

# Three Weird Tricks For Managing Data

Ideas from Marmot



# First Weird Trick

Add Loading Notes to MARC Records



## Add a note to MARC records about how and when they were loaded

- Which load profile loaded a particular record?
- When was it loaded? When was it overlaid?
- Which library did the load?

## Add a line to each load profile that inserts a note into each record

- Put a different note in each profile to identify that loader in the records it creates or overlays.
- This is the line:
  - |995|T|0|0|b|m|0|y|N|0| |aLoaded with m2btab.sky in this month
- NOTE: That is the only line creating a 995. Do not have another line loading incoming 995s. If you did that you would duplicate them all if one of your records were reloaded, such as from an authority vendor.
- That will insert the note “Loaded with m2btab.sky in this month” into the 995 of every record it loads.
- You can use a different field if you have stuff in your 995s.

# Now you will get notes like this in your records

Bib Record · b57218766

File Edit View Go Tools Reports Window Help

Insert Save View Edit Summary Verify Export Print Close

**b57218766** Last Updated: 01-24-2018 Created: 01-03-2018 Revisions: 4

runs away and eventually finds shelter with Fish, the old Kickapoo Indian who camps by the river. Wishing Caswell Dead is a historical mystery about the evil that hides within a village, one girl who is determined to save herself and her child, and a violent murder no one wants to solve"--|c  
Provided by publisher.

SUBJECT 655 7 Detective and mystery fiction.|2lcgft

OTHER TITLE (q) 776 0 8 |iOnline version:|aStoltey, Patricia, author.|tWishing Caswell dead|bFirst edition.|d  
Waterville, Maine : Five Star Publishing, a part of Cengage Learning, Inc., [2017]|z9781432834388|w  
(DLC) 2017026815

LOAD NOTE 995 Loaded with m2btabsky in this month ←

b57218766 Edit Mode (INS)

## Protect the note from overlay

- Protect the field
  - @ov\_protect="b=Vm(995)"
- Now it will create a new note with each overlay.
- This creates a history of this record.

# Each overlay adds another one because of @ov\_protect

Bib Record · b57218766

File Edit View Go Tools Reports Window Help

Insert Save Save All View Edit Summary Verify Export Print Close

b57218766 Last Updated: 01-24-2018 Created: 01-03-2018 Revisions: 4

Wishing Caswell Dead is a historical mystery about the evil that hides within a village, one girl who is determined to save herself and her child, and a violent murder no one wants to solve"--|c  
Provided by publisher.

SUBJECT 655 7 Detective and mystery fiction.|2lcgft

OTHER TITLE (q) 776 0 8 |iOnline version:|aStoltey, Patricia, author.|tWishing Caswell dead|bFirst edition.|d  
Waterville, Maine : Five Star Publishing, a part of Cengage Learning, Inc., [2017]|z9781432834388|w  
(DLC) 2017026815

LOAD NOTE 995 Loaded with m2btab.ltiac in this month ←

LOAD NOTE 995 Loaded with m2btab.sky in this month ←

b57218766 Edit Mode (INS)

“In this month” will be wrong next month

- Do a Global Update to fix it on the first of every month.





On the first day of each month, gather the new records

Boolean Search

Review File Name:

Store Record Type:

Range

Classic

Enhanced

JSON

Term	Operator	Type	Field	Condition	Value A	Value B
1		BIBLIOGRAPHIC	MARC Tag 995	has	in this month	

BIBLIOGRAPHIC MARC Tag 995 has "in this month"

Group

Ungroup

Insert Line

Search Use Existing Search Retrieve Saved Query Save Save As Close



# Global Update to change “this month” to last month’s date

Change Variable-length Field(Editing Command) ✕

Change field

Use Displayed Field  Match Case

Match Whole Field  Match Whole Subfield

	Field Group Tag	MARC Tag	Ind1	Ind2	Data
Find	<any>	995	<any>	<any>	this month
Replace	<no change>	<no change>	<no change>	<no change>	2018.01

# Now they look like this

Bib Record - b57218766

File Edit View Go Tools Reports Window Help

Insert Save Save All View Edit Summary Verify Export Print Close

**b57218766** Last Updated: 01-25-2018 Created: 01-03-2018 Revisions: 5

into prostitution by Caswell, Jo Mae discovers she is pregnant and vows to escape. when Caswell is injured by a near lightning hit, he becomes more dangerous, and more hated. The flawed residents of the Village of Sangamon harbor many secrets. Caswell knows them all. Will he tell? Jo Mae runs away and eventually finds shelter with Fish, the old Kickapoo Indian who camps by the river. Wishing Caswell Dead is a historical mystery about the evil that hides within a village, one girl who is determined to save herself and her child, and a violent murder no one wants to solve"--lc  
Provided by publisher.

SUBJECT	655		7	Detective and mystery fiction.  2 c gft
OTHER TITLE (q)	776	0	8	iOnline version:  aStoltey, Patricia, author.  tWishing Caswell dead  bFirst edition.  d Waterville, Maine : Five Star Publishing, a part of Cengage Learning, Inc., [2017]  z9781432834388  w(DLC) 2017026815
LOAD NOTE	995			Loaded with m2btab.ltiac in 2018.01
LOAD NOTE	995			Loaded with m2btab.sky in 2018.01

b57218766 Edit Mode (INS)

# Eventually a record may look like this

Bib Record · b53642004

File Edit View Go Tools Reports Window Help

Insert Save View Edit Summary Verify Export Print Close

**b53642004** Last Updated: 01-24-2018 Created: 02-06-2017 Revisions: 807

LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.ltiac in this month
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.ltiac in 2017.10
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.ing in 2017.09
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.ltiac in 2017.09
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.b in 2017.04
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.b in 2017.03
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.b in 2017.02
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.b in 2017.02
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.b in 2017.02
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.ing in 2017.02
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.ven in 2017.02
LOAD NOTE	995	<input type="checkbox"/>	<input type="checkbox"/>	Loaded with m2btab.click in 2017.09

b53642004 Edit Mode (INS)

## How do you know who loaded it?

- OCLC puts a code in the 994 field.
- Use a translation map to make it readable.

# The 994 from OCLC

Bib Record - b53642004

File Edit View Go Tools Reports Window Help

Insert Save View Edit Summary Verify Export Print Close

b53642004 Last Updated: 01-24-2018 Created: 02-06-2017 Revisions: 807

MISC	908			.o10503468
MISC	994			C0 bCDF
LOAD NOTE	995			Loaded with m2btab.ltiac in this month
LOAD NOTE	995			Loaded with m2btab.ltiac in 2017.10
LOAD NOTE	995			Loaded with m2btab.ing in 2017.09
LOAD NOTE	995			Loaded with m2btab.ltiac in 2017.09
LOAD NOTE	995			Loaded with m2btab.b in 2017.04
LOAD NOTE	995			Loaded with m2btab.b in 2017.03
LOAD NOTE	995			Loaded with m2btab.b in 2017.02

b53642004 Edit Mode (INS)

## Change the line in the loader to call a translation map

- 994|995|+|0|0|b|m|0|y|N|0|%map="m2bmap.id“
- Take the 994 and create a 995 based on the translation map.
- This replaces the note line for OCLC loaders.



## Here's the map

@delimiter=~

@case=true

@stop\_on\_map=true

|aC0|bCOM~|aExported from Connexion by CMU and loaded with m2btabs.b in this month

|aC0|bBV#~|aExported from Connexion by Salida and loaded with m2btabs.b in this month

|aC0|bCDF~|aExported from Connexion by FLC Reed and loaded with m2btabs.b in this month

|a92|bCOM~|aOCLC offline update by CMU and loaded with m2btabs.b in this month

|a92|bBV#~|aOCLC offline update by Salida and loaded with m2btabs.b in this month

|a92|bCDF~|aOCLC offline update by FLC Reed and loaded with m2btabs.b in this month

.\*~|aLoaded with m2btabs.b in this month





## Different maps for each loader

- The map includes the name of the loader, so you need a different one for each loader that loads OCLC records.



# Now the note tells you who did the load

Bib Record · b23292635


File Edit View Go Tools Reports Window Help

Insert Save Save All View Edit Summary Verify Export Print Close

**b23292635** Last Updated: 01-25-2018 Created: 11-19-2003 Revisions: 1594

MISC	938			YBP Library Services bYANK n1596717
MISC	994			C0 bHFU
LOAD NOTE	995	0	0	Exported from Connexion by Pitkin and loaded with m2btabs.b in this month
LOAD NOTE	995			Loaded with m2btabs.ltiac in this month
LOAD NOTE	995			Loaded with m2btabs.b in 2017.12
LOAD NOTE	995			Loaded with m2btabs.b in 2017.10
LOAD NOTE	995			Loaded with m2btabs.ltiac in 2017.08
LOAD NOTE	995			Loaded with m2btabs.ing in 2017.07
LOAD NOTE	995			...

b23292635 Edit Mode (OVR)



## Put the 995 in it's own Field Group

- You can toggle by Field Group in Global Update to see statistics.



# Breakdown in Global Update like this:

Sierra · Marmot Library Network · Lloyd Chittenden

File Edit View Go Tools Reports Admin Window Help

sierra FUNCTION Global Update

Edit Remove Print

1. Select records

2. Command input

3. Preview

Statistics

1. Select records

Record Type

BIBLIOGRAPHIC  AUTHORITY  PATRON  ITEM  ORDER  CHECKIN  COURSE

Review Review file: 490. lhc new records (23) (BIBLIOGRAPHIC) Search

	#		LOAD NOTE	# ENTRIES
<input checked="" type="checkbox"/>	1	b	99500 aExported from Connexion by Bud Werner and loaded with m2btab.b in this month	5
<input checked="" type="checkbox"/>	2	b	99500 aExported from Connexion by CCU and loaded with m2btab.b in this month	1
<input checked="" type="checkbox"/>	3	b	99500 aExported from Connexion by CMU and loaded with m2btab.b in this month	1
<input checked="" type="checkbox"/>	4	b	99500 aExported from Connexion by FLC Reed and loaded with m2btab.b in this month	1
<input checked="" type="checkbox"/>	5	b	99500 aExported from Connexion by Pitkin and loaded with m2btab.b in this month	6
<input checked="" type="checkbox"/>	6	b	995  aLoaded with m2btab.acq2 in 2017.04	1
<input checked="" type="checkbox"/>	7	b	995  aLoaded with m2btab.b in 2017.01	1
<input checked="" type="checkbox"/>	8	b	995  aLoaded with m2btab.b in 2017.02	1
<input checked="" type="checkbox"/>	9	b	995  aLoaded with m2btab.b in 2017.03	3
<input checked="" type="checkbox"/>	10	b	995  aLoaded with m2btab.b in 2017.04	8
<input checked="" type="checkbox"/>	11	b	995  aLoaded with m2btab.b in 2017.05	2

118 Selected Entries limited to 14 results Total : 39 Groups, 118 Entries

Questions about the loading notes process?



# Second Weird Trick

Scan Records Directly into a Review File



## Load files into a review file with a barcode scanner

- What do you do when you have a cart of books you need to put in a review file?
- There's nothing they have in common except they are on the same truck.
- You want to scan the barcodes to put them directly into a review file.



## Enhanced Mode Create Lists:

Boolean Search ✕

Review File Name:

Store Record Type:  ▼

Range ▼ Start  Stop

Enhanced

JSON

Enhanced

ITEM	BARCODE	
in		+ 🔗 ✕
		+ ✕



# Use the 'in' operator

Boolean Search

Review File Name:

Store Record Type:

Range

Enhanced

JSON

ITEM	BARCODE	
in		+ 🔗 ✕
		+ 🔗 ✕
		+ ✕

# Scan the first barcode

Boolean Search

Review File Name:

Store Record Type:

Range

Enhanced

JSON

ITEM	BARCODE	
in		+ 🔗 ✕
	9700000005963	+ ✕

# Press TAB then SPACE

Boolean Search ✕

Review File Name:

Store Record Type:  ▼

Range ▼      Start       Stop

Enhanced

JSON

ITEM	BARCODE		
in		+	✕
	9700000005963	+	✕
		+	✕

Scan the next barcode, repeat

Boolean Search

Review File Name:

Store Record Type:

Range

Enhanced

JSON

ITEM	BARCODE		
in		+	+
	9700000005963	+	+
	9700000005964	+	+

Can I just scan, without using the keyboard?

YES!

Program your scanner to add the tab and space after each scan



## Scanner programing guide

- A scanner programming guide is attached on the website
- It may not work for all scanners




# Print out the guide and scan the barcodes

FILE TOOLS VIEW Programming a barcode scanner to add barcodes to Create Lists\_0-1.docx [Read-Only] - Word

Programming a barcode scanner to add barcodes to Create Lists


Scan each of these barcodes in sequence

**Recall Defaults**




This will reset the scanner to factory default settings

**Enter/Exit Configuration Mode**



**Configurable Suffix #1**



SCREENS 1-2 OF 7 140%



\*Note: this will not work with all scanners – your results may vary

Now you don't need the keyboard between scans

Boolean Search

Review File Name:

Store Record Type:

Range

Enhanced

JSON

ITEM	BARCODE		
in		+	+
	9700000005963	+	+
	9700000005964	+	+



Delete the empty line at the end, or you will get all the records with an empty barcode field

**Boolean Search**

Review File Name:

Store Record Type:

Range

**Enhanced**

**JSON**

ITEM	BARCODE	
in		+ [link] X
9700000005963		+ X
		+ X

**Search Use Existing Search Retrieve Saved Query Save Save As Close**


# Remember to reset your scanner to defaults when you are done

FILE TOOLS VIEW Programming a barcode scanner to add barcodes to Create Lists\_0-1.docx [Read-Only] - Word

Programming a barcode scanner to add barcodes to Create Lists

Scan each of these barcodes in sequence


**Recall Defaults**



3 9 9 9 9 9 8


This will reset the scanner to factory default settings

**Enter/Exit Configuration Mode**



3 9 9 9 9 9 9

**Configurable Suffix #1**



3 9 0 4 5 0 0

SCREENS 1-2 OF 7 140%

Questions about scanning directly into a review file?



# Third Weird Trick

Analyze Your Data With OpenRefine





- OpenRefine is free data wrangling software.
- Like a spreadsheet or database, but works with existing data.
- Just Google it and download.
- Operates in your browser.
- It will tax the speed of your browser.
- Faster browser actually makes a difference.

## These projects are small examples of what OpenRefine can do

- OpenRefine is extremely powerful.
- This is a tiny portion of what you might do with it.
- These particular projects are probably not interesting to you.
- They may give you an idea of what OpenRefine would allow you to do with your own data.

## Simple OpenRefine project: What is in my 902 fields?

- I found that the most common local field in my system was 902.
- These seemed to be data left over from some ancient project on my system, but what are they? Can I use them? Can I delete them?
- They seem to be numbers like 071205 and 060630.
- Some other weird notes and things.

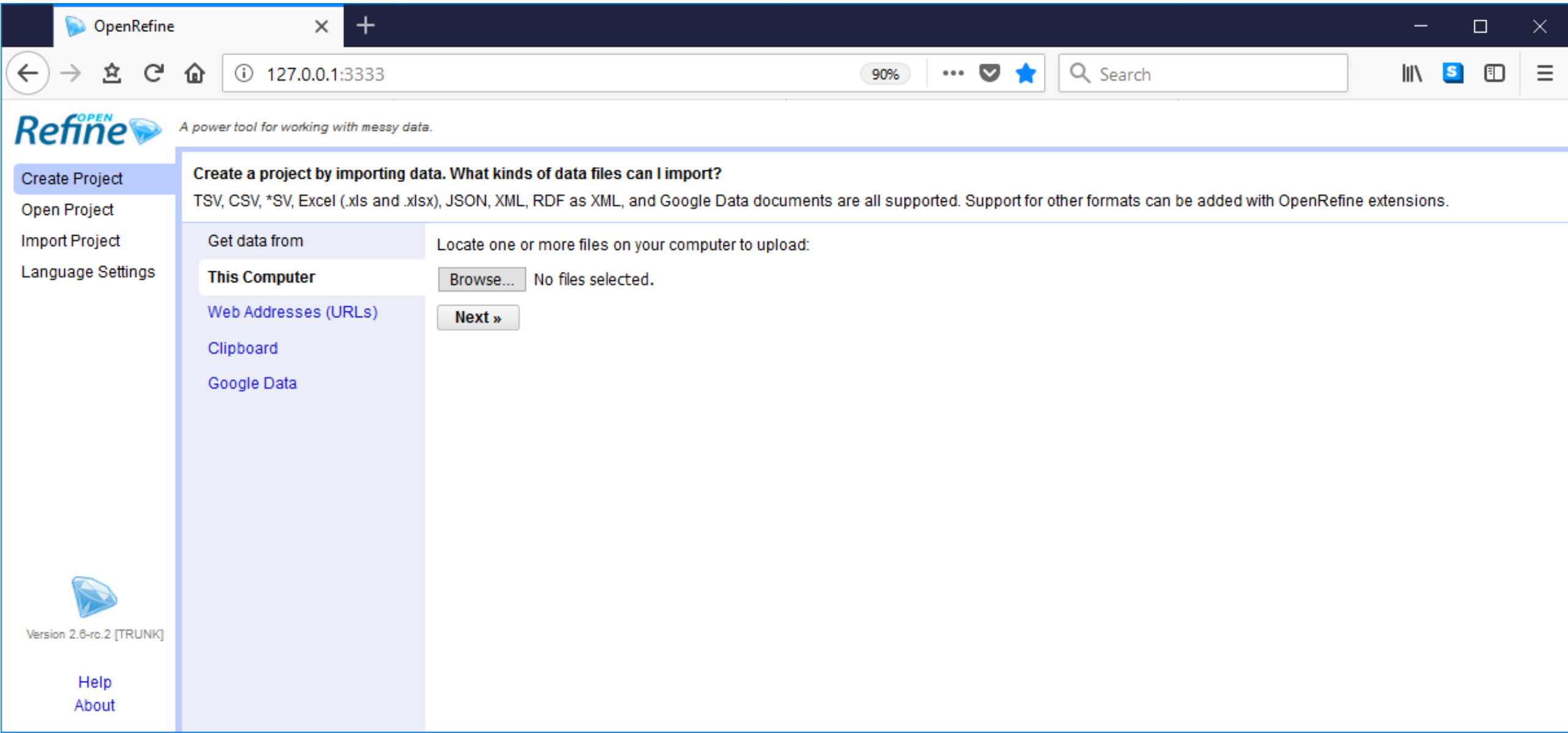
## Exported the 902 fields from Sierra into a tab delimited file

```
902.tab - Notepad
File Edit Format View Help
"RECORD #(BIBLIO)"      "902"
"b10000185"            "060630"
"b10000288"            "060630"
"b10000331"            "060713"
"b10000355"            "060713"
"b10000616"            "060713"
"b10000677"            "060712"
"b10000690"            "060630"
"b10000719"            "060630"
"b10000756"            "060713"
"b1000077x"            "060714"
"b10000781"            "060713"
"b10000811"            "060630"
"b10000823"            "060630"
"b10000847"            "060712"
"b10000884"            "060630"
"b1000094x"            "060713"
"b10000951"            "060713"
"b10000999"            "060714"
"b10001001"            "060630"
```





# Start OpenRefine in a browser



The screenshot shows a web browser window with the OpenRefine interface. The browser's address bar shows the URL `127.0.0.1:3333`. The OpenRefine logo and tagline "A power tool for working with messy data." are visible at the top left. A sidebar on the left contains navigation links: "Create Project", "Open Project", "Import Project", "Language Settings", "Help", and "About". The main content area is titled "Create a project by importing data. What kinds of data files can I import?" and lists supported formats: TSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. Below this, there are three options for getting data: "This Computer" (selected), "Web Addresses (URLs)", and "Google Data". The "This Computer" option includes a "Browse..." button and the text "No files selected." A "Next »" button is also present.

OpenRefine

127.0.0.1:3333

**Refine** OPEN  
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. Support for other formats can be added with OpenRefine extensions.


Get data from

**This Computer**  No files selected.

Web Addresses (URLs)

Clipboard

Google Data

  
Version 2.6-rc.2 [TRUNK]


Help  
About

Click Browse... to find the file

The screenshot shows a web browser window with the OpenRefine application running at 127.0.0.1:3333. The page title is 'OpenRefine' and the URL is '127.0.0.1:3333'. The main content area is titled 'Create a project by importing data. What kinds of data files can I import?' and lists supported formats: TSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. Below this, there is a section for 'Get data from' with four options: 'This Computer', 'Web Addresses (URLs)', 'Clipboard', and 'Google Data'. The 'This Computer' option is selected, and a 'Browse...' button is highlighted with a red box. To the right of the 'Browse...' button, it says 'No files selected.' Below the 'Browse...' button is a 'Next »' button. The left sidebar contains navigation links: 'Create Project', 'Open Project', 'Import Project', 'Language Settings', 'Help', and 'About'. The version number 'Version 2.6-rc.2 [TRUNK]' is displayed at the bottom left.

OpenRefine

127.0.0.1:3333

**Refine** OPEN  A power tool for working with messy data.

Create Project

Open Project

Import Project

Language Settings

Version 2.6-rc.2 [TRUNK]

Help

About

**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. Support for other formats can be added with OpenRefine extensions.

Get data from

**This Computer**

Web Addresses (URLs)

Clipboard

Google Data

Locate one or more files on your computer to upload:

**Browse...** No files selected.

**Next »**

# Click Next »

The screenshot shows the OpenRefine web interface in a browser window. The browser's address bar shows the URL 127.0.0.1:3333. The page title is 'OpenRefine'. The main content area is titled 'Create a project by importing data. What kinds of data files can I import?' and lists supported formats: TSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. Below this, there are four options for 'Get data from': 'This Computer', 'Web Addresses (URLs)', 'Clipboard', and 'Google Data'. The 'This Computer' option is selected, and it shows a 'Browse...' button and a file named '902.tab'. A red box highlights the 'Next »' button. The left sidebar contains navigation links: 'Create Project', 'Open Project', 'Import Project', 'Language Settings', 'Help', and 'About'. The bottom left corner shows the version 'Version 2.6-rc.2 [TRUNK]'.

OpenRefine

127.0.0.1:3333

90%

Search

**Refine** OPEN A power tool for working with messy data.

Create Project

Open Project

Import Project

Language Settings

Get data from

Locate one or more files on your computer to upload:

This Computer

Browse... 902.tab

Next »

Web Addresses (URLs)

Clipboard

Google Data

Version 2.6-rc.2 [TRUNK]

Help

About

# Usually the defaults work when creating a project

OpenRefine

127.0.0.1:3333

90%

Search

Refine *A power tool for working with messy data.*

Create Project « Start Over Configure Parsing Options Project name 902 tab Create Project »

	RECORD #(BIBLIO)	902
1.	b10000185	060630
2.	b10000288	060630
3.	b10000331	060713
4.	b10000355	060713
5.	b10000616	060713
6.	b10000677	060712
7.	b10000690	060630

Parse data as

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

MARC files

RDF/N3 files

XML files

Open Document Format spreadsheets (ods)

Character encoding

Columns are separated by

commas (CSV)

tabs (TSV)

custom \t

Escape special characters with \

Ignore first 0 line(s) at beginning of file

Parse next 1 line(s) as column headers

Discard initial 0 row(s) of data

Load at most 0 row(s) of data

Parse cell text into numbers, dates, ...

Quotation marks are used to enclose cells containing column separators

Store blank rows

Store blank cells as nulls

Store file source (file names, URLs) in each row

Update Preview

Version 2.8-rc.2 [TRUNK]

Help About

It recognizes this is a tab file

# 640,000 rows of data ready to play with

The screenshot shows the OpenRefine web interface in a browser window. The browser tab is titled "902 tab - OpenRefine". The address bar shows the URL "127.0.0.1:3333/project?project=2122872729378". The page header includes the OpenRefine logo, the text "902 tab", a "Permalink" link, and buttons for "Open...", "Export", and "Help". Below the header, there is a "Facet / Filter" section with "Undo / Redo" options. The main content area displays a table with 640,172 rows. The table has three columns: "All", "RECORD #(BIBL)", and "902". The first 20 rows of the table are visible, showing record numbers 1 through 20, their corresponding BIBL numbers, and the value in the "902" column. A sidebar on the left contains a section titled "Using facets and filters" with instructions and a link to "Watch these screencasts".

902 tab - OpenRefine

127.0.0.1:3333/project?project=2122872729378

90% Search

Open... Export Help

Refine 902 tab Permalink

Facet / Filter Undo / Redo

640172 rows

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

« first < previous 1 - 50 next > last »

All	RECORD #(BIBL)	902
☆	1. b10000185	060630
☆	2. b10000288	060630
☆	3. b10000331	060713
☆	4. b10000355	060713
☆	5. b10000616	060713
☆	6. b10000677	060712
☆	7. b10000690	060630
☆	8. b10000719	060630
☆	9. b10000756	060713
☆	10. b1000077x	060714
☆	11. b10000781	060713
☆	12. b10000811	060630
☆	13. b10000823	060630
☆	14. b10000847	060712
☆	15. b10000884	060630
☆	16. b1000094x	060713
☆	17. b10000951	060713
☆	18. b10000999	060714
☆	19. b10001001	060630
☆	20. b10001037	060713

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

# Create a text facet

902 tab - OpenRefine

127.0.0.1:3333/project?project=2122872729378

90% Search

Refine OPEN 902 tab Permalink

Open... Export Help

Facet / Filter Undo / Redo

640172 rows

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

« first < previous 1 - 50 next > last »

Extensions:

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

	All	RECORD #(BIBL)	902
1.	b10000185		
2.	b10000288		
3.	b10000331		
4.	b10000355		
5.	b10000616		
6.	b10000677		
7.	b10000690		
8.	b10000719		
9.	b10000756		
10.	b1000077x		
11.	b10000781		
12.	b10000811	060630	
13.	b10000823	060630	
14.	b10000847	060712	
15.	b10000884	060630	
16.	b1000094x	060713	
17.	b10000951	060713	
18.	b10000999	060714	
19.	b10001001	060630	
20.	b10001037	060713	

Facet

- Text facet
- Numeric facet
- Timeline facet
- Scatterplot facet
- Custom text facet...
- Custom Numeric Facet...
- Customized facets

Text filter

Edit cells

Edit column

Transpose

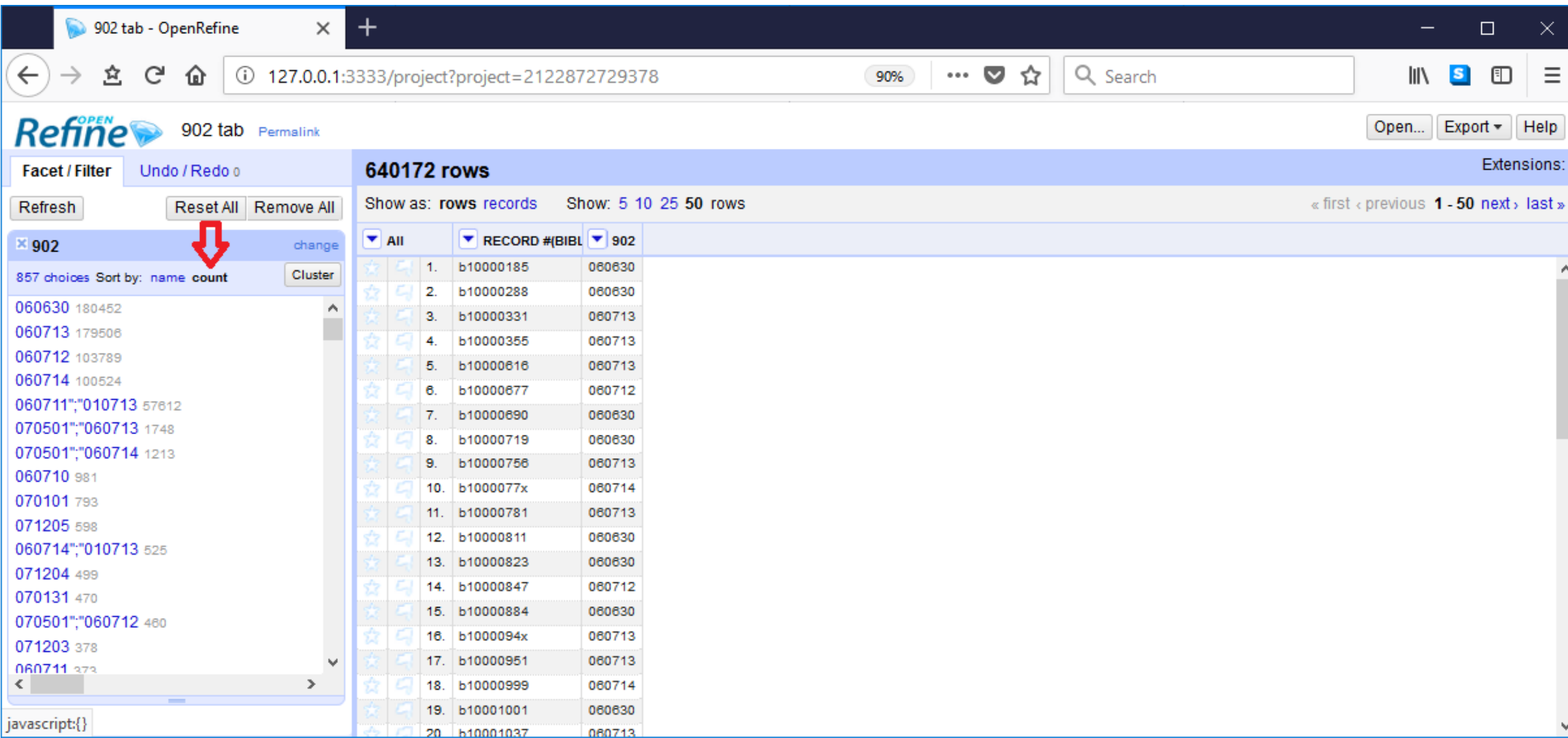
Sort...

View

Reconcile

javascript: {}

# Sort the facet by count to see most common data



The screenshot shows the OpenRefine web interface. The browser address bar displays the URL `127.0.0.1:3333/project?project=2122872729378`. The main interface shows a facet for the value '902' with 857 choices. The facet is sorted by 'count', and a red arrow points to this sort option. The facet list shows the following items:

- 060630 180452
- 060713 179506
- 060712 103789
- 060714 100524
- 060711";"010713 57612
- 070501";"060713 1748
- 070501";"060714 1213
- 060710 981
- 070101 793
- 071205 598
- 060714";"010713 525
- 071204 499
- 070131 470
- 070501";"060712 460
- 071203 378
- 060711 373

The main table displays 640172 rows, showing the first 20 records. The table columns are 'All', 'RECORD #(BIBL)', and '902'. The first 20 rows are:

All	RECORD #(BIBL)	902
1.	b10000185	060630
2.	b10000288	060630
3.	b10000331	060713
4.	b10000355	060713
5.	b10000616	060713
6.	b10000677	060712
7.	b10000690	060630
8.	b10000719	060630
9.	b10000756	060713
10.	b1000077x	060714
11.	b10000781	060713
12.	b10000811	060630
13.	b10000823	060630
14.	b10000847	060712
15.	b10000884	060630
16.	b1000094x	060713
17.	b10000951	060713
18.	b10000999	060714
19.	b10001001	060630
20.	b10001037	060713

Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

640172 rows

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

Extensions: « first < previous 1 - 50 next > last »

902 change

857 choices Sort by: name count Cluster

060630	180452
060713	179508
060712	103789
060714	100524
060711";"010713	57612
070501";"060713	1748
070501";"060714	1213
060710	981
070101	793
071205	598
060714";"010713	525
071204	499
070131	470
070501";"060712	460
071203	378
060711	373

All	RECORD #(BIBL)	902
1.	b10000185	060630
2.	b10000208	060630
3.	b10000331	060713
4.	b10000355	060713
5.	b10000616	060713
6.	b10000677	060712
7.	b10000690	060630
8.	b10000719	060630
9.	b10000756	060713
10.	b1000077x	060714
11.	b10000781	060713
12.	b10000811	060630
13.	b10000823	060630
14.	b10000847	060712
15.	b10000884	060630
16.	b1000094x	060713
17.	b10000951	060713
18.	b10000999	060714
19.	b10001001	060630
20.	b10001037	060713

060630 appears in 180,452 records



Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

902 change

857 choices Sort by: name count Cluster

060630	180452
060713	179506
060712	103789
060714	100524
060711";"010713	57612
070501";"060713	1748
070501";"060714	1213
060710	981
070101	793
071205	598
060714";"010713	525
071204	499
070131	470
070501";"060712	460
071203	378
060711	373

640172 rows

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

Extensions: « first < previous 1 - 50 next > last »

All	RECORD #(BIBL)	902
☆ ↶	1. b10000185	060630
☆ ↶	2. b10000288	060630
☆ ↶	3. b10000331	060713
☆ ↶	4. b10000385	060713
☆ ↶	5. b10000616	060713
☆ ↶	6. b10000677	060712
☆ ↶	7. b10000690	060630
☆ ↶	8. b10000719	060630
☆ ↶	9. b10000756	060713
☆ ↶	10. b1000077x	060714
☆ ↶	11. b10000781	060713
☆ ↶	12. b10000811	060630
☆ ↶	13. b10000823	060630
☆ ↶	14. b10000847	060712
☆ ↶	15. b10000884	060630
☆ ↶	16. b1000094x	060713
☆ ↶	17. b10000951	060713
☆ ↶	18. b10000999	060714
☆ ↶	19. b10001001	060630
☆ ↶	20. b10001037	060713

57,512 records have a 902 with 060711 and another 902 with 010713

## Have you figured out what the numbers are yet?

- Dates. We think they are codes for the date someone did something to these records.
- Assuming that's the case, we can find the newest date.
- We could just sort the 902 column, but what about those multiple fields?
- We can split them out.



# Split multi-valued cells...

The screenshot shows the OpenRefine web interface. The browser address bar displays `127.0.0.1:3333/project?project=2122872729378`. The interface includes a search bar, navigation buttons, and a main data table with 640172 rows. A facet for '902' is active on the left, showing 857 choices. A context menu is open over a cell in the 'RECORD #(BIBL)' column, with the 'Split multi-valued cells...' option highlighted. The table data includes columns for 'All', 'RECORD #(BIBL)', and '902'.

All	RECORD #(BIBL)	902
639866.	b30912957	
639852.	b30868920	
639864.	b30908504	
639859.	b30895789	
639862.	b30901236	
33055.	b10907737	
633263.	b26931552	
633275.	b2693243x	
633292.	b26933640	
633333.	b26936744	
633346.	b26937736	
633349.	b2693811x	071205
633364.	b26939915	071205
633365.	b26940061	071205
633373.	b26940395	071205
633374.	b26940401	071205
633377.	b26940528	071205
633394.	b26942124	071205
633402.	b26942458	071205
633417.	b26943049	071205

The separator is “;”

The screenshot shows the OpenRefine web interface. The browser tab is "902 tab - OpenRefine" and the address bar shows "127.0.0.1:3333/project?project=2122872729378". The main area displays a table with 640172 rows. A modal dialog is open in the center, asking "What separator currently separates the values?" with a text input field containing ";" and "OK" and "Cancel" buttons. A context menu is also visible over the table, with options like "Split multi-valued cells...", "Join multi-valued cells...", and "Cluster and edit...".

902 tab - OpenRefine

127.0.0.1:3333/project?project=2122872729378

Refine 902 tab Permalink

Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

640172 rows

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

Extensions: « first < previous 1 - 50 next > last »

902 change

857 choices Sort by: name count Cluster

All	RECORD #(BIBL)	902
639866	b30912957	
639852	b30888920	
639864	b30908504	
639859	b30895789	
639862	b30901236	
33055	b10907737	
633263	b26931552	
633275	b2693243x	
633292	b26933640	
633333	b26936744	
633346	b26937736	
633349	b2693811x	071205
633364	b26939915	071205
633365	b26940061	071205
633373	b26940395	071205
633374	b26940401	071205
633377	b26940528	071205
633394	b26942124	071205
633402	b26942458	071205
633417	b26943049	071205

View Split multi-valued cells... Reconcile Join multi-valued cells... Cluster and edit...

What separator currently separates the values?

;

OK Cancel

javascript: {}

Now every 902 is in its own row

The screenshot shows the OpenRefine web interface. At the top, a browser window is titled "902 tab - OpenRefine" with the address bar showing "127.0.0.1:3333/project?project=2122872729378". The OpenRefine logo and "902 tab" are visible in the top left. On the right, there are buttons for "Open...", "Export", and "Help".

The main interface is divided into a left sidebar and a main table area. The sidebar contains a "Facet / Filter" section for "902" with "914 choices" and a "Cluster" button. Below this is a list of facet values with their counts, such as "060713 181477".

The main table area shows "702844 rows" and "Show as: rows records". It includes a "Show:" dropdown set to "5 10 25 50 rows" and a "Sort" dropdown. The table has columns for "All", "RECORD #(BIBL)", and "902". The data rows show individual records with their respective IDs and values.

All	RECORD #(BIBL)	902
501065.		20020224001642.0
497782.		20020108093813.0
500457.		20020107000000.0
497779.		20020104000000.0
497776.		20011222000000.0
492570.		20011127000000.0
494877.		20011118000000.0
493201.		20011110235218.0
501326.		20011106000000.0
499902.		20011105102629.0
491261.		20011104000000.0
491531.		20011104000000.0
491530.		20011030000000.0
490516.		20011020235414.0
494217.		20011016000000.0
491257.		20011012000000.0
491259.		20011005000000.0
490608.		20011002000000.0
488962.		20010928000000.0
488445.		20010925000000.0



Scroll down a couple pages and the most recent is 2009

The screenshot shows the OpenRefine web interface in a browser window. The browser's address bar displays the URL `127.0.0.1:3333/project?project=2122872729378`. The page title is "902 tab - OpenRefine". The main content area shows a table of records with 702844 rows. The table is sorted by year, and the most recent year shown is 2009. The table has three columns: "RECORD #(BIBL)", "902", and "year". The "902" column contains values like "980803", "980802", "980801", and "091217". The "year" column contains values like "2009", "2008", "2007", "2006", "2005", "2004", "2003", "2002", "2001", "2000", "1999", "1998", "1997", "1996", "1995", "1994", "1993", "1992", "1991", "1990", "1989", "1988", "1987", "1986", "1985", "1984", "1983", "1982", "1981", "1980", "1979", "1978", "1977", "1976", "1975", "1974", "1973", "1972", "1971", "1970", "1969", "1968", "1967", "1966", "1965", "1964", "1963", "1962", "1961", "1960", "1959", "1958", "1957", "1956", "1955", "1954", "1953", "1952", "1951", "1950", "1949", "1948", "1947", "1946", "1945", "1944", "1943", "1942", "1941", "1940", "1939", "1938", "1937", "1936", "1935", "1934", "1933", "1932", "1931", "1930", "1929", "1928", "1927", "1926", "1925", "1924", "1923", "1922", "1921", "1920", "1919", "1918", "1917", "1916", "1915", "1914", "1913", "1912", "1911", "1910", "1909", "1908", "1907", "1906", "1905", "1904", "1903", "1902", "1901", "1900", "1899", "1898", "1897", "1896", "1895", "1894", "1893", "1892", "1891", "1890", "1889", "1888", "1887", "1886", "1885", "1884", "1883", "1882", "1881", "1880", "1879", "1878", "1877", "1876", "1875", "1874", "1873", "1872", "1871", "1870", "1869", "1868", "1867", "1866", "1865", "1864", "1863", "1862", "1861", "1860", "1859", "1858", "1857", "1856", "1855", "1854", "1853", "1852", "1851", "1850", "1849", "1848", "1847", "1846", "1845", "1844", "1843", "1842", "1841", "1840", "1839", "1838", "1837", "1836", "1835", "1834", "1833", "1832", "1831", "1830", "1829", "1828", "1827", "1826", "1825", "1824", "1823", "1822", "1821", "1820", "1819", "1818", "1817", "1816", "1815", "1814", "1813", "1812", "1811", "1810", "1809", "1808", "1807", "1806", "1805", "1804", "1803", "1802", "1801", "1800", "1799", "1798", "1797", "1796", "1795", "1794", "1793", "1792", "1791", "1790", "1789", "1788", "1787", "1786", "1785", "1784", "1783", "1782", "1781", "1780", "1779", "1778", "1777", "1776", "1775", "1774", "1773", "1772", "1771", "1770", "1769", "1768", "1767", "1766", "1765", "1764", "1763", "1762", "1761", "1760", "1759", "1758", "1757", "1756", "1755", "1754", "1753", "1752", "1751", "1750", "1749", "1748", "1747", "1746", "1745", "1744", "1743", "1742", "1741", "1740", "1739", "1738", "1737", "1736", "1735", "1734", "1733", "1732", "1731", "1730", "1729", "1728", "1727", "1726", "1725", "1724", "1723", "1722", "1721", "1720", "1719", "1718", "1717", "1716", "1715", "1714", "1713", "1712", "1711", "1710", "1709", "1708", "1707", "1706", "1705", "1704", "1703", "1702", "1701", "1700", "1699", "1698", "1697", "1696", "1695", "1694", "1693", "1692", "1691", "1690", "1689", "1688", "1687", "1686", "1685", "1684", "1683", "1682", "1681", "1680", "1679", "1678", "1677", "1676", "1675", "1674", "1673", "1672", "1671", "1670", "1669", "1668", "1667", "1666", "1665", "1664", "1663", "1662", "1661", "1660", "1659", "1658", "1657", "1656", "1655", "1654", "1653", "1652", "1651", "1650", "1649", "1648", "1647", "1646", "1645", "1644", "1643", "1642", "1641", "1640", "1639", "1638", "1637", "1636", "1635", "1634", "1633", "1632", "1631", "1630", "1629", "1628", "1627", "1626", "1625", "1624", "1623", "1622", "1621", "1620", "1619", "1618", "1617", "1616", "1615", "1614", "1613", "1612", "1611", "1610", "1609", "1608", "1607", "1606", "1605", "1604", "1603", "1602", "1601", "1600", "1599", "1598", "1597", "1596", "1595", "1594", "1593", "1592", "1591", "1590", "1589", "1588", "1587", "1586", "1585", "1584", "1583", "1582", "1581", "1580", "1579", "1578", "1577", "1576", "1575", "1574", "1573", "1572", "1571", "1570", "1569", "1568", "1567", "1566", "1565", "1564", "1563", "1562", "1561", "1560", "1559", "1558", "1557", "1556", "1555", "1554", "1553", "1552", "1551", "1550", "1549", "1548", "1547", "1546", "1545", "1544", "1543", "1542", "1541", "1540", "1539", "1538", "1537", "1536", "1535", "1534", "1533", "1532", "1531", "1530", "1529", "1528", "1527", "1526", "1525", "1524", "1523", "1522", "1521", "1520", "1519", "1518", "1517", "1516", "1515", "1514", "1513", "1512", "1511", "1510", "1509", "1508", "1507", "1506", "1505", "1504", "1503", "1502", "1501", "1500", "1499", "1498", "1497", "1496", "1495", "1494", "1493", "1492", "1491", "1490", "1489", "1488", "1487", "1486", "1485", "1484", "1483", "1482", "1481", "1480", "1479", "1478", "1477", "1476", "1475", "1474", "1473", "1472", "1471", "1470", "1469", "1468", "1467", "1466", "1465", "1464", "1463", "1462", "1461", "1460", "1459", "1458", "1457", "1456", "1455", "1454", "1453", "1452", "1451", "1450", "1449", "1448", "1447", "1446", "1445", "1444", "1443", "1442", "1441", "1440", "1439", "1438", "1437", "1436", "1435", "1434", "1433", "1432", "1431", "1430", "1429", "1428", "1427", "1426", "1425", "1424", "1423", "1422", "1421", "1420", "1419", "1418", "1417", "1416", "1415", "1414", "1413", "1412", "1411", "1410", "1409", "1408", "1407", "1406", "1405", "1404", "1403", "1402", "1401", "1400", "1399", "1398", "1397", "1396", "1395", "1394", "1393", "1392", "1391", "1390", "1389", "1388", "1387", "1386", "1385", "1384", "1383", "1382", "1381", "1380", "1379", "1378", "1377", "1376", "1375", "1374", "1373", "1372", "1371", "1370", "1369", "1368", "1367", "1366", "1365", "1364", "1363", "1362", "1361", "1360", "1359", "1358", "1357", "1356", "1355", "1354", "1353", "1352", "1351", "1350", "1349", "1348", "1347", "1346", "1345", "1344", "1343", "1342", "1341", "1340", "1339", "1338", "1337", "1336", "1335", "1334", "1333", "1332", "1331", "1330", "1329", "1328", "1327", "1326", "1325", "1324", "1323", "1322", "1321", "1320", "1319", "1318", "1317", "1316", "1315", "1314", "1313", "1312", "1311", "1310", "1309", "1308", "1307", "1306", "1305", "1304", "1303", "1302", "1301", "1300", "1299", "1298", "1297", "1296", "1295", "1294", "1293", "1292", "1291", "1290", "1289", "1288", "1287", "1286", "1285", "1284", "1283", "1282", "1281", "1280", "1279", "1278", "1277", "1276", "1275", "1274", "1273", "1272", "1271", "1270", "1269", "1268", "1267", "1266", "1265", "1264", "1263", "1262", "1261", "1260", "1259", "1258", "1257", "1256", "1255", "1254", "1253", "1252", "1251", "1250", "1249", "1248", "1247", "1246", "1245", "1244", "1243", "1242", "1241", "1240", "1239", "1238", "1237", "1236", "1235", "1234", "1233", "1232", "1231", "1230", "1229", "1228", "1227", "1226", "1225", "1224", "1223", "1222", "1221", "1220", "1219", "1218", "1217", "1216", "1215", "1214", "1213", "1212", "1211", "1210", "1209", "1208", "1207", "1206", "1205", "1204", "1203", "1202", "1201", "1200", "1199", "1198", "1197", "1196", "1195", "1194", "1193", "1192", "1191", "1190", "1189", "1188", "1187", "1186", "1185", "1184", "1183", "1182", "1181", "1180", "1179", "1178", "1177", "1176", "1175", "1174", "1173", "1172", "1171", "1170", "1169", "1168", "1167", "1166", "1165", "1164", "1163", "1162", "1161", "1160", "1159", "1158", "1157", "1156", "1155", "1154", "1153", "1152", "1151", "1150", "1149", "1148", "1147", "1146", "1145", "1144", "1143", "1142", "1141", "1140", "1139", "1138", "1137", "1136", "1135", "1134", "1133", "1132", "1131", "1130", "1129", "1128", "1127", "1126", "1125", "1124", "1123", "1122", "1121", "1120", "1119", "1118", "1117", "1116", "1115", "1114", "1113", "1112", "1111", "1110", "1109", "1108", "1107", "1106", "1105", "1104", "1103", "1102", "1101", "1100", "1099", "1098", "1097", "1096", "1095", "1094", "1093", "1092", "1091", "1090", "1089", "1088", "1087", "1086", "1085", "1084", "1083", "1082", "1081", "1080", "1079", "1078", "1077", "1076", "1075", "1074", "1073", "1072", "1071", "1070", "1069", "1068", "1067", "1066", "1065", "1064", "1063", "1062", "1061", "1060", "1059", "1058", "1057", "1056", "1055", "1054", "1053", "1052", "1051", "1050", "1049", "1048", "1047", "1046", "1045", "1044", "1043", "1042", "1041", "1040", "1039", "1038", "1037", "1036", "1035", "1034", "1033", "1032", "1031", "1030", "1029", "1028", "1027", "1026", "1025", "1024", "1023", "1022", "1021", "1020", "1019", "1018", "1017", "1016", "1015", "1014", "1013", "1012", "1011", "1010", "1009", "1008", "1007", "1006", "1005", "1004", "1003", "1002", "1001", "1000", "999", "998", "997", "996", "995", "994", "993", "992", "991", "990", "989", "988", "987", "986", "985", "984", "983", "982", "981", "980", "979", "978", "977", "976", "975", "974", "973", "972", "971", "970", "969", "968", "967", "966", "965", "964", "963", "962", "961", "960", "959", "958", "957", "956", "955", "954", "953", "952", "951", "950", "949", "948", "947", "946", "945", "944", "943", "942", "941", "940", "939", "938", "937", "936", "935", "934", "933", "932", "931", "930", "929", "928", "927", "926", "925", "924", "923", "922", "921", "920", "919", "918", "917", "916", "915", "914", "913", "912", "911", "910", "909", "908", "907", "906", "905", "904", "903", "902".

RECORD #(BIBL)	902	year
678744.	980803	2009
503449.	980802	2009
517591.	980802	2009
545611.	980802	2009
652023.	980802	2009
503370.	980801	2009
520371.	980801	2009
529900.	980801	2009
14289. b10437149	091217	2009
33802. b10907737	091216	2009
83138. b12129185	090925	2009
78699. b12072473	090630	2009
86690. b12173058	090630	2009
86840. b12174865	090630	2009
695914. b26931552	071205	2009
695926. b2693243x	071205	2009
695943. b26933640	071205	2009
695984. b26936744	071205	2009
695997. b26937736	071205	2009
696000. b2693811x	071205	2009

## Upshot

- Based on the breakdown in OpenRefine we determined these are old data. Nobody has added these to our system since December of 2009.
- We decide it is ok to delete them.



## Looking for duplicates

- Our 001 and 019 are not in the same index, so Headings Reports won't see those duplicates.
- We were thinking about reindexing to fix this, but we didn't know how big the problem was.





We used SQL to pull out the 001 and 019 fields into a CSV

```
SELECT
```

```
    'b' || vt.record_num || 'a' as Record_num,  
    vt.marc_tag,  
    unnest(string_to_array(vt.field_content, '|a')) AS Data
```

```
FROM
```

```
    sierra_view.varfield_view as vt
```

```
WHERE
```

```
    vt.marc_tag = '019'
```

```
OR
```

```
    vt.marc_tag = '001'
```

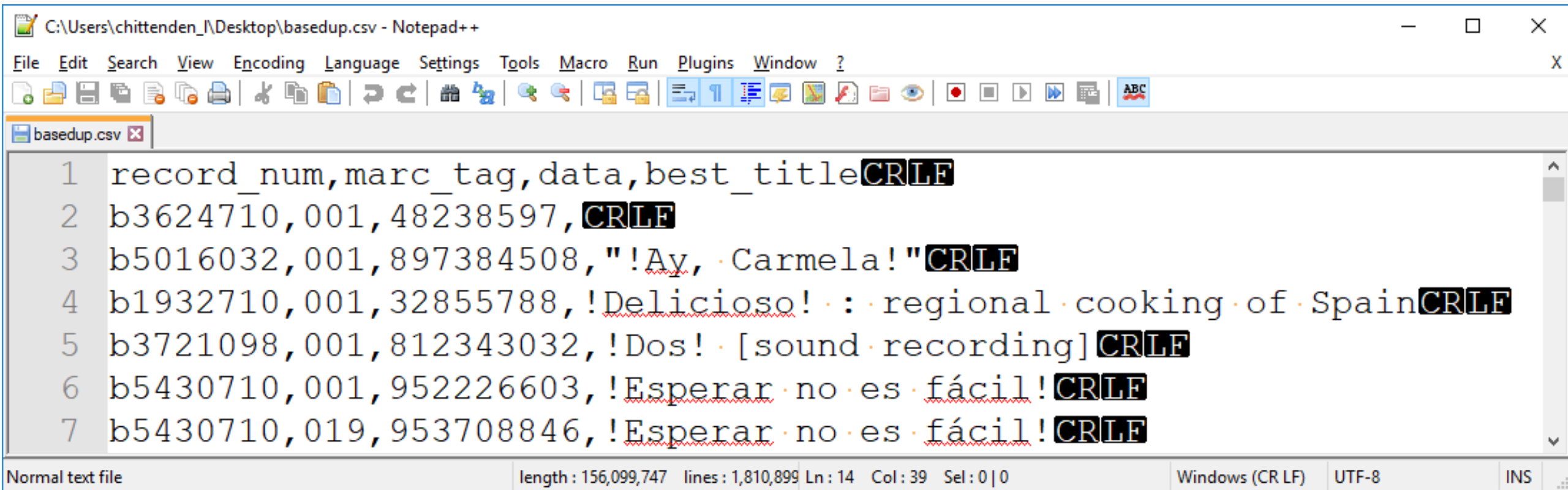
```
ORDER BY
```

```
    vt.record_num
```



## That produced this file

- Columns for: Record number, MARC tag, Number, Title



```
1 record_num,marc_tag,data,best_titleCRLF
2 b3624710,001,48238597,CRLF
3 b5016032,001,897384508,"!Ay, Carmela!"CRLF
4 b1932710,001,32855788,!Delicioso! : regional cooking of SpainCRLF
5 b3721098,001,812343032,!Dos! [sound recording]CRLF
6 b5430710,001,952226603,!Esperar no es fácil!CRLF
7 b5430710,019,953708846,!Esperar no es fácil!CRLF
```

Normal text file length : 156,099,747 lines : 1,810,899 Ln : 14 Col : 39 Sel : 0 | 0 Windows (CR LF) UTF-8 INS

Click Browse... to find the file

The screenshot shows the OpenRefine web interface in a browser window. The browser's address bar displays the URL 127.0.0.1:3333. The page header includes the OpenRefine logo and the tagline "A power tool for working with messy data." A left-hand navigation menu contains the following items: "Create Project" (highlighted), "Open Project", "Import Project", and "Language Settings". The main content area is titled "Create a project by importing data. What kinds of data files can I import?" and lists supported formats: TSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. Below this, the "Get data from" section is active, showing a list of options: "This Computer", "Web Addresses (URLs)", "Clipboard", and "Google Data". The "This Computer" option is selected, and a "Browse..." button is highlighted with a red rectangle. To the right of the "Browse..." button, the text "No files selected." is visible. Below the "Browse..." button is a "Next »" button. The bottom left corner of the page shows a diamond icon and the text "Version 2.6-rc.2 [TRUNK]", with "Help" and "About" links below it.

OpenRefine

127.0.0.1:3333

**Refine** OPEN A power tool for working with messy data.

Create Project

Open Project

Import Project

Language Settings

Version 2.6-rc.2 [TRUNK]

Help

About

**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. Support for other formats can be added with OpenRefine extensions.

Get data from

**This Computer**

Web Addresses (URLs)

Clipboard

Google Data

Locate one or more files on your computer to upload:

**Browse...** No files selected.

Next »

Click Next »

The screenshot shows the OpenRefine web interface in a browser window. The browser's address bar displays the URL `127.0.0.1:3333`. The page header features the OpenRefine logo and the tagline "A power tool for working with messy data." On the left side, there is a navigation menu with the following items: "Create Project" (highlighted), "Open Project", "Import Project", and "Language Settings". At the bottom of the menu, there is a diamond icon and the text "Version 2.6-rc.2 [TRUNK]", along with "Help" and "About" links. The main content area is titled "Create a project by importing data. What kinds of data files can I import?" and lists supported formats: TSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents. Below this, there are three options for "Get data from": "This Computer", "Web Addresses (URLs)", and "Google Data". The "This Computer" option is selected, and the text "Locate one or more files on your computer to upload:" is displayed. A "Browse..." button is highlighted with a blue box, and the filename "basedup.csv" is shown next to it. A red arrow points to the filename. Below the "Browse..." button, the "Next »" button is highlighted with a red box.

# Many options for types of files to load:

The screenshot shows the OpenRefine web interface. The browser address bar shows the URL 127.0.0.1:3333. The page title is 'OpenRefine' and the subtitle is 'A power tool for working with messy data.' The main content area is titled 'Configure Parsing Options' and shows a table of data with columns 'record\_num', 'marc\_tag', 'data', and 'best\_title'. The project name is 'basedup csv'. The 'Parse data as' section is set to 'CSV / TSV / separator-based files'. The 'Character encoding' is set to an empty field. The 'Columns are separated by' section has 'commas (CSV)' selected. The 'Escape special characters with \ ' section is empty. The 'Update Preview' button is visible. The bottom left corner shows the version 'Version 2.6-rc.2 [TRUNK]' and links for 'Help' and 'About'.

OpenRefine

127.0.0.1:3333

Refine <sup>OPEN</sup> A power tool for working with messy data.

Create Project « Start Over Configure Parsing Options Project name basedup csv Create Project »

	record_num	marc_tag	data	best_title
1.	b3624710	001	48238597	
2.	b5016032	001	897384508	!Ay, Carmela!
3.	b1932710	001	32855788	!Delicioso! : regional cooking of Spain
4.	b3721098	001	812343032	!Dos! [sound recording]
5.	b5430710	001	952226603	!Esperar no es fácil!

Parse data as

Character encoding

Update Preview

CSV / TSV / separator-based files

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

MARC files

RDF/N3 files

XML files

Open Document Format spreadsheets (.ods)

Columns are separated by

commas (CSV)

tabs (TSV)

custom , \_\_\_\_\_

Escape special characters with \

Ignore first 0 line(s) at beginning of file

Parse next 1 line(s) as column headers

Discard initial 0 row(s) of data

Load at most 0 row(s) of data

Parse cell text into numbers, dates, ...

Quotation marks are used to enclose cells containing column separators

Store blank rows

Store blank cells as nulls

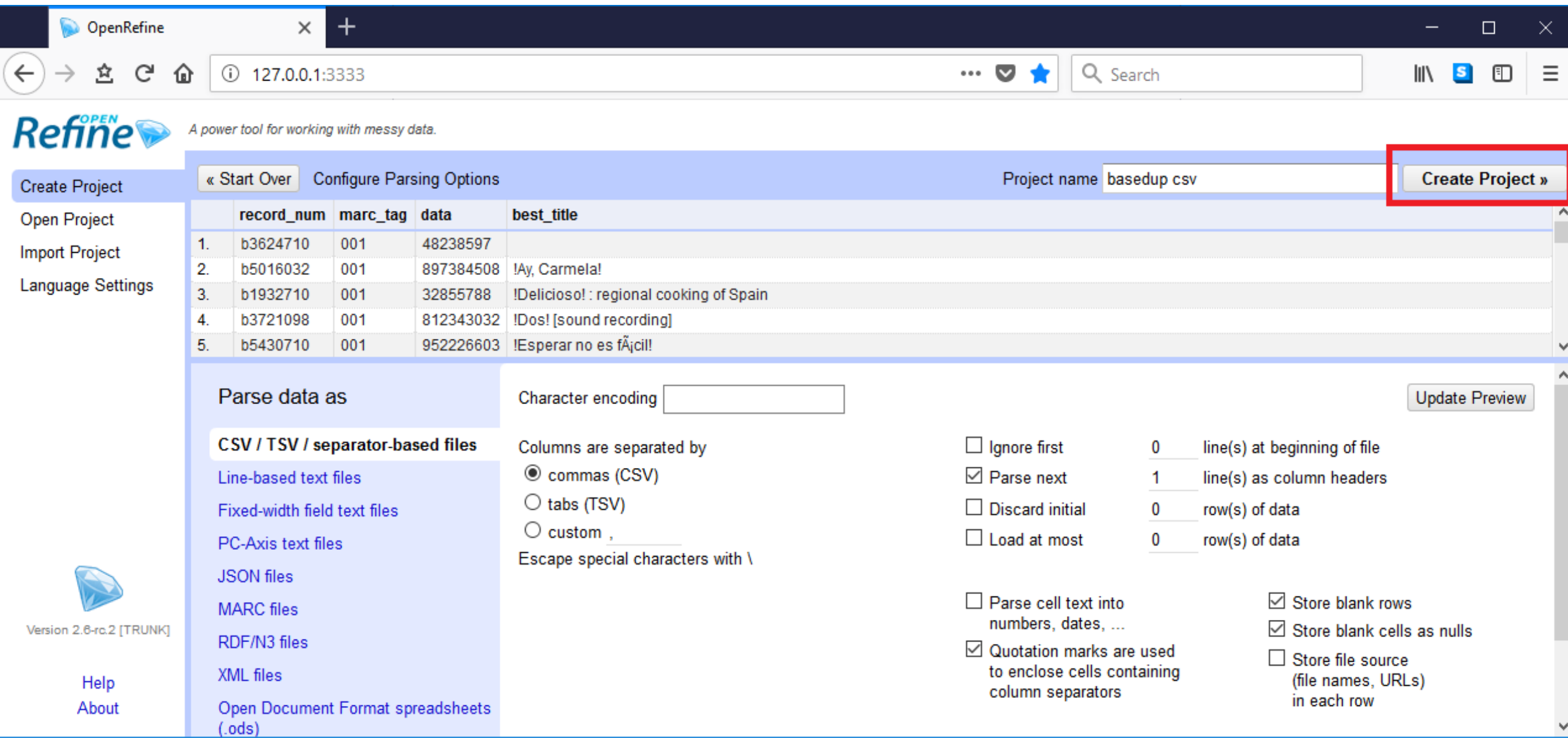
Store file source (file names, URLs) in each row

Version 2.6-rc.2 [TRUNK]

Help

About

It sees that we are loading a CSV, so just click Create Project »



OpenRefine

127.0.0.1:3333

Refine *A power tool for working with messy data.*

Create Project

« Start Over Configure Parsing Options Project name basedup csv **Create Project »**

	record_num	marc_tag	data	best_title
1.	b3624710	001	48238597	
2.	b5016032	001	897384508	!Ay, Carmela!
3.	b1932710	001	32855788	!Delicioso! : regional cooking of Spain
4.	b3721098	001	812343032	!Dos! [sound recording]
5.	b5430710	001	952226603	!Esperar no es fãcil!

Parse data as

**CSV / TSV / separator-based files**

Line-based text files

Fixed-width field text files

PC-Axis text files

JSON files

MARC files

RDF/N3 files

XML files

Open Document Format spreadsheets (.ods)

Character encoding

Columns are separated by

commas (CSV)

tabs (TSV)

custom , \_\_\_\_\_

Escape special characters with \

Ignore first 0 line(s) at beginning of file

Parse next 1 line(s) as column headers

Discard initial 0 row(s) of data

Load at most 0 row(s) of data

Parse cell text into numbers, dates, ...

Quotation marks are used to enclose cells containing column separators

Store blank rows

Store blank cells as nulls

Store file source (file names, URLs) in each row

Update Preview

Version 2.6-rc.2 [TRUNK]

Help

About

Here's the data, all 1.8 million rows.

The screenshot shows the OpenRefine web interface. The browser address bar displays the URL `127.0.0.1:3333/project?project=2441104465189`. The page title is "basedup csv" with a "Permalink" link. The main content area shows a table with 1810897 rows. The table has columns for "record\_num", "marc\_tag", "data", and "best\_title". The first 50 rows are visible, showing various record numbers and titles. A sidebar on the left contains a "Facet / Filter" section with "Undo / Redo" options and a "Using facets and filters" guide. The guide explains that facets and filters are used to select subsets of data and provides a link to "Watch these screencasts".

Facet / Filter Undo / Redo 0

1810897 rows

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 50 next > last »

All	record_num	marc_tag	data	best_title	
☆	33.	b4703784	001	12154435	"---and ladies of the club"
☆	34.	b4764669	001	50853617	"---first do no harm" [videorecording]
☆	35.	b2296747	001	52871940	"--The burden of psychiatric conditions has been heavily underestimated" [electronic resource] : the impact of mental illness on society.
☆	36.	b4740366	001	4710532	"--We are the living proof--" : the justice model for corrections
☆	37.	b4764151	001	28827124	"--and they laid them to rest in the little plot beside the Pecos" : final report on the relocation of old Seven Rivers Cemetery, Eddy County, New
☆	38.	b1209929	001	30959525	"--and they laid them to rest in the little plot beside the Pecos" [microform] : final report on the relocation of old Seven Rivers Cemetery, Eddy
☆	39.	b1308106	001	2788555	"...Of gods and ghosts." [sound recording]
☆	40.	b3773005	019	777264089	"10" [videorecording]
☆	41.	b3773005	001	705362648	"10" [videorecording]
☆	42.	b4770635	001	812539120	"101" Strings play Victor Herbert favorites [sound recording].
☆	43.	b1076366	001	26539506	"1812" Overture, op. 49 ; Marche slave, op 31 [sound recording]
☆	44.	b4763004	001	13225493	"1812" overture [sound recording] : op. 49 ; Capriccio italien : op. 45 ; Cossack dance : from Mazeppa
☆	45.	b3531059	019	732905383	"21" : the story of Roberto Clemente : a graphic novel
☆	46.	b3531059	001	712176585	"21" : the story of Roberto Clemente : a graphic novel
☆	47.	b3531059	019	756701011	"21" : the story of Roberto Clemente : a graphic novel
☆	48.	b2908107	001	31294899	"25 Ä@tudes cÄ@Äbres" [sound recording] ; Fantasy, opus 28, for flute & harp
☆	49.	b1372926	001	25810188	"70% by '92" : safety belt program idea sampler
☆	50.	b1318625	001	4582038	"88" grand old songs : containing 88 old favorites that will live forever! For voice piano and uke

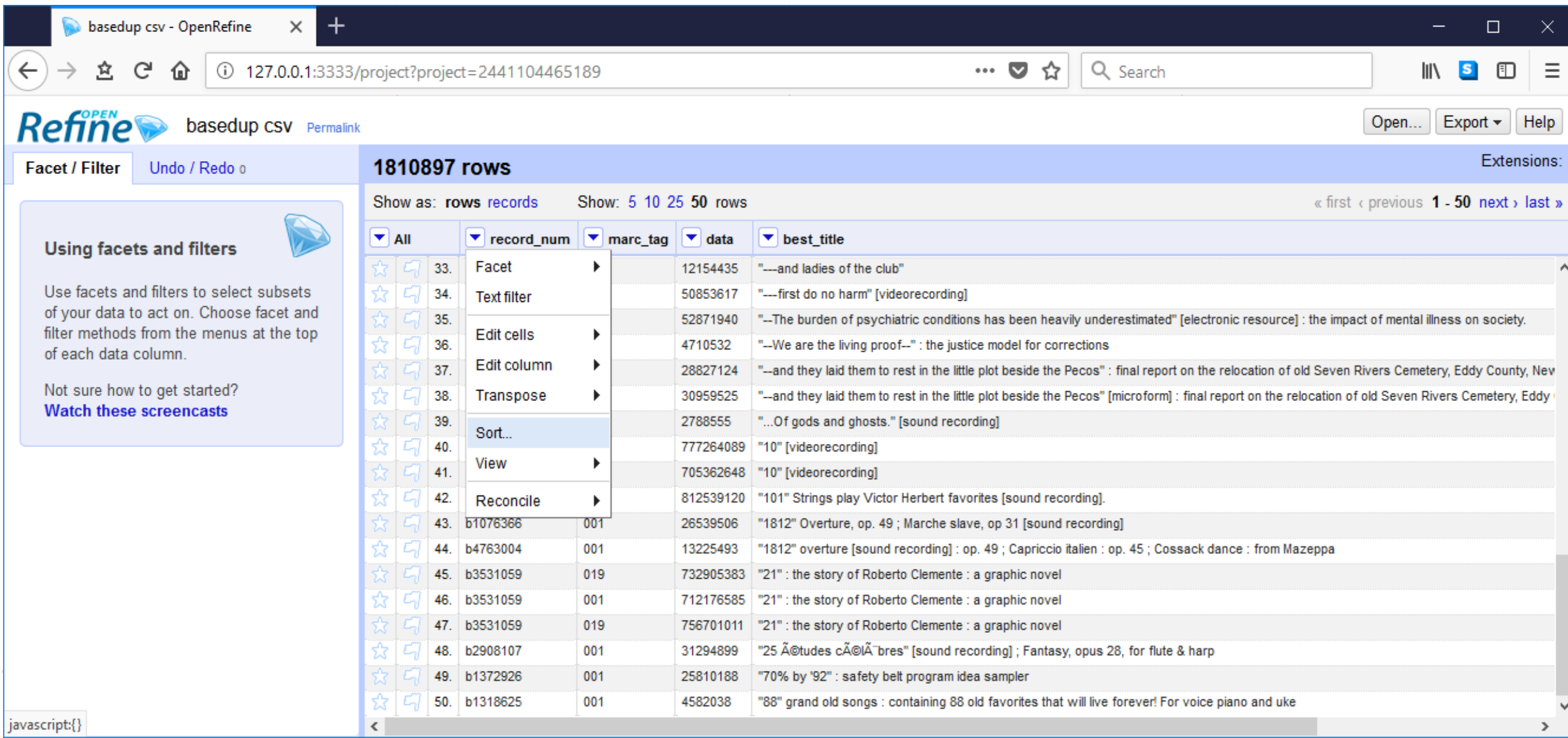
Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)



# You can sort a column



The screenshot shows the OpenRefine web interface. The browser address bar displays the URL `127.0.0.1:3333/project?project=2441104465189`. The page title is "basedup csv" with a "Permalink" link. The main content area shows a table with 181,089 rows. The table has columns: "All", "record\_num", "marc\_tag", "data", and "best\_title". A context menu is open over the "record\_num" column header, with the "Sort..." option highlighted. The table rows show various records with their respective IDs and titles.

**Using facets and filters**

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

**1810897 rows**

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

« first < previous 1 - 50 next > last »

All	record_num	marc_tag	data	best_title
☆	33.	Facet	12154435	"---and ladies of the club"
☆	34.	Text filter	50853617	"---first do no harm" [videorecording]
☆	35.	Edit cells	52871940	"--The burden of psychiatric conditions has been heavily underestimated" [electronic resource] : the impact of mental illness on society.
☆	36.	Edit column	4710532	"--We are the living proof--" : the justice model for corrections
☆	37.	Edit column	28827124	"--and they laid them to rest in the little plot beside the Pecos" : final report on the relocation of old Seven Rivers Cemetery, Eddy County, New
☆	38.	Transpose	30959525	"--and they laid them to rest in the little plot beside the Pecos" [microform] : final report on the relocation of old Seven Rivers Cemetery, Eddy
☆	39.	Sort...	2788555	"...Of gods and ghosts." [sound recording]
☆	40.	View	777264089	"10" [videorecording]
☆	41.	Reconcile	705362648	"10" [videorecording]
☆	42.		812539120	"101" Strings play Victor Herbert favorites [sound recording].
☆	43.	b1076366 001	26539506	"1812" Overture, op. 49 ; Marche slave, op 31 [sound recording]
☆	44.	b4763004 001	13225493	"1812" overture [sound recording] : op. 49 ; Capriccio italien : op. 45 ; Cossack dance : from Mazeppa
☆	45.	b3531059 019	732905383	"21" : the story of Roberto Clemente : a graphic novel
☆	46.	b3531059 001	712176585	"21" : the story of Roberto Clemente : a graphic novel
☆	47.	b3531059 019	756701011	"21" : the story of Roberto Clemente : a graphic novel
☆	48.	b2908107 001	31294899	"25 Ã©tudes cÃ©lÃ©bres" [sound recording] ; Fantasy, opus 28, for flute & harp
☆	49.	b1372926 001	25810188	"70% by '92" : safety belt program idea sampler
☆	50.	b1318625 001	4582038	"88" grand old songs : containing 88 old favorites that will live forever! For voice piano and uke



# Filter

The screenshot shows the OpenRefine web interface. The browser address bar displays the URL `127.0.0.1:3333/project?project=2441104465189`. The page title is `basedup csv`. The main content area shows a table with 1810897 rows. A context menu is open over the 'data' column, listing actions such as Facet, Text filter, Edit cells, Edit column, Transpose, Sort..., View, and Reconcile. The table columns are 'All', 'record\_num', 'marc\_tag', 'data', and 'best\_title'. The 'best\_title' column contains various text entries, some starting with exclamation marks. A sidebar on the left contains a section titled 'Using facets and filters' with a diamond icon and a paragraph of text. The bottom left corner shows a small code editor with `javascript: {}`.

Facet / Filter Undo / Redo 0

1810897 rows

Show as: rows records Show: 5 10 25 50 rows Extensions: « first < previous 1 - 50 next > last »

All	record_num	marc_tag	data	best_title
☆	1.	b3624710		7
☆	2.	b5016032		08 !Ay, Carmela!
☆	3.	b1932710		8 !Delicioso! : regional cooking of Spain
☆	4.	b3721098		32 !Dos! [sound recording]
☆	5.	b5430710		03 !Esperar no es fácil!
☆	6.	b5430710		46 !Esperar no es fácil!
☆	7.	b4921402		50 !Esplendores! Splendors of Mexico [electronic resource (video)]
☆	8.	b1520707		6 !Estoy enojado contigo!
☆	9.	b1719322		!Hablamos! Puerto Ricans speak.
☆	10.	b5161192		91 !No dejes que la paloma conduzca el autobus!
☆	11.	b1749677	001 03052440	!Oh, hermoso mundo!
☆	12.	b1322531	001 5734839	!Oh, los gringos!
☆	13.	b5283789	001 875314590	!QuÃ© confusiÃ³n!
☆	14.	b5283789	019 876821923	!QuÃ© confusiÃ³n!
☆	15.	b4742695	001 15075356	!QuÃ© gitano! : gypsies of southern Spain
☆	16.	b1956546	001 16215729	!QuÃ© gitano! : gypsies of southern Spain
☆	17.	b2638120	001 61325859	!Vamamos con Pancho Villa! [videorecording] = Let's go with Pancho Villa!
☆	18.	b1695108	001 06000387	!Vas bien! : Spanish reading keys

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

javascript: {}

# Create a facet

The screenshot shows the OpenRefine web interface. At the top, there's a browser window titled 'basedup csv - OpenRefine' with the URL '127.0.0.1:3333/project?project=2441104465189'. The main header includes the 'Refine' logo, the file name 'basedup csv', and buttons for 'Open...', 'Export', and 'Help'. Below the header, there's a 'Facet / Filter' section with 'Undo / Redo' options. A sidebar on the left contains a 'Using facets and filters' guide. The main area displays a table with 1810897 rows. A context menu is open over the 'data' column, showing options like 'Facet', 'Text filter', 'Edit cells', 'Edit column', 'Transpose', 'Sort...', 'View', and 'Reconcile'. The 'Facet' option is selected, and a sub-menu is visible with options: 'Text facet', 'Numeric facet', 'Timeline facet', 'Scatterplot facet', 'Custom text facet...', 'Custom Numeric Facet...', 'Customized facets', and 'Facet by error'. A second menu is also visible, listing various facet types: 'Word facet', 'Duplicates facet', 'Numeric log facet', '1-bounded numeric log facet', 'Text length facet', 'Log of text length facet', 'Unicode char-code facet', 'Facet by error', and 'Facet by blank'. The table data includes columns for 'record\_num', 'marc\_tag', 'data', and 'best\_title'.

Facet / Filter Undo / Redo

1810897 rows

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

All	record_num	marc_tag	data	best_title
☆	1.	b3624710	001	
☆	2.	b5016032	001	
☆	3.	b1932710	001	
☆	4.	b3721098	001	
☆	5.	b5430710	001	
☆	6.	b5430710	019	
☆	7.	b4921402	001	
☆	8.	b1520707	001	
☆	9.	b1719322	001	
☆	10.	b5161192	001	
☆	11.	b1749677	03052440	!Oh, hermoso mundo!
☆	12.	b1322531	5734839	!Oh, los gringos!
☆	13.	b5283789	875314590	!QuÃ© confusiÃ³n!
☆	14.	b5283789	876821923	!QuÃ© confusiÃ³n!
☆	15.	b4742695	15075356	!QuÃ© gitano! : gypsies of southern Spain
☆	16.	b1956546	16215729	!QuÃ© gitano! : gypsies of southern Spain
☆	17.	b2638120	61325859	!Vamamos con Pancho Villa! [videorecording] = Let's go with Pancho Villa!
☆	18.	b1695108	06000387	!Vas bien! : Spanish reading keys

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

Word facet

Duplicates facet

Numeric log facet

1-bounded numeric log facet

Text length facet

Log of text length facet

Unicode char-code facet

Facet by error

Facet by blank

Text facet

Numeric facet

Timeline facet

Scatterplot facet

Custom text facet...

Custom Numeric Facet...

Customized facets

Facet

Text filter

Edit cells

Edit column

Transpose

Sort...

View

Reconcile

# Here's a duplicate facet on the 'data' column

The screenshot shows the OpenRefine web interface. The browser address bar displays `127.0.0.1:3333/project?project=2485249919267`. The main header shows the project name "basedup csv" and a "Permalink" button. On the right, there are "Open...", "Export", and "Help" buttons. The left sidebar contains a "Facet / Filter" section with "Undo / Redo" and buttons for "Refresh", "Reset All", and "Remove All". A facet for the "data" column is active, showing two choices: "false" (1783639) and "true" (27258). The facet is highlighted with a red box. The main table displays 1810897 rows, with the first 10 rows visible. The columns are "All", "record\_num", "marc\_tag", "data", "best\_title", and "Column 5". The "data" column contains values like "48238597", "897384508", "32855788", "812343032", "952226603", "953708846", "100003250", "37678296", "211170", and "641530891".

basedup csv - OpenRefine

127.0.0.1:3333/project?project=2485249919267

Refine OPEN basedup csv Permalink

Open... Export Help

Facet / Filter Undo / Redo

Refresh Reset All Remove All

1810897 rows

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

Extensions: « first < previous 1 - 10 next > last »

data change

2 choices Sort by: name count

false 1783639

true 27258

Facet by choice counts

All	record_num	marc_tag	data	best_title	Column 5
☆	1.	b3624710	001	48238597	
☆	2.	b5016032	001	897384508	!Ay, Carmela!
☆	3.	b1932710	001	32855788	!Delicioso! : regional cooking of Spain
☆	4.	b3721098	001	812343032	!Dos! [sound recording]
☆	5.	b5430710	001	952226603	!Esperar no es fácil!
☆	6.	b5430710	019	953708846	!Esperar no es fácil!
☆	7.	b4921402	001	100003250	!Esplendores! Splendors of Mexico [electronic resource (video)]
☆	8.	b1520707	001	37678296	!Estoy enojado contigo!
☆	9.	b1719322	001	211170	!Hablamos! Puerto Ricans speak.
☆	10.	b5161192	001	641530891	!No dejes que la paloma conduzca el autobus!

There are 27,258 dups and 1,783,639 unique in the data column

basedup csv - OpenRefine

127.0.0.1:3333/project?project=2485249919267

Refine OPEN basedup csv Permalink

Open... Export Help

Facet / Filter Undo / Redo

Refresh Reset All Remove All

1810897 rows

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

Extensions: « first < previous 1 - 10 next > last »

**data** change

2 choices Sort by: name count

false 1783639

true 27258

Facet by choice counts

All	record_num	marc_tag	data	best_title	Column 5
☆	1.	b3624710	001	48238597	
☆	2.	b5016032	001	897384508	!Ay, Carmela!
☆	3.	b1932710	001	32855788	!Delicioso! : regional cooking of Spain
☆	4.	b3721098	001	812343032	!Dos! [sound recording]
☆	5.	b5430710	001	952226603	!Esperar no es fÁcil!
☆	6.	b5430710	019	953708846	!Esperar no es fÁcil!
☆	7.	b4921402	001	100003250	!Esplendores! Splendors of Mexico [electronic resource (video)]
☆	8.	b1520707	001	37678296	!Estoy enojado contigo!
☆	9.	b1719322	001	211170	!Hablamos! Puerto Ricans speak.
☆	10.	b5161192	001	641530891	!No dejes que la paloma conduzca el autobus!

# Duplicate facet and text facet

The screenshot shows the OpenRefine web interface. The browser address bar displays '127.0.0.1:3333/project?project=2485249919267'. The page title is 'basedup csv'. The main content area shows a table with 1810897 rows. Two facets are highlighted with red boxes:

- data facet:** Shows 2 choices: 'false' (1783639) and 'true' (27258). It is sorted by name and count.
- marc\_tag facet:** Shows 2 choices: '001' (1304067) and '019' (506830). It is sorted by name and count.

The main table displays the following data:

	All	record_num	marc_tag	data	best_title	Column 5
1.	b3624710	001	48238597			
2.	b5016032	001	897384508	!Ay, Carmela!		
3.	b1932710	001	32855788	!Delicioso! : regional cooking of Spain		
4.	b3721098	001	812343032	!Dos! [sound recording]		
5.	b5430710	001	952226603	!Esperar no es fÁcil!		
6.	b5430710	019	953708846	!Esperar no es fÁcil!		
7.	b4921402	001	100003250	!Esplendores! Splendors of Mexico [electronic resource (video)]		
8.	b1520707	001	37678296	!Estoy enojado contigo!		
9.	b1719322	001	211170	!Hablamos! Puerto Ricans speak.		
10.	b5161192	001	641530891	!No dejes que la paloma conduzca el autobus!		

# Facets interact. Select 019, and duplicates facet only counts those dups

The screenshot shows the OpenRefine web interface. The browser address bar displays '127.0.0.1:3333/project?project=2485249919267'. The main header shows 'Refine basedup csv' with a 'Permalink' and buttons for 'Open...', 'Export', and 'Help'. The interface is divided into a left sidebar for facets and a main table area.

**Facet / Filter** (Undo / Redo 0)

Refresh    Reset All    Remove All

**data** (change)    2 choices    Sort by: name count

- false 485422 ←
- true 21408

Facet by choice counts

**marc\_tag** (change invert reset)    2 choices    Sort by: name count    Cluster

- 001 1304067
- 019 506830 ←    exclude

Facet by choice counts

**506830 matching rows** (1810897 total)

Show as: rows records    Show: 5 10 25 50 rows    Extensions:    « first < previous 1 - 10 next > last »

	All	record_num	marc_tag	data	best_title
☆	6.	b5430710	019	953708846	!Esperar no es fÃ¡cil
☆	14.	b5283789	019	876821923	!QuÃ© confusiÃ³n!
☆	27.	b3787463	019	2335609	"... to form a more perfect union ...": justice for American women : report of the National Commission on the Observance of International Women's Day
☆	28.	b3787463	019	317547769	"... to form a more perfect union ...": justice for American women : report of the National Commission on the Observance of International Women's Day
☆	40.	b3773005	019	777264089	"10" [videorecording]
☆	45.	b3531059	019	732905383	"21" : the story of Roberto Clemente : a graphic novel
☆	47.	b3531059	019	756701011	"21" : the story of Roberto Clemente : a graphic novel
☆	56.	b1359755	019	2779780	"A glimmer of their own beauty" : Black sounds of the twenties.
☆	57.	b1359755	019	761987	"A glimmer of their own beauty" : Black sounds of the twenties.
☆	65.	b3603650	019	747782195	"A secret to be buried" : the diary and life of Emily Hawley Gillespie, 1858-1888

# Now we have all the dups, in either field

The screenshot shows the OpenRefine web interface. The browser address bar displays `127.0.0.1:3333/project?project=2485249919267`. The main header shows the project name "basedup csv" and a "Permalink" button. The interface is divided into a left sidebar for facets and filters, and a main table area.

**Facet / Filter:** The "data" facet is active, showing two choices: "false" with 1783639 items and "true" with 27258 items. A red arrow points to the "true" count. The facet is sorted by "name" and "count".

**Table:** The table displays 27258 matching rows (out of 1810897 total). The columns are: "All", "record\_num", "marc\_tag", "data", "best\_title", and "Column 5". The "data" column contains values like "001" and "019". The "best\_title" column contains titles such as "Blaze finds the trail", "A treasury of horse stories", "Death of the Thresher", "Dinosaur time", "Latin America", "Mine for keeps", "Constitutional history of England", and "The lord of the rings".

All	record_num	marc_tag	data	best_title	Column 5
<input type="checkbox"/>	213910.	b3808574	001	00102248	Blaze finds the trail
<input type="checkbox"/>	213911.	b4260777	001	00102248	Blaze finds the trail
<input type="checkbox"/>	213912.	b4966398	001	00102248	Blaze finds the trail
<input type="checkbox"/>	53406.	b2647866	001	00626214	A treasury of horse stories,
<input type="checkbox"/>	53407.	b1880668	001	00626214	A treasury of horse stories,
<input type="checkbox"/>	380316.	b2224108	001	00711149	Death of the Thresher.
<input type="checkbox"/>	380317.	b1491372	001	00711149	Death of the Thresher.
<input type="checkbox"/>	408031.	b3921837	001	00942820	Dinosaur time
<input type="checkbox"/>	408036.	b1016130	001	00942820	Dinosaur time.
<input type="checkbox"/>	812501.	b1912455	001	09389862	Latin America
<input type="checkbox"/>	812502.	b2173406	001	09389862	Latin America
<input type="checkbox"/>	916454.	b1152612	019	1003685	Mine for keeps
<input type="checkbox"/>	916457.	b2344211	019	1003685	Mine for keeps.
<input type="checkbox"/>	332073.	b4689310	001	10047064	Constitutional history of England
<input type="checkbox"/>	332076.	b1731157	019	10047064	Constitutional history of England,
<input type="checkbox"/>	1551988.	b2214944	001	10051170	The lord of the rings
<input type="checkbox"/>	1635353.	b1562246	019	10051170	The two towers,
<input type="checkbox"/>	176541.	b1235707	001	10064554	Barbed wit & malicious humor
<input type="checkbox"/>	176543.	b1760893	019	10064554	Barbed wit & malicious humor



# I want to export the record numbers of the duplicates

The screenshot shows the OpenRefine web interface. The browser address bar displays `127.0.0.1:3333/project?project=2485249919267`. The main header shows the project name "basedup csv" and a "Permalink" button. Below the header, there are buttons for "Facet / Filter", "Undo / Redo", "Refresh", "Reset All", and "Remove All".

The main content area displays a table with the following columns: "All", "record\_num", "marc\_tag", "data", and "best\_title". The table contains 27258 rows, with the first few rows being duplicates. A red arrow points to the "true" filter option in the "data" facet, which is currently selected. Another red arrow points to the "Custom tabular exporter..." option in the "Export" dropdown menu.

All	record_num	marc_tag	data	best_title	
<input type="checkbox"/>	558.	b1245887	001	7891139	"If elected ... " : unsuccessful candidates for the presidency, 1796-1968
<input type="checkbox"/>	560.	b1251654	019	7891139	"If elected--" : unsuccessful candidates for the presidency, 1796-1968
<input type="checkbox"/>	590.	b4796035	019	844052204	"Islam" means peace : understanding the muslim principle of nonviolence toda
<input type="checkbox"/>	592.	b3607174	019	844052204	"Islam" means peace [electronic resource] : understanding the muslim princip
<input type="checkbox"/>	771.	b4779012	019	48094256	"Mr. Charles W. Chesnutt's stories"
<input type="checkbox"/>	772.	b5070835	019	48094256	"Mr. Charles W. Chesnutt's stories"
<input type="checkbox"/>	788.	b4796156	019	732959267	"My nerves are bad" : Puerto Rican women managing mental illness and HIV
<input type="checkbox"/>	791.	b3608519	019	732959267	"My nerves are bad" [electronic resource] : Puerto Rican women managing m
<input type="checkbox"/>	1032.	b4778715	019	81063989	"Speaking of operations--"
<input type="checkbox"/>	1033.	b5069152	019	81063989	"Speaking of operations--"
<input type="checkbox"/>	1034.	b4778715	019	611339991	"Speaking of operations--"
<input type="checkbox"/>	1035.	b5069152	019	611339991	"Speaking of operations--"
<input type="checkbox"/>	1134.	b4617684	001	53369171	"The Utes must go!" : American expansion and the removal of a people
<input type="checkbox"/>	1135.	b2493026	001	53369171	"The Utes must go!" : American expansion and the removal of a people
<input type="checkbox"/>	1167.	b1826577	001	11796942	"The good war" : an oral history of World War Two
<input type="checkbox"/>	1168.	b1055076	019	11796942	"The good war" : an oral history of World War Two
<input type="checkbox"/>	1267.	b4789090	019	786277855	"To serve a larger purpose" : engagement for democracy and the transformation of higher education
<input type="checkbox"/>	1269.	b3607330	019	786277855	"To serve a larger purpose" [electronic resource] : engagement for democracy and the transformation of higher education



# Select only record\_num column

The image shows a screenshot of the OpenRefine web interface. A dialog box titled "Custom Tabular Exporter" is open, showing options for exporting data. The "Content" tab is selected, and the "record\_num" column is chosen for export. The "Options for record\_num" section is expanded, showing various output options for reconciled cells and unmatched cells. The "record\_num" column is selected in the "Select and Order Columns to Export" list. The "Options for record\_num" section is expanded, showing options for outputting entity names, IDs, or links, and for handling unmatched cells. The "record\_num" column is selected in the "Select and Order Columns to Export" list. The "Options for record\_num" section is expanded, showing options for outputting entity names, IDs, or links, and for handling unmatched cells.

basedup csv - OpenRefine

127.0.0.1:3333/project?project=2485249919267

### Custom Tabular Exporter

Content Download Upload Option Code

Select and Order Columns to Export

- record\_num
- marc\_tag
- data
- best\_title
- Column 5

Options for record\_num

For reconciled cells, output

- Matched entity's name
- Matched entity's ID
- Link to matched entity's page
- Cell's content

Output nothing for unmatched cells

- ISO 8601, e.g., 2011-08-24T18:36:10+08:00
- Short locale format
- Long locale format
- Custom
- Medium locale format
- Full locale format

Use local time zone  Omit time  [Help](#)

Select All De-select All

- Output column headers
- Output empty rows (ie all cells null)
- Ignore facets and filters and export all rows

# Use Import File function in Create Lists

Import Records (into review file #20)

Review File Name: 019 dups

Store Record Type: BIBLIOGRAPHIC b

Choose File

b1245887  
b1251654  
b4796035  
b3607174  
b1770010

Import Cancel

# Now you have a list of dups to clean up

Boolean Review File:019 dups

File Tools

Boolean Review File:019 dups

Add Edit Remove Dedupe Close

	Record	Description
<input type="checkbox"/>	1 b18265777	"The good war" : an oral history of World War Two / Studs Terkel.
<input type="checkbox"/>	2 b10550768	"The good war" : an oral history of World War Two / Studs Terkel.
<input type="checkbox"/>	3 b29339303	(500) days of Summer [videorecording] / Fox Searchlight Pictures presents a Watermark production ; produced by Jessica Tuchinsky, Mark Waters, Mason ...
<input type="checkbox"/>	4 b34146192	(500) days of Summer [videorecording] / Fox Searchlight Pictures presents a Watermark production ; director of photography, Eric Steelberg ; produced by J...
<input type="checkbox"/>	5 b13217446	200 songs in three volumes for voice and piano / Schubert ; [edited by] Sergius Kagen ; English translations and notes by Gerard Mackworth-Young.
<input type="checkbox"/>	6 b47189460	200 songs in three volumes, for voice and piano. / [Selected by] Sergius Kagen. English translations and notes by Gerard Mackworth-Young. High.
<input type="checkbox"/>	7 b43634321	24 songs for voice and piano / Mahler.
<input type="checkbox"/>	8 b17944934	24 songs for voice and piano (high) / Mahler.
<input type="checkbox"/>	9 b15786407	A Christmas carol / Charles Dickens ; [illustrated by] Roberto Innocenti.
<input type="checkbox"/>	10 b47767789	A Christmas carol / by Charles Dickens.
<input type="checkbox"/>	11 b13655802	A Guide for the management, analysis, and interpretation of occupational mortality data / Nina Lalich ... [and others].
<input type="checkbox"/>	12 b25442958	A Guide for the management, analysis, and interpretation of occupational mortality data / Nina Lalich ... [and others].
<input type="checkbox"/>	13 b54276275	A Multi-Industrial Linkages Approach to Cluster Building in East Asia [electronic resource] : Targeting the Agriculture, Food, and Tourism Industry / edited by A...
<input type="checkbox"/>	14 b54085068	A Multi-Industrial Linkages Approach to Cluster Building in East Asia: Targeting the Agriculture, Food, and Tourism Industry.
<input type="checkbox"/>	15 b12771788	A Rocket in my pocket : the rhymes and chants of young Americans / compiled by Carl Withers ; illustrated by Susanne Suba.
<input type="checkbox"/>	16 b15652919	A Rocket in my pocket : the rhymes and chants of young Americans / compiled by Carl Withers ; illustrated by Susanne Suba.
<input type="checkbox"/>	17 b20897583	A Thurber carnival / by James Thurber.
<input type="checkbox"/>	18 b12123018	A Thurber carnival / by James Thurber.
<input type="checkbox"/>	19 b35976494	A dictionary of contemporary history, 1945 to the present [electronic resource] / Duncan Townson.
<input type="checkbox"/>	20 b34162549	A dictionary of contemporary history, 1945 to the present [electronic resource] / Duncan Townson.
<input type="checkbox"/>	21 b13415281	A dictionary of the Choctaw language / by Cyrus Byington. Ed. by John R. Swanton and Henry S. Halbert.
<input type="checkbox"/>	22 b47134264	A dictionary of the Choctaw language / Edited by John R. Swanton and Henry S. Halbert.

## Learn more about OpenRefine

- <https://data-lessons.github.io/library-openrefine/>



Questions about OpenRefine?

