- Recent changes
- Random item
- Query Service
- Nearby
- Help
- Donate

- Tools
- What links here
 - Related changes
 - Special pages
 - Permanent link
 - Page information
 - Concept URI
 - Cite this page

property (P2306) < ETICHETTA

qualifier to define a property constraint in combination with P2302 (property constraint), or to limit the scope of Q44292881 (wikidata statement) [edit](#)

LIBERAMENTE
MODIFICABILE
(cit.)

allowed qualifier < ALIAS

^ DESCRIZIONE

[In more languages](#) Configure

Language	Label	Description	Also known as
English	property	qualifier to define a property constraint in combination with P2302 (property constraint), or to limit the scope of Q44292881 (wikidata statement)	allowed qualifier
Italian	proprietà	qualificatore per definire un vincolo di proprietà in combinazione con P2302	
French	propriété	qualificateur pour définir une contrainte pour une propriété en combinaison avec P2302	
Sardinian	No label defined	No description defined	

[All entered languages](#)

Data type

Property

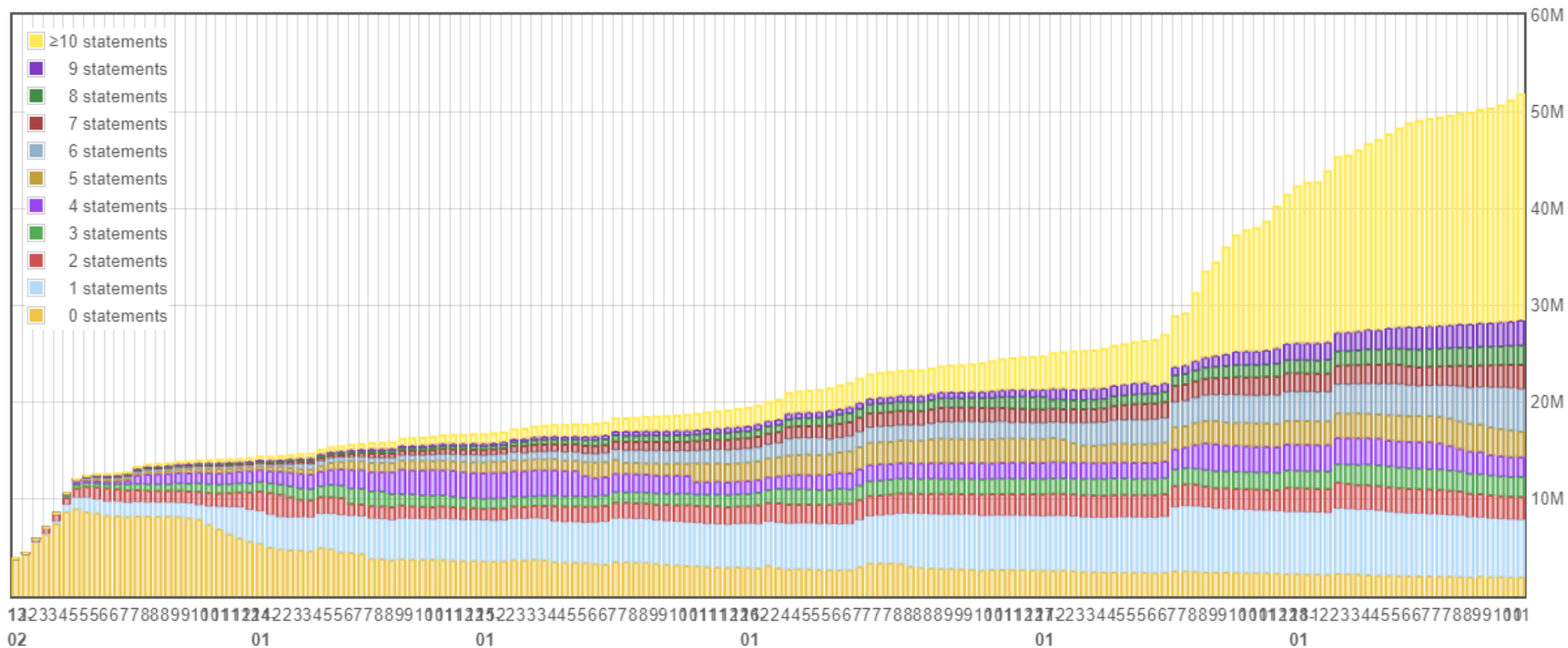
Statements

instance of	Wikidata property to indicate a constraint 0 references	edit + add reference
	Wikidata qualifier 0 references	edit + add reference

Statements per item

Table

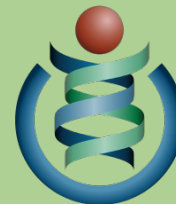
Plot



DIFFICOLTÀ

PROGETTO

Relativamente
facile



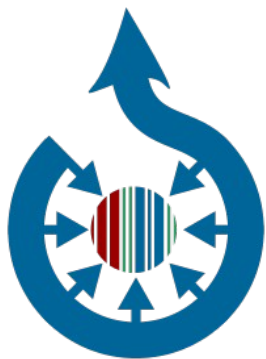
Qualche
complicazione



DANGER ZONE



Come risolvere questo problema?



Structured data – una *nuova istanza di Wikibase* che verrà aggiunta a Commons per gestire i dati dei file multimediali

https://commons.wikimedia.org/wiki/Commons:Structured_data

Lexicographical data – un *ampliamento di Wikidata* dedicata ai lemmi/lessemi di Wikizionario

https://www.wikidata.org/wiki/Wikidata:Lexicographical_data



(1)



L'obiettivo di Structured Data

Aggiungere metadati in una modalità **strutturata, multilingua e leggibile dalle macchine**

per rendere più facile **cercare, trovare, organizzare e riutilizzare**
i file di Wikimedia Commons

File:Van Gogh - Starry Night - Google Art Project.jpg

From Wikimedia Commons, the free media repository

File history File usage on Commons File usage on other wikis Metadata

Download all sizes
Use this file on the web
Use this file on a wiki
Email a link to this file
Information about reusing

Moon
0 references

star
0 references

Alpines
0 references

cloud
0 references

collection
Museum of Modern Art
1 reference

inventory number
472.1941
collection Museum of Modern Art
1 reference

location of final assembly
Saint-Rémy-de-Provence
1 reference

File page



Wikitext
Templates
Categories

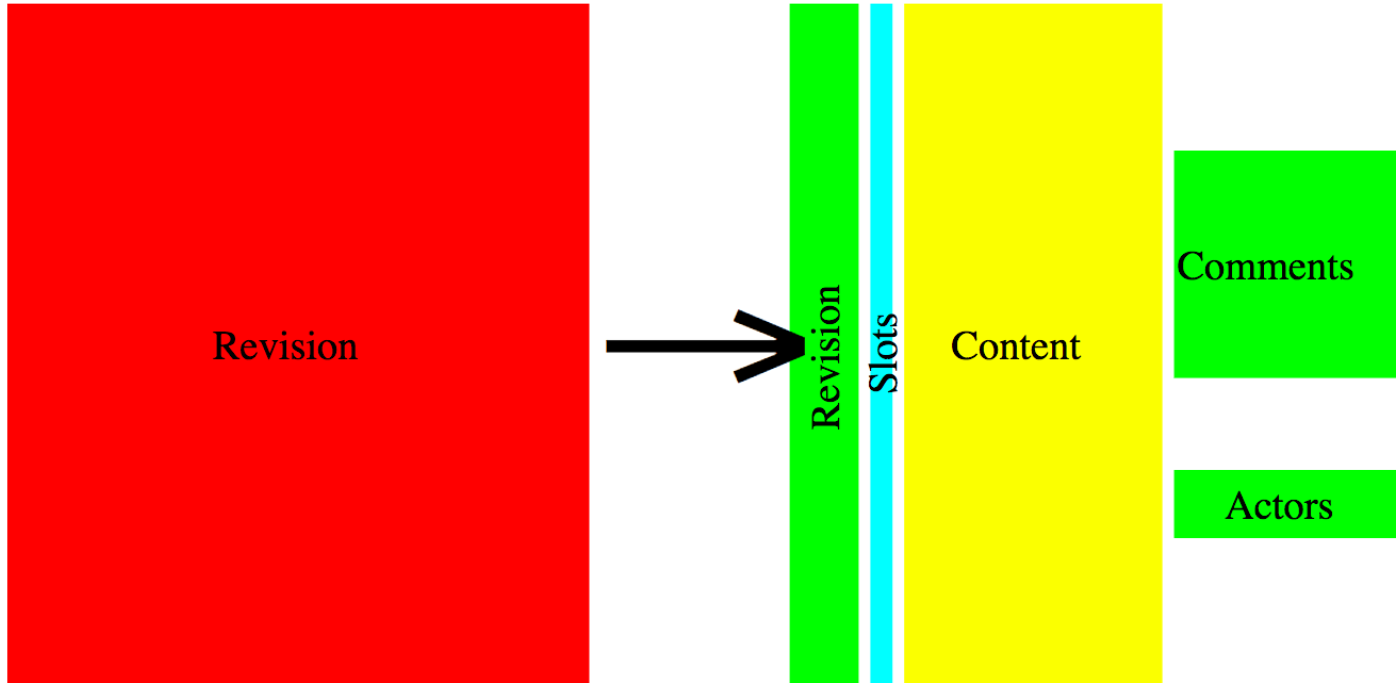
Data section

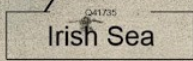
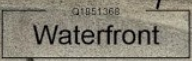
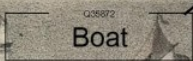
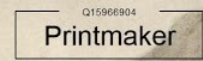
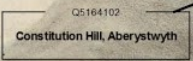
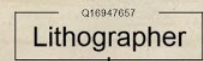
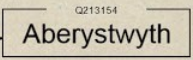
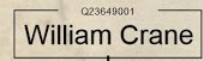
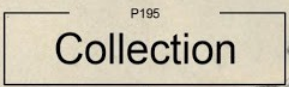
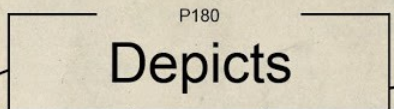
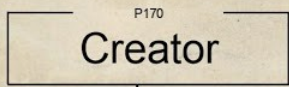
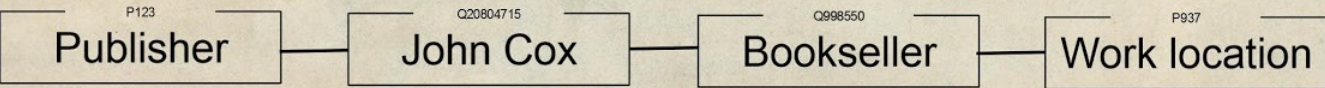
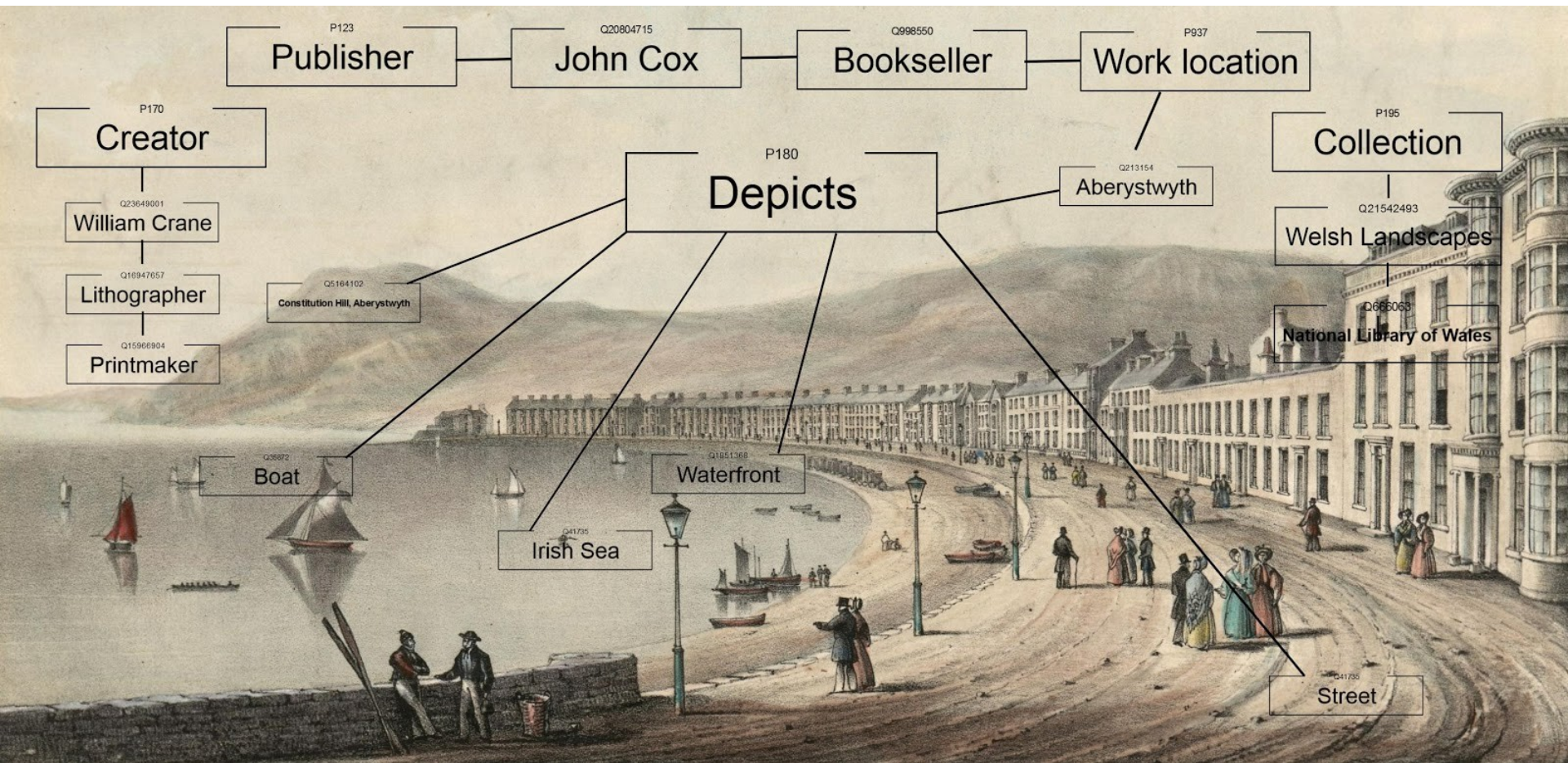


Statements
Properties
Items

The Starry Night (1889) by Vincent van Gogh, Museum of Modern Art. From Google Art Project, Public Domain

Le “multi-content revisions”

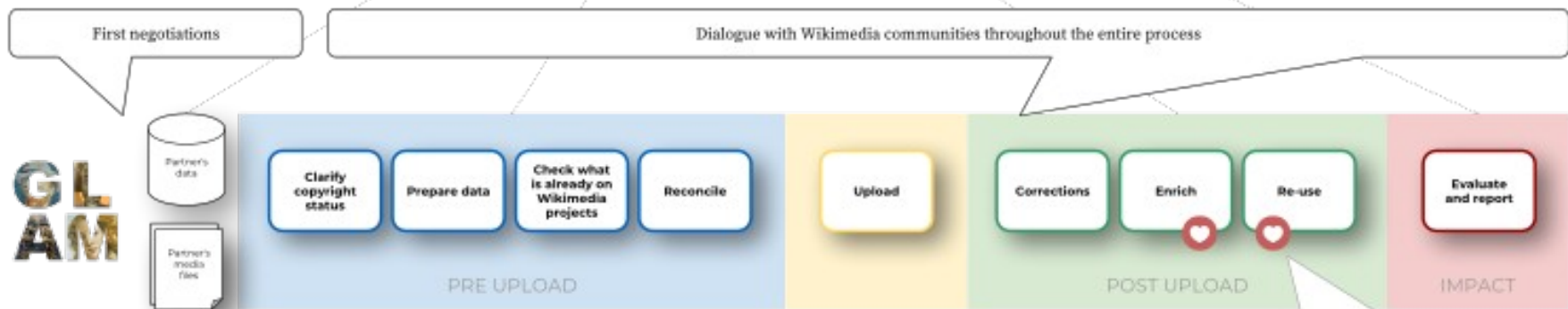




I problemi spesso riscontrati in fase di analisi

- Le nostre metriche *non* sono le metriche che contano
- I metadati spesso vanno rivisti *prima*
- Wikimedia Commons non è *davvero* multilingua
- *MANCA LA DOCUMENTAZIONE!*

Per approfondire: [meta:Research:GLAM and Commons](#)



Workflow

for data and media partnerships
with Wikidata and with (Structured Data on) Wikimedia Commons



Version 2 of Data and Media Partnerships Workflow, by Sandra Puzosnier - User:SandraP (WMP), July 2018
Square GLAM logo CC BY-SA 3.0/Reported by Preston stone after a work by User:Busby, Elegant Themes Line-Style Icons (male and female profile) and Circle Icon Heart by Kenney Sing / Elegant Themes, GNU General Public License v2.

Le “multi-lingual captions”


Una “descrizione” più breve del file e più facilmente traducibile da un utente “casuale”

Mockup: [Pam Drouin](#) / [Wikimedia Commons](#) / [CC-BY-SA 4.0](#) / [GFDL](#)

Foto: [Pamjunipero](#) / [Wikimedia Commons](#) / [CC-BY-SA 4.0](#)

File: [Downtown Seattle sunset.jpg](#)
From Wikimedia Commons, the free media repository

[Download](#) [Use this file](#) [Use this file](#) [Email a link](#) [Information](#)



Size of this preview: 600 × 600 pixels. Other resolutions: 320 × 240 pixels | 640 × 480 pixels | 1,024 × 768 pixels | 1,280 × 960 pixels | 4,032 × 3,024 pixels.
This file has annotations. Move the mouse pointer over the image to see them.

[Add a note](#)

[Original file](#) (4,032 × 3,024 pixels, file size: 2.44 MB, MIME type: image/jpeg); ([request rotation](#)); ZoomViewer: [flash/no flash](#)

[Open in Media Viewer](#)

Media info

Multilingual captions

English Seattle skyline view from Smith Tower with Mount Rainier in background at sunset Seattle skyline view from Smith Tower with Mount Rainier in background at sunset Seattle skyline view from Smith Tower with Mount Rainier in background at sunset Seattle skyline view from Smith Tower with Mount Rainier in background at sunset

[See more languages](#)

Summary [\[edit \]](#)

Description	English: Sunset view of Mt. Rainier from Smith Tower
Date	23 October 2017, 18:04:29

Description

Español: Vista de la [Basílica de Nuestra Señora de Licheń](#), situada en la pequeña población de Licheń Stary, cerca de [Konin](#), [Voivodato de Gran Polonia](#), Polonia. La basílica, diseñada por [Barbara Bielecka](#), fue construida entre 1994 y 2005. Con una nave centra de 120 metros de longitud, 77 m de anchura y 98 m de altura, así como una torre de 141,5 m de altura, es la iglesia de mayor tamaño en Polonia y una de las mayores del mundo.

English: View at dusk of the [Basilica of Our Lady of Sorrows](#), Queen of Poland, located in the village of [Licheń Stary](#), near [Konin](#), [Greater Poland Voivodeship](#), [Poland](#). The basílica, designed by [Barbara Bielecka](#) was built between 1994 and 2004. With a 120 metres (390 ft) long and 77 metres (253 ft) wide nave, a 98 metres (322 ft) high central dome, and a 141.5 metres (464 ft) high tower, it is Poland's largest church and one of the [largest churches in the world](#).

Polski: Widok o zmierzchu na [bazylikę Najświętszej Maryi Panny Licheńskiej](#) w [Licheniu Starym](#) koło [Konina](#), w [województwie wielkopolskim](#), w [Polsce](#). Bazylikę zaprojektowaną przez [Barbarę Bielecką](#) zbudowano w latach 1994–2004. [Nawa główna](#) o długości 120 m i szerokości 77 m, [kopała](#) o wysokości 103,5 m oraz wieża o wysokości 141,5 m czynią ją największym kościołem w Polsce i jednym z największych na świecie.

Multilingual captions

English

Seattle skyline view from Smith Tower with Mount Rainier in background at sunset

French



Vue sur la skyline de Seattle|

[+ Add a caption](#)[Cancel](#)[Publish changes](#)



This double-humped camel with tiger-faced saddlebag raises its head and bares its teeth.

Download this file

- Re-use information
- Share on Internet
- Use on a Wiki

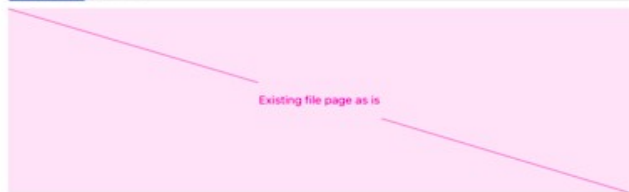
Uploader	Type
Username	image / jpg

Public domain
Free for use by anyone for any purpose without restriction under copyright law. Full license

Direct URL
https://commons.wikimedia.org/wiki/File:Chinese_Camel_-_Walters_492383_-_Profile.jpg

Attribution for re-use **REQUIRED**
Username / Wikimedia Commons / "Chinese Ca...

File information Metadata



Metadata

[back to file info]

depicts P180 • Edit

Search to add value

statuette Q9351 • Make primary

- Usage justification — public domain
- Usage jurisdiction — United States of America

camel Q7375 • Make primary

What's metadata?

Learn more

usage rights P180 • Edit

GNU Free Documentation License v1.2 Q49497 • Make primary

- Obligations — attribution, share alike
- Usage justification — OTRS
- Author — Shacho0822

Creative Commons Attribution-Share Alike 3.0 Unported Q49497 • Make primary

- Obligations — attribution, share alike
- Usage justification — OTRS
- Author — Walters Art Museum
- Attribution text — Walters Art Museum
- Attribution link — <http://thewalters.org>

depicts

P180 • Edit

Search to add value from Wikidata

statuette

Q9351 • Make primary

- Usage justification — public domain
- Usage jurisdiction — United States of America

camel

Q7375 • Make primary

usage rights

P### • Edit

GNU Free Documentation License v1.2

Q##### • Make primary

- Obligations — attribution, share alike
- Usage justification — OTRS
- Author — Shacho0822

Creative Commons Attribution-Share Alike 3.0 Unported

Q##### • Make primary

- Obligations — attribution, share alike
- Usage justification — OTRS
- Author — Walters Art Museum
- Attribution text — Walters Art Museum
- Attribution link — <http://thewalters.org>

Il principale avviso dai GLAM

Dobbiamo imparare a distinguere fra:

- 1) I metadati che si applicano al **file**
- 2) I metadati che si applicano all'**opera rappresentata nel file**

Cosa possiamo fare noi per aiutare?

- Iscriverti alla [newsletter](#)
- Aiutare a trovare esempi di (ri)utilizzo dei dati
- Contribuire a [definire le proprietà](#) che verranno usate
- Contribuire/realizzare script o tool per i caricamenti massivi
- ...ditecelo voi!

(2)



L'obiettivo di Lexicographical Data

Aggiungere metadati in una modalità **strutturata, multilingua e leggibile dalle macchine**

per rendere più facile **cercare, trovare, organizzare e riutilizzare** i lessemi di Wikizionario

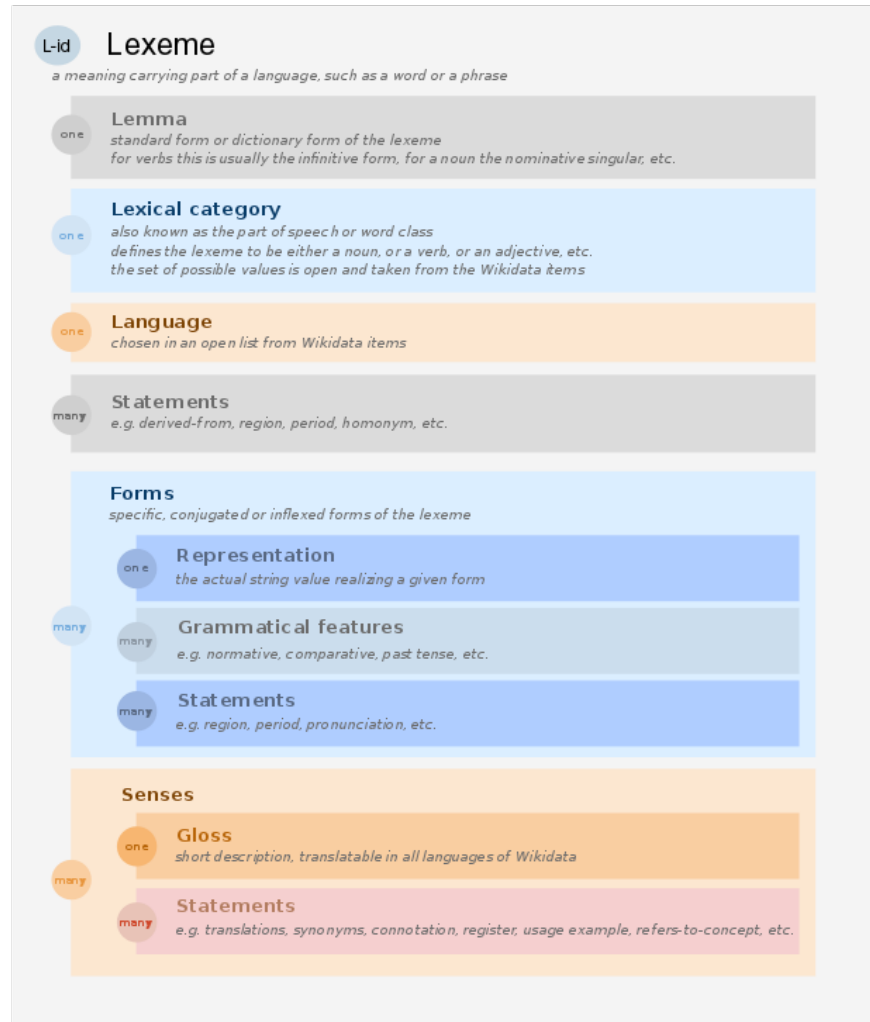
...ma siamo sicuri che serve?

Sì!

Wikizionario è molto simile a com'era Wikipedia 6 anni fa, ossia:

- un sacco di informazioni presenti solo *in formato testo*, replicate x numero di volte per x numero di versioni
- un sacco di persone che vorrebbero riutilizzare queste informazioni e non riesce a farlo (**un esempio?**)

- **Lessema (Lxxx)** – una parola o una locuzione con uno o più significati (esempio)
- **Lemma** – forma standard da dizionario
- **Categoria lessicale** – sostantivo, verbo, avverbio, ecc.
- **Altre informazioni** (lingua, derivazione, sinonimi, omonimi, contrari, pronuncia, ecc.)
- **Forme (Fxxx)** – maschile/femminile, singolare/plurale, tempi, ecc.
- **Sensi (Sxxx)** – descrizioni brevi del lessema



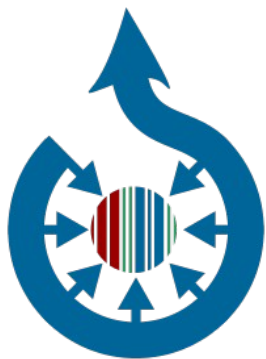
Che cosa si prevede di farci?

- Condividere il carico di lavoro fra le comunità
- Aprirsi ai ricercatori e/o a nuovi utenti
- Creare nuove applicazioni interattive basate su Wikizionario
 - dizionari/tesauri/soggettari specifici o generali
 - raccolte di esercizi o giochi per imparare le lingue
 - tool per l'analisi del testo o dell'etimologia
 - ...ditecelo voi!

Alcuni link per iniziare a lavorare

- [Special:NewLexeme](#) – il form di base
- [Lexeme Forms](#) – un tool semplice per creare più velocemente i lessemi
- [Ordia](#) – query e statistiche su LD in tempo reale

Grazie per l'attenzione! Domande?



https://commons.wikimedia.org/wiki/Commons:Structured_data

https://www.wikidata.org/wiki/Wikidata:Lexicographical_data

