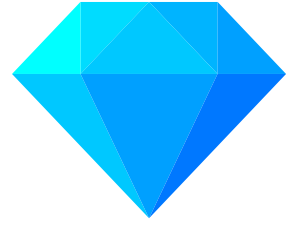


Einführung in OpenRefine: Datenaufbereitung für Kulturerbe-Metadaten

Sebastian Wolff
Kulturpool – NHM Wien
03.06.2026

Was ist OpenRefine?



"A power tool for working with messy data."

- OpenSource Software zum Bearbeiten von Daten (ehem. Google Refine)
- Läuft und bearbeitet Daten lokal
- Aktive Community und Weiterentwicklung
- Erweiterbar mit verschiedenen Extensions
- Aktuellste Version: OpenRefine 3.10.0

Disclaimer

Für die Erstellung der Beispiele wurde eine adaptierte Version von Metadaten von Kulturpool-Objekten verwendet. Die Inhalte wurden so verändert, dass sie typische Anwendungsfälle und Bearbeitungsschritte in OpenRefine besser veranschaulichen. Sie entsprechen daher nicht den Originaldaten.

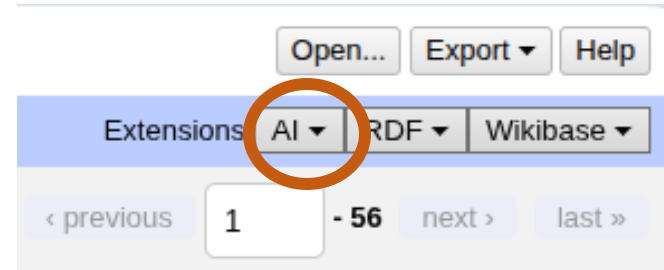
Extensions

- Vielzahl an Extensions verfügbar
- Installation:

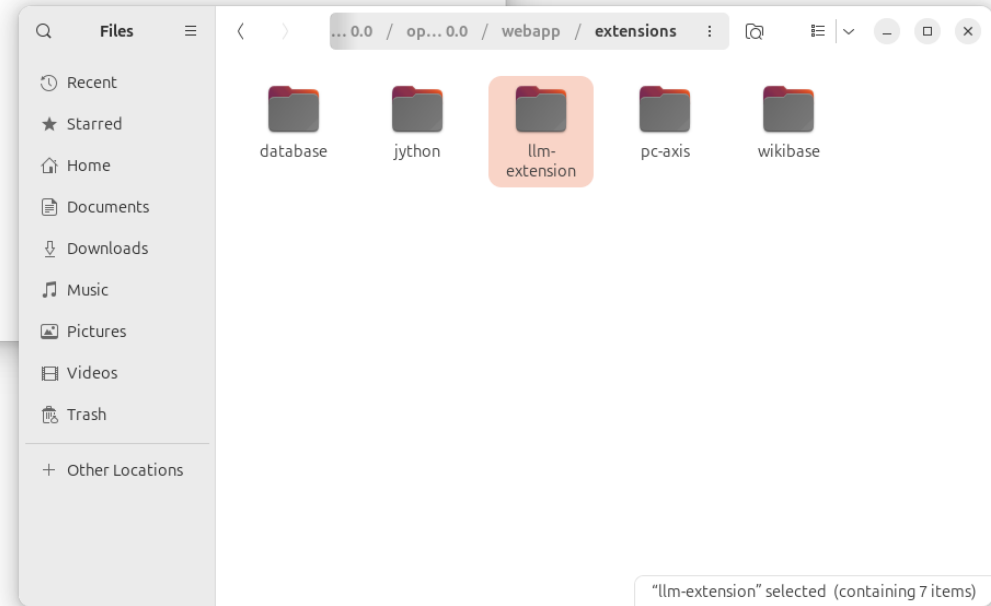
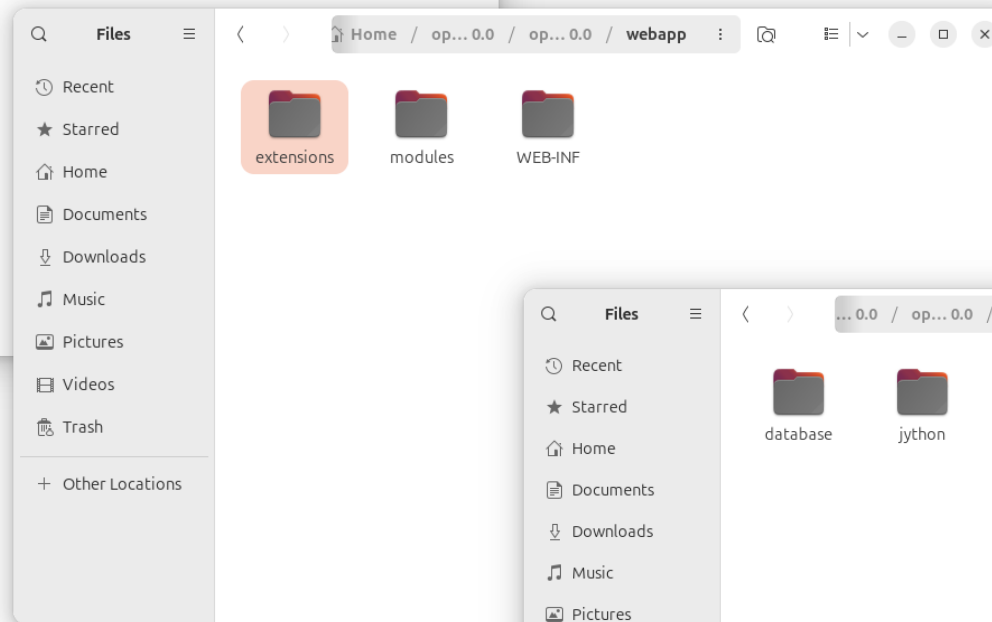
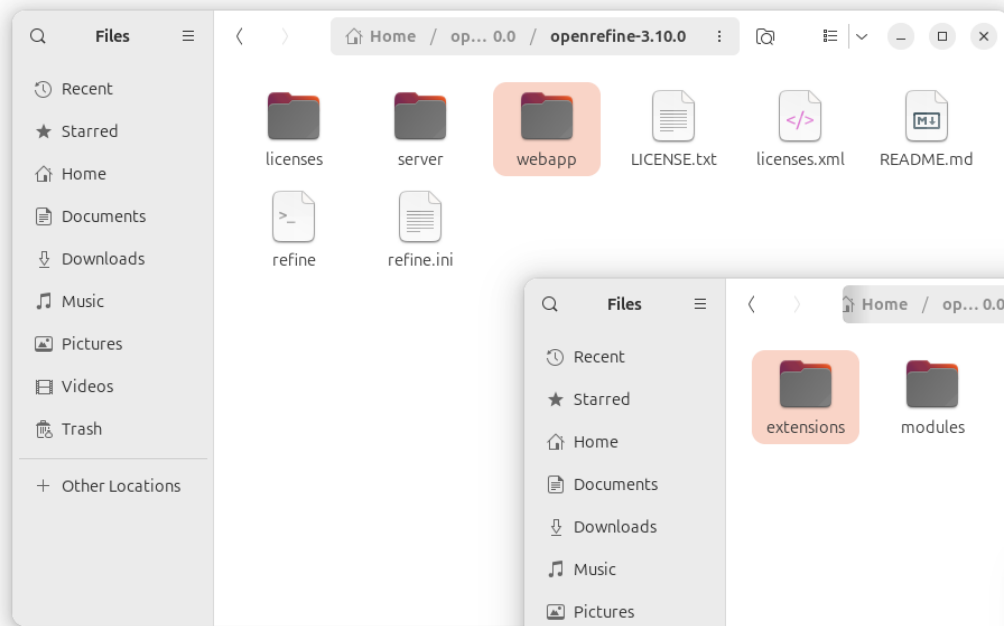
Download von ZIP-File

Entpacken des ZIP-Files in webapp/extensions

Kontrolle in OpenRefine



Neue Funktionen bestätigen
die korrekte Installation



OpenRefine Basics – Reihen und Records

- Tabellarische Darstellung der Daten als Spalten und Reihen bzw. Records
- Zellen können aufgeteilt oder vereinigt werden
- Freie Wahl an Separators – dabei aber aufpassen
- Blank Down und Fill down
- Undo/Redo
- Zahlen, Datum und Boolean - grün

56 rows					
Show as: rows records Show: 5 10 25 50 100 500 1000 rows					
All	ID	Title	Title_language	dcType	
☆ ↗	1.	https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho	Brühl bei Weitra		
☆ ↗	2.	https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho	Frauen*Spaziergänge Innensichten	lecture	Choose new match
☆ ↗	3.			lecture	Choose new match
☆ ↗	4.			lecture	Choose new match
☆ ↗	5.			lecture	Choose new match
☆ ↗	6.			lecture	Choose new match
☆ ↗	7.				
☆ ↗	8.				
☆ ↗	9.				
☆ ↗	10.				
☆ ↗	11.				
☆ ↗	12.				
☆ ↗	13.				
☆ ↗	14.				
☆ ↗	15.				
☆ ↗	16.				
☆ ↗	17.	https://id.kulturpool.at/467f4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de	Klebebände Choose new match
☆ ↗	18.	https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsornat	de	print Choose new match
☆ ↗	19.				
☆ ↗	20.	https://id.kulturpool.at/f4c03c5f-9663-42f1-84f9-3e4829a88dbb/cho	Waldlandschaft mit Fels	de	Klebebände Choose new match
☆ ↗	21.				
☆ ↗	22.	https://id.kulturpool.at/14e72067-c98a-486d-a984-fd1af0cf3296/cho	Gedeon	de	print Choose new match
☆ ↗	23.	https://id.kulturpool.at/9b964b04-c1de-491d-840a-1fb987b6b41d/cho	Speculum concionatorum, in septem libros distributum	la	Buch Choose new match
☆ ↗	24.	https://id.kulturpool.at/e2acf681-b57f-426e-a55d-7d74ebfe0fa8/cho	Deutsche Grammatik für Gymnasien, Realschulen und ähnliche Lehranstalten	de	Buch Choose new match
☆ ↗	25.				

20 records					
Show as: row records Show: 5 10 25 50 100 500 1000 records					
All	ID	Title	Title_language	dcType	
☆ ↗	1.	https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho	Brühl bei Weitra		
☆ ↗	2.	https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho	Frauen*Spaziergänge Innensichten	lecture	Choose new match
☆ ↗				lecture	Choose new match
☆ ↗				lecture	Choose new match
☆ ↗				lecture	Choose new match
☆ ↗				lecture	Choose new match
☆ ↗				lecture	Choose new match
☆ ↗					
☆ ↗					
☆ ↗					
☆ ↗					
☆ ↗					
☆ ↗					
☆ ↗					
☆ ↗					
☆ ↗	3.	https://id.kulturpool.at/467f4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de	Klebebände Choose new match
☆ ↗	4.	https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsornat	de	print Choose new match
☆ ↗					
☆ ↗	5.	https://id.kulturpool.at/f4c03c5f-9663-42f1-84f9-3e4829a88dbb/cho	Waldlandschaft mit Fels	de	Klebebände Choose new match
☆ ↗					
☆ ↗	6.	https://id.kulturpool.at/14e72067-c98a-486d-a984-fd1af0cf3296/cho	Gedeon	de	print Choose new match
☆ ↗	7.	https://id.kulturpool.at/9b964b04-c1de-491d-840a-1fb987b6b41d/cho	Speculum concionatorum, in septem libros distributum	la	Buch Choose new match
☆ ↗	8.	https://id.kulturpool.at/e2acf681-b57f-426e-a55d-7d74ebfe0fa8/cho	Deutsche Grammatik für Gymnasien, Realschulen und ähnliche Lehranstalten	de	Buch Choose new match
☆ ↗					

Umschalten zwischen "rows" und "records" Ansicht

20 records

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

All	ID	Title	Title_language	dcType
☆ ↗	1. https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho			
☆ ↗	2. https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho			
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗	3. https://id.kulturpool.at/467f4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de	Klebebände Choose new match
☆ ↗	4. https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsort	de	print Choose new match
☆ ↗	5. https://id.kulturpool.at/f4c03c5f-9663-42f1-84f9-3e4829a88dbb/cho	Waldlandschaft mit Fels	de	Klebebände Choose new match

20 records

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

All	ID	Title	Title_language	dcType
☆ ↗	1. https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho			
☆ ↗	2. https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho			
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗				
☆ ↗	3. https://id.kulturpool.at/467f4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de	Klebebände Choose new match
☆ ↗	4. https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsort	de	print Choose new match
☆ ↗	5. https://id.kulturpool.at/f4c03c5f-9663-42f1-84f9-3e4829a88dbb/cho	Waldlandschaft mit Fels	de	Klebebände Choose new match

Aufteilen bzw. Vereinigen von Zellen

OpenRefine OpenRefine Webinar Demo [Permalink](#)

Facet / Filter **Undo / Redo 57 / 57** <

Extract... Apply...

Filter

"Klebebandseite"

- ☒ 38. Match item Klebebände (Q113285123) for 2 cells containing "Klebebandseite" in column dcType
- + 39. Create new column dcType_WikidataID based on column dcType by filling 33 rows with cell.recon.match.id
- ☒ 40. Discard recon judgments and clear recon data for 16 cells in column Creator

20 records

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

	All	ID
☆	1.	https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho
☆	2.	https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho
☆		
☆		
☆		
☆		
☆		
☆		

Unendliches Undo / Redo zum Rückgängigmachen von Aktionen

OpenRefine Basics - Spalten

- Spalten können einfach neu geordnet werden **!Vorsicht bei Records!**
- Spalten können vereinigt oder getrennt werden
- Sortieren um leere Felder o.Ä. zu erkennen
- Sortieren muss extra permanent gemacht werden
- Viele Optionen über Drop-Down Menüs

20 records

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

▼ All	▼ ID	▼ Title	▼ Title_language
☆ ↗ 1.	15a973-157a-412b-b554-	Brühl bei Weitra	
☆ ↗ 2.	02019e-f45c-46ec-86be-	Frauen*Spaziergänge Innensichten	
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗			
☆ ↗ 3.	https://id.kulturpool.at/467ff4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de
☆ ↗ 4.	https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsornat	de
☆ ↗			

20 records

Show as: rows records

Show: 5 10 25 50 100 500 1000 records

All	ID	Title	Title_language
 	1. https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho	Brühl bei Weitra	
 	2. https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho	Frauen*Spaziergänge Innensichten	
 			
 			
 			
 			
 			
 			
 			
 			
 			
 			
 			
 	3. https://id.kulturpool.at/467ff4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de
 	4. https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsornat	de
 			

Sort by ID

Sort cell values as

☒ text ☐ case-sensitive
 ☐ numbers
 ☐ dates
 ☐ booleans

☒ a - z ☐ z - a

Position blanks and errors

☐ Valid values
 ☐ Errors
 ☐ Blanks

Drag and drop to re-order

OK

Cancel

Sortieren der Reihen anhand der Werte

Faceting und Filtering

- Facetten als Grundpfeiler der Datenbearbeitung
- Verschiedene Arten von Facetten vorhanden oder Möglichkeit eigene zu erstellen
- Liste von Werten welche bearbeitet werden können
- Textfilter für suchen und filtern
- Datenbereinigung
- Subset "Erstellung" zur Bearbeitung

Facet / Filter

Undo / Redo 57 / 57

Refresh

Reset all

Remove all

Title

change

21 choices

Sort by: name count

Cluster

Brühl bei Weitra 1

Cleome rotundifolia (Mart. & Zucc.) Iltis 1

Deutsche Grammatik für Gymnasien, Realschulen und ähnliche Lehranstalten 1

Frauen*Spaziergänge | Innensichten 1

Gedeon 1

Geschichte des Grafen Wilhelm von Holland, römischen Königs. 1

Histoire De La Gaule, Par M. Serpette De Marincourt, Avocat A La Court

20 records

Show as: rows records

Show: 5 10 25 50 100 500 1000 records

All	ID	Title	Title_language
☆ 7	1. https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho		
☆ 7	2. https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho		
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7			
☆ 7	3. https://id.kulturpool.at/467ff4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de
☆ 7	4. https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsornat	de
☆ 7			

Facet

Text filter

Edit cells

Edit column

Transpose

Sort...

Hide / Show

Reconcile

Extract using AI

Text facet

Numeric facet

Timeline facet

Scatterplot facet...

Custom text facet...

Custom numeric facet...

Customized facets

Facetten ermöglichen es nach Texten oder anderen Werten zu filtern

Clustering

- OpenRefine verfügt über verschiedene Clustering Methoden
- Möglichkeit eigene Algorithmen z.B. mit Python zu integrieren
- Ähnliche Daten gruppieren und vereinheitlichen

Facet / Filter Undo / Redo 57 / 57

Refresh Reset all Remove all

20 records

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

Cluster

21 choices Sort by: name count

Brühl bei Weitra 1
Cleome rotundifolia (Mart. & Zucc.) Itlis 1
Deutsche Grammatik für Gymnasien, Realschulen und ähnliche Lehranstalten 1
Frauen*Spaziergänge | Innensichten 1
Gedeon 1
Geschichte des Grafen Wilhelm von Holland, römischen Königs. 1
Histoire De La Gaule, Par M. Serpette De Marincourt, Avocat A La Court

All	ID	Title	Title_language
1.	https://id.kulturpool.at/0815a973-157a-412b-b554-58f8c321fdc0/cho		
2.	https://id.kulturpool.at/8002019e-f45c-46ec-86be-e265708e466d/cho		
3.	https://id.kulturpool.at/467ff4a8-ed4c-46c6-8e14-3c94df9ca166/cho	Wüstenwanderung Moses	de
4.	https://id.kulturpool.at/983a6ca0-21d4-4bc8-8573-5a8af7925c06/cho	Kaiser Leopold II. im kaiserlichen Krönungsornat	de
5.	https://id.kulturpool.at/f4c03c5f-9663-42f1-84f9-3e4829a88dbb/cho	Waldlandschaft mit Fels	de

Facet

Text filter

Edit cells

Edit column

Transpose

Sort...

Hide / Show

Reconcile

Extract using AI

Cluster and edit...

Replace...

Es gibt zwei Optionen um das Cluster-Menü zu öffnen

Gymnasien, Realschulen und ähnliche
Lehranstalten 1

Frauen*Sp
Innensicht
Gedeon 1
Geschicht
von Hollar
Histoire D
Serpette D

Choose new match

Cluster and edit column "Subject"

Find groups of different cell values that might be other representations of the same thing. For example, "New York" and "new york" likely refer to the same concept and just differ by capitalization, and "Gödel" and "Godel" probably refer to the same person. [Find out more...](#)

Method **Key collision** Keying function **Metaphone3** Manage clustering fuctions

☐ Auto-update 2 clusters found

Merge?	Values in cluster	New cell value	Cluster size	Row count
☒	☒ theorie ☒ theory	theorie	2	2
☐	☐ art contemporain ☐ arte contemporânea	art contemporain	2	2

Average length of choices
6 — 17

Length variance of choices
0.5 — 1

Select all Deselect all Export clusters **Merge selected & re-cluster** Merge selected & Close Close

Methode und Keying function beeinflussen die erstellten Cluster

Transformations

- Kann verwendet werden für:
 - Spalten aufteilen
 - Datenstandardisierung
 - Daten aus Texten extrahieren
- Kann in GREL, Python, oder Clojure geschrieben werden
- Häufige Transformationen sind bereits implementiert

GREL

- General Refine Expression Language
- Ressourcen mit Rezepten für häufigen Transformationen:
<https://github.com/OpenRefine/OpenRefine/wiki/Recipes>
- Unterstützt Regular Expression mit Anpassungen

Reconciliation Basics

- Gibt die Möglichkeit über Values oder über IDs zu reconcilen
- Anhand der Ergebnisse der Reconciliation können wir neue Spalten hinzufügen
- Möglichkeit Properties zu übergeben und Typen auszuwählen

Reconciliation Service hinzufügen

- Auswahl an verfügbaren Services: <https://reconciliation-api.github.io/testbench/0.2/#/>
- Man kann auch APIs anspielen – user-agent dabei wichtig
- RDF Extension bietet Funktionalitäten für APIs und SPARQL Endpoints

Reconcile column Title

Select a service to reconcile against:

- ☒ Wikidata reconcil.link (en)
https://wikidata.reconci.link/en/api
- ☐ Wikidata reconcil.link (de)
https://wikidata.reconci.link/de/api
- ☐ Bionomia
https://api.bionomia.net/reconcile
- ☐ VIAF
http://localhost:8080/reconcile/viaf
- ☐ ORCID - Smart Names Mode
http://localhost:8080/reconcile/orcid/smartnames
- ☐ DBpedia
http://127.0.0.1:3333/extension/rdf-extension/services/dbpedia
- ☐ Getty Vocabularies Reconciliation Service
https://services.getty.edu/vocab/reconcile/
- ☐ World Historical Gazetteer Reconciliation Service
https://whgazetteer.org/reconcile
- ☐ GeoNames Reconciliation Service from FornPunkt
https://fornpunkt.se/apis/reconciliation/geonames
- ☐ GND reconciliation for OpenRefine
https://lobid.org/gnd/reconcile
- ☐ GND reconciliation for OpenRefine
https://reconcile.gnd.network

Add standard service...

Discover services...

Next »

Es können neue Services hinzugefügt werden

Typen grenzen die Suche ein und Properties liefern zusätzliche Informationen

Reconcile column Title

Reconcile each cell to an entity of one of these types:

- ☒ human
Q5
- ☐ human biblical figure
Q20643955
- ☐ rocking stone
Q1061002
- ☐ entity
Q35120

Also use relevant details from other columns:

Column

ID

Title_language

dcType

dcType_WikidataID

dcType_language

dcType_ID

dcType_gndID

dcType_aatID

As property

☐ Reconcile against
type:

☐ Reconcile against no particular type

☒ Auto-match candidates with high confidence

Maximum number of candidates to return

Back

Start reconciling...

Cancel

Vorschläge überprüfen

- Score wird von Services berechnet und nicht von OpenRefine – Individuell verschieden
- Reconciliation verfügt je nach Service über Auto-Matching
- Manuelles korrigieren oder auswählen ist meistens notwendig

Predigt

- ☒ ☒ Predigt (100)
- ☒ ☒ Predigt (100)
- ☒ ☒ Bergpredigt (78)
- ☒ ☒ Create new item

[See more](#) | [Search for match](#)

Deutsch

[Choose new match](#)

Grammatik

- ☒ ☒ Grammatik (100)
- ☒ ☒ Grammatiker (100)

)

- ☒ ☒ Grammatik (100)
- ☒ ☒ Create new item

[See more](#) | [Search for match](#)

Wilhelm

- ☒ ☒ Wilhelm II. (100)
- ☒ ☒ Wilhelm (100)
- ☒ ☒ Wilhelm (100)
- ☒ ☒ Create new item

[See more](#) | [Search for match](#)

Latein

[Choose new match](#)

Rede

- ☒ ☒ Graph (100)
- ☒ ☒ Hexencredo (100)

)

- ☒ ☒ Réde (100)
- ☒ ☒ Create new item

[See more](#) | [Search for match](#)

AI Extension

- Ermöglicht es LLMs direkt anzuspielen
- Kann verwendet werden für:
 - Datenbereinigung und Standardisierung
 - Übersetzung
 - Entitätserkennung
 - Klassifizierung (e.g. Sentiment Analysis)
 - Fakten oder Konzeptextrahierung



Manage LLM Providers

This extension supports the openai chat completion model and can be defined and used in any project.

Label	Model	Server URL	Actions
Google: Gemini 2.0 Flash	google/gemini-2.0-flash-001	https://openrouter.ai/api/v1/chat/completions	Edit Delete
OpenAI: GPT-5 Mini	openai/gpt-5-mini	https://openrouter.ai/api/v1/chat/completions	Edit Delete

This extension supports Chat Completion API services (/v1/chat/completions) from various LLM providers. See our provider setup guide for configuration details, parameter documentation for optimizing your prompts and results.

- [LLM Provider setup guide](#)
- [Feature working guide](#)

[Add LLM Provider](#) [Close](#)

Neue LLM Provider benötigen ein Label, eine URL, einen Modelnamen und einen API Key

LLM Provider

Label:

Server URL:

Model:

API Key:

Temperature:

Top-P:

Seed:

Max tokens:

Wait Time (milliseconds):

[Help](#) [Test service](#) [Cancel](#) [Save](#)

Show as: [rows](#) **records** Show: [5](#) [10](#) **[25](#)** [50](#) [100](#) [500](#) [1000](#) records

- Facet ▶
- 🔍 Text filter
- Edit cells ▶
- Edit column ▶
- Transpose ▶
- Sort...
- Hide / Show ▶
- Reconcile ▶
- Extract using AI**

Extract using AI from column Title

Column

Add new

Title_AI_Translation

LLM Provider

Google: Gemini 2.0 Flash

Response format

Text

Describe what needs to be done

Translate the provided title to english.

Provide JSON Schema

Enabled when Response Format is set to JSON Schema

Preview

History

Starred

value

Brühl bei Weitra

response

Brühl near Weitra

Help

Generate Preview

RDF Extensions

- Ermöglicht es APIs, SPARQL Endpoints und RDF-Files als Reconciliation Services hinzuzufügen
- Daten im RDF Format exportieren
- Vokabularien/Ontologien importieren

Vor- und Nachteile

VORTEILE

- Lokale Bearbeitung der Daten
- Original-Daten werden nicht verändert
- Aktive Community und OpenSource
- Flexibel und breites Anwendungsgebiet
- Erweiterbar mit Extensions
- Einfache Anwendung

NACHTEILE

- Abhängig von lokaler Rechenleistung
- Reconciliation und LLM Abfrage abhängig von externen Services
- Kein direkter JSON Export

Weitere Ressourcen

- OpenRefine Manual: <https://openrefine.org/docs>
- OpenRefine Forum and Helpdesk: <https://forum.openrefine.org/>
- Extensions: <https://openrefine.org/extensions>
- Liste externer Ressourcen:
<https://github.com/OpenRefine/OpenRefine/wiki/External-Resources>
- OpenRefine Youtube Channel: <https://www.youtube.com/@OpenRefineProject>

Einführung in OpenRefine: Datenaufbereitung für Kulturerbe-Metadaten

Sebastian Wolff
Kulturpool – NHM Wien
03.06.2026