Introduction to data cleaning

Vessela Ensberg

Associate Director, Data Management Program

How you don't want your data to be

Location	Soil sample weight	Species	Count	Coordinates
Dutton Island	256 g	Great egret	7	38.076976, -121.972561
Honker Bay	0.356 kg	Heron	9	38.070764, -121.916514
New York Slough	345 g	Goliath heron	3	38°01'55.2"N 121°51'40.2"W

Goals of today's workshop

Start using open refine

Clean data in a reproducible way

UCDAVIS
LIBRARY

References

- http://www.datacarpentry.org/OpenRefine-ecology-lesson/
- https://data-lessons.github.io/library-openrefine/
- Verborgh, Ruben, and Wilde, Max De. Using OpenRefine, Packt Publishing, 2013. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/ucdavis/detail.action?docID=1389316. (free eBook access for UC Davis users through the library)