**MASON — Cleaning**

<0 figures>

**Cleaning and Exploring Your Data with Open Refine**

Mason, I.

This three-hour workshop introduces participants to Open Refine, which is a powerful tool for cleaning, normalization, and exploration of 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.

**Contact Information**

Ingrid Mason (on behalf of Intersect Australia, tutorial instructors TBC)

ingrid.mason@intersect.org.au http://intersect.org.au/contact

**Research Interests**

eResearch, digital humanities

**Area of Expertise**

The eResearch analysts and software developers at Intersect Australia deliver training and tutorials to support uptake of ubiquitous and domain-specific technologies employed in data-intensive research. This training is regularly delivered to academics and higher degree students in its member universities in Australia. A common area of expertise shared across the team is data, information and technology management, informatics and data modelling, through to technical evaluation and software engineering to enable the development, application, and uptake of technologies in support of research.

**Target Audience**

Attendees accustomed to using ubiquitous tools on their desktop computer, e.g., Excel, and interested in building up further skills to process their data. Mastering tasks such as

• 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 websites: 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 coordinates.

**Participant Number**

30 participants

**Technical Requirements**

Attendees bring a laptop they have administration access to (able to download software) and robust wireless must be available.

**Course Outline**

Open Refine is a powerful free tool for exploring, normalizing, and cleaning up datasets. In this 2.5-hour 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 that 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.

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.

For more information, see http://intersect.org.au/course-­resources.