

# **OpenRefine in 10 minutes**

**Richard Vankoningsveld, Technology  
and Support Librarian**

**May 2018**

# What is OpenRefine?

- A free, open source, powerful tool for working with messy data
- An 'alternative' to Excel
- A great tool for exploring tabular data
- Also a great tool for cleaning up that data

# OpenRefine history and help

- Developed by Google in 2010
- Now open source and maintained by volunteers
- Web: <http://openrefine.org/>
- Installers for Win, Mac,  
Linux: <http://openrefine.org/download.html>
- Codebase: <https://github.com/OpenRefine/OpenRefine>
- Contribution guide: <http://openrefine.org/community.html>

# What can OpenRefine do?

- Import data in various formats
  - Excel, CSV, JSON, XML and more
- Analyse data
  - Faceting, searching
- Edit data
  - Cluster similar values; transform cells and columns
- Extend data
  - From web/database apis
- Export data
- Automate repeat jobs

# Why use OpenRefine?

- Easier and more reliable than Excel for data work
- Can handle very large datasets
- Doesn't change your data without telling you
- Supports regular expressions
- Jython/Python support

# Potential use cases

- Catalogue data checking/cleaning/editing
- Log data exploration/analysis
- Converting data formats

# Want to learn more?

- Library Carpentry  
lesson: <https://librarycarpentry.github.io/lc-open-refine/>
- Contact me:
  - @richyvk
  - richard.vankoningsveld@sclqld.org.au