

Structured Data on Commons

♦ editing and upload ♦

with OpenRefine

What we're working on and what's coming



Sandra Fauconnier
Eugene Egbe
Joey Salazar
Antonin Delpeuch



🤔 OpenRefine? 🤨

<https://openrefine.org/> 

<https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine>

- A power tool to clean messy data, popular in a diverse range of communities. It has been serving the needs of journalists, librarians, Wikimedians and scientists for more than 10 years, and is taught in many curricula and workshops around the world.
- Heavily used for Wikidata batch editing (see <https://www.wikidata.org/wiki/Special:Tags>)
- Also supports Wikibase as of OpenRefine 3.5 beta2
- The world's only open source project with its own emoji:  



Vanderkrogt MnM - OpenRefine

https://hub.paws.wmcloud.org/user/Spinster/openrefine/project?project=2596636803043

OpenRefine Vanderkrogt MnM [Permalink](#) Open... Export Help

Facet / Filter 82 matching rows (17247 total) Schema Issues 4 Preview Extensions: Wikidata

Undo / Redo 1199 / 1199 Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 25 next > last »

Refresh Reset All Remove All

country change invert reset

2 choices Sort by: **name** count Cluster

France 2 exclude

Netherlands 82 Facet by choice counts

municipality invert reset

gou case sensitive regular expression

	All	external_id	external_url	title_orig	desc_en	desc_nl	municipality	province	country	countryn
	640.	FR25eh	http://vanderkrogt.net/statuses/object.php?record=FR25eh	De Nonnen fan it Nijkleaster	statue by Annet Haring in Scharnegoutum, Netherlands	standbeeld door Annet Haring in Scharnegoutum, Nederland	Scharnegoutum	Friesland	Netherlands Choose new match	Nederland
	7194.	ZE08aq	http://vanderkrogt.net/statuses/object.php?record=ZE08aq	Johannes Vijverberg	statue by Miems van Citters in Noordgouwe, Netherlands	standbeeld door Miems van Citters in Noordgouwe, Nederland	Noordgouwe	Zeeland	Netherlands Choose new match	Nederland
	8099.	ZH21aa	http://vanderkrogt.net/statuses/object.php?record=ZH21aa	Cornelis & Frederik de Houtman <input checked="" type="checkbox"/> Create new item Search for match	statue by Xav. Stracké in Gouda, Netherlands	standbeeld door Xav. Stracké in Gouda, Nederland	Gouda	Zuid-Holland	Netherlands Choose new match	Nederland
	8100.	ZH21ab	http://vanderkrogt.net/statuses/object.php?record=ZH21ab	Gheraert Leeu (de boekdrukker) <input checked="" type="checkbox"/> Create new item Search for match	statue by Roel Bendijk in Gouda, Netherlands	standbeeld door Roel Bendijk in Gouda, Nederland	Gouda	Zuid-Holland	Netherlands Choose new match	Nederland
	8101.	ZH21ac	http://vanderkrogt.net/statuses/object.php?record=ZH21ac	A.A. van Bergen Ijzendoorn <input checked="" type="checkbox"/> Create new item Search for match	statue by unknown artist in Gouda, Netherlands	standbeeld door unknown artist in Gouda, Nederland	Gouda	Zuid-Holland	Netherlands Choose new match	Nederland
	8102.	ZH21ad	http://vanderkrogt.net/statuses/object.php?record=ZH21ad	Dirck Volkertsz. Coornhert <input checked="" type="checkbox"/> gablestone of Dirck Volkertsz Coornhert (87) <input checked="" type="checkbox"/> Create new item Search for match	statue by Theo van Reijn in Gouda, Netherlands	standbeeld door Theo van Reijn in Gouda, Nederland	Gouda	Zuid-Holland	Netherlands Choose new match	Nederland
	8103.	ZH21ae	http://vanderkrogt.net/statuses/object.php?record=ZH21ae	Dirck en Wouter Crabeth <input checked="" type="checkbox"/> Create new item Search for match	statue by Jo Uiterwaal in Gouda, Netherlands	standbeeld door Jo Uiterwaal in Gouda, Nederland	Gouda	Zuid-Holland	Netherlands Choose new match	Nederland
	8104.	ZH21af	http://vanderkrogt.net/statuses/object.php?record=ZH21af	Desiderius Erasmus <input checked="" type="checkbox"/> Bust of Desiderius Erasmus (87) <input checked="" type="checkbox"/> Desiderius Erasmus (86) <input checked="" type="checkbox"/> Erasmus (67) <input checked="" type="checkbox"/> Desiderius Erasmus of Rotterdam, 1467/1469-1536, Scholar, Theologian, and Satirist (27) <input checked="" type="checkbox"/> Create new item Search for match	statue by Hildo Krop in Gouda, Netherlands	standbeeld door Hildo Krop in Gouda, Nederland	Gouda	Zuid-Holland	Netherlands Choose new match	Nederland

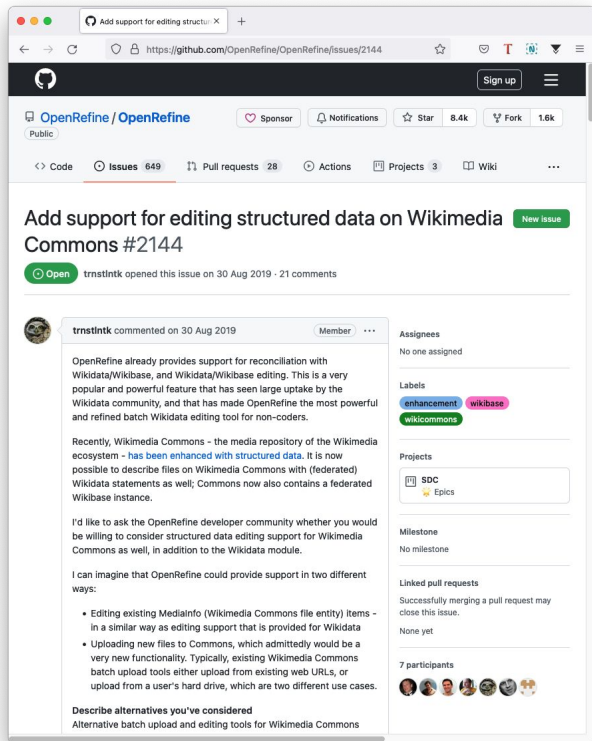


Wouldn't it be cool if OpenRefine would also edit the structured data of Commons files?

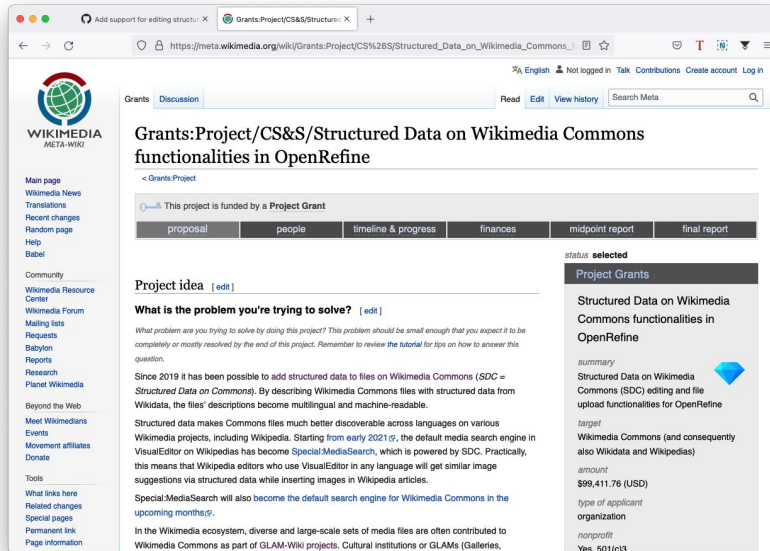
can i haz powerfulz batch file upload with SDC?

Maybe we can finally sunset the GLAMwiki Toolset! 🤖

Structured Data on Commons (SDC) functionalities in OpenRefine (2021-22)



The screenshot shows a GitHub issue titled "Add support for editing structured data on Wikimedia Commons #2144". The issue was opened by user `trnstntk` on August 30, 2019. The issue description discusses the need for OpenRefine to support editing structured data on Wikimedia Commons, mentioning the current limitations of Wikidata/Wikibase editing and the potential for a federated Wikidata statements as well. The issue also mentions that Commons now also contains a federated Wikibase instance. The issue is currently open and has 21 comments.



The screenshot shows a Wikimedia Grants page for the project "Grants:Project/CS&S/Structured Data on Wikimedia Commons functionalities in OpenRefine". The project is funded by a Project Grant. The page includes a navigation menu with options like "proposal", "people", "timeline & progress", "finances", "midpoint report", and "final report". The "Project idea" section describes the problem: "What is the problem you're trying to solve?" and provides a detailed description of the project's goals and impact. The project summary includes the amount of the grant (\$99,411.76 USD) and the type of applicant (organization).

- 👉 Feature requested in August 2019
- 👉 Wikimedia grant awarded in June 2021
- 👉 Development: July 2021-June 2022



What are we building? When will it be ready?



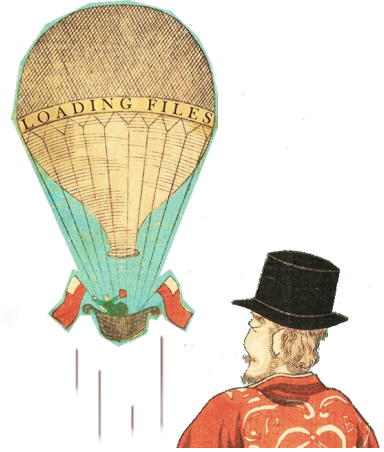
1 Reconciliation

 Now-ish 



2 Edit existing files' structured data

 End January 2022



3 Upload new Commons files with structured data

 End June 2022



Reconciliation, you say? 🤔

✨ The magic happens here ✨

title
File:LamidoGrandMosque.jpg
File:Foumban - mosque entry.jpg
File:MosquéeFadil.jpg
File:Yaoundé Palais des Congrès.jpg
File:Mosquée Bafia.jpg
File:LT 37 Mosqué central de new-bell.jpg
File:LT 35 Mosqué central de new-bell.jpg
File:LT 36 Mosqué central de new-bell.jpg
File:LT 38 Mosqué central de new-bell.jpg
File:LT 39 Mosqué central de new-bell.jpg
File:LT 40 Mosqué central de new-bell.jpg
File:LT 41 Mosqué central de new-bell.jpg
File:LT 42 Mosqué central de new-bell.jpg
File:LT 43 Mosqué central de new-bell.jpg
File:Mosquée de Ngaoundéré, 2006.jpg
File:Mosquée centrale de Foumban.JPG
File:Mosquée centrale de New-Bell.JPG
File:Mosquée foumban.JPG
File:Bamum mosque, in Cameroon (IMP-DEFAP CMCFGB-CP042 2).jpg
File:Mosque on Ndawara tea plantation.JPG

Commons
reconciliation >

(List of) Commons
file names

title
File:LamidoGrandMosque.jpg
Choose new match
File:Foumban - mosque entry.jpg
Choose new match
File:MosquéeFadil.jpg
Choose new match
File:Yaoundé Palais des Congrès.jpg
Choose new match
File:Mosquée Bafia.jpg
Choose new match
File:LT 37 Mosqué central de new-bell.jpg
Choose new match
File:LT 35 Mosqué central de new-bell.jpg
Choose new match
File:LT 36 Mosqué central de new-bell.jpg
Choose new match
File:LT 38 Mosqué central de new-bell.jpg
Choose new match
File:LT 39 Mosqué central de new-bell.jpg
Choose new match
File:LT 40 Mosqué central de new-bell.jpg
Choose new match
File:LT 41 Mosqué central de new-bell.jpg
Choose new match
File:LT 42 Mosqué central de new-bell.jpg
Choose new match
File:LT 43 Mosqué central de new-bell.jpg

Data extension >

M-ids

```

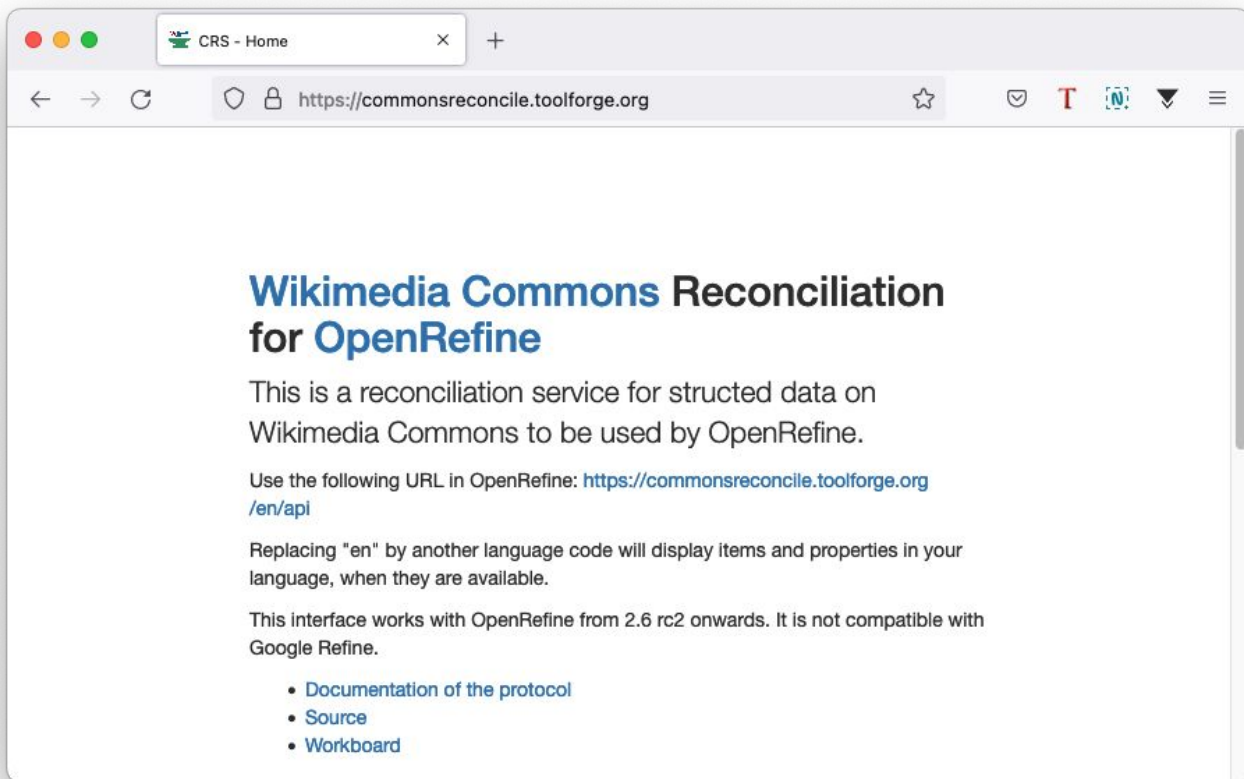
=={{int:filedesc}}==
{{Information
|description={{#if:|La mosquée centrale de la ville d'Ebolowa vue de face}}
|date=2019-03-13
|source=Collection Groupe scolaire Rosa Verenini au Cameroun
|author=Collection Groupe scolaire Rosa Verenini
|permission=
|other versions=
}}

```

Extract wikitext

mosque
Choose new match
Bonamousadi Mosque
Choose new match
Douala
Choose new match
motor car
Choose new match
road
Choose new match
mosque
Choose new match
Bonamousadi Mosque

Extract structured
data



 <https://commonsreconcile.toolforge.org/> 

What are we building? When will it be ready?



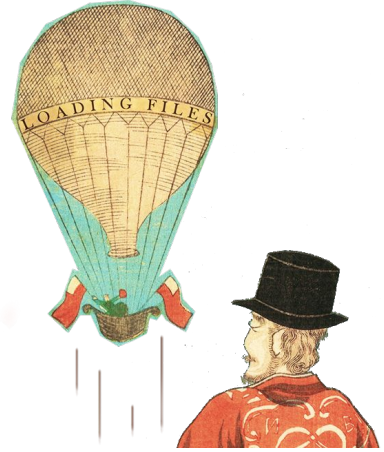
1 Reconciliation

 Now-ish 



2 Edit existing files' structured data

 End January 2022



3 Upload new Commons files with structured data

 End June 2022

Test, participate, stay in touch!

- Timeline and monthly reports on Meta: <https://w.wiki/4JbP>
- Info page on Commons:
<https://commons.wikimedia.org/wiki/Commons:OpenRefine>
 - Or <https://w.wiki/4JbR> in short :-)
 - With links to Phabricator, GitHub, signup on wiki

Questions?