OpenRefine

What is OpenRefine?

*"A powerful **tool** for working with messy data"*

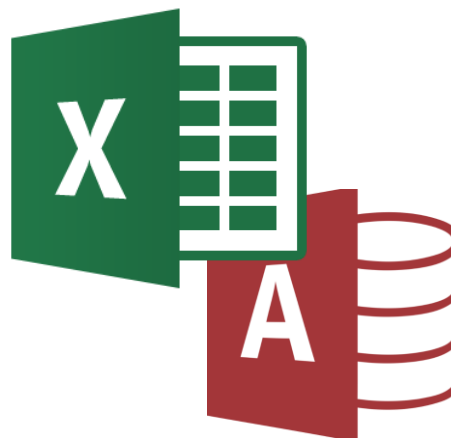Faceting

Clustering

Reconciling

What it is not?

Database

Excel-like experience

Feature comparison

Database

- Infrastructure and programing skills to edit
- Absence of easy visualization

Excel

- Usually one cell editing
- Useful to document data and perform operations
- Data is not always visible
- Deficient visualization

OpenRefine

- Multiple cell editing
- Easy exploration and transformation
- Interactive visualization

Feature comparison

OpenRefine

- Multiple cell editing
- Easy exploration and transformation
- Interactive visualization

- Free
- Desktop app (offline)
- Infinite undo function
- Use of APIs
- Several type files to export/import

- Faceting / Filters
- Use of regular expressions
- Extensions
- Large community of developers (extensions, tutorials, etc)

Practical session

DATA CLEANING AND
STANDARDIZATION

Conventions

- Formulas (copy-paste)

  `Cell.recon.match.id`

- Commands in OpenRefine

  Edit column

- Column names

  nameRecon

- Useful hyperlinks

  www.gbif.org

- Column menu

Questions ?

Exercises