

---

# INTRODUCTION TO DATA CLEANING

WITH OPENREFINE & WORD CLOUDS

Meg Miller, GIS & Data Visualization Librarian  
Natalie Vielfaure, Digital Curation Archivist  
University of Manitoba Libraries



# SESSION BREAKDOWN



Software Overview



Data Cleaning: What is it?



Benefits



Hands-on Exercises



OpenRefine

What is it?  
Uses



Voyant Tools

What is it?  
Word Clouds

## SOFTWARE OVERVIEW



## DATA CLEANING: WHAT IS IT?

- Improving the overall quality of your data.

Manitoba

MB

Man.

Manitoba

Manitoab

mb

## RESOLVING INCONSISTENCIES

- Variant spelling
- Inconsistent case
- Duplication

# FORMATTING



UNIT OF MEASUREMENTS



DATES, TEXT, NUMBERS



COLUMNS / ROWS



FORMAT MIGRATION

# NULL VALUES

- No value
- 0 value
- Errors
  - Examples:
    - =MODE.MULT(A2:G2) → #NULL!
    - =SUM(A2:A5) → #NAME?



## DATA CLEANING: BENEFITS

**Efficient  
research**

**Easier  
conversion**

**Secondary  
use**



**[bit.ly/orf\\_2022](https://bit.ly/orf_2022)**

**HANDS ON EXERCISES**