

OpenRefine Steps

Creating a Project

1. Open OpenRefine.
2. Click “Create Project” in the left menu.
3. In the center menu, select “Web Addresses (URLs).” Then copy the URL to the “Circulation Data” spreadsheet on the LibGuide and paste it in the URL blank in OpenRefine. (NOTE: To copy the link, right-click the file name and choose “Copy Link Address.”)

Refine A power tool for working with messy data.

Create Project
Open Project
Import Project
Language Settings

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Sup formats can be added with OpenRefine extensions.

Get data from: Enter one or more web addresses (URLs) pointing to data to download:

Add Another URL Next >

Web Addresses (URLs)
Clipboard
Google Data

4. Press “Next” underneath the URL you pasted to go to the next steps.
5. On the new page, select “Excel files” under **Parse data as**. From here, you should see the data on display in a preview window.

Refine A power tool for working with messy data.

Create Project
Open Project
Import Project
Language Settings

« Start Over Configure Parsing Options Project name: ld.php Create Project »

CHARGE_LOCATION_NAME	TITLE	PATRON_GROUP_NAME	GOV_ITEM_TYPE	CHARGE_DATE_TIME
* AFA Circulation Desk	9H on rigor / edited by Richard Burdett and Wilfried Wang.	Undergraduate Student	2 Hour Reserve	2016-10-13T18:19:09Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-10-09T15:22:52Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-11-02T19:15:53Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-11-03T12:33:10Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-12-01T00:56:10Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-02T11:49:58Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-15T12:35:41Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-22T10:34:02Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-23T14:47:11Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-26T20:19:32Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-10-04T18:31:15Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-10-06T14:14:15Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-10-07T16:52:53Z
* SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-10-11T12:14:11Z

Worksheets to Import

Worksheet name	Rows	Parse next	Ignore first	Store blank cells as nulls	Store file source
_Circ Stats by Library_Patron	23451 rows	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Line-based text files
CSV / TSV / separator-based files
Fixed-width field text files
PC-Axis text files
JSON files
MARC files
RDF/N3 files
XML files
Open Document Format spreadsheets (.ods)
RDF/XML files
Excel files

Version 2.6-rc-2 [TRUNK]
Help
About

- Now, in the top right, give the project a name. (I will use “utacircdata.”) Then press “Create Project.”
- Now that you have created a project, OpenRefine will display the first 10 rows while also displaying the total number of rows in large font at the top.

 utacircdata [Permalink](#)

Undo / Redo 0

23450 rows

Show as: **rows** records Show: 5 10 25 50 rows

<input type="checkbox"/> All	<input type="checkbox"/> CHARGE_LOCATION	<input type="checkbox"/> TITLE
<input type="checkbox"/>	1. * AFA Circulation Desk	9H on rigor / edited by Richard Burdett and Wilfried
<input type="checkbox"/>	2. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	3. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	4. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	5. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	6. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	7. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	8. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	9. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part
<input type="checkbox"/>	10. * SEL Circulation Desk	AASHTO: LRFD Bridge Design Specifications (Part

Facets and filters



Facets and filters to select subsets of data to act on. Choose facet and modes from the menus at the top of the data column.

How to get started?
[View screencasts](#)

Cleaning Library Names and Headers

1. Next to **CHARGE_LOCATION_NAME**, click the dropdown arrow.
2. Select “Facet” then “Text Facet.” All unique values for that column will show on the left.

The screenshot shows the Refine interface for a dataset named 'utacircdata'. The main table has 23450 rows. The 'CHARGE_LOCATION_NAME' column is selected, and its facet menu is open. The menu options include 'Facet', 'Text filter', 'Edit cells', 'Edit column', 'Transpose', 'Sort...', 'View', and 'Reconcile'. The 'Facet' option is selected, and a sub-menu is open showing 'Text facet' as the chosen option. On the left, the facet results show 3 choices: 'AFA Circulation Desk' (2630), 'CEN Circulation Desk' (16957), and 'SEL Circulation Desk' (3863).

3. Change the library branches to their full names: Architecture & Fine Arts Library, Central Library, and Science & Engineering Library. To do this, hover your mouse over a name, and click “edit” which appears next to it. Then type the name to replace it.
4. Next, press the down arrow by **CHARGE_LOCATION_NAME** again.
5. This time, click “Edit Column” and “Rename this column.” In the blank that appears, type “Location Name.”
6. Do the same for **CHARGE_DATE_TIME**. Rename it to “Charge Date.”

All	Location Name	TITLE	PATRON_GROUP_I	GOV_ITEM_TYP	Charge Date
1.	Architecture & Fine Arts Library	9H on rigor / edited by Richard Burdett and Wilfried Wang.	Undergraduate Student	2 Hour Reserve	2016-10-13T18:19:09Z
2.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-10-09T15:22:52Z
3.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-11-02T19:15:53Z
4.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-11-03T12:33:10Z
5.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Faculty	2 Hour Reserve	2016-12-01T00:56:10Z
6.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-02T11:49:58Z
7.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-15T12:35:41Z
8.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-22T10:34:02Z
9.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-23T14:47:11Z
10.	Science & Engineering Library	AASHTO: LRFD Bridge Design Specifications (Part I)	Graduate Student	2 Hour Reserve	2016-09-26T20:19:32Z

How your data should look at this point.

Grouping Reserve Items

- In the column **GOV_ITEM_TYPE**, click the down arrow and select “Edit column” then “Add column based on this column.”

ows « first < previous 1 - 10 next >

	PATRON_GROUP	GOV_ITEM_TYP	Reserve Length
edited by Richard Burdett and Wilfried	Undergraduate Student	Facet	2 Hour Reserve
D Bridge Design Specifications (Part	Faculty	Text filter	2 Hour Reserve
D Bridge Design Specifications (Part	Faculty	Edit cells	2 Hour Reserve
D Bridge De		Split into several columns...	2 Hour Reserve
D Bridge De		Add column based on this column...	2 Hour Reserve
D Bridge De		Add column by fetching URLs...	2 Hour Reserve
D Bridge De		Rename this column	2 Hour Reserve
D Bridge De		Remove this column	2 Hour Reserve
D Bridge De		Move column to beginning	2 Hour Reserve
D Bridge De		Move column to end	2 Hour Reserve
D Bridge De		Move column left	2 Hour Reserve
D Bridge De		Move column right	2 Hour Reserve

- Give the new column the title “Reserve Length” next to **New Column Name**.
- In the expression, type `chomp(value, " Reserve")`

Add column based on column GOV_ITEM_TYPE

New column name:

set to blank
 store error
 copy value from original column

Expression: Language: No syntax error.

Preview History Starred Help

row	value	chomp(value, " Reserve")
1.	2 Hour Reserve	2 Hour
2.	2 Hour Reserve	2 Hour
3.	2 Hour Reserve	2 Hour
4.	2 Hour Reserve	2 Hour
5.	2 Hour Reserve	2 Hour
6.	2 Hour Reserve	2 Hour
7.	2 Hour Reserve	2 Hour

OK Cancel

10. The Preview should display the checkout time for reserve items. If it looks like the image above, press "Okay."
11. In the column **GOV_ITEM_TYPE**, click the down arrow and select "Facet" then "Text Facet." All unique values for that column will show on the left.
12. Next, you are going to change all of the different type of Reserve items (e.g., 2 Hour Reserve and 24 Hour Reserve) to Reserve. To do this, click "Cluster."
13. In the pop-up menu, change **Method** to "Nearest Neighbor." Then, next to the two clusters, type Reserve.

Cluster & Edit column "Reserve Length"

This feature helps you find groups of different cell values that might be alternative representations of the same concept. For example, "New York" and "New York" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Gödel" are also likely to refer to the same concept.

Method Distance Function Radius Block C

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
3	2628	<ul style="list-style-type: none"> • 2 Hour Reserve (1617 rows) • 24 Hour Reserve (698 rows) • 3 Hour Reserve (313 rows) 	<input type="checkbox"/>	<input type="text" value="Reserve"/>
2	137	<ul style="list-style-type: none"> • 7 Day Reserve (93 rows) • 3 Day Reserve (44 rows) 	<input type="checkbox"/>	<input type="text" value="Reserve"/>

14. Check the boxes under **Merge?**, then press "Merge Selected & Close."
15. Now, when viewing facets in **GOV_ITEM_TYPE**, Reserve should be a facet.

GOV_ITEM_TYPE change

22 choices Sort by: **name** count Cluster

- Microsoft Surface 846
- Non-Circulating 14
- Periodical 180
- Popular Reading 351
- Reserve 2765 edit include
- Robot 22
- Score 26
- Tech Lending Items 35
- Video 339
- Webcams 110

Facet by choice counts

This will create one large Reserve category.

Exporting Data

16. Click "Export" in the top right. Choose "Excel (.xls)." This will save the spreadsheet to your computer.