

Hunting for lost churches on Wikimedia projects with OpenRefine

a personal project - Marco Chemello (Wikimedia Italia) 2023

short URL:

<https://w.wiki/8WHC>



Photo: [Wikiqita in Calabria primavera 2022 partecipanti f05](#)
by [Marco Chemello \(Wikimedia Italia\)](#) 2022
CC BY-SA 4.0 on Wikimedia Commons

Wiki Loves Monuments: a lot of church buildings in Italy...

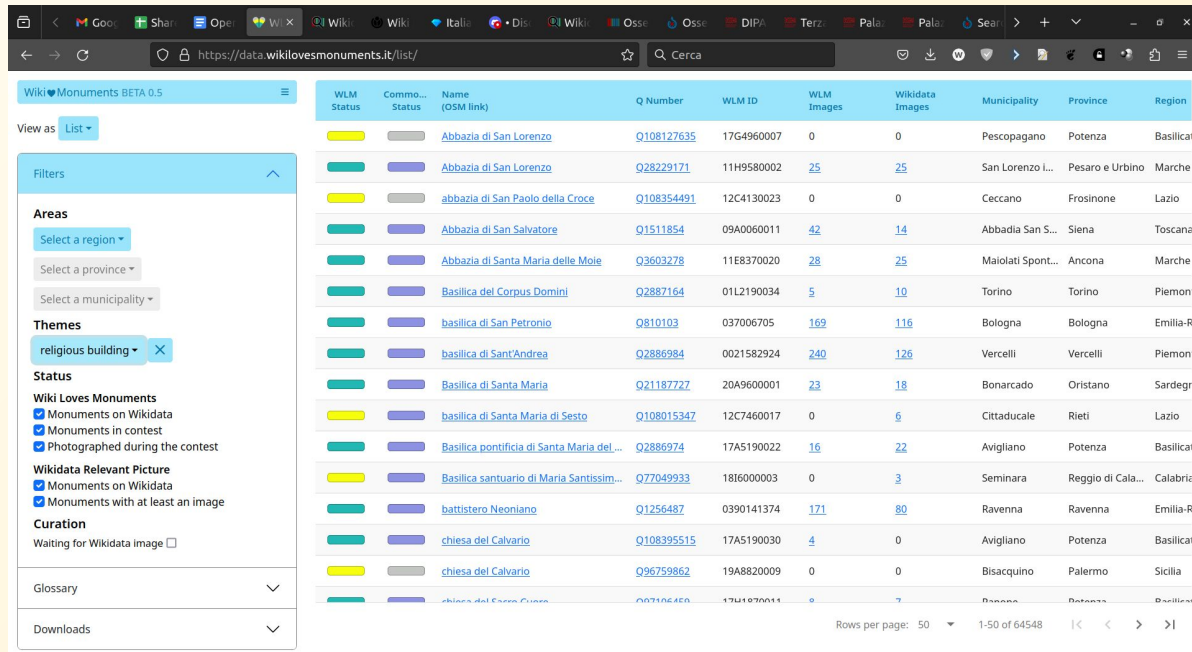
From the results of the 2023 edition of Wiki Loves Monuments Italy:

64 555 religious buildings in Italy present in the WLM lists

The list is based
on Wikidata
(now the most comprehensive
database on this topic!)

...but there are still
many gaps to fill!

https://wiki.wikimedia.it/wiki/Wiki_Loves_Monuments/2023



WLM Status	Commo... Status	Name (OSM link)	Q Number	WLM ID	WLM Images	Wikidata Images	Municipality	Province	Region
Yellow	Grey	Abbazia di San Lorenzo	Q108127635	17G4960007	0	0	Pescopagano	Potenza	Basilica
Green	Purple	Abbazia di San Lorenzo	Q28229171	11H9580002	25	25	San Lorenzo I...	Pesaro e Urbino	Marche
Yellow	Grey	abbazia di San Paolo della Croce	Q108354491	12C4130023	0	0	Ceccano	Frosinone	Lazio
Green	Purple	Abbazia di San Salvatore	Q1511854	09A0060011	42	14	Abbadia San S...	Siena	Toscana
Green	Purple	Abbazia di Santa Maria delle Moie	Q3603278	11E8370020	28	25	Malotati Spont...	Ancona	Marche
Green	Purple	Basilica del Corpus Domini	Q2887164	01L2190034	5	10	Torino	Torino	Piemon
Green	Purple	basilica di San Petronio	Q810103	037006705	169	116	Bologna	Bologna	Emilia-R
Green	Purple	basilica di Sant'Andrea	Q2886984	0021582924	240	126	Vercelli	Vercelli	Piemon
Green	Purple	Basilica di Santa Maria	Q21187727	20A9600001	23	18	Bonarcado	Oristano	Sardegr
Yellow	Purple	basilica di Santa Maria di Sesto	Q108015347	12C7460017	0	6	Cittaducale	Rieti	Lazio
Green	Purple	Basilica pontificia di Santa Maria del...	Q2886974	17A5190022	16	22	Avigliano	Potenza	Basilica
Yellow	Purple	Basilica santuario di Maria Santissim...	Q77049933	18I6000003	0	3	Seminara	Reggio di Cala...	Calabri
Green	Purple	battistero Neoniano	Q1256487	0390141374	171	80	Ravenna	Ravenna	Emilia-R
Green	Purple	chiesa del Calvario	Q108395515	17A5190030	4	0	Avigliano	Potenza	Basilica
Yellow	Grey	chiesa del Calvario	Q96759862	19A8820009	0	0	Bisacquino	Palermo	Sicilia
Green	Purple	chiesa del Gesù Gesù	Q97106460	17A1070011	0	7	Prosciutto	Palermo	Basilica




Still a lot of Commons categories not linked to Wikidata...

Category: **Beata Vergine Maria Immacolata (Rome)**

From Wikimedia Commons, the free media repository

Media in category "Beata Vergine Maria Immacolata (Rome)"

The following 3 files are in this category, out of 3 total.

		
Roma (Z. Giustiniana) - Chiesa di Maria Immacolata 1.JPG 2,272 × 1,704; 579 KB	Roma (Z. Giustiniana) - Chiesa di Maria Immacolata 2.JPG 1,704 × 2,272; 613 KB	Roma (Z. Giustiniana) - Chiesa di Maria Immacolata 3.JPG 1,704 × 2,272; 560 KB

Categories: Churches in Zona La Giustiniana (LIV) | Immaculate Conception churches in Lazio

This page was last edited on 29 March 2012, at 06:32.

Files are available under licenses specified on their description page. All structured data from the file namespace is available under the Creative Commons CC0 license; all unstructured text is available under the Creative Commons Attribution-ShareAlike license; additional terms may apply. By

no infobox = no Wikidata

Still a lot of Commons categories not linked to Wikidata...

Many of the building are really present on Wikidata with an item, so we need to create a link...

... but how many other are not in Wikidata?

The screenshot shows a browser window displaying the Wikimedia Commons category page for "Beata Vergine Assunta (San Salvatore Monferrato)". The page includes a sidebar with navigation options like "Main page", "Upload file", and "Tools". The main content area shows the category name, a search bar, and a list of two images. A red circle highlights a message that says "NO WIKIDATA ID FOUND!" with links to search for the item on Wikidata or create a new item.

Category: **Discussion** | View | Edit | History | Search Wikimedia Commons



Category:Beata Vergine Assunta (San Salvatore Monferrato)

View Wikimedia Commons in italiano

From Wikimedia Commons, the free media repository

Media in category "Beata Vergine Assunta (San Salvatore Monferrato)"

The following 2 files are in this category, out of 2 total.

	
San Salvatore Monferrato-chiesa bv assunta-campanile.jpg 2,448 × 3,264; 1.16 MB	San Salvatore Monferrato-chiesa bv assunta-facciata.jpg 2,448 × 3,264; 1.47 MB

NO WIKIDATA ID FOUND!
[Search for Beata Vergine Assunta \(San Salvatore Monferrato\) on Wikidata](#)
[Create new Wikidata item](#)
[Upload media](#)

Categories: [Churches in San Salvatore Monferrato](#) | [Churches of the Assumption in Piedmont](#)
Non-topical/index: Uses of Wikidata Infobox with no item

Collecting the data (roughly) performing some search on Wikimedia Commons

Do you know better ways to do so?

(yes, there are better ways to try!)

The screenshot shows a browser window displaying the Wikimedia Commons search results for the query `-"churches in Italy" -incategory:"Uses of Wikidata Infobox" -"in art" -"little italy"`. The search results page includes a sidebar with navigation links, a search bar, and a list of search results. Red annotations highlight specific elements:

- A red circle highlights the search query in the search bar.
- A red arrow points from the text "much to filter out!" to the search results count "Results 1 - 100 of 65,449".
- A red arrow points from the text "bad results" to the first search result, "Category:Confessionals in Sicily".
- A red arrow points from the text "good" to the second search result, "Category:Churches of the Purification of Mary in Friuli-Venezia Giulia".

Search results:

- Category:Confessionals in Sicily**
Regions of **Italy**: Apulia · Emilia-Romagna · Liguria · Lombardy · Piedmont · **Sicily** · Trentino-South Tyrol · Tuscany · **Veneto**...
1 member (0 subcategories, 1 file) - 04:51, 3 December 2023
- Category:Churches of the Purification of Mary in Friuli-Venezia Giulia**
Regions of **Italy**: ...
2 members (2 subcategories, 0 files) - 22:04, 7 October 2023
- Category:San Francesco (Nicosia (Italy))**
2 members (0 subcategories, 2 files) - 17:16, 13 October 2023
- Category:San Sebastiano (Augusta (Italy))**
1 member (0 subcategories, 1 file) - 21:30, 23 September 2023

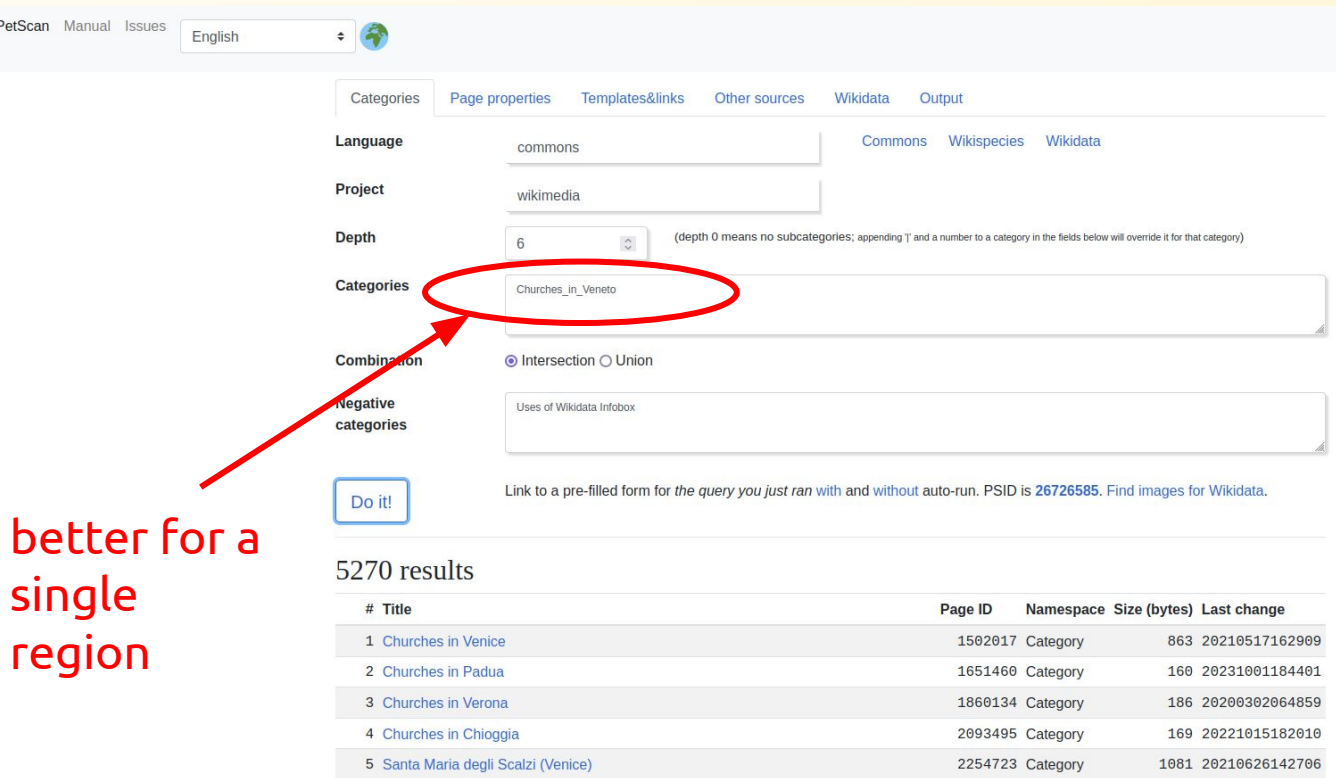
much to filter out!

bad results

good

Collecting the data (roughly) using PetScan

yes, PetScan could
do it better...
but it doesn't work with
big lists,
and/or with too much
filters



The screenshot shows the PetScan interface with the following settings:

- Language: commons
- Project: wikimedia
- Depth: 6
- Categories: Churches_in_Veneto
- Combination: Intersection
- Negative categories: Uses of Wikidata Infobox

A red circle highlights the 'Categories' field, and a red arrow points from the text 'better for a single region' to it.

5270 results

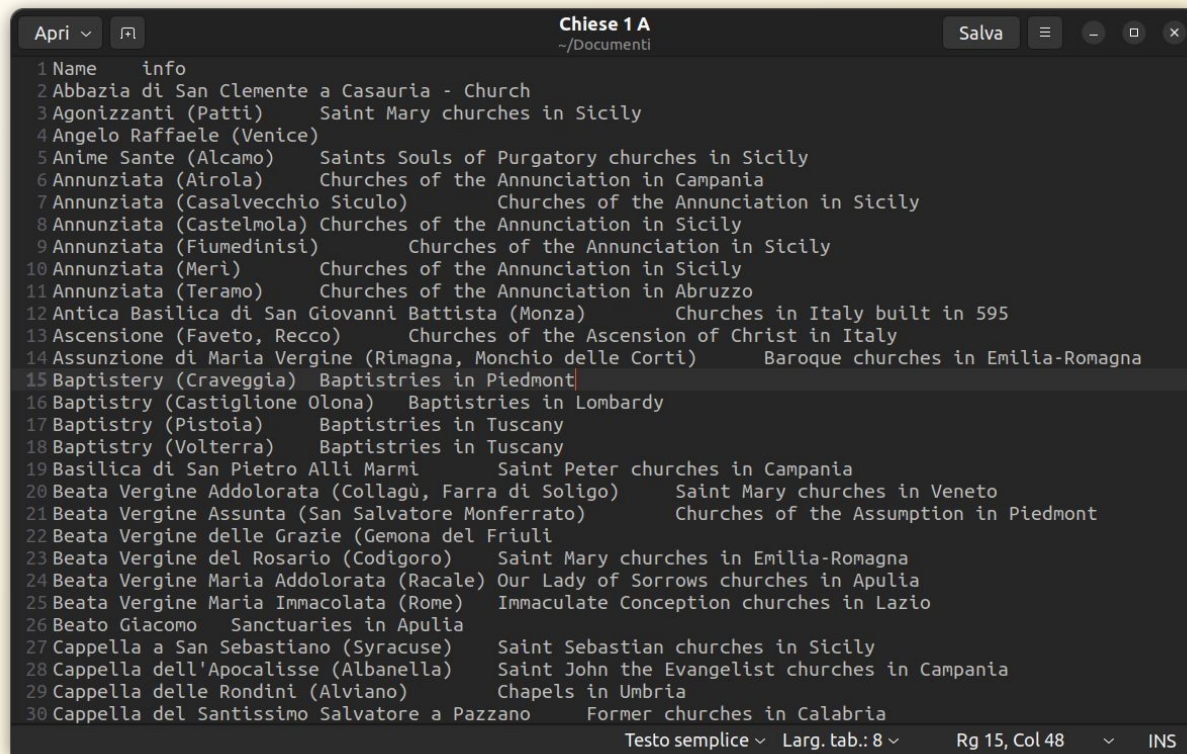
#	Title	Page ID	Namespace	Size (bytes)	Last change
1	Churches in Venice	1502017	Category	863	20210517162909
2	Churches in Padua	1651460	Category	160	20231001184401
3	Churches in Verona	1860134	Category	186	20200302064859
4	Churches in Chioggia	2093495	Category	169	20221015182010
5	Santa Maria degli Scalzi (Venice)	2254723	Category	1081	20210626142706

better for a
single
region

Filtering the data (roughly)

pasting, ordering and manually filtering the list of categories

excluding categories
of paintings, interiors,
architectural elements etc.



```
Chiese 1 A
~/Documenti
Salva

1 Name      info
2 Abbazia di San Clemente a Casauria - Church
3 Agonizzanti (Patti)      Saint Mary churches in Sicily
4 Angelo Raffaele (Venice)
5 Anime Sante (Alcamo)    Saints Souls of Purgatory churches in Sicily
6 Annunziata (Airola)     Churches of the Annunciation in Campania
7 Annunziata (Casalvecchio Siculo)    Churches of the Annunciation in Sicily
8 Annunziata (Castelmola) Churches of the Annunciation in Sicily
9 Annunziata (Fiumedinisi) Churches of the Annunciation in Sicily
10 Annunziata (Merì)      Churches of the Annunciation in Sicily
11 Annunziata (Teramo)    Churches of the Annunciation in Abruzzo
12 Antica Basilica di San Giovanni Battista (Monza) Churches in Italy built in 595
13 Ascensione (Faveto, Recco) Churches of the Ascension of Christ in Italy
14 Assunzione di Maria Vergine (Rimagna, Monchio delle Corti) Baroque churches in Emilia-Romagna
15 Baptistery (Craveggia) Baptistries in Piedmont
16 Baptistry (Castiglione Olona) Baptistries in Lombardy
17 Baptistry (Pistoia)    Baptistries in Tuscany
18 Baptistry (Volterra)  Baptistries in Tuscany
19 Basilica di San Pietro Alli Marmi      Saint Peter churches in Campania
20 Beata Vergine Addolorata (Collagù, Farra di Soligo) Saint Mary churches in Veneto
21 Beata Vergine Assunta (San Salvatore Monferrato) Churches of the Assumption in Piedmont
22 Beata Vergine delle Grazie (Gemona del Friuli)
23 Beata Vergine del Rosario (Codigoro)    Saint Mary churches in Emilia-Romagna
24 Beata Vergine Maria Addolorata (Racale) Our Lady of Sorrows churches in Apulia
25 Beata Vergine Maria Immacolata (Rome)  Immaculate Conception churches in Lazio
26 Beato Giacomo          Sanctuaries in Apulia
27 Cappella a San Sebastiano (Syracuse)    Saint Sebastian churches in Sicily
28 Cappella dell'Apocalisse (Albanella)    Saint John the Evangelist churches in Campania
29 Cappella delle Rondini (Alviano)        Chapels in Umbria
30 Cappella del Santissimo Salvatore a Pazzano Former churches in Calabria

Testo semplice v Larg. tab.: 8 v Rg 15, Col 48 v INS
```



importing & deriving informations on OpenRefine (with a lot of fixing) from category name

- Commons category
- name (label)
- description (English / Italian)
- city
- location
- titular saint

The screenshot shows the OpenRefine interface with a table of 13 rows of data. The table has columns for label-EN, label-IT, city, location, desc-IT, desc-EN, category, and URL. A sidebar on the left provides instructions on using facets and filters. The bottom of the interface shows a search bar and navigation controls.

	label-EN	label-IT	city	location	desc-IT	desc-EN	category	URL	Info
1.	Abbey church of San Clemente a Casauria	chiesa dell'abbazia di San Clemente a Casauria	Castiglione a Casauria	Abbazia di San Clemente a Casauria	edificio religioso nel comune italiano di Castiglione a Casauria	religious building in Abbazia di San Clemente a Casauria, Castiglione a Casauria, Italy	Abbazia di San Clemente a Casauria - Church	https://commons.wikimedia.org/wiki/Category:Abbazia_di_San_Clemente_a_Casauria...Church	
2.	Agonizzanti	chiesa degli Agonizzanti	Patti		edificio religioso nel comune italiano di Patti	religious building in Patti, Italy	Agonizzanti (Patti)	https://commons.wikimedia.org/wiki/Category:Agonizzanti_(Patti)	Saint Mary churches in Sicily
3.	Angelo Raffaele	chiesa dell'Angelo Raffaele	Venice		edificio religioso nel comune italiano di Venice	religious building in Venice, Italy	Angelo Raffaele (Venice)	https://commons.wikimedia.org/wiki/Category:Angelo_Raffaele_(Venice)	
4.	Anime Sante	chiesa delle Anime Sante	Alcamo		edificio religioso nel comune italiano di Alcamo	religious building in Alcamo, Italy	Anime Sante (Alcamo)	https://commons.wikimedia.org/wiki/Category:Anime_Sante_(Alcamo)	Saints Souls of Purgatory churches in Sicily
5.	Annunziata	chiesa di Maria Santissima Annunziata	Airola		edificio religioso nel comune italiano di Airola	religious building in Airola, Italy	Annunziata (Airola)	https://commons.wikimedia.org/wiki/Category:Annunziata_(Airola)	Churches of the Annunciation in Campania
6.	Annunziata	chiesa di Maria Santissima Annunziata	Casalvechio Siculo		edificio religioso nel comune italiano di Casalvechio Siculo	religious building in Casalvechio Siculo, Italy	Annunziata (Casalvechio Siculo)	https://commons.wikimedia.org/wiki/Category:Annunziata_(Casalvechio_Siculo)	Churches of the Annunciation in Sicily
7.	Annunziata	chiesa di Maria Santissima Annunziata	Castelmola		edificio religioso nel comune italiano di Castelmola	religious building in Castelmola, Italy	Annunziata (Castelmola)	https://commons.wikimedia.org/wiki/Category:Annunziata_(Castelmola)	Churches of the Annunciation in Sicily
8.	Annunziata	chiesa di Maria Santissima Annunziata	Flumedinisi		edificio religioso nel comune italiano di Flumedinisi	religious building in Flumedinisi, Italy	Annunziata (Flumedinisi)	https://commons.wikimedia.org/wiki/Category:Annunziata_(Flumedinisi)	Churches of the Annunciation in Sicily
9.	Annunziata	chiesa di Maria Santissima Annunziata	Merì		edificio religioso nel comune italiano di Merì	religious building in Merì, Italy	Annunziata (Merì)	https://commons.wikimedia.org/wiki/Category:Annunziata_(Merì)	Churches of the Annunciation in Sicily
10.	Annunziata	chiesa di Maria Santissima Annunziata	Teramo		edificio religioso nel comune italiano di Teramo	religious building in Teramo, Italy	Annunziata (Teramo)	https://commons.wikimedia.org/wiki/Category:Annunziata_(Teramo)	Churches of the Annunciation in Abruzzo
11.	Ancient Basilica of San Giovanni Battista	antica basilica di San Giovanni Battista	Monza		edificio religioso nel comune italiano di Monza	religious building in Monza, Italy	Antica Basilica di San Giovanni Battista (Monza)	https://commons.wikimedia.org/wiki/Category:Antica_Basilica_di_San_Giovanni_Battista_(Monza)	Churches in Italy built in 595
12.	Ascensione	chiesa dell'Ascensione	Recco	Faveto	edificio religioso nel comune italiano di Recco in localita Faveto	religious building in Faveto, Recco, Italy	Ascensione (Faveto, Recco)	https://commons.wikimedia.org/wiki/Category:Ascensione_(Faveto_Recco)	Churches of the Ascension of Christ in Italy
13.	Assunzione di Maria Vergine	chiesa dell'Assunzione di Maria Vergine	Monchio delle Corti	Rimagna	edificio religioso nel comune italiano di Monchio delle Corti in localita Rimagna	religious building in Rimagna (Rimagna, Monchio delle Corti), Italy	Assunzione di Maria Vergine (Rimagna, Monchio delle Corti)	https://commons.wikimedia.org/wiki/Category:Assunzione_di_Maria_Vergine_(Rimagna_Monchio_delle_Corti)	Baroque Churches in Emilia-Romagna

reconciling to Wikidata with existing items of churches (tried different ways to match better, as the name variants are a lot)

		Search for match				
		24.	Beata Vergine Maria Immacolata	chiesa della Beata Vergine Maria Immacolata edit	Immacolata Concezione Choose new match	Immaculate Conception of Mary
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Chiesa della Beata Vergine Maria Immacolata (95)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Chiesa della Beata Vergine Maria Immacolata (94)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa della Beata Vergine Maria Immacolata (91)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa di Santa Maria Immacolata (90)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa della Beata Vergine Immacolata (89)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa dell'Immacolata (87)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa della Beata Maria Vergine Immacolata di Lourdes (84)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa di Santa Maria Immacolata (84)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa della Beata Vergine Immacolata (82)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Basilica dell'Immacolata Concezione (81)			
			<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> chiesa della Madonna di			

creating missing items of churches and hamlets and updating many more

513 matching rows (681 total) Schema Issues Preview

Target Wikibase instance: Wikidata

The schema below specifies how your tabular data will be transformed into Wikidata edits.

Start from an existing schema: Churches2 Save new...

label-EN	label-IT	NamedAfter 1	Saint-EN	NamedAfter 2	NamedAfter 3	city	location	NewLocationDesc	NewLocationDesc-EN	NewLocationCategory	desc-IT	desc-EN	category	URL	info
----------	----------	--------------	----------	--------------	--------------	------	----------	-----------------	--------------------	---------------------	---------	---------	----------	-----	------

label-IT remove

Terms

Label	it	label-IT	<input type="checkbox"/>	override if present	remove
Description	it	desc-IT	<input type="checkbox"/>	override if present	remove
Alias	it	label-IT	<input type="checkbox"/>	override if present	remove
Label	en	label-EN	<input type="checkbox"/>	override if present	remove
Description	en	desc-EN	<input type="checkbox"/>	override if present	remove
Alias	en	NamedAfter 1	<input type="checkbox"/>	override if present	remove
Alias	en	Saint-EN	<input type="checkbox"/>	override if present	remove

+ add term

Statements

Commons category	category	remove	configure	+ add qualifier	paste reference
▼ 1 references					
copy		remove			
imported from	Wikimedia Commo	remove			
Commons cat	category	remove			
last update	2023-12-02	remove			

results (after 3 months)

6500+ categories found (with no link in Wikidata)

3000+ new items created on Wikidata
(previously unknown religious buildings = +5%)

updated 3500+ existing items on Wikidata
(category link, descriptions, aliases, locations,
titular saints)

(as of 5 Jan. 2024)



Photo: [Rovine Condolima Monterosso](#)
[Calabro Wikigita Calabria 2022 f02](#)
by [Marco Chemello \(Wikimedia Italia\)](#) 2022
CC BY-SA 4.0 on Wikimedia Commons

Thank you all!

Presentation made by [Marco Chemello](#) in 2023
[Wikimedia Italia](#) - associazione per la diffusione della conoscenza libera.

Texts distributed in [CC0 1.0 Universal](#).

Images by the author are distributed with CC BY-SA licenses (see captions).

Some trademarks and logos cited are not owned by Wikimedia Italia.

marco.chemello@wikimedia.it

www.wikimedia.it

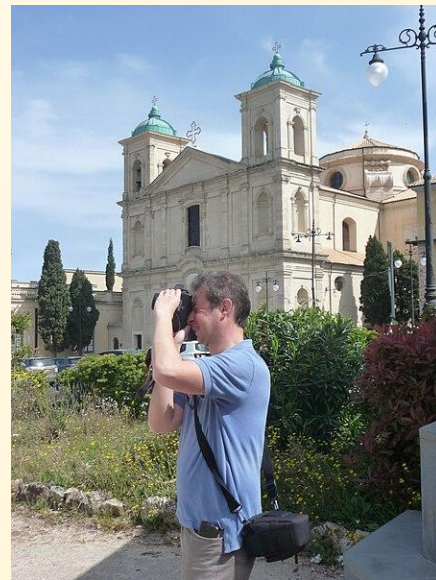
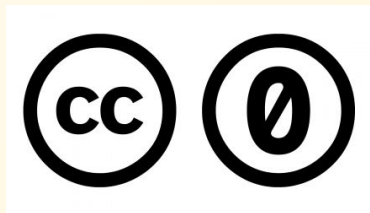


Photo: [Wikigita in Calabria primavera 2022 partecipanti f00](#) by [Marco Chemello \(Wikimedia Italia\)](#) 2022

CC BY-SA 4.0 on Wikimedia Commons

Marco Chemello

GLAM specialist and trainer
(Wikimedia Italia)



An active contributor to Wikipedia since 2004, he is one of the administrators of the Italian edition of the free encyclopedia since 2005. As a Wikipedian in residence he collaborated on many OpenGLAM projects for sharing the contents of cultural institutions, including the National Museum of Science and Technology Leonardo da Vinci in Milan and the Central Institute for Archives (ICAR) of the Italian Ministry of Culture (MiC). Having joined the Wikimedia Italia staff as a trainer and GLAM specialist, he was a Wikimedian in residence at the BEIC Foundation, the Academy of Sciences of Turin, the Ricordi Historical Archive of Milan, the Polo del '900. He collaborates with various universities, including the Politecnico and the Statale of Milan; since 2021 he has been among the first two Wikipedians in residence at an Italian university, the University of Padua. With Unipd he created 3 MOOCs for students, teachers and cultural institutions [on EduOpen](#).