OPEN REFINE: MINI CHEATSHEET

1
• Create project
o Select file
o Make sure you have chosen CSV format

2
FORMAT YOUR CELLS:
o Publication number column:(scroll down menu)– Edit Cells – Transform
- Value replace: value.replace(" ","") (define what needs to be replaced, by what)
- Title column: edit cells common transforms , to title case: turns the cell content in capital letters
- Export, comma separated value – save as yourtitle.csv
NEW MASTER DATASET

3
SPLITTING DATASET:
? Applicants column: Split multi-valued cells. Define what is the separator: ;; (it is the standard separator on the lens). Result: column split on the applicant names now (which used to be in a cell as Applicant A;;Applicant B). We can now use this column only to count our applicants (although you can fill out the rest of the cells with rest of the entries)

4
CLUSTERING & MERGING (e.g.: applicants)
Applicants – Facet – Text Facet - Cluster: key collision method, fingerprint. Result: merge suggestions (clusters=groups of applicants that have been grouped by the algorithm together as possibly the same entry – in which case we merge and treat them as the same applicant). Merge and recluster, changing the algorithm ("function"). You can rename the clusters. Export – comma separated values -save now these clustered applicants

5
ATTENTION!

DO NOT SPLIT FURTHER COLUMNS (e.g. inventors) WHEN YOU ALREADY SPLITTED OTHER FIELDS. Go back to your core dataset ("undo" until you see your initial number of rows. Now you can continue)