Le nettoyage de données!
Clean your Data!

Constance Crompton
Chaire de recherche du Canada en humanités numériques
Canada Research Chair in Digital Humanities

OpenRefine

Dans ces diapos

- OpenRefine: Instructions d'installation
- OpenRefine: des instructions de base
- OpenRefine: instructions avancées (APIs)
- Ressources supplémentaires
- Merci!

In this deck

- OpenRefine installation instructions
- OpenRefine basics
- OpenRefine: going further (APIs)
- Further Resources
- Thank you!

OpenRefine

OpenRefine est une application de nettoyage de données qui s'exécute dans votre logiciel de navigation (Firefox ou Chrome).

Instructions d'installation et vidéos tutorielles : https://openrefine.org/ (choisissez la version applicable - mac, win, ou linux)

Suivez les instructions d'installation

Pour ouvrir OpenRefine, double-cliquez sur l'application à l'endroit où votre installation l'a enregistrée (probablement dans Applications)

Pour fermer OpenRefine dans Windows : Control-C

Pour fermer OpenRefine sur un Mac : Cliquez sur l'application OpenRefine dans la dock, puis appelez Quit

OpenRefine is an data cleaning application that runs in your browser (Firefox or Chrome).

Download and install OpenRefine.

Installation instructions and tutorial videos: https://openrefine.org/ (choose mac or win versions (or linux if that is what you are running))

Follow the installation instructions

To open OpenRefine, double click on the application wherever your installation saved it (likely in Applications)

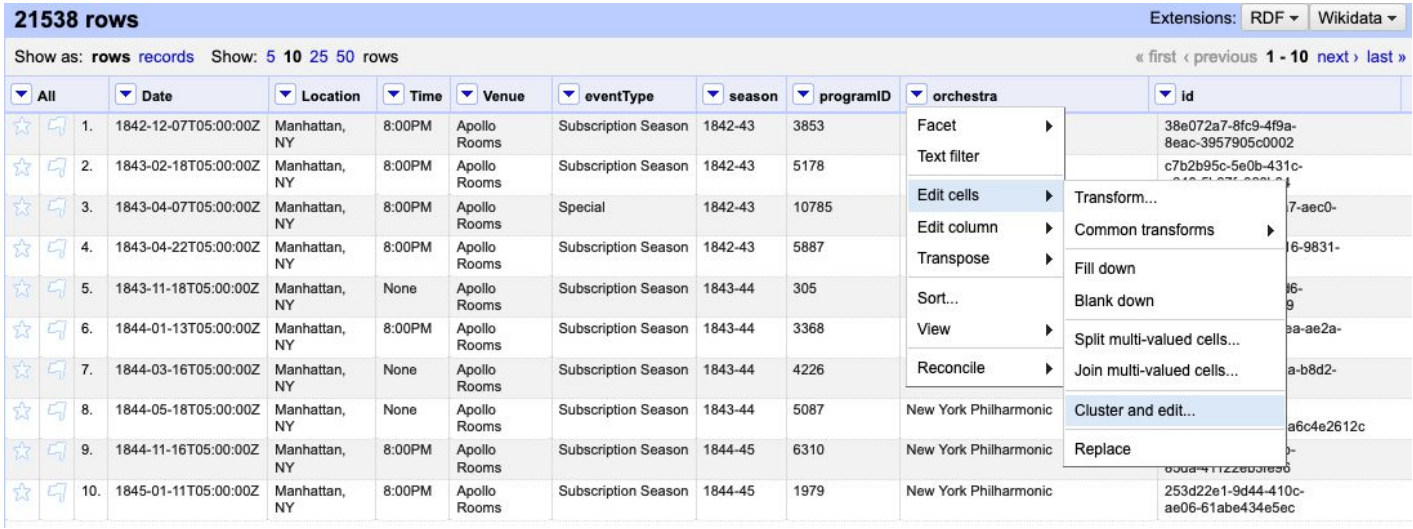
To shut down OpenRefine in Windows: Control-C

To shut down OpenRefine on a Mac: Click the OpenRefine app in the dock, invoke Quit

OpenRefine: des instructions de base/basics

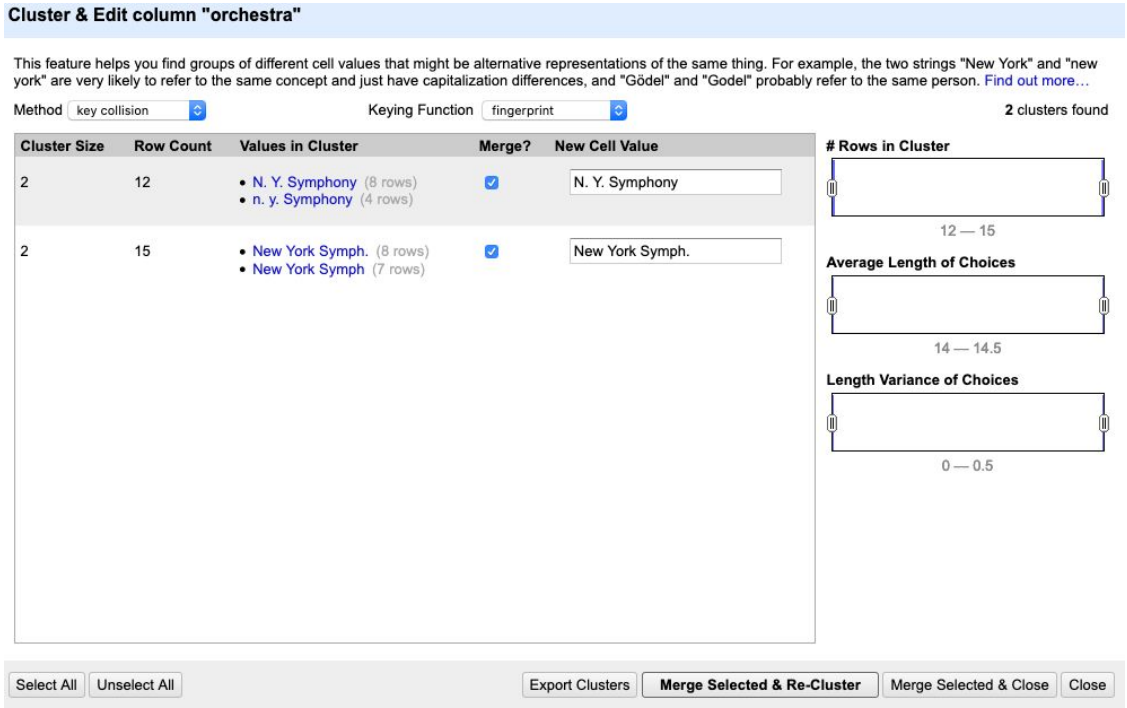
Télécharger
https://www.dropbox.com/s/ysz6f3th0eojvbo/concerts_cleaning_nettoyage.csv?dl=0

Download
https://www.dropbox.com/s/ysz6f3th0eojvbo/concerts_cleaning_nettoyage.csv?dl=0



Column Arrow> Edit Cells> Cluster and Edit

OpenRefine

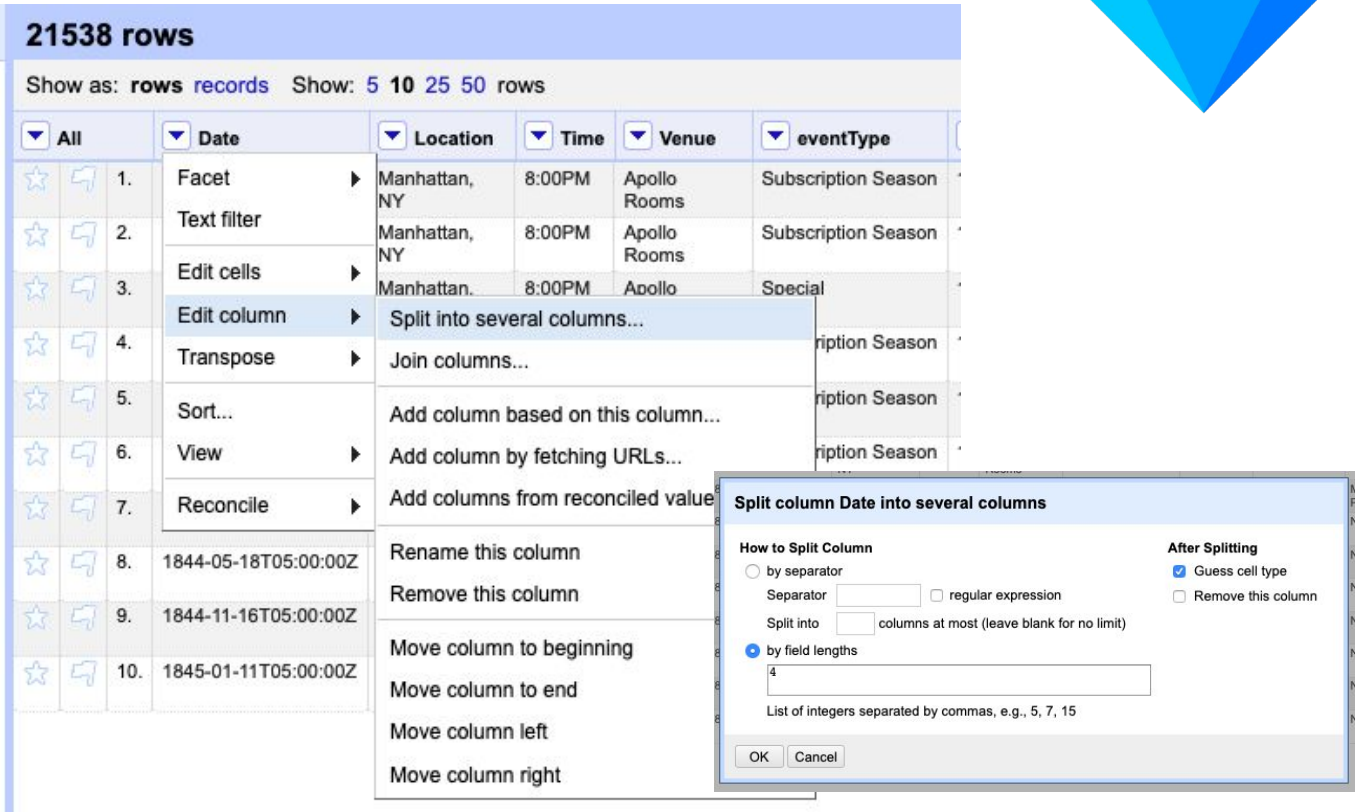


Column Arrow> Edit Cells> Cluster and Edit

OpenRefine

Column Arrow> Edit Column> Split into several columns

… et après renommer la colonne
… and then rename the column

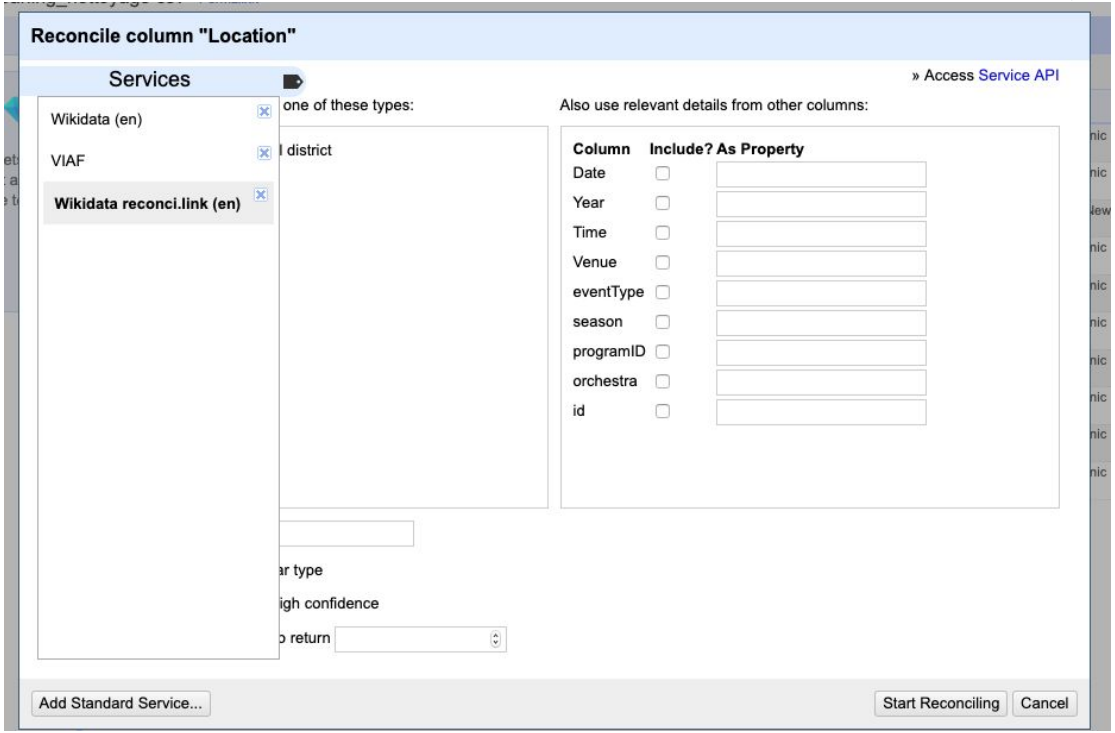


OpenRefine avec APIs / with APIs

Wikidata: https://www.wikidata.org

Column Arrow (Time)> Reconcile> Start Reconciling
https://wikidata.reconci.link/en/api

Time of day

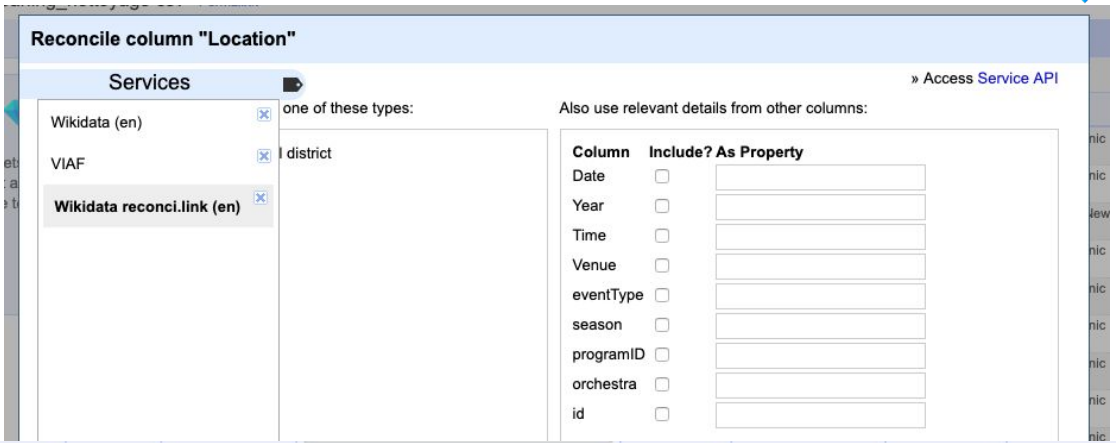


OpenRefine avec APIs / with APIs

Wikidata: https://www.wikidata.org

Column Arrow (Time)> Reconcile> Start Reconciling
https://wikidata.reconci.link/en/api

Time of day



OpenRefine avec APIs / with APIs

Télécharger
https://www.dropbox.com/s/y2x4mrqe1a5fu47/craikpeople.xlsx?dl=0

Let's populate a whole spreadsheet with Wikidata

Column Arrow> Edit Columns> Add Columns from reconciled values

GREL *VIAF ID Column Arrow> Edit Cells> Transform*

Download
https://www.dropbox.com/s/y2x4mrqe1a5fu47/craikpeople.xlsx?dl=0

Remplissons une feuille de calcul entière avec des Wikidata

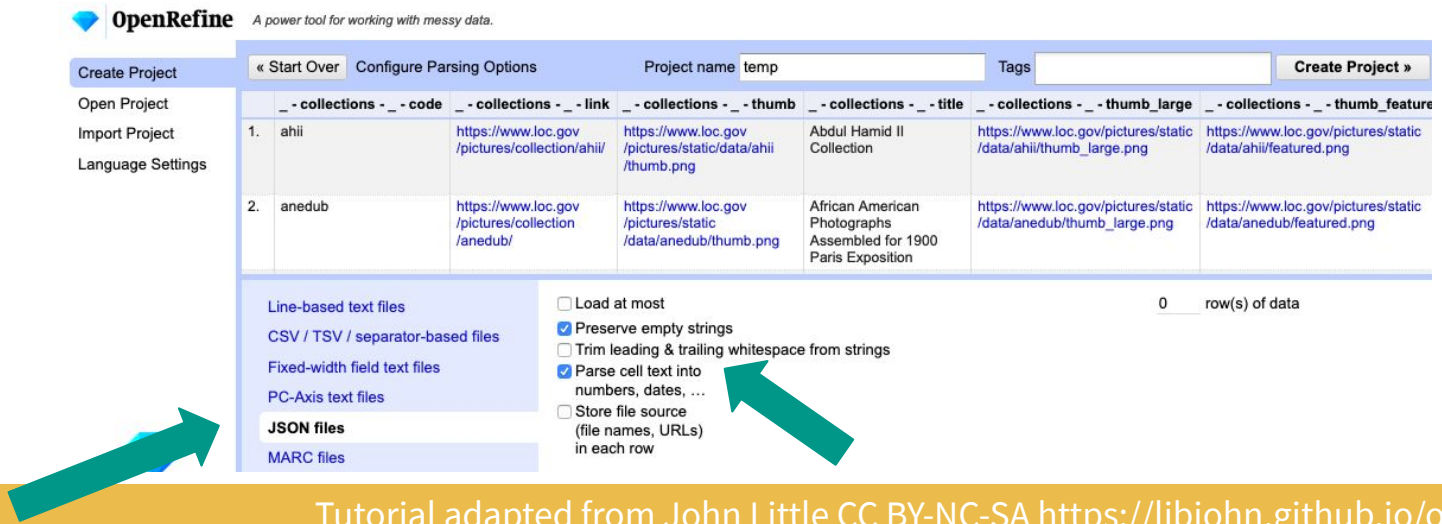
Column Arrow> Edit Columns> Add Columns from reconciled values

GREL *VIAF ID Column Arrow> Edit Cells> Transform*

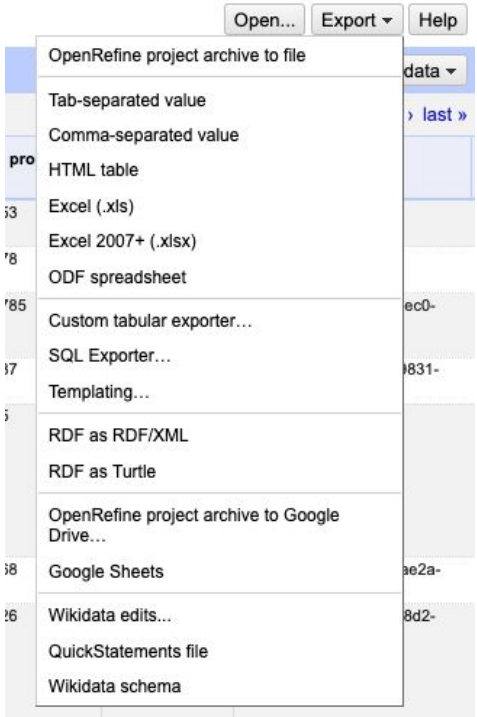
OpenRefine avec APIs / with APIs

Library of Congress API instructions
https://libraryofcongress.github.io/data-exploration/requests.html

Notre API call: https://www.loc.gov/pictures/?q=canada&fo=json



OpenRefine exporter/ export



Ressources supplémentaires | Further Resources

Data Carpentry EN: https://datacarpentry.org/2015-09-21-Genentech/lessons/02-starting-with-OpenRefine.html

Amadine Oricheta FR: https://datahist.hypotheses.org/200

John Little EN: https://libjohn.github.io/openrefine/

Miriam Posner EN:
http://miriamposner.com/classes/dh101f17/tutorials-guides/data-manipulation/get-started-with-openrefine

Programming Historian (Seth Van Hooland FR):
https://programminghistorian.org/fr/lecons/nettoyer-ses-donnees-avec-openrefine

Programming Historian (Evan Peter Williamson EN):
https://programminghistorian.org/en/lessons/fetch-and-parse-data-with-openrefine

Merci - restez en contact
Thank you - be in touch

constance.crompton@uottawa.ca | @clkcrompton

À l'atelier prochain #Boîte à outils en SHN: https://dhsite.org/ateliers-boite-a-outils-numeriques/
See you at the next #DHToolbox: https://dhsite.org/workshops-digital-humanities-toolbox/