

# OpenRefine

<https://librarycarpentry.org/lc-open-refine/>

# Today's tools and references

- Zoom interaction
  - Chat, React, and Speak!
- These slides
- OpenRefine software
- The "doaj-article-sample.csv" data file
- The "OpenRefine" lesson

**What is OpenRefine?**



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. Support for other formats can be added with OpenRefine extensions.

Get data from

Locate one or more files on your computer to upload:

**This Computer**

Choose Files no files selected

Web Addresses (URLs)

Next »

Clipboard

Database

Google Data



Version 3.4.1 [437dc4d]

Preferences

Help

About

# Resolve inconsistencies

## Data you have

London

London]

London,]

london

## Desired data

London

London

London

London

# Split data

Address in single field	Institution	Library name	Address 1	Address 2	Town/City	Region	Country	Postcode
University of Wales, Llyfrgell Thomas Parry Library, Llanbadarn Fawr, ABERYSTWYTH, Ceredigion, SY23 3AS, United Kingdom	University of Wales	Llyfrgell Thomas Parry Library	Llanbadarn Fawr		Aberystwyth	Ceredigion	United Kingdom	SY23 3AS

# Pull data from other sources

<b>Data you have</b>	<b>Date of Birth from VIAF (Virtual International Authority File)</b>	<b>Date of Death from VIAF (Virtual International Authority File)</b>
Braddon, M. E. (Mary Elizabeth)	1835	1915
Rossetti, William Michael	1829	1919
Prest, Thomas Peckett	1810	1879

# Our topics

- Importing data into OpenRefine
- Layout of OpenRefine
- Faceting and filtering
- Clustering
- Transformations
- Exporting datasets
- Advanced OpenRefine functions



# Importing data into OpenRefine

Start a project.

# Layout of OpenRefine

Separate records from rows.

# **Faceting and filtering**

Group same values together to filter and edit data.

# Exercise

Use a text facet for the "Licence" column to answer these questions:

- What is the most common License in the file?
- How many articles in the file don't have a license assigned?

# Exercise

Create a Text facet on the "Language" column and correct the variation in the EN and English values.

# Clustering

Group together similar values to identify inconsistencies.

# Transformations

Use the power of GREL.

# Exercise

Use Facets and GREL to put the titles in Title Case:

1. Facet by the "Publisher" column.
2. Select "Akshantala Enterprises" and "Society of Pharmaceutical Technocrats."
  - a. Tip: Select multiple values in the facet with the `include` link.
3. Click the dropdown menu on the "Title" column.
4. Choose: `Edit cells > Transform...`
5. In the Expression box type the function: `value.toTitlecase()`
6. Look at the Preview pane to see the effect.
7. Click OK



# OpenRefine data types

- String
- Number
- Date
- Boolean
- Array

# **Exporting data**

Save to one of many formats.

# Advanced OpenRefine functions

<https://librarycarpentry.org/lc-open-refine/13-looking-up-data/index.html>