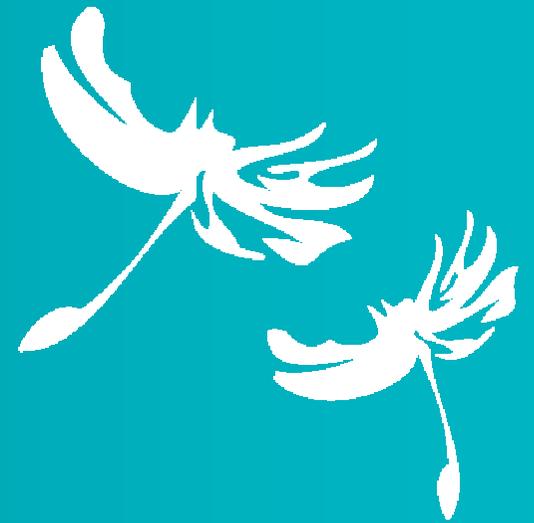




OpenRefine



# Links



<http://openrefine.org/>

Vielzahl an Tutorials: <https://github.com/OpenRefine/OpenRefine/wiki/External-Resources>

Empfehlung:

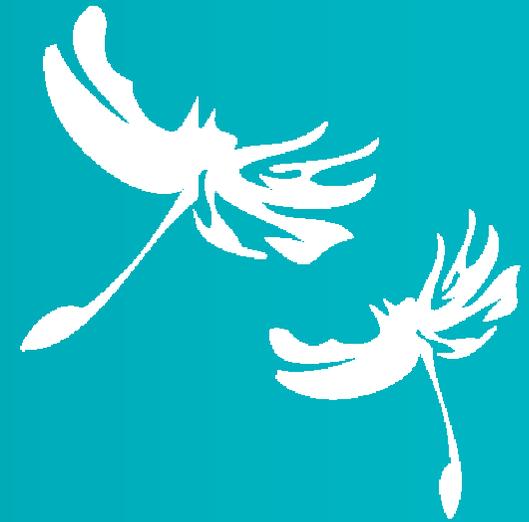
Seth van Hooland , Ruben Verborgh and Max De Wilde , "Cleaning Data with OpenRefine," Programming Historian, (2013-08-05), <http://programminghistorian.org/lessons/cleaning-data-with-openrefine>

Die [Übersetzung von Google Translate](#) ist ganz brauchbar.

Die im Tutorial benannte Ausgangsdatei **phm-collection.tsv** ist unter folgender Adresse erreichbar: <http://data.freeyourmetadata.org/powerhouse-museum/phm-collection.tsv>

\*googleRefine für MIDAS-XML unter: <http://openrefine.org/download.html>

# Datenbereinigung



# Facetten



Browser window showing a Refine interface for data exploration. The browser address bar shows a URL: `127.0.0.1:3333/project?project=1804735246426`. The Refine interface includes a search bar with the query `lohmeier github summersdhool`.

The main interface displays a table with 7838 rows. The table columns are: `All`, `Status`, `a9998 [Bearbeitu]`, `a8450 [Aufnahm]`, `a8450 a8491 [Atr]`, `a8450 a8490 [For]`, `a8450 a8470 [Au]`, `a8450 a8494 [Au]`, `a8450 a8497 [Zu]`, and `a84`. The first five rows of the table are:

1	veröffentlicht	Welsichten-Stilfried-Stöver-2017-KW25-fertig	Facet	Text facet	le_ses-svr_0000001	1871/1883	Kauf	Albu
2	veröffentlicht	Welsichten-Stilfried-Stöver-2017-KW25-fertig	Facet	Text facet	le_ses-svr_0000002	1871/1883	Kauf	Albu
3	veröffentlicht	Welsichten-Stilfried-Stöver-2017-KW25-fertig	Postiv	Stellfried und Comp.	le_ses-svr_0000003	1871/1883	Kauf	Albu
4	veröffentlicht	Welsichten-Stilfried-Stöver-2017-KW25-fertig	Postiv	Stellfried und Comp.	le_ses-svr_0000004	1871/1883	Kauf	Albu colo
5	veröffentlicht	Welsichten-Stilfried-Stöver-2017-KW25-fertig	Postiv	Stilfried und Comp.	le_ses-svr_0000005	1871/1883	Kauf	Albu colo

A context menu is open over the first row, showing options: `Facet`, `Text filter`, `Edit cells`, `Edit column`, `Transpose`, `Sort...`, `View`, and `Reconcile`. The `Facet` sub-menu is expanded, showing options: `Text facet`, `Numeric facet`, `Timeline facet`, `Scatterplot facet`, `Custom text facet...`, `Custom Numeric Facet...`, and `Customized facets`.

On the left side, a facet panel for `a8450 [Aufnahme Medientyp] 1` is visible, showing 4 choices: `Diapositiv 31`, `Negativ 6699`, `Positiv 1105`, and `Postiv 3`. The facet is sorted by `name count`.

# Zellentransformationen

7838 rows Extensions:

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

▼ All ▼ Status ▼ a9998 [Bearbeit] ▼ a8450 [Aufnahme Medientyp] 1 ▼ a8450 [Aufnahme Medientyp] 2 ▼ a8450 [Aufnahme Medientyp] 3 ▼ a8450 [Aufnahme Medientyp] 4 ▼ a8450 [Aufnahme Medientyp] 5 ▼ a8450 [Aufnahme Medientyp] 6 ▼ a8450 [Aufnahme Medientyp] 7 ▼ a8450 [Aufnahme Medientyp] 8

1.	veröffentlicht	Wellsichten-Stilfried-Stöver-2017-KW25-fertig
2.	veröffentlicht	Wellsichten-Stilfried-Stöver-2017-KW25-fertig

Facet  
Text filter  
Edit cells  
Edit column  
Transpose  
Sort...  
View  
Reconcil

### Custom text transform on column a8450 [Aufnahme Medientyp] 1

Expression Language **General Refine Expression Language (GREL)**

```
value.replace("Postitiv", "Positiv")
```

No syntax error.

**Preview** History Starred Help

row	value	value.replace("Postitiv", "Positiv")
1.	Positiv	Positiv
2.	Positiv	Positiv
3.	Postitiv	Positiv
4.	Postitiv	Positiv
5.	Postitiv	Positiv
6.	Positiv	Positiv

On error  keep original  set to blank  store error  Re-transform up to  times until no change

OK Cancel

# Zeile ergänzen | Werte bearbeiten



7838 rows											Extensions:
Show as: rows records		Show: 5 10 25 50 rows				« first ‹ previous 1 - 10 next › last »					
All	Status	a9998 [Bearbeit	a8450 [Aufnahm	a8450 a8491 [At	a8450 a8490 [Fot	a8450 a8470 [Au	URL	a8450 a8494 [Au	a844		
☆	🗨	1.	veröffentlicht	Weltsichten-Stillfried-Stöver-2017-KW25-fertig	Positiv	Stillfried und Comp.		le_ses-svr_0000001	<a href="http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000001.jpg">http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000001.jpg</a>	1871/1883	Kauf
☆	🗨	2.	veröffentlicht	Weltsichten-Stillfried-Stöver-2017-KW25-fertig	Positiv	Stellfried und Comp.		le_ses-svr_0000002	<a href="http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000002.jpg">http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000002.jpg</a>	1871/1883	Kauf
☆	🗨	3.	veröffentlicht	Weltsichten-Stillfried-Stöver-2017-KW25-fertig	Positiv	Stellfried und Comp.		le_ses-svr_0000003	<a href="http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000003.jpg">http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000003.jpg</a>	1871/1883	Kauf
☆	🗨	4.	veröffentlicht	Weltsichten-Stillfried-Stöver-2017-KW25-fertig	Positiv	Stellfried und Comp.		le_ses-svr_0000004	<a href="http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000004.jpg">http://fotothek.slub-dresden.de/fotos/le_ses-svr_0000004.jpg</a>	1871/1883	Kauf

# Mergen und Clustern

## Cluster & Edit column "a55df [Schlagwort]"

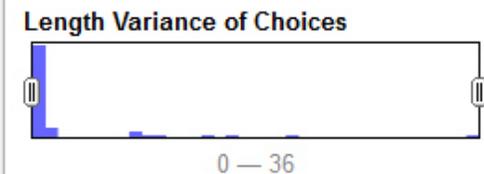
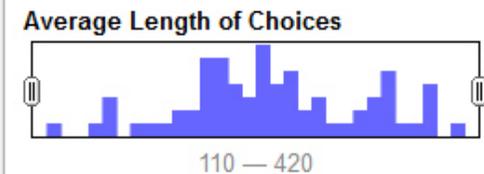
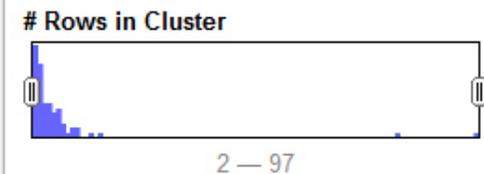
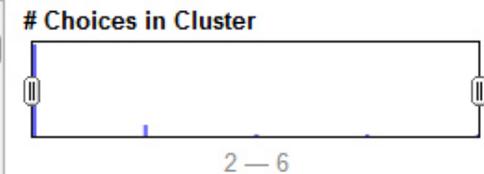
This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer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

Method key collision

Keying Function fingerprint

71 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
6	16	<ul style="list-style-type: none"> <li>Gebirge &amp; mountain &amp; Berg &amp; Landschaft &amp; landscape &amp; Polen &amp; Poland &amp; Slowakei &amp; Slovakia &amp; Europa &amp; Europe &amp; Osteuropa &amp; East Europe &amp; Karpaten &amp; Carpathian Mountains &amp; Hohe Tatra &amp; High Tatra &amp; Exkursion &amp; excursion &amp; Geographie &amp; geography (7 rows)</li> <li>Berg &amp; mountain &amp; Gebirge &amp; Landschaft &amp; landscape &amp; Polen &amp; Poland &amp; Slowakei &amp; Slovakia &amp; Europa &amp; Europe &amp; Osteuropa &amp; East Europe &amp; Karpaten &amp; Carpathian Mountains &amp; Hohe Tatra &amp; High Tatra &amp; Exkursion &amp; excursion &amp; Geographie &amp; geography (4 rows)</li> <li>Landschaft &amp; landscape &amp; Berg &amp; mountain &amp; Gebirge &amp; Polen &amp; Poland &amp; Slowakei &amp; Slovakia &amp; Europa &amp; Europe &amp; Osteuropa &amp; East Europe &amp; Karpaten &amp; Carpathian Mountains &amp; Hohe Tatra &amp; High Tatra &amp; Exkursion &amp; excursion &amp; Geographie &amp; geography (2 rows)</li> <li>Berg &amp; mountain &amp; Gebirge &amp; landscape &amp; Landschaft &amp; Polen &amp; Poland &amp; Slowakei &amp; Slovakia &amp; Europa &amp; Europe &amp; Osteuropa &amp; East Europe &amp; Karpaten &amp; Carpathian Mountains &amp; Hohe Tatra &amp; High Tatra &amp; Exkursion &amp; excursion &amp; Geographie &amp; geography (1 rows)</li> <li>Gebirge &amp; Berg &amp; mountain &amp; Landschaft &amp;</li> </ul>	<input type="checkbox"/>	Gebirge & mountain & Berg & Lan



Select All Unselect All

Export Clusters

Merge Selected & Re-Cluster

Merge Selected & Close

Close

# Join & Split

7838 7838 rows Extensions: Extensionen

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

Titel	a5230 [Sachbezeichnung]	a55df [Schlagwort]	a599a [Anmerkung]	a599a a599e [Freigegeben]	a599a [Anmerkung]	a599a a599e [Freigegeben]	a99d3 [APS Arc]	a99d1 [APS Reit]	a99d1 [APS Reit]
Foto	Japan   Asien   Asia   Ostasien   East Asia   Mann   man   Portrait   portrait   Männerportrait   portrait of a man   Bildnis   Gruppenportrait   group portrait   Ainu (Ethnie)   Ainu people (ethnic group)	Findbucheintrag	Ainu, 3 Männer Japan				Weltsichten & DFG-Weltsichten & wsneu7	Fotos	
Foto	Japan   Asien   Asia   Ostasien   East Asia   Mann   man   Esel   donkey   Tier   animal   Zaun   fence	Findbucheintrag	Japaner zu Pferde Japan				Weltsichten & DFG-Weltsichten & wsneu7	Fotos	
Foto	Japan   Asien   Asia   Ostasien   East Asia   Mann   man   Bambus   bamboo   Arbeiter   worker   Wald   forest   Bambuswald   bamboo forest   Werkzeug   tool	Findbucheintrag	Bambusdickicht bei Homoko Japan				Weltsichten & DFG-Weltsichten & wsneu7	Fotos	
Foto	Japan   Asien   Asia   Ostasien   East Asia   Mann   man   Portrait   portrait   Männerportrait   portrait of a man   Bildnis   Sandalen   sandals   Waraji   Kopftuch   headscarf   Arbeiter   worker	Findbucheintrag	Arbeiter Japan	<a href="#">edit</a>			Weltsichten & DFG-Weltsichten & wsneu7	Fotos	

# Join & Split



	a55df [Schlagw								
Japan	Asien	Asia	Ostasien	East Asia	Mann	man	Portrait	portrait	
Japan	Asien	Asia	Ostasien	East Asia	Mann	man	Esel	donkey	
Japan	Asien	Asia	Ostasien	East Asia	Mann	man	Bambus	bamboo	
Japan	Asien	Asia	Ostasien	East Asia	Mann	man	Portrait	portrait	
Japan	Asien	Asia	Ostasien	East Asia	Frau	woman	Portrait	portrait	
Japan	Asien	Asia	Ostasien	East Asia	Kreuzigung	crucifixion	Hinrichtung	execution	

# Bearbeitung von Spalten

2017-KW25-fertig

### Re-order / Remove Columns

Drag columns to re-order

- Status
- a9998 [Bearbeitungsstand]
- a8450 [Aufnahme Medientyp] 1
- a8450 a8491 [Atelier Verlag] 1
- a8450 a8490 [Fotograf] 1
- a8450 a8470 [Aufnahmenr.] 1
- URL
- a8450 a8494 [Aufnahmedatum] 1
- a8450 a8497 [Zugangsart] 1
- a8450 a8496 [Zugangsdatum] 1
- a8450 a8499 [Erhaltung
- a8450 a8476 [Alte Aufnahmenr.] 1
- a8450 a8498 [Aufnahmeherkunft] 1
- a8450 a8555 [Bildkarte Rückseite] 1
- a8450 a8478 [Aufnahmeart techn.] 1

Drop columns here to remove

- a8450 a8596 [Begleitinformation] 1

OK Cancel

Export Help

Extensions:

10 next last

a8494 [Au a84

Kauf

Kauf

Kauf

Kauf

Kauf

Kauf

Kauf



# Zellen zusammenführen

Refine OPEN xlsx XML xlsx Permalink

Facet

Use

Use of y  
filter  
of e

Not  
Wa

New column name

set to blank  store error  copy value from original column

Expression  Language

No syntax error.

Preview History Starred Help

row	value	cells[\"a5108 [Standort] 1\"].value + \" \" + cells[\"a5108 [Standort] 2\"].value
1.	Japan	Japan Malta
2.	Japan	Error: Cannot retrieve field from null
3.	Japan	Error: Cannot retrieve field from null
4.	Japan	Error: Cannot retrieve field from null
5.	Japan	Error: Cannot retrieve field from null
6.	Japan	Error: Cannot retrieve field from null

OK Cancel

Open... Export . Help

ensions:  
last »  
5df [Schl

# Wenn-Dann-Beziehungen

ubsets  
cet and  
the top

dreier bariger und  
mit Fellmänteln

### Add column based on column a5108 [Standort] 1

New column name

set to blank  store error  copy value from original column

Expression  Language

```
if(isBlank(cells["a5108 [Standort] 1"].value),cells["a5108 [Standort] 2"].value,cells["a5108 [Standort] 1"].value)
```

No syntax error.

**Preview** History Starred Help

row	value	if(isBlank(cells["a5108 [Standort] 1"].value),cells["a5108 [Standort] 2"].value,cells["a5108 [Standort] 1"].value)
1.	Japan	Japan
2.	Japan	Japan
3.	Japan	Japan
4.	Japan	Japan
5.	Japan	Japan
6.	Japan	Japan

OK Cancel

(Haritsuke,  
japanisch: 磯)  
Verbrecher. Bei der  
japanischen Art der  
Kreuzigung wurden  
dem zum Tode  
verurteilten

## 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.

- Edit single cell on row 3, column a8450 [Aufnahme Medientyp] 1
- Edit single cell on row 4, column a8450 [Aufnahme Medientyp] 1
- Edit single cell on row 5, column a8450 [Aufnahme Medientyp] 1
- Edit single cell on row 3, column a8450 [Aufnahme Medientyp] 1
- Edit single cell on row 5, column a8450 [Aufnahme Medientyp] 1
- Text transform on cells in column a8450 [Aufnahme Medientyp] 1 using expression `grel:value.replace("Positiv", "Positiv")`
- Create column URL at index 6 based on column a8450 a8470 [Aufnahmenr.] 1 using expression `grel:"http://fotothek.slub-dresden.de/fotos/" + value + ".jpg"`
- Split multi-valued cells in column a55df [Schlagwort]
- Join multi-valued cells in column a55df [Schlagwort]
- Split column a55df [Schlagwort] by separator
- Move column aob26 a2700 [Bauwerkname] 1 to positio

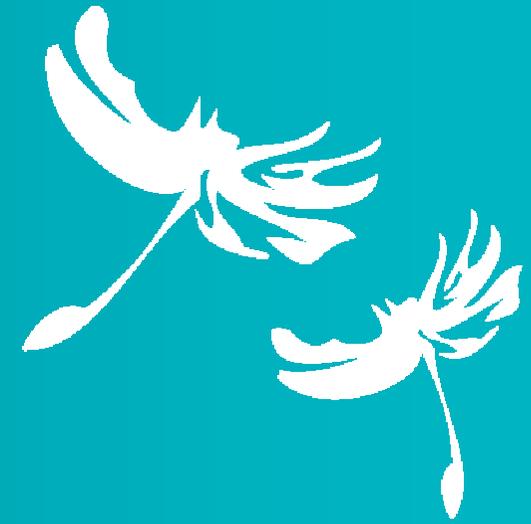
Select All

Unselect All

Close

```
{
  "op": "core/multivalued-cell-join",
  "description": "Join multi-valued cells",
  "columnName": "a55df [Schlagwort]",
  "keyColumnName": "Status",
  "separator": " | "
},
{
  "op": "core/column-split",
  "description": "Split column a55df [Schl",
  "engineConfig": {
    "mode": "row-based",
    "facets": []
  },
  "columnName": "a55df [Schlagwort]",
  "guessCellType": true,
  "removeOriginalColumn": true,
  "mode": "separator",
  "separator": " | ",
  "regex": false,
  "maxColumns": 0
},
{
  "op": "core/column-move",
  "description": "Move column aob26 a2700",
  "columnName": "aob26 a2700 [Bauwerkname]",
  "index": 78
}
]
```

# Datenkonversion





# XML zu xlsx



```
[
  {
    "op": "core/column-rename",
    "description": "Rename column obj - lvl to lvl",
    "oldColumnName": "obj - lvl",
    "newColumnName": "lvl"
  },
  {
    "op": "core/column-rename",
    "description": "Rename column obj - a5000 to a5000 [Objektnummer]",
    "oldColumnName": "obj - a5000",
    "newColumnName": "a5000 [Objektnummer]"
  },
  {
    "op": "core/column-rename",
    "description": "Rename column obj - Status to Status",
    "oldColumnName": "obj - Status",
    "newColumnName": "Status"
  },
  {
    "op": "core/column-rename",
    "description": "Rename column obj - a9998 to a9998 [Bearbeitungsstand]",
    "oldColumnName": "obj - a9998",
    "newColumnName": "a9998 [Bearbeitungsstand]"
  },
  {
    "op": "core/multivalued-cell-join",
    "description": "Join multi-valued cells in column a9998 [Bearbeitungsstand]",
    "columnName": "a9998 [Bearbeitungsstand]",
    "keyColumnName": "lvl",
    "separator": " | "
  },
  {
    "op": "core/column-rename",
    "description": "Rename column obj - a8450 to a8450 [Aufnahme Medientyp]",
    "oldColumnName": "obj - a8450",
    "newColumnName": "a8450 [Aufnahme Medientyp]"
  },
  {
    "op": "core/multivalued-cell-join",
    "description": "Join multi-valued cells in column a8450 [Aufnahme Medientyp]",
    "columnName": "a8450 [Aufnahme Medientyp]",
    "keyColumnName": "lvl",
    "separator": " | "
  },
  {
    "op": "core/column-split",
    "description": "Split column a8450 [Aufnahme Medientyp] by separator",
    "engineConfig": {
      "mode": "record-based",
      "facets": []
    }
  },
],
```

# Vielen Dank für Ihre Aufmerksamkeit. Gibt es Fragen?

Karolin Schmahl  
Fachstelle Mediathek-Fotografie  
Fachstelle Mediathek-Ton  
karolin.schmahl@slub-dresden.de  
+49 (0) 351 4677-631

<http://www.deutsche-digitale-bibliothek.de>

<https://pro.deutsche-digitale-bibliothek.de>

[www.facebook.com/ddbkultur](http://www.facebook.com/ddbkultur)

[www.twitter.com/ddbkultur](http://www.twitter.com/ddbkultur)

<https://www.deutsche-digitale-bibliothek.de/user/newsletter>