

LAS 6292: QA/QC 2 - OpenRefine

updated: 2021-02-24

Objectives and Competencies:

OpenRefine (formerly Google Refine) is a powerful, free, and open source tool that is used to work with and clean messy data. By the end of this lesson students will:

1. Be able to import a dataset into OpenRefine, make changes to the dataset and its structure, and export the revised Dataset.
2. Learn how to automatically track changes made and export the record of changes

Pre-class Preparation (Students):

1. Install OpenRefine on your computer and verify it works by following the [instructions here](#).
2. Scan the [basic workflow](#) for what we will be doing.
3. Read about OpenRefine and see some quick videos of [how it works here](#)

Lessons & Resources Used in Class

1. The Data Carpentry Lesson we did with Dr. Ye: [OpenRefine for the Social Sciences](#)
2. The repository with the data and metadata used in the lesson: [SAFI_Survey](#)
3. URL for accessing OpenRefine if a new tab/window doesn't open: <http://127.0.0.1:3333/>
4. Guides for writing OpenRefine commands with General Refine Expression Language
 - [OpenRefine GREL](#)
 - a great [GREL Guide](#) from the Univ Illinois
 - EVEN BETTER: [OpenRefine cheatsheets](#), including GREL commands

Other OpenRefine tools and tutorials

1. [Cleaning Data with OpenRefine](#): a tutorial from *The Programming Historian* website
2. Data Carpentry Lesson: [OpenRefine for Ecologists](#) # Sources
3. Cleaning Data with OpenRefine [Video Tutorial No. 1](#) and [Video Tutorial No. 2](#)

UF Library Workshops

1. The schedule of [UF Library Workshops](#)
 - [Tidy Tuesday with Hao](#): practice data manipulation, clean-up, and analysis using the tidyverse with Dr. Hao Ye.
 - An [upcoming OpenRefine workshop](#) by Dr. Hao Ye (March 30,2021)