

Cleaning and exploring your data with Open Refine

An Intersect Training Course – Duration ~2.5hrs

Course Outline

Open Refine is a powerful free tool for exploring, normalising and cleaning up datasets. In this tutorial we'll work through the various features of Refine, including importing data, faceting, clustering, and calling into remote APIs, by working on a fictional but plausible humanities research project.

We will start with a research question in mind and use the features of Refine to gain insights and to find answers. The research question relates to NSW police stations — finding out what we can about where they are located, their heritage status, and the kinds of archival records State Records NSW holds on them.

During this course you will be shown how to get map coordinates by creating direct batch queries to Google maps.

Possible uses

- Cleaning up messy data: If you have a text file with semi-structured data, you can edit it using facets, transformations and clustering to make the data cleanly structured.
- Transformation of data: Converting data to different formats, normalising and the de-normalising of data.
- Parsing of data from web sites: Open Refine contains a URL fetch feature.
- Adding data to your dataset by fetching it from web-services with json: A good example of this is used for geocoding addresses to geographic co-ordinates.

What will be covered?

Participants will learn:

- How to install Open Refine
- How to create a project
- How to organise your data
- Ways to explore your data
- How to work with APIs
- How to export a dataset