

Cleaning data with OpenRefine



To download, to watch introductory videos:

<http://openrefine.org/>

To find documentation, FAQ, how-to:

<https://github.com/OpenRefine/OpenRefine/wiki>

What can Refine do for you?

- Clean data
 - Typos
 - Duplicates
 - Standardize data for mapping to relational database
- Enhance data
 - Georeferencing
 - Taxonomy
- Explore data

What kind of file can I upload?

- TSV, CSV, *SV, Excel (.xls and .xlsx),
- JSON
- XML
- RDF as XML
- Wiki markup
- Google Data document

Tips and Tutorials (Part 1)

Openrefine presentation by Deb Paul, from iDigBio Georeferencing Train-the-trainers 2 (Day5), includes PPT and sample CSV:

https://www.idigbio.org/wiki/index.php/GWG_Second_Train_the_Trainers_Workshop

<https://www.idigbio.org/sites/default/files/workshop-presentations/ttt2/OpenRefineTTT22013.pptx>

<https://www.idigbio.org/sites/default/files/workshop-presentations/ttt2/OpenRefineTTT2Demo.csv>

Entertaining, informative ppt and other tips and links by journalist Tom Meagher:

<http://tinyurl.com/jwpfu3s>

<http://blog.tommeagher.com/2012/06/clean-data-is-the-best-weapon-against-the-planet-of-the-apes/>

<http://blog.tommeagher.com/2013/02/more-tips-for-using-openrefine/>

An OpenRefine Walkthrough by Javier Otegui, UC Boulder:

http://bit.ly/BITW13_OpenRefine

Tips and Tutorials (Part 2)

Getting started with OpenRefine and some recipes for spreadsheets, by Angella Thompson and Liath Appleton at UT Austin:

<https://wikis.utexas.edu/pages/viewpage.action?pageId=46631837>

<https://wikis.utexas.edu/display/specify6/Google+Refine+recipes+for+spreadsheet+prep>

Tutorial, with associated CSV:

http://enipedia.tudelft.nl/wiki/OpenRefine_Tutorial

Tutorial/walkthrough

<http://schoolofdata.org/handbook/recipes/cleaning-data-with-refine/>

Website

<http://freeyourmetadata.org/>

How to get lat-long from google, how to parse JSON:

<http://blog.ouseful.info/2013/02/20/geocoding-using-the-google-maps-geocoder-via-openrefine/>