

Sätt igång med **OpenRefine**



 **Introduktion** 

 **Vad och varför** 

 **Steg för steg med Alicia** 

 **Eget arbete** 



 **Vi dokumenterar tillsammans** 

<https://w.wiki/85fu>



Om oss



<https://w.wiki/85fu>

Wikimedia Sverige

En ideell förening som arbetar för att alla människor ska ha fri tillgång till mänsklighetens samlade kunskap.

<https://wikimedia.se>



<https://w.wiki/85fu>

Wikimedia Sverige

Fri kunskap för en bättre värld. Att hjälpa människor och organisationer att skapa och bevara fri kunskap, och att göra den enkelt tillgänglig för alla.

<https://w.wiki/85fu>



Wikimedia Sverige

Vi hjälper kulturarvsinstitutioner jobba med Wikimediaplattformarna, Länkade Öppna Data och fri mjukvara.

Projektet *Användbara auktoriteter för datadriven samlingsforskning* med SHM och Nationalmuseum:

<https://shm.se/blog-article/slutfas-anvandbara-auktoriteter/>

<https://w.wiki/85fu>



Mjukvaran

<https://w.wiki/85fu>



OpenRefine

Ett **gratis** verktyg byggt på **öppen källkod** för att analysera och städa upp i stora datasamlingar.



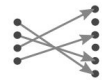
Faceting

Drill through large datasets using facets and apply operations on filtered views of your dataset.



Clustering

Fix inconsistencies by merging similar values thanks to powerful heuristics.



Reconciliation

Match your dataset to external databases via reconciliation services.



Infinite undo/redo

Rewind to any previous state of your dataset and replay your operation history on a new version of it.



Privacy

Your data is cleaned on your machine, not in some dubious data laundering cloud.



Wikibase

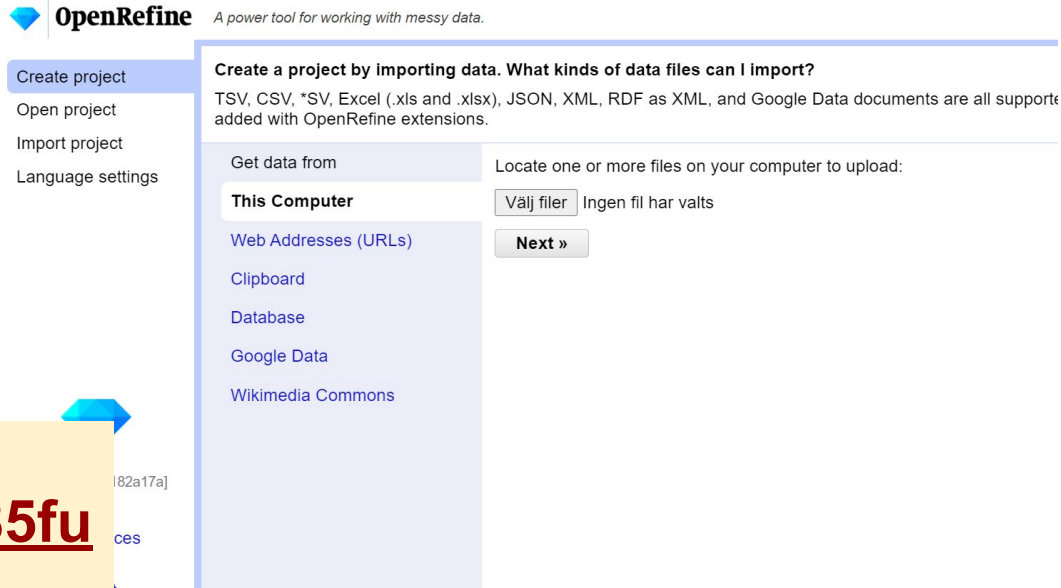
Contribute to Wikidata, the free knowledge base anyone can edit, and other Wikibase instances.

Steg för steg



<https://w.wiki/85fu>

Dataimport, projekt



OpenRefine *A power tool for working with messy data.*

Create project
Open project
Import project
Language settings

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported with OpenRefine extensions.

Get data from: Locate one or more files on your computer to upload:

- This Computer** Ingen fil har valts
- Web Addresses (URLs)
- Clipboard
- Database
- Google Data
- Wikimedia Commons

<https://w.wiki/85fu>

Första städningen

1 Re-order / Remove columns

5

st
e
s

Drag columns to re-order

Drop columns here to remove

WikidataID	sourcing circumstances
ref_url	Internal_ID
Namn	
Gender	
nationality	
Födelseår	
Dödsår	
ViafID	

<https://w.wiki/85fu>

Dataanalys

Facet / Filter Undo / Redo 35 / 35

Refresh Reset all Remove all

nationality change
6 choices Sort by: name count Cluster

- amerikansk 1
- dansk 1
- svensk 10
- Sverige 6
- tysk 9
- österrikisk 1
- (blank) 2

Facet by choice counts

30 rows

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

	Namn	Gender	nationality	Födelseår
	Karl Bergsten Choose new match	man	svensk	1869
is? tld=10045&viewType=detailView	Casper Tamm Choose new match	man	Sverige Choose new match	1874-10-20
	Ivar Kreuger	man	svensk	1880-03-02
	Knut Tillberg	man	svensk	1860-03-04
is? tld=10389&viewType=detailView	Hans Reiff Choose new match	man		1580
is? tld=10466&viewType=detailView	Clary Josephson Choose new match	kvinn	Sverige Choose new match	1872-03-13

<https://w.wiki/85fu>

Filter

Facet / Filter Undo / Redo 35 / 35

Refresh Reset all Remove all

Födelseår invert reset

186

case sensitive regular expression

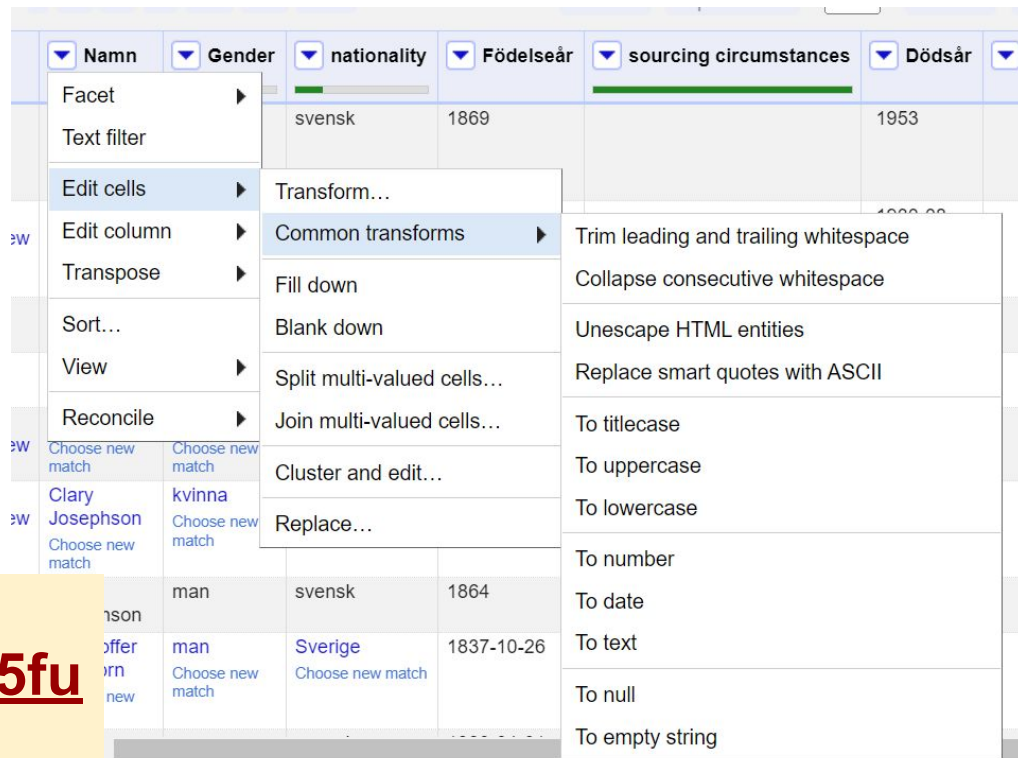
5 matching rows (30 total)

Show as: **rows** records Show: 5 **10** 25 50 100 500 1000 rows « first < p

internal_ID	ref_url	Namn	Gender	nationality	Födelseår	source
		Karl Bergsten <small>Choose new match</small>	man	svensk	1869	
7		Knut Tillberg	man	svensk	1860-03-04	
7		Erik Josephson	man	svensk	1864	
3		Henric Tamm	man	svensk	1869-05-31	
1		Carl Frisk	man	svensk	1865-11-20	

<https://w.wiki/85fu>

Enkla transformationer



The screenshot shows a data table with columns: Namn, Gender, nationality, Födelseår, sourcing circumstances, and Dödsår. A context menu is open over a cell, displaying options for editing and transforming data. The 'Common transforms' submenu is expanded, showing various text and data manipulation options.

Namn	Gender	nationality	Födelseår	sourcing circumstances	Dödsår
Clary Josephson	kvinnna	svensk	1864		
offer	man	Sverige	1837-10-26		

- Facet
- Text filter
- Edit cells
 - Transform...
- Edit column
 - Common transforms
 - Trim leading and trailing whitespace
 - Collapse consecutive whitespace
 - Unescape HTML entities
 - Replace smart quotes with ASCII
 - To titlecase
 - To uppercase
 - To lowercase
 - To number
 - To date
 - To text
 - To null
 - To empty string
 - Fill down
 - Blank down
- Sort...
- View
 - Split multi-valued cells...
- Reconcile
 - Join multi-valued cells...
 - Cluster and edit...
 - Replace...

<https://w.wiki/85fu>

Redigering

1761-01-24 1
1774 1
1785 1
1786 1
1797 1
1798-07-29 1
1801 1
1803 1
1804 1
1805 1

The screenshot shows a web application interface with a list of dates on the left and a modal dialog for editing a date. The modal dialog is open over the date '1798-07-29'. The dialog contains a text input field with the date '1798-07-29' entered, and two buttons: 'Apply' and 'Cancel'. Below the buttons are the keyboard shortcuts 'Enter' and 'Esc'. The background shows a table with columns for dates and names, and a search bar at the top.

Date	Name
1761-01-24	
1774	
1785	
1786	
1797	
1798-07-29	Casper Tamm
1801	
1803	
1804	
1805	

Search: /eMuseumPlus?&module=artist&objectId=10045&viewType=detailView

Choose new match

Apply Cancel

Enter Esc

1798-07-29

1798-07-29

Choose new match

Ivar

Choose new match

Josephson

Choose new match

Erik

<https://w.wiki/85fu>

Dela – slå ihop

Split column Födelseår into several columns

How to split column

by separator

Separator regular expression

Split into columns at most (leave blank for no limit)

by field lengths

List of integers separated by commas, e.g., 5, 7, 15

After Splitting

Guess cell type

Remove this column

OK Cancel

<https://w.wiki/85fu>

Fram och tillbaka

The screenshot shows a software interface for data transformation. At the top, there are tabs for 'Facet / Filter' and 'Undo / Redo 35 / 35', along with a '30 rows' indicator. Below these are buttons for 'Extract...' and 'Apply...'. A 'Filter' input field is present. The main area displays a list of 9 steps:

0. Create project
1. Text transform on 30 cells in column Internal_ID: value.toString()
2. Text transform on 30 cells in column Internal_ID: grel:value.split(".")[0]
3. Text transform on 22 cells in column WikidataID: grel:value.split("/")[-1]
4. Use values as reconciliation identifiers for 22 cells in column WikidataID
5. Save Wikibase schema
6. Perform Wikibase edits
7. Reconcile cells in column Namn to type Q5
8. Star row 1
9. Reconcile cells in column Namn to type Q5

On the right side, there are partial labels for rows: 'Show as: rc', '/eMuseumPlu:', 'le=artist&objectl', and 'le=artist&objectl'.

<https://w.wiki/85fu>

Dataexport

Export custom tabular

Content Download Upload Option Code

Select and order columns to export

Options for **WikidataID**

For reconciled cells, output

WikidataID

Internal_ID

ref_url

Namn

Matched entity's name

Matched entity's ID

Link to matched entity's page

Cell's content

Output nothing for unmatched cells

ISO 8601, e.g., 2011-08-24T18:36:10+08:00

Short locale format

Medium locale format

Long locale format

Full locale format

Custom [Help](#)

Use local time zone

Omit time

tances

<https://w.wiki/85fu>

GREL



<https://w.wiki/85fu>

General Refine Expression Language

Ett skriptspråk som påminner om **JavaScript**.

Bygger på **variabler** och möjliggör komplex datamanipulation.

<https://w.wiki/85fu>

General Refine Expression Language

Resurser:

<https://openrefine.org/docs/manual/grel>

<https://github.com/OpenRefine/OpenRefine/wiki/Recipes>

<https://w.wiki/85fu>

Delning

Custom text transform on column Födelseår

Expression Language General Refine Expression Language (GREL) ▾

```
value.split("-")[0]
```

No syntax error.

Preview History Starred Help

row	value	value.split("-")[0]
1.	1869	1869
2	1874-10-20	1874
	380-03-02	1880
	360-03-04	1860
	580	1580

<https://w.wiki/85fu>

Ersättning

Custom text transform on column Födelseår

Expression

Language

General Refine Expression Language (GREL) ▾

```
value.replace("-", ", ")
```

No syntax err

Preview

History

Starred

Help

row	value	value.replace("-", ", ")
-----	-------	--------------------------

1.	1869	1869
----	------	------

	1874-10-20	1874, 10, 20
--	------------	--------------

	1880-03-02	1880, 03, 02
--	------------	--------------

	1860-03-04	1860, 03, 04
--	------------	--------------

	1880	1580
--	------	------

<https://w.wiki/85fu>

Komma åt andra celler

Custom text transform on column Namn

Expression Language No syntax error

```
value + ", " + cells["nationality"].value
```

Preview History Starred Help

row	value	value + ", " + cells["national ...
1.	Karl Bergsten	Karl Bergsten, svensk
	Casper Tamm	Casper Tamm, Sverige
	Ivar Kreuger	Ivar Kreuger, svensk
	Knut Tillberg	Knut Tillberg, svensk

<https://w.wiki/85fu>

Kopiera från annat projekt

Add column based on column archivesID

New column name

On error set to blank store error copy value from original column

Expression Language No syntax error.

Preview History Starred Help

row	value	cell.cross("archives_sample", ...
1.	232	https://aspace.myinstitution.edu/public/repositories/2/resources/232/
2.	12	https://aspace.myinstitution.edu/public/repositories/2/resources/12/
		https://aspace.myinstitution.edu/public/repositories/2/resources/8593/
		Error: java.lang.IndexOutOfBoundsException: Index: 0, Size: 0
		https://aspace.myinstitution.edu/public/repositories/2/resources/934/
		https://aspace.myinstitution.edu/public/repositories/2/resources/546
		Error: java.lang.IndexOutOfBoundsException: Index: 0, Size: 0

<https://w.wiki/85fu>

Eget arbete



<https://w.wiki/85fu>

TACK!



WIKIMEDIA
SVERIGE