

# Basic Finding Aid Clean-Up Using OpenRefine

Presented by: Bailey Hoffner, Metadata and Collections  
Management Archivist

# What is OpenRefine?

---

An open-source,  
browser-based tool for  
cleaning up messy data

<https://openrefine.org>



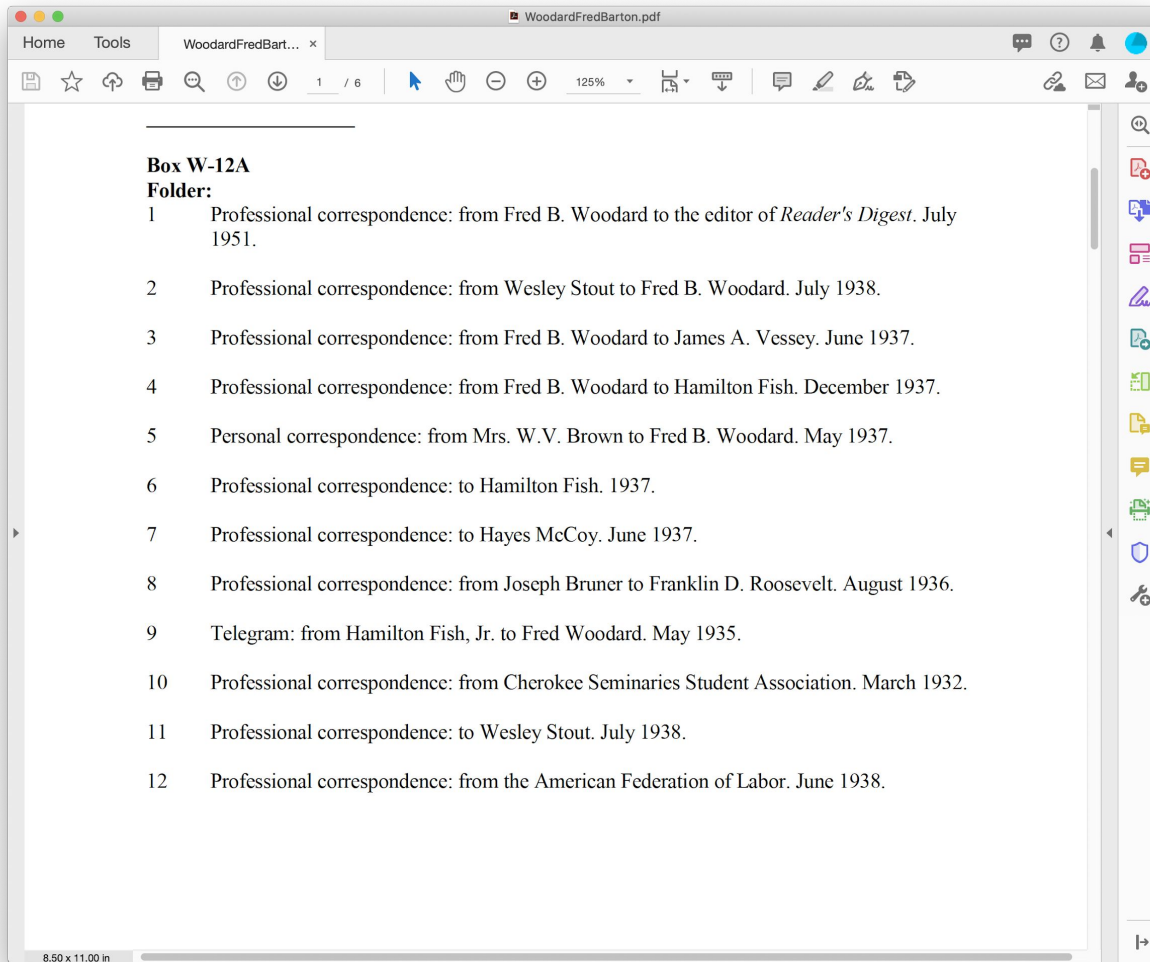
**OpenRefine**

# Why OpenRefine for Archivists?

- low-touch approach to transforming legacy finding aids
- batch fix human error
- encourages understanding of finding aids as iterative documents
- structured data = more options in the future



# Goal: Structured Data



training-ex-xlsx.xls - Compatibility Mode

Home Insert Draw Page Layout Formulas Data Review View Acrobat Tell me

	A	B	C	D	E	F
1	Folder	Title	Date	subject		
2	1	Professional correspondence: from Fred B. Woodard to the editor of	July 1951.	Archer, Ada (Jones)		
3	2	Professional correspondence: from Wesley Stout to Fred B. Woodard	July 1938.	Walker, Tim		
4	3	Professional correspondence: from Fred B. Woodard to James A. V	June 1937.	Wilson, Florence		
5	4	Professional correspondence: from Fred B. Woodard to Hamilton Fish	December 1937.	Cherokee Indians		
6	5	Personal correspondence: from Mrs. W.V. Brown to Fred B. Woodard	May 1937.	Cherokee Male Seminary		
7	6	Professional correspondence: to Hamilton Fish.	1937.	Cherokee Language		
8	7	Professional correspondence: to Hayes McCoy.	June 1937.	Sanders, Soggy		
9	8	Professional correspondence: from Joseph Bruner to Franklin D. Ro	August 1936.	Alberty, Gerry		
10	9	Telegram: from Hamilton Fish, Jr. to Fred Woodard.	May 1935.	Boudinot, Elias		
11	10	Professional correspondence: from Cherokee Seminaries Student A	March 1932.	Watie, Buck		
12	11	Professional correspondence: to Wesley Stout.	July 1938.	Duncan, DeWitt Clinton		
13	12	Professional correspondence: from the American Federation of Lab	June 1938.	Tahlequah		
14	13	Professional correspondence: from Hamilton Fish to Fred B. Wooda	1938.	Indian Schools		
15	14	Professional correspondence: from W.P. Tomson to Fred Woodard	1938.	Cherokee Female Seminary		
16	15	Professional correspondence: from National Legislative representati	June 1938.	McSpadden, Florence		
17	16	Professional correspondence: from L.M. Bailey.	September-October	Mayes, Patsy		
18	17	Professional correspondence: to Hamilton Fish.	1938-39.	Adair, Bluiie		
19	18	Professional correspondence: from Fred B. Woodard to Benard Hef	August 1938.	Miller, Mattie		
20	19	Professional correspondence: from Jas A. Veasey to Fred B. Wood	August 1938.	Starr, Lulu		
21	20	Professional correspondence: concerning the Keep America Out of	1936-1939.	Prather, Ella		
22	21	Professional correspondence: from Fred B. Woodard to Florence Bc	October 1943.	Thompson, Babe		
23	22	Professional correspondence: from Fred B. Woodard to Ralph Mitch	June 1941.	Schimpsher, Bunt		

training ex xlsx +

Ready 204%



# PDF → Spreadsheet

	A	B	C
1	data	subject	
2	1 Professional correspondence: from Fred B. Woodard to the editor of <i>Reader's Digest</i> . July 1951.	Archer, Ada (Jones)	
3	2 Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	Walker, Tim	
4	3 Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	Wilson, Florence	
5	4 Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	Cherokee Indians	
6	5 Personal correspondence: from Mrs. W.V. Brown to Fred B. Woodard. May 1937.	Cherokee Male Seminary	
7	6 Professional correspondence: to Hamilton Fish. 1937.	Cherokee Language	
8	7 Professional correspondence: to Hayes McCoy. June 1937.	Sanders, Soggy	
9	8 Professional correspondence: from Joseph Bruner to Franklin D. Roosevelt. August 1936.	Alberty, Gerry	
10	9 Telegram: from Hamilton Fish, Jr. to Fred Woodard. May 1935.	Boudinot, Elias	
11	10 Professional correspondence: from Cherokee Seminaries Student Association. March 1932.	Watie, Buck	
12	11 Professional correspondence: to Wesley Stout. July 1938.	Duncan, DeWitt Clinton	
13	12 Professional correspondence: from the American Federation of Labor. June 1938.	Tahlequah	
14	13 Professional correspondence: from Hamilton Fish to Fred B. Woodard. 1938.	Indian Schools	
15	14 Professional correspondence: from W.P. Tomson to Fred Woodard and from Fred Woodard to Hamilton Fish. 1938.	Cherokee Female Seminary	
16	15 Professional correspondence: from National Legislative representatives. June 1938.	McSpadden, Florence	
17	16 Professional correspondence: from L.M. Bailey. September-October 1939.	Mayes, Patsy	
18	17 Professional correspondence: to Hamilton Fish. 1938-39.	Adair, Bluie	
19	18 Professional correspondence: from Fred B. Woodard to Benard Hefferman. August 1938.	Miller, Mattie	
20	19 Professional correspondence: from Jas A. Veasey to Fred B. Woodard. August 1938.	Starr, Lulu	
21	20 Professional correspondence: concerning the Keep America Out of War Committee and flyers. 1936-1939.	Prather, Ella	

# OpenRefine: Initial Formatting Options

The screenshot shows the OpenRefine web interface. At the top, the browser address bar shows '127.0.0.1'. The OpenRefine logo and tagline 'A power tool for working with messy data.' are visible. The main interface is divided into several sections:

- Project Information:** Project name is 'training ex.xlsx'. There are buttons for '« Start Over', 'Configure Parsing Options', and 'Create Project »'.
- Data Table:** A table with two columns: 'data' and 'subject'. It contains 14 rows of professional correspondence records.
- Parse data as:** A section for selecting the file format. Under 'Excel files', 'training-ex.xlsx#Sheet1' is selected with 85 rows. Other options include JSON files, CSV/TSV, Fixed-width field text files, PC-Axis text files, MARC files, JSON-LD files, RDF/N3 files, and RDF/N-Triples files.
- Import Options:** Includes 'Worksheets to Import' (Select All, Unselect All), 'Ignore first 0 line(s) at beginning of file', 'Parse next 1 line(s) as column headers', 'Discard initial 0 row(s) of data', and 'Load at most 0 row(s) of data'.
- Storage Options:** Includes 'Store blank rows' (checked), 'Store blank cells as nulls' (checked), and 'Store file source (file names, URLs) in each row' (unchecked).
- Footer:** Version 3.3 [58b839b], Preferences, Help, and About links.

# OpenRefine: Trim Whitespace

The screenshot shows the OpenRefine web interface. The browser address bar displays '127.0.0.1'. The page title is 'OpenRefine training ex.xlsx'. The interface includes a 'Facet / Filter' sidebar on the left with a 'Using facets and filters' tip. The main area shows a table with 84 rows. The 'data' column is selected, and the 'Edit cells' menu is open, with 'Common transforms' and 'Trim leading and trailing whitespace' highlighted. The table has columns for 'data' and 'subject'. The 'subject' column contains names like 'Archer, Ada (Jones)', 'Walker, Tim', 'Wilson, Florence', etc.

	data	subject
1.	Facet	Archer, Ada (Jones)
2.	Text filter	Walker, Tim
3.	Edit cells	Wilson, Florence
4.	Edit column	Cherokee Indians
5.	Transpose	Cherokee Male Seminary
6.	Sort...	Cherokee Language
7.	View	Sanders, Soggy
8.	Reconcile	Alberty, Gerry
9.		Boudinot, Elias
10.		Watie, Buck
11.	11 Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest. July 1951.	Duncan, DeWitt Clinton
12.	12 Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	Tahlequah
13.	13 Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	Indian Schools
14.	14 Professional correspondence: from W.P. Tomson to Fred B. Woodard. August 1900.	Cherokee Female Seminary
15.	15 Professional correspondence: from National Legislative Committee on Indian Affairs to Fred B. Woodard. August 1938.	McSpadden, Florence
16.	16 Professional correspondence: from L.M. Bailey. September 1938.	Mayes, Patsy
17.	17 Professional correspondence: to Hamilton Fish. December 1937.	Adair, Bluite
18.	18 Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	Miller, Mattie
19.	19 Professional correspondence: from Jas A. Veasey to Fred B. Woodard. August 1900.	Starr, Lulu
20.	20 Professional correspondence: concerning the Keep America Out of War Committee and flyers. 1936-1939.	Prather, Ella
21.	21 Professional correspondence: from Fred B. Woodard to Florence Booker. October 1943.	Thompson, Babe
22.	22 Professional correspondence: from Fred B. Woodard to Ralph Mitchel. June 1941.	Schimpsher, Bunt
23.	23 Professional correspondence: from Harve Tibbott to Mrs. John Edminston.	Duncan, Mae
24.	24 Personal correspondence: from Ralph to Fred. June 1941.	Carter, Ella
25.	25 Personal correspondence: from Fred B. Woodard. 1937-1942.	Nash, Fannie
26.	26 Professional correspondence: from Grant Foreman to Fred B. Woodard. October 1943.	Ingram
27.	27 Professional correspondence: from John Kane to Fred B. Woodard. 1943-45.	Morgan, Llwellyn



# OpenRefine: Trim Whitespace

OpenRefine training ex.xlsx Permalink

Text transform on 84 cells in column data: value.trim() Undo

Open... Export Help

Facet / Filter Undo / Redo 1 / 1

84 rows Extensions: Wikidata

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

All	data	subject
☆	1. 1 Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest. July 1951.	Archer, Ada (Jones)
☆	2. 2 Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	Walker, Tim
☆	3. 3 Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	Wilson, Florence
☆	4. 4 Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	Cherokee Indians
☆	5. 5 Personal correspondence: from Mrs. W.V. Brown to Fred B. Woodard. May 1937.	Cherokee Male Seminary
☆	6. 6 Professional correspondence: to Hamilton Fish. 1937.	Cherokee Language
☆	7. 7 Professional correspondence: to Hayes McCoy. June 1937.	Sanders, Soggy
☆	8. 8 Professional correspondence: from Joseph Bruner to Franklin D. Roosevelt. August 1936.	Alberty, Gerry
☆	9. 9 Telegram: from Hamilton Fish, Jr. to Fred Woodard. May 1935.	Boudinot, Elias
☆	10. 10 Professional correspondence: from Cherokee Seminaries Student Association. March 1932.	Watie, Buck
☆	11. 11 Professional correspondence: to Wesley Stout. July 1938.	Duncan, DeWitt Clinton
☆	12. 12 Professional correspondence: from the American Federation of Labor. June 1938.	Tahlequah
☆	13. 13 Professional correspondence: from Hamilton Fish to Fred B. Woodard. 1938.	Indian Schools
☆	14. 14 Professional correspondence: from W.P. Tomson to Fred Woodard and from Fred Woodard to Hamilton Fish. 1938.	Cherokee Female Seminary
☆	15. 15 Professional correspondence: from National Legislative representatives. June 1938.	McSpadden, Florence
☆	16. 16 Professional correspondence: from L.M. Bailey. September-October 1939.	Mayes, Patsy
☆	17. 17 Professional correspondence: to Hamilton Fish. 1938-39.	Adair, Bluie
☆	18. 18 Professional correspondence: from Fred B. Woodard to Benard Heffernan. August 1938.	Miller, Mattie
☆	19. 19 Professional correspondence: from Jas A. Veasey to Fred B. Woodard. August 1938.	Starr, Lulu
☆	20. 20 Professional correspondence: concerning the Keep America Out of War Committee and flyers. 1936-1939.	Prather, Ella
☆	21. 21 Professional correspondence: from Fred B. Woodard to Florence Booker. October 1943.	Thompson, Babe
☆	22. 22 Professional correspondence: from Fred B. Woodard to Ralph Mitchel. June 1941.	Schimpsher, Bunt
☆	23. 23 Professional correspondence: from Harve Tibbott to Mrs. John Edminston.	Duncan, Mae
☆	24. 24 Personal correspondence: from Ralph to Fred. June 1941.	Carter, Ella
☆	25. 25 Personal correspondence: from Fred B. Woodard. 1937-1942.	Nash, Fannie
☆	26. 26 Professional correspondence: from Grant Foreman to Fred B. Woodard. October 1943.	Ingram
☆	27. 27 Professional correspondence: from John Kane to Fred B. Woodard. 1943-45.	Morgan, Llwellyn

# OpenRefine: Split Columns

OpenRefine training ex.xlsx Permalink

Open... Export Help

Facet / Filter Undo / Redo 2 / 2

84 rows Extensions: Wikidata

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

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	data	subject
1.	Facet	Archer, Ada (Jones)
2.	Text filter	Walker, Tim
3.	Edit cells	Wilson, Florence
4.	Edit column	Cherokee Indians
5.	Transpose	Cherokee Male Seminary
6.	Sort...	Cherokee Language
7.	View	Sanders, Soggy
8.	Reconcile	Alberty, Gerry
9.		Boudinot, Elias
10.		Watie, Buck
11.	11 Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest. July 1951.	Duncan, DeWitt Clinton
12.	12 Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	Tahlequah
13.	13 Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	Indian Schools
14.	14 Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	Cherokee Female Seminary
15.	15 Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	McSpadden, Florence
16.	16 Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	Mayes, Patsy
17.	17 Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	Adair, Bluite
18.	18 Professional correspondence: from Fred B. Woodard to Benard Heffernan. August 1938.	Miller, Mattie
19.	19 Professional correspondence: from Jas A. Veasey to Fred B. Woodard. August 1938.	Starr, Lulu
20.	20 Professional correspondence: concerning the Keep America Out of War Committee and flyers. 1936-1939.	Prather, Ella
21.	21 Professional correspondence: from Fred B. Woodard to Florence Booker. October 1943.	Thompson, Babe
22.	22 Professional correspondence: from Fred B. Woodard to Ralph Mitchell. June 1941.	Schimpsher, Bunt
23.	23 Professional correspondence: from Harve Tibbott to Mrs. John Edminston.	Duncan, Mae
24.	24 Personal correspondence: from Ralph to Fred. June 1941.	Carter, Ella
25.	25 Personal correspondence: from Fred B. Woodard. 1937-1942.	Nash, Fannie
26.	26 Professional correspondence: from Grant Foreman to Fred B. Woodard. October 1943.	Ingram
27.	27 Professional correspondence: from John Kane to Fred B. Woodard. 1943-45.	Morgan, Llwellyn

# OpenRefine: Split Columns

The screenshot shows the OpenRefine web interface. At the top, it says "OpenRefine training ex.xlsx" with a "Permalink" button. On the right, there are "Open...", "Export", and "Help" buttons. Below the header, there's a "Facet / Filter" section on the left and a "84 rows" section on the right. The "84 rows" section shows a table with columns "All", "data", and "subject". The "data" column contains text entries, and the "subject" column contains names and locations. A dialog box titled "Split column data into several columns" is open in the center. It has two main sections: "How to Split Column" and "After Splitting". In "How to Split Column", the "by separator" option is selected. The "Separator" field is empty, and the "regular expression" checkbox is unchecked. The "Split into" field is set to "2" columns at most. In "After Splitting", the "Guess cell type" and "Remove this column" checkboxes are checked. At the bottom of the dialog, there are "OK" and "Cancel" buttons.

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](#)

84 rows

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

Extensions: Wikidata

« first < previous 1 - 50 next > last »

▼ All ▼ data ▼ subject

1.	1 Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest. July 1951.	Archer, Ada (Jones)
2.	2 Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	Walker, Tim
3.	3 Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	Wilson, Florence
		Cherokee Indians
		Cherokee Male Seminary
		Cherokee Language
		Sanders, Soggy
		Alberty, Gerry
		Boudinot, Elias
		Watie, Buck
		Duncan, DeWitt Clinton
		Tahlequah
		Indian Schools
		Fish. 1938. Cherokee Female Seminary
		McSpadden, Florence
		Mayes, Patsy
		Adair, Bluie
		Miller, Mattie
		Starr, Lulu
		Prather, Ella
		Thompson, Babe
		Schimpsher, Bunt
		Duncan, Mae
		Carter, Ella
		Nash, Fannie
		Ingram
		Morgan, Llwellyn

**Split column data into several columns**

**How to Split Column**

by separator

Separator   regular expression

Split into  columns at most (leave blank for no limit)

by field lengths

List of integers separated by commas, e.g., 5, 7, 15

**After Splitting**

Guess cell type

Remove this column

OK Cancel



# OpenRefine: Split Columns

OpenRefine training ex.xlsx Permalink

Split 84 cell(s) in column data into several columns by separator Undo

Open... Export Help

Facet / Filter Undo / Redo 3 / 3

84 rows Extensions: Wikidata

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

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	data 1	data 2	subject
1.	1	Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest. July 1951.	Archer, Ada (Jones)
2.	2	Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	Walker, Tim
3.	3	Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	Wilson, Florence
4.	4	Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	Cherokee Indians
5.	5	Personal correspondence: from Mrs. W.V. Brown to Fred B. Woodard. May 1937.	Cherokee Male Seminary
6.	6	Professional correspondence: to Hamilton Fish. 1937.	Cherokee Language
7.	7	Professional correspondence: to Hayes McCoy. June 1937.	Sanders, Soggy
8.	8	Professional correspondence: from Joseph Bruner to Franklin D. Roosevelt. August 1936.	Alberty, Gerry
9.	9	Telegram: from Hamilton Fish, Jr. to Fred Woodard. May 1935.	Boudinot, Elias
10.	10	Professional correspondence: from Cherokee Seminaries Student Association. March 1932.	Watie, Buck
11.	11	Professional correspondence: to Wesley Stout. July 1938.	Duncan, DeWitt Clinton
12.	12	Professional correspondence: from the American Federation of Labor. June 1938.	Tahlequah
13.	13	Professional correspondence: from Hamilton Fish to Fred B. Woodard. 1938.	Indian Schools
14.	14	Professional correspondence: from W.P. Tomson to Fred Woodard and from Fred Woodard to Hamilton Fish. 1938.	Cherokee Female Seminary
15.	15	Professional correspondence: from National Legislative representatives. June 1938.	McSpadden, Florence
16.	16	Professional correspondence: from L.M. Bailey. September-October 1939.	Mayes, Patsy
17.	17	Professional correspondence: to Hamilton Fish. 1938-39.	Adair, Bluite
18.	18	Professional correspondence: from Fred B. Woodard to Benard Hefferman. August 1938.	Miller, Mattie
19.	19	Professional correspondence: from Jas A. Veasey to Fred B. Woodard. August 1938.	Starr, Lulu
20.	20	Professional correspondence: concerning the Keep America Out of War Committee and flyers. 1936-1939.	Prather, Ella
21.	21	Professional correspondence: from Fred B. Woodard to Florence Booker. October 1943.	Thompson, Babe
22.	22	Professional correspondence: from Fred B. Woodard to Ralph Mitchel. June 1941.	Schimpsher, Bunt
23.	23	Professional correspondence: from Harve Tibbott to Mrs. John Edminston.	Duncan, Mae
24.	24	Personal correspondence: from Ralph to Fred. June 1941.	Carter, Ella
25.	25	Personal correspondence: from Fred B. Woodard. 1937-1942.	Nash, Fannie
26.	26	Professional correspondence: from Grant Foreman to Fred B. Woodard. October 1943.	Ingram
27.	27	Professional correspondence: from John Kane to Fred B. Woodard. 1943-45.	Morgan, Llwellyn

# OpenRefine: Cluster & Edit

The screenshot shows the OpenRefine web interface. At the top, the browser address bar displays '127.0.0.1'. The OpenRefine logo and 'training ex.xlsx' are visible. The main area shows a table with 84 rows. The table has columns for 'All', 'data 1', 'data 2', and 'subject'. A context menu is open over the 'subject' column, listing options such as 'Facet', 'Text filter', 'Edit cells', 'Edit column', 'Transpose', 'Sort...', 'View', 'Reconcile', 'Cluster and edit...', and 'Replace'. The 'Cluster and edit...' option is highlighted. On the left, a sidebar titled 'Using facets and filters' provides instructions on how to use facets and filters to select subsets of data.

**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	data 1	data 2	subject
1.	1	Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest. July 1951.	
2.	2	Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	
3.	3	Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	
4.	4	Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1938.	
5.	5	Personal correspondence: from Mrs. W.V. Brown to Fred B. Woodard. May 1937.	
6.	6	Professional correspondence: to Hamilton Fish. 1937.	
7.	7	Professional correspondence: to Hayes McCoy. June 1937.	
8.	8	Professional correspondence: from Joseph Bruner to Franklin D. Roosevelt. August 1938.	
9.	9	Telegram: from Hamilton Fish, Jr. to Fred Woodard. May 1935.	
10.	10	Professional correspondence: from Cherokee Seminaries Student Association. May 1938.	
11.	11	Professional correspondence: to Wesley Stout. July 1938.	Duncan, DeWitt Clinton
12.	12	Professional correspondence: from the American Federation of Labor. June 1938.	Tahlequah
13.	13	Professional correspondence: from Hamilton Fish to Fred B. Woodard. 1938.	Indian Schools
14.	14	Professional correspondence: from W.P. Tomson to Fred Woodard and from Fred Woodard to Hamilton Fish. 1938.	Cherokee Female Seminary
15.	15	Professional correspondence: from National Legislative representatives. June 1938.	McSpadden, Florence
16.	16	Professional correspondence: from L.M. Bailey. September-October 1939.	Mayes, Patsy
17.	17	Professional correspondence: to Hamilton Fish. 1938-39.	Adair, Bluite
18.	18	Professional correspondence: from Fred B. Woodard to Benard Hefferman. August 1938.	Miller, Mattie
19.	19	Professional correspondence: from Jas A. Veasey to Fred B. Woodard. August 1938.	Starr, Lulu
20.	20	Professional correspondence: concerning the Keep America Out of War Committee and flyers. 1936-1939.	Prather, Ella
21.	21	Professional correspondence: from Fred B. Woodard to Florence Booker. October 1943.	Thompson, Babe
22.	22	Professional correspondence: from Fred B. Woodard to Ralph Mitchel. June 1941.	Schimpsher, Bunt
23.	23	Professional correspondence: from Harve Tibbott to Mrs. John Edminston.	Duncan, Mae
24.	24	Personal correspondence: from Ralph to Fred. June 1941.	Carter, Ella
25.	25	Personal correspondence: from Fred B. Woodard. 1937-1942.	Nash, Fannie
26.	26	Professional correspondence: from Grant Foreman to Fred B. Woodard. October 1943.	Ingram
27.	27	Professional correspondence: from John Kane to Fred B. Woodard. 1943-45.	Morgan, Llwellyn

# OpenRefine: Cluster & Edit

The screenshot shows the OpenRefine interface with a modal dialog titled "Cluster & Edit column 'subject'". The dialog includes a description of the clustering feature, configuration options for the clustering method (nearest neighbor, levenshtein), radius (1.0), and block characters (6). It displays two clusters found. The first cluster has a size of 2 and a row count of 2, containing "Fort Gibson" and "Fort Gibsson". The second cluster has a size of 2 and a row count of 5, containing "Tahlequah" and "Talequah". The dialog also features sliders for "# Rows in Cluster" (set to 2-5) and "Average Length of Choices" (set to 8.5-11.5). At the bottom, there are buttons for "Select All", "Unselect All", "Export Clusters", "Merge Selected & Re-Cluster", "Merge Selected & Close", and "Close".

**Cluster & Edit column "subject"**

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more...](#)

Method: nearest neighbor | levenshtein | Radius: 1.0 | Block Chars: 6 | 2 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
2	2	<ul style="list-style-type: none"><li>Fort Gibson (1 rows)</li><li>Fort Gibsson (1 rows)</li></ul> <a href="#">Browse this cluster</a>	<input type="checkbox"/>	Fort Gibson
2	5	<ul style="list-style-type: none"><li>Tahlequah (4 rows)</li><li>Talequah (1 rows)</li></ul>	<input type="checkbox"/>	Tahlequah

# Rows in Cluster: 2 — 5

Average Length of Choices: 8.5 — 11.5

Select All | Unselect All | Export Clusters | Merge Selected & Re-Cluster | Merge Selected & Close | Close



# OpenRefine: Regular Expressions

The screenshot shows the OpenRefine web interface. At the top, the browser address bar shows '127.0.0.1'. The OpenRefine header includes the logo, the text 'OpenRefine training ex.xlsx', a 'Permalink' link, and buttons for 'Open...', 'Export', and 'Help'. Below the header, there are tabs for 'Facet / Filter' and 'Undo / Redo 4 / 5'. The main area displays '84 rows' and a table with columns: 'All', 'data 1', 'data 2', and 'subject'. A context menu is open over the 'data 2' column, listing actions such as 'Facet', 'Text filter', 'Edit cells', 'Edit column', 'Transpose', 'Sort...', 'View', 'Reconcile', 'Rename this column', 'Remove this column', 'Move column to beginning', 'Move column to end', 'Move column left', and 'Move column right'. The 'Edit column' option is highlighted. A sidebar on the left contains a 'Using facets and filters' section with instructions and a link to 'Watch these screencasts'. The table data includes rows with dates and names, such as 'Archer, Ada (Jones)' and 'Walker, Tim'.

All	data 1	data 2	subject
1.	1	ence: from Fred B. Woodard to the editor of Reader's Digest. July	Archer, Ada (Jones)
2.	2	ence: from Wesley Stout to Fred B. Woodard. July 1938.	Walker, Tim
3.	3	ence: from Fred B. Woodard to James A. Vessey. June 1937.	Wilson, Florence
4.	4	ember 1937.	Cherokee Indians
5.	5	1937.	Cherokee Male Seminary
6.	6		Cherokee Language
7.	7		Sanders, Soggy
8.	8	August 1936.	Alberty, Gerry
9.	9		Boudinot, Elias
10.	10	Professional correspondon. March 1932.	Watie, Buck

# OpenRefine: Regular Expressions

OpenRefine training ex.xlsx Permalink

Facet / Filter Undo / Redo 4 / 5 84 rows Extensions: Wikidata

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](#)

**Add column based on column data 2**

New column name

On error  set to blank  store error  copy value from original column

Expression  Language  No syntax error.

Preview History Starred Help

row	value	value.partition(/(((January) ( ...
1.	Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest. July 1951.	July 1951.
2.	Professional correspondence: from Wesley Stout to Fred B. Woodard. July 1938.	July 1938.
3.	Professional correspondence: from Fred B. Woodard to James A. Vessey. June 1937.	June 1937.
4.	Professional correspondence: from Fred B. Woodard to Hamilton Fish. December 1937.	December 1937.

OK Cancel

# OpenRefine: Join Columns

OpenRefine training ex.xlsx Permalink

Facet / Filter Undo / Redo 8 / 8

Refresh Reset All Remove All

84 rows

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

Extensions: Wikidata

« first < previous 1 - 50 next > last »

	All	data 1	data 2	date 2	date	subject
1.	1	1	Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest.		July 1951.	Archer, Ada (Jones)
2.	2	2	Professional correspondence: from Wesley Stout to Fred B. Woodard.		July 1938.	Walker, Tim
3.	3	3	Professional correspondence: from Fred B. Woodard to James A. Vessey.		June 1937.	Wilson, Florence
4.	4	4	Professional correspondence: from Fred B. Woodard to Hamilton Fish.		December 1937.	Cherokee Indians
5.	5	5	Personal correspondence: from Mrs. W.V. Brown to Fred B. Woodard.		May 1937.	Cherokee Male Seminary
6.	6	6	Professional correspondence: to Hamilton Fish.	1937.		Cherokee Language
7.	7	7	Professional correspondence: to Hayes McCoy.		June 1937.	Sanders, Soggy
8.	8	8	Professional correspondence: from Joseph Bruner to Franklin D. Roosevelt.		August 1936.	Alberty, Gerry
9.	9	9	Telegram: from Hamilton Fish, Jr. to Fred Woodard.		May 1935.	Boudinot, Elias
10.	10	10	Professional correspondence: from Cherokee Seminaries Student Association.		March 1932.	Watie, Buck
11.	11	11	Professional correspondence: to Wesley Stout.		July 1938.	Duncan, DeWitt Clinton
12.	12	12	Professional correspondence: from the American Federation of Labor.		June 1938.	Tahlequah
13.	13	13	Professional correspondence: from Hamilton Fish to Fred B. Woodard.	1938.		Indian Schools
14.	14	14	Professional correspondence: from W.P. Tomson to Fred Woodard and from Fred Woodard to Hamilton Fish.	1938.		Cherokee Female Seminary
15.	15	15	Professional correspondence: from National Legislative representatives.		June 1938.	McSpadden, Florence
16.	16	16	Professional correspondence: from L.M. Bailey.		September-October 1939.	Mayes, Patsy
17.	17	17	Professional correspondence: to Hamilton Fish.	1938-39.		Adair, Blueie
18.	18	18	Professional correspondence: from Fred B. Woodard to Benard Heffeman.		August 1938.	Miller, Mattie
19.	19	19	Professional correspondence: from Jas A. Veasey to Fred B. Woodard.		August 1938.	Starr, Lulu
20.	20	20	Professional correspondence: concerning the Keep America Out of War Committee and flyers.	1936-1939.		Prather, Ella
21.	21	21	Professional correspondence: from Fred B. Woodard to Florence Booker.		October 1943.	Thompson, Babe
22.	22	22	Professional correspondence: from Fred B. Woodard to Ralph Mitchel.		June 1941.	Schimpsher, Bunt
23.	23	23	Professional correspondence: from Harve Tibbott to Mrs. John Edminston.			Duncan, Mae
24.	24	24	Personal correspondence: from Ralph to Fred.		June 1941.	Carter, Ella
25.	25	25	Personal correspondence: from Fred B. Woodard.	1937-1942		Nash, Fannie



# OpenRefine: Join Columns

The screenshot shows the OpenRefine web interface. At the top, it displays 'OpenRefine training ex.xlsx' and 'Permalink'. The main area shows a table with 84 rows. A context menu is open over the table, with 'Join columns...' selected. The table has columns for 'All', 'data 1', 'data 2', 'date 2', 'date', and 'subject'. The 'subject' column contains various names and dates.

	All	data 1	data 2	date 2	date	subject
1.	1	Professional correspondence: from Fred B. Woodard to the editor of Reader's Digest.			Facet	rcher, Ada (Jones)
2.	2	Professional correspondence: from Wesley Stout to Fred B. Woodard.			Text filter	walker, Tim
3.	3	Professional correspondence: from Fred B. Woodard to James A. Vessey.			Edit cells	wilson, Florence
4.	4	Professional correspondence: from Fred B. Woodard to Hamilton Fish.			Edit column	Cherokee Indians
5.	5	Personal correspondence: from Mrs. W.V. Brown to Fred			Transpose	Cherokee Male Seminary
6.	6	Professional correspondence: to Hamilton Fish.			Sort...	Cherokee Language
7.	7	Professional correspondence: to Hayes McCoy.			View	anders, Soggy
8.	8	Professional correspondence: from Joseph Bruner to Fr			Reconcile	oudinot, Elias
9.	9	Telegram: from Hamilton Fish, Jr. to Fred Woodard.				March 1932.
10.	10	Professional correspondence: from Cherokee Seminarie				July 1938.
11.	11	Professional correspondence: to Wesley Stout.				Duncan, DeWitt Clinton
12.	12	Professional correspondence: from the American Feder				June 1938.
13.	13	Professional correspondence: from Hamilton Fish to Fre				Tahlequah
14.	14	Professional correspondence: from W.P. Tomson to Fred to Hamilton Fish.				Indian Schools
15.	15	Professional correspondence: from National Legislative				Cherokee Female Seminary
16.	16	Professional correspondence: from L.M. Bailey.				June 1938.
17.	17	Professional correspondence: to Hamilton Fish.		1938-39.		McSpadden, Florence
18.	18	Professional correspondence: from Fred B. Woodard to Benard Heffernan.			August 1938.	September-October 1939.
19.	19	Professional correspondence: from Jas A. Veasey to Fred B. Woodard.			August 1938.	Mayes, Patsy
20.	20	Professional correspondence: concerning the Keep America Out of War Committee and flyers.		1936-1939.		Adair, Bluye
21.	21	Professional correspondence: from Fred B. Woodard to Florence Booker.			October 1943.	Miller, Mattie
22.	22	Professional correspondence: from Fred B. Woodard to Ralph Mitchel.			June 1941.	Starr, Lulu
23.	23	Professional correspondence: from Harve Tibbott to Mrs. John Edminston.				Prather, Ella
24.	24	Personal correspondence: from Ralph to Fred.			June 1941.	Thompson, Babe
25.	25	Personal correspondence: from Fred B. Woodard.		1937-1942		Schimpsher, Bunt
						Duncan, Mae
						Carter, Ella
						Nash, Fannie

# OpenRefine: Join Columns

The screenshot shows the OpenRefine web interface with a 'Join columns' dialog box open. The background shows a table with columns for 'date', 'date 2', and 'subject'. The dialog box has two main sections: 'Select and order columns to join' and 'Select options'.

**Select and order columns to join:**

- date
- data 1
- data 2
- date 2
- subject

**Select options:**

- Separator between the content of each column:
- Enter one or more characters, or keep blank to join the columns without separator.
- Replace nulls with...
- Enter one or more characters, or keep blank to replace nulls with blank strings.
- Skip nulls.
- In separator and nulls substitutes, use \n for new lines, \t for tabulation, \\n for \n, \\t for \t.
- Write result in selected column.
- Write result in new column named...
- Delete joined columns.

Buttons: Select All, De-select All, OK, Cancel

# OpenRefine: JSON Export & Reuse

Live Walk Through!



# OpenRefine: JSON Export & Reuse

**Extract Operation History**

Extract and save parts of your operation history as JSON that you can apply to this or other projects in the future.

- Text transform on cells in column data using expression value.trim()
- Text transform on cells in column data using expression value.replace(/s+/, ' ')
- Split column data by separator
- Mass edit cells in column subject
- Create column date at index 2 based on column data 2 using expression grel:value.partition((((January))(February))(March))(April))(May))(June))(July))(August))(September))(October))(November))(December))(circa).\*)/)[1]
- Text transform on cells in column data 2 using expression grel:value.partition((((January))(February))(March))(April))(May))(June))(July))(August))(September))(October))(November))(December))(circa).\*)/)[0].trim()
- Create column date 2 at index 2 based on column data 2 using expression grel:value.partition(/(d/d/d.d.\*)/)[1]
- Text transform on cells in column data 2 using expression grel:value.partition(/(d/d/d.d.\*)/)[0].trim()
- Create column Date at index 4 based on column date using expression join ([coalesce(cells[ date ].value,""),coalesce(cells[ date 2 ].value,""),")
- Remove column date 2
- Remove column date 1

Select All Unselect All

Close

```
{
  "op": "core/text-transform",
  "engineConfig": {
    "facets": [],
    "mode": "row-based"
  },
  "columnName": "data",
  "expression": "value.trim()",
  "onError": "keep-original",
  "repeat": false,
  "repeatCount": 10,
  "description": "Text transform on cells in column d
},
{
  "op": "core/text-transform",
  "engineConfig": {
    "facets": [],
    "mode": "row-based"
  },
  "columnName": "data",
  "expression": "value.replace(/s+/, ' ')",
  "onError": "keep-original",
  "repeat": false,
  "repeatCount": 10,
  "description": "Text transform on cells in column d
},
{
  "op": "core/column-split",
  "engineConfig": {
    "facets": [],
    "mode": "row-based"
  },
  "columnName": "data",
  "guessCellType": true,

```

# Thank You

---

Bailey Hoffner  
Metadata and Collections  
Management Archivist  
baileys@ou.edu