

# OpenRefine en Wikibase: Status en plannen 2022

Wikibase kennisgroep, 22 juli 2022

Sandra Fauconnier [sandra@openrefine.org](mailto:sandra@openrefine.org)



**OpenRefine**



The screenshot shows a web browser window with the URL [https://en.wikiversity.org/wiki/OpenRefine\\_to\\_Wikibase:\\_Data\\_Upload\\_Pipeline](https://en.wikiversity.org/wiki/OpenRefine_to_Wikibase:_Data_Upload_Pipeline). The page features the Wikiversity logo on the left and a navigation menu. The main content area is titled "OpenRefine to Wikibase: Data Upload Pipeline" and includes a "Resource" tab, a "Discuss" tab, and a search bar. The text describes guidelines for data upload, mentioning tools like OpenRefine, Wikibase, and Kompakt. A table of contents is visible at the bottom of the main content area.

OpenRefine to Wikibase: Data Upload Pipeline

Guidelines by *Lozana Rossenova and Lucia Sohmen*

This guidelines document was prepared at TIB's [Open Science Lab](#) (OSL) alongside the development of an MVP (minimum viable product) for the enrichment of 3D media files with semantic metadata within the context of Task Area 1: Data Capture and Enrichment of the [NFDI4Culture](#) consortium. The MVP builds on several existing FOSS tools:

- [OpenRefine](#), a data cleaning, reconciliation and batch upload tool;
- [Wikibase](#), a suite of services developed by Wikimedia Germany; it combines the ability to handle large volumes of data points within a linked open data environment, which includes sophisticated data querying and extraction services via a dedicated SPARQL endpoint;
- [Kompakt](#), a browser-based open-source 3D- and multimedia viewer Kompakt with built-in collaborative annotation features.

The integrated suite of tools follows FAIR principles and facilitates linking 3D-objects and annotations, and their cultural context (including historical people and places, geo-location and capture-technology metadata), to the broader semantic web and various national and international authority records (GND, VIAF and more).

This guide focuses on the first two tools part of the toolchain: OpenRefine and Wikibase. The following sections present a complete workflow: from starting to collect and structure data in spreadsheets, to transforming and reconciling data via OpenRefine, and finally to reviewing the different options for uploading your data to a Wikibase instance.

Contents [hide]	
1	Data collection
1.1	Entity simplicity
1.2	Thinking in triples
1.3	One entity per field
1.4	One sheet per entity type

[https://en.wikiversity.org/wiki/OpenRefine\\_to\\_Wikibase:\\_Data\\_Upload\\_Pipeline](https://en.wikiversity.org/wiki/OpenRefine_to_Wikibase:_Data_Upload_Pipeline)



**OpenRefine**

# NFDI grant Wikibase development in OpenRefine

NFDI = German National Research Data Infrastructure

Now till end December 2022

2 (3) deliverables:

1. Batch upload and batch editing of media files in Wikibases through OpenRefine
2. Batch editing of data stored in custom (non-Wikidata) data types in Wikibases through OpenRefine

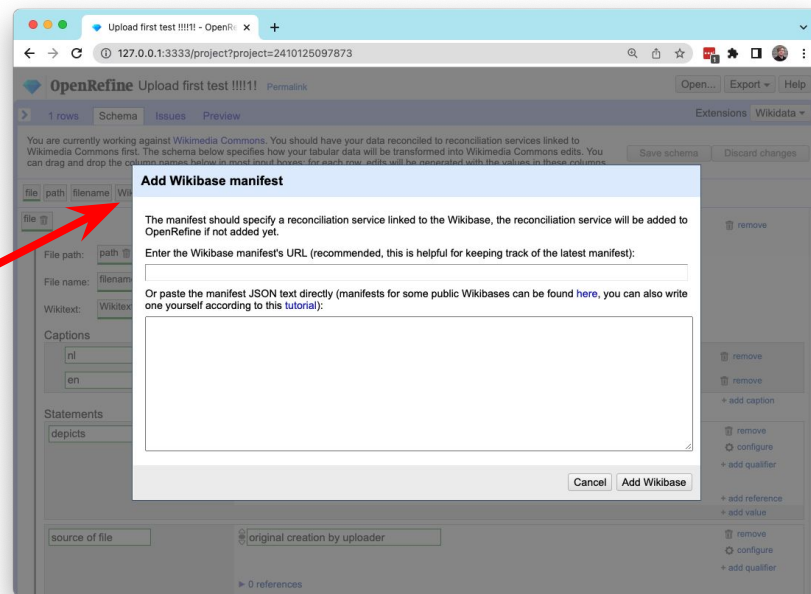
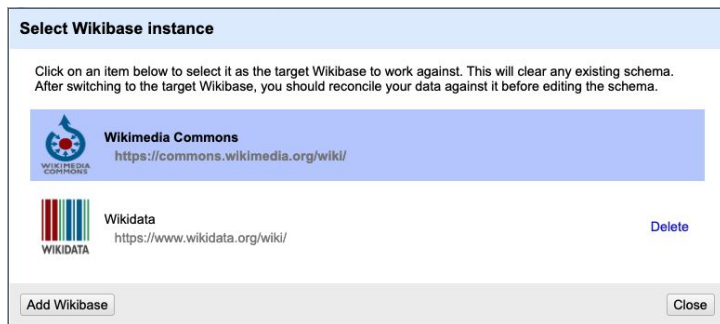
E.g. local media file; EDTF data type

(+ Dissemination to NFDI stakeholders, i.e. documentation + masterclass)

Who: Antonin Delpuch (development), Sandra Fauconnier (management) + Lozana Rossenova as NFDI4Culture liaison

Grant is (hopefully) renewable!

# Media file upload: current status (Wikimedia Commons upload)



(SDC batch editing in OpenRefine 3.6;  
Media file upload from upcoming version 3.7)

Upload first test !!!!! - OpenRefine

127.0.0.1:3333/project?project=2410125097873

OpenRefine Upload first test !!!!! Permalink

Open... Export Help

1 rows Schema Issues Preview Extensions Wikidata

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first < previous 1 of 1 page next > last »

	file	path	filename	Wikitext	P180
1.	Gazelle Kwikstep folding bike.jpg <small>Choose new match</small>	/Users/fokky/Desktop/Gazelle Kwikstep folding bike.jpg	Gazelle Kwikstep folding bike	<pre> {{Information}} =={{int:license-header}}== {{self:cc-by-4.0}}  [[Category:Gazelle vintage bicycles]] [[Category:Folding bicycles]] </pre>	Gazelle Kwikstep <small>Choose new match</small>



Upload first test !!!!! - OpenRefine

127.0.0.1:3333/project?project=2410125097873

OpenRefine Upload first test !!!!! Permalink

Open... Export Help

1 rows Schema Issues Preview Extensions Wikidata

You are currently working against Wikimedia Commons. You should have your data reconciled to reconciliation services linked to Wikimedia Commons first. The schema below specifies how your tabular data will be transformed into Wikimedia Commons edits. You can drag and drop the column names below in most input boxes: for each row, edits will be generated with the values in these columns.

Save schema Discard changes

file	path	filename	Wikitext	P180
file				

File path: path

File name: filename

Wikitext: Wikitext

Statements

depicts: Gazelle Kwikstep

source of file: original creation by uploader

Uploaded to Wikimedia Commons as  
[https://commons.wikimedia.org/wiki/File:Gazelle\\_Kwikstep\\_folding\\_bike.jpeg](https://commons.wikimedia.org/wiki/File:Gazelle_Kwikstep_folding_bike.jpeg)



OpenRefine

# Reconciliation and Wikibases

A [Python web service](#) is currently required to expose a reconciliation endpoint for a Wikibase instance - <https://github.com/wetneb/openrefine-wikibase>

Integration in Docker image: would be great to have; WMDE has it on its roadmap, but will not be developed soon (yet)

Documentation for Wikibase managers: <https://openrefine-wikibase.readthedocs.io/>

Can your reconciliation service be used by external parties? Then add it to the reconciliation testbench! Instructions at the bottom - <https://reconciliation-api.github.io/testbench/#/>



**OpenRefine**

What do you, as Wikibase administrator, need from OpenRefine?

How can we all help OpenRefine to make that happen?