

Open Refine

LOGICIEL POUR LA QUALIFICATION DES DONNÉES HISTORIQUES

Présentation

- Développé à l'origine sous le titre de Freebase Gridworks par Metaweb Technologies.
- Metaweb est racheté par Google en 2010, le logiciel est renommé Google Refine.
- En 2012, Google Refine devient Open Refine.
- Première version en 2010, dernière version stable janvier 2022, version 3.5.2
- Le site pour l'installation et la documentation : https://openrefine.org/

Installation

- Sur le site d'Open Refine, allez dans la rubrique Downloads et téléchargez le fichier ZIP
- Décompressez le fichier ZIP
- Ouvrez le fichier ZIP et sélectionnez l'application Open Refine
- Une fenêtre invite de commande s'ouvre, laissez l'invite de commande ouverte sinon
 Open Refine ne fonctionne pas.
- Une fenêtre du navigateur s'ouvre automatiquement sur la page d'accueil du logiciel

Ouvrir un fichier

- Créer un projet
- Ouvrir un projet
- Importer un projet
- Langue
- Formats comme le CSV, TSV, Excel JSON sont acceptées pour créer un projet.
- On peut également connecter une base de donnée à *Open Refine*



Un outil puissant pour travailler avec des données désordonnées.

Créer un projet

Ouvrir un projet

Importer un projet

Langue

Créer un projet en important des données. Quelles sorte Les documents de type TSV, CSV, *SV, Excel (.xls and .xlsx)

Récupérer les données à partir de

Cet ordinateur

Adresses web (URLs)

Presse-papier

Database

Google Data

Chercher un ou plusieur

Choisir des fichiers | A

Suivant »

Ouvrir un fichier (exemple)

- Choississez Adresses web (URLs)
- Ecrire l'URL pointant vers les données : https://tinyurl.com/n5ktws4n
- Vérifiez que les données s'affichent correctement
- Créez votre projet



Adel

Unknown

Active Civil

3

Unknown

Ferme

Badens

Unknown

Gräfin von

Unknown

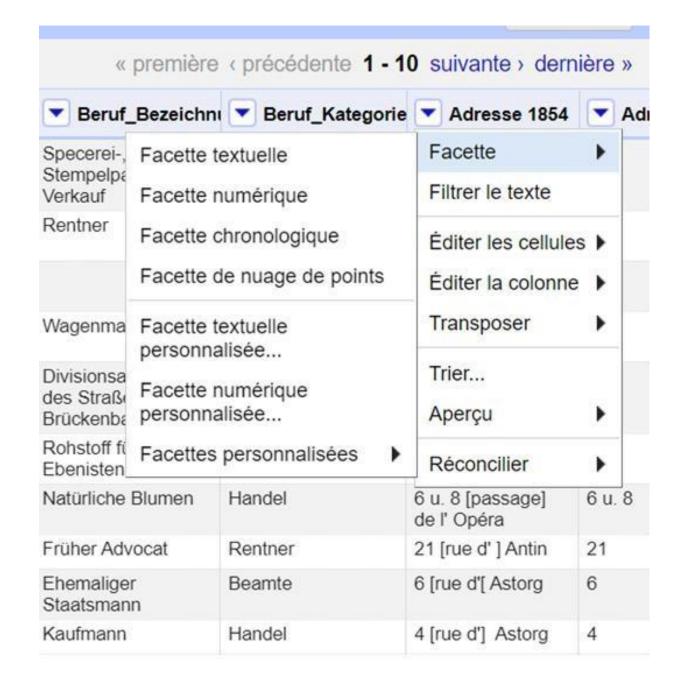
Les principaux onglets

- **Facette/Filtre** permet de voir les filtres que vous avez activés
- Défaire/Refaire permet d'annuler des actions ou d'enregistrer des actions
- Ouvrir pour ouvrir un autre fichier
- **Exporter** pour enregistrer votre fichier dans différents formats
- Aide vous redirige sur la pageGithub d'Open Refine



Exploration des données

- Les Facettes permettent de voir les différents choix dans une colonne de données.
- Filtrer le texte, permet de rechercher des données spéficiques dans votre fichier
- **Trier**, permet de choisir comment organiser ses données
- Apercu pour cacher ou afficher des colonnes



Transformations des données

- Les Facettes permettent de grouper les données qui sont similaires avec l'option groupe
- Transformer permet de nettoyer
 ses données avec le langage GREL
 qui est le langage du logiciel
- Transformations courantes

 permet d'effectuer un premier

 nettoyage des données avec des

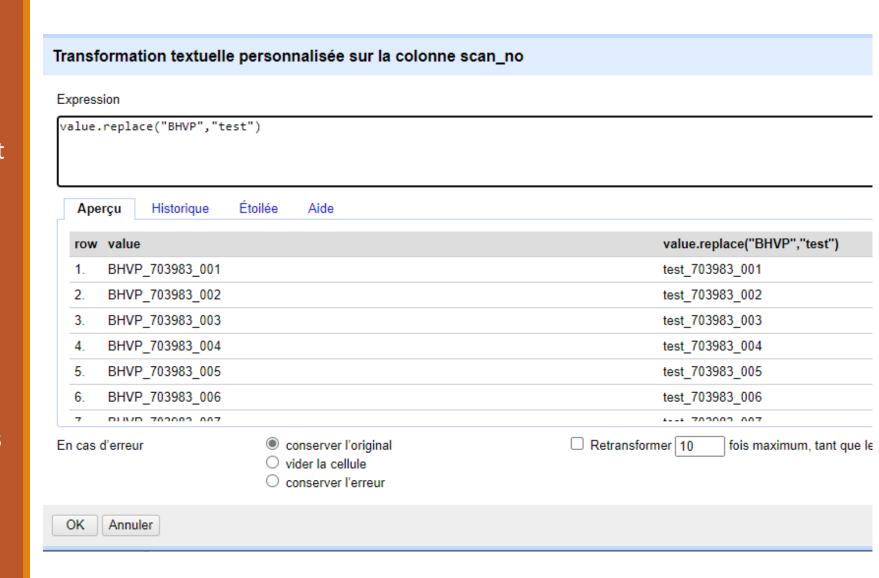
 options déjà définies par *Open Refine*

SX Permalien



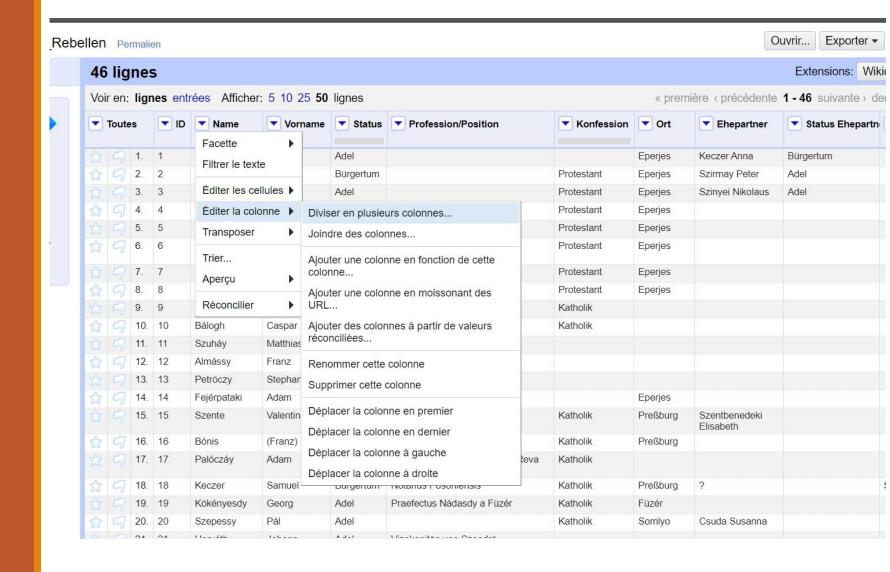
Transformations des données (suite)

- Apercu permet de voir le résultat de la fonction GREL avant de l'exécuter
- Historique sont les fonctions qui ont déjà été utilisées
- Étoilée permet de voir les fonctions favorites
- Aide donne des élements de syntaxe pour écrire des fonctions
- value.replace permet de remplacer ou de supprimer des termes



Transformations des données (suite)

- Pour transformer les données, on peut éditer la colonne
- Diviser en plusieurs colonnes
- Joindre des colonnes
- Ajouter des colonnes à partir de valeurs réconciliées



Réconciliation des données

- Dans la colonne souhaitée,
 sélectionnez réconciliation
 puis démarrer la réconciliation
- Choisissez un service de réconciliation
- Ensuite, choisissez un **type** d'entité
- Démarrer la réconciliation

 pour faire correspondre vos

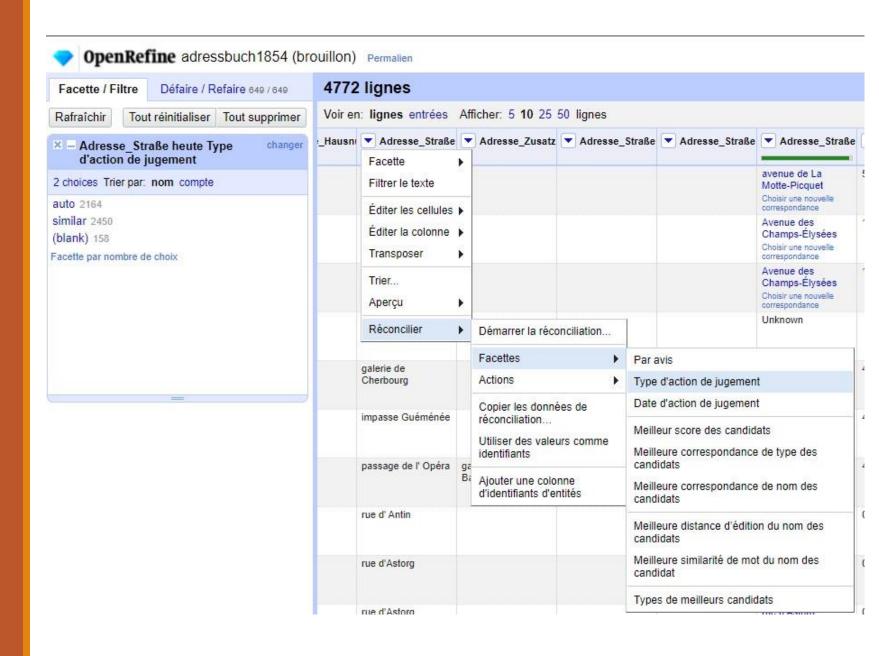
 données

https://reconciliation-api.github.io/testbench/

Services						» Accès Serv	vice API
Services		Réconcilier chaque cellule avec une entité de l'un de ces	types:	Utiliser également les détails p	ertinents d	es autres colonnes:	
Wikidata (en)	×	street	A.	Colonne	Inclure	? Comme propriété	^
Getty Vocabularies	≥ OIS	Q79007		Person_ID			
Reconciliation Service		O Dead end street Q12731		Person_Name_Vorname			
OpenLibrary	×	O avenue		Person_Name_Nachname			
		Q7543083		Person_Name_Namenszusa	tz 🗆		
Geonames Reconciliation Service VIAF	×	O walkway Q13634881		Person_Geschlecht			
		watercolor painting		Beruf_Bezeichnung			
	×	Q18761202		Beruf_Kategorie			
	×	O tourist attraction		Adresse 1854			
GND reconciliation for OpenRefine		O bus stop		Adresse_Hausnummer			
		Q953806		Adresse Zusatz			
GND reconciliation for OpenRefine	×	O thoroughfare Q83620		Adresse Straße 2			
		ocovered passages of Paris	•	Adresse Straße 3			-
Getty Vocabularies Reconciliation Service	×	Réconcilier avec le type:	3				
Wikibase Registry Test (en)		Réconcilier sans type particulier					
	×	Correspondance automatique des valeurs candidates	S				
Wikidata reconci.link (de)	×	Nombre maximal de candidats renvoyés					
Wikidata reconci.link (fr)	×						
Trinidata 1000ffoliilik (II)		Ajouter un service standard				Démarrer la réconciliation	Annul

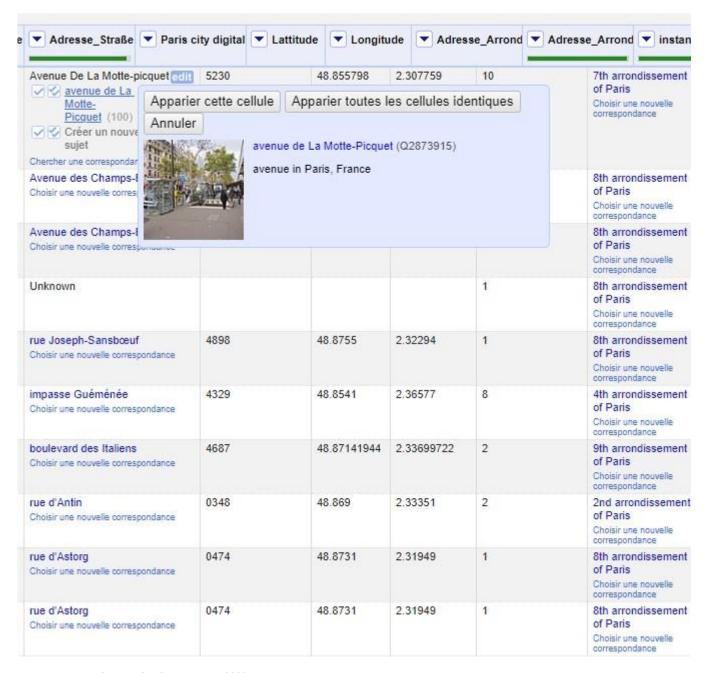
Réconciliation des données (suite)

- Type d'action de jugement permet de voir les données qui ont été réconciliées
- Auto sont les données qui ont trouvé une correspondance
- Similar, Open Refine nous laisse le choix pour faire correspondre les données
- None / Unconcilied , les données ne sont pas réconciliées
- **Blank**, la cellule est vide



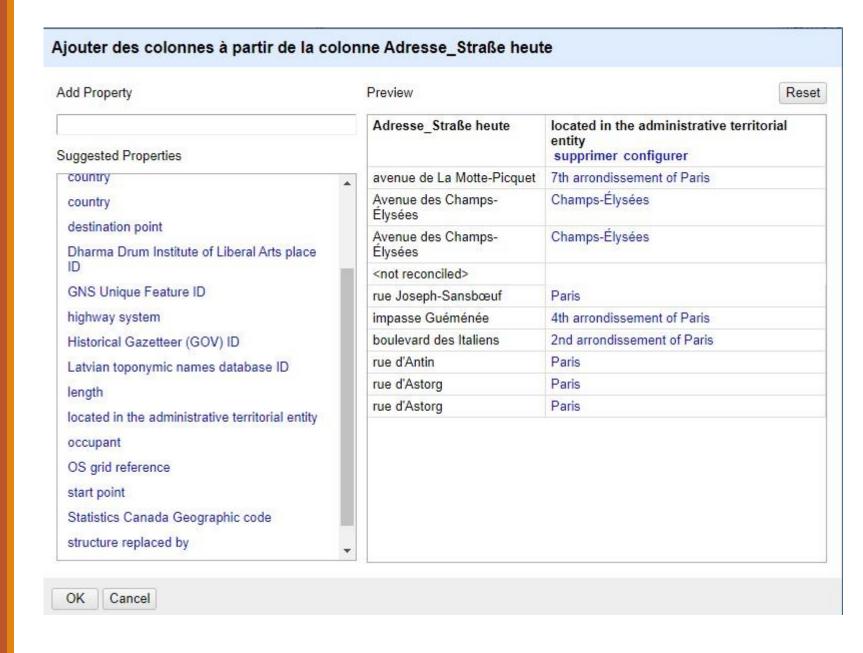
Réconciliation des données (suite)

- Apparier cette cellule, permet de faire correspondre uniquement cette cellule avec cette source de données externe
- identiques, permet de faire correspondre toutes les cellules qui ont le même nom avec cette source de données externe



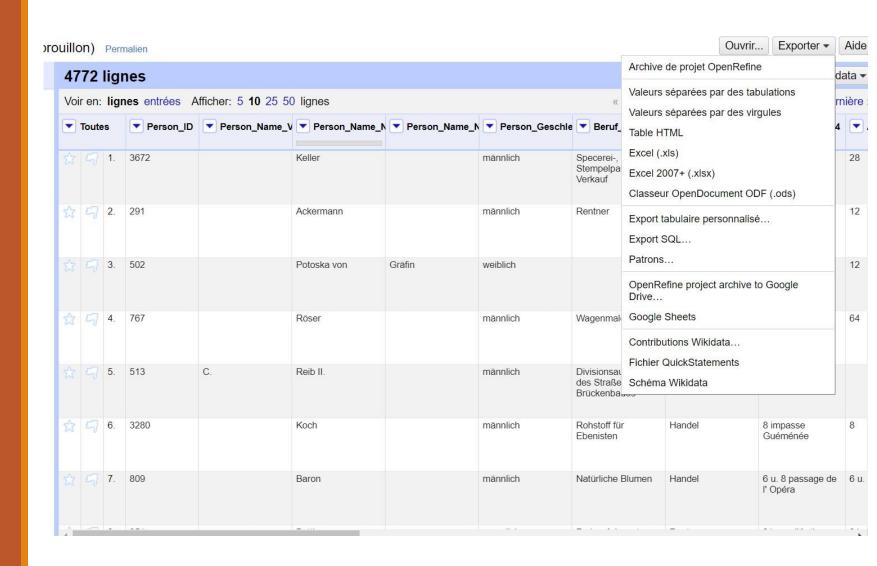
Ajouter des colonnes à partir de valeurs réconciliées

- Allez dans éditer la colonne >
 ajouter des colonnes à partir de
 valeurs réconciliés
- Cette option permet d'ajouter de nouvelles colonnes dans votre tableau de données avec la réconciliation faite précédemment
- Choisir la propriété qui correspond aux données que vous souhaitez afficher



Export des données

- Archive de projet Open Refine génère un projet Open Refine au format .tar
- Valeurs séparées par des tabulations/virgules génèrent des fichiers TSV et CSV
- Table HTML construit un tableau de données au format HTML qu'on peut insérer dans une page web
- Excel et Excel 2007 génère des tableaux pour différentes versions d'Excel
- ODF pour un classeur manipulable sous Libre Office ou Open Office Calc
- Patrons pour un format JSON
- Avec google drive en archive ou google sheets.



Liens utiles:

- •Seth van Hooland, Ruben Verborgh, et Max De Wilde, «Nettoyer ses données avec OpenRefine,» traduction par Sybille Clochet, *The Programming Historian en français* 1 (2019), https://doi.org/10.46430/phfr0004.
- Mathieu Saby, Nettoyer et préparer ses données avec Open Refine, 6/11/2020,
 https://msaby.gitlab.io/tutoriel-openrefine/index.html
- Documentation Open Refine, https://docs.openrefine.org/