

Cas pratique : d'un inventaire index PDF à un inventaire XML EAD (ou tout autre chose)

id	pre nom	cote	naissance_date	naissance_annee	naissance_lieu	coordonnees_geographiques	domicile	ph.
bastille_0	François Julien	F/1dIV/29			Authon-du-Perche	48.1905555555556;0.8505		
bastille_1	Jean Pierre	F/1dIV/29	25 mars 1768	1768	Hestroff	49.2622222222222;6.4347222222222	Merten (Moselle)	fusilier
bastille_2	Jean Nicolas	F/1dIV/32/2	28 janvier 1765	1765	Paris	48.5926444444444;2.3513585858586	214 faubourg Saint-Antoine - Paris (Seine)	
bastille_3	Jean François Joseph	F/1dIV/29	12 janvier 1762	1762	Ornat			
bastille_4	Jean Baptiste Duchesney	F/1dIV/29					3 rue de Sévres - Paris (Seine)	
bastille_5	Gustave	F/1dIV/29	6 novembre 1763	1763	Berville-en-Moselle	49.2713585858586;6.6477777777778	25 rue Mauffrard - Paris (Seine)	grenadier
bastille_6	François	F/1dIV/32/2	3 février 1763	1763	Marleville	47.9755555555556;5.9875	4 rue des Trois Coursures - Paris (Seine)	
bastille_7	Joseph	F/1dIV/29			Mollkirch	48.5063585858586;7.3852777777778	Mollkirch (Bas-Rhin)	
bastille_8	Jean	F/1dIV/32/2	17 novembre 1747	1747	Grosbliedestroff	49.1503888888889;7.0277777777778	25 rue de la Croix - Paris (Seine)	garde fran

ALEXANDRE (Jean Pierre).
Né le 25 mars 1768 à Hestroff (Moselle).
Domicile : Merten (Moselle).
Profession : préposé des douanes.
Au moment des faits : Fusilier.
Récompense obtenue : médaille.

F/1dIII/29.

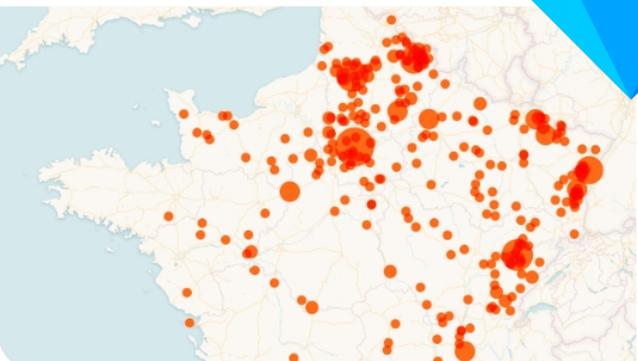
ARNOLDE (Jean).
Né le 17 novembre 1747 à Grosbliedestroff (Moselle).
Domicile : 25 rue de la Croix à Paris.
Profession : pompier de Paris.
Au moment des faits : Garde française.

F/1dIII/32/2.

ARNOLDE ou ARNOULD (Jean).
Domicile : 18 Grande-Rue à La Chapelle-Saint-Denis (Seine).
Récompense obtenue : secours.

ALISSE (Jean Nicolas).
Né le 28 janvier 1765 à Paris (Seine).
Domicile : 214 faubourg Saint-Antoine à Paris.
Profession : serrurier en métiers.
Récompense obtenue : pension de 250 francs.

F/1dIII/32/2.



```
<!-- DASTILLE_U -->
<unitid type="cote-de-consultation">F/1dIII/29</unitid>
<unitid>Dossier</unitid>
<unittitle>
<persname>AGUILLE, François Julien</persname>
</unittitle>
<unitdate calendar="gregorian" era="ce" normal="1830-01-01/1833-12-31">[1830-1833]</
<physdesc><extent>2 pièces</extent></physdesc>
</did>
<scopecontent>
<p>Date de naissance : non renseigné</p>
<p>Lieu de naissance : Authon (Eure-et-Loir)</p>
<p>Domicile : non renseigné</p>
<p>Profession : non renseigné</p>
<p>Profession au moment des faits : non renseigné</p>
<p>Récompense : non renseigné</p>
<p>Remarques : réclame le secours de 500 francs, qui lui a été accordé.</p>
</scopecontent>
</e>
```

À l'origine : un inventaire PDF

Vainqueurs de la Bastille F¹⁴III 29 à 32¹

3

Inventaire index PDF de 55 pages,
avec 2 colonnes.

Mise en forme typographique,
entête de page, numéro de page.

Contenu textuel structuré.

A

AGUILLÉ (François, Julien).

Né à Authon (Eure-et-Loir).

Remarques : Réclame le secours de 500 frs,
qui lui a été accordé. 2 pièces.

F/1dIII/29.

ALEXANDRE (Jean Pierre).

Né le 25 mars 1768 à Hestroff (Moselle).

Domicile : Merten (Moselle).

Profession : préposé des douanes.

Au moment des faits : Fusilier.

Récompense obtenue : médaille.

F/1dIII/29.

ALISSE (Jean Nicolas).

Né le 28 janvier 1765 à Paris (Seine).

Domicile : 214 faubourg Saint-Antoine à
Paris.

Profession : serrurier en métiers.

Récompense obtenue : pension de 250
francs.

F/1dIII/32/2.

ANDRÉ (François).

Né le 3 février 1763 à Martinville (Vosges).

Domicile : 4 rue des Trois Couronnes à
Paris.

Profession : tourneur en cuivre.

F/1dIII/32/2.

ARNOLD (Joseph).

Né à Mollkirch (Bas-Rhin).

Domicile : idem.

Récompense obtenue : médaille.

F/1dIII/29.

ARNOLD (Jean).

Né le 17 novembre 1747 à Grosbliederstroff
(Moselle).

Domicile : 25 rue de la Croix à Paris.

Profession : pompier de Paris.

Au moment des faits : Garde française.

F/1dIII/32/2.

ARNOLD ou ARNOULD (Jean).

Domicile : 18 Grande-Rue à La Chapelle-
Saint-Denis (Seine).

Récompense obtenue : secours.

Sommaire

1. Transformation du contenu en tableur structuré (diapos 4 à 27)
2. Harmonisation et enrichissement (diapos 28 à 41)
3. Exporter en XML EAD (diapos 42 à 46)
4. Des pistes pour aller plus loin... (diapos 47 à 49)

1. Transformation du contenu en tableur structuré

1.1. Copier le contenu texte

Copier le contenu du PDF

Coller dans Notepad++ (voire le Bloc notes)

Supprimer les parties introductives

7	Inventaire-index	1	A
8	par Monique Mayeur	2	AGUILLÉ (François, Julien).
9	secrétaire de documentation	3	Né à Authon (Eure-et-Loir).
10	2002	4	Remarques : Réclame le secours de 500 frs,
11	2	5	qui lui a été accordé. 2 pièces.
12	Vainqueurs de la Bastille F 1d III 29 à 32 1	6	F/1dIII/29.
13	GÉNÉRALITÉS	7	ALEXANDRE (Jean Pierre).
14	dossier n° 1. Lettres et pétitions en faveur des Vainqu	8	Né le 25 mars 1768 à Hestroff (Moselle).
15	dossier n° 2. Exécution de l'ordonnance royale du 8 mai	9	Domicile : Merten (Moselle).
16	dossier n° 3. Exécution de l'ordonnance royale du 1er n	10	Profession : préposé des douanes.
17	dossier n° 4. Préparation de la loi du 26 avril 1833 -	11	Au moment des faits : Fusilier.
18	dossier n° 5. Préparation de la loi du 26 avril 1833, c	12	Récompense obtenue : médaille.
19	des Archives du royaume.	13	F/1dIII/29.
20	dossier n° 6. Exécution de la loi du 26 avril 1833.	14	ALISSE (Jean Nicolas).
21	dossier n° 7. Demandes présentées par des veuves et enf	15	Né le 28 janvier 1765 à Paris (Seine).
22	de rejets.	16	Domicile : 214 faubourg Saint-Antoine à
23	dossier n° 8. Notes relatives aux Vainqueurs de la Bast	17	Paris.
24	dossier n° 9. Sauveurs et conservateurs du Trésor de la	18	Profession : serrurier en métiers.
25	F/1dIII/29 .	19	Récompense obtenue : pension de 250
26	2	20	francs.
27	3	21	F/1dIII/32/2.
28	Vainqueurs de la Bastille F 1d III 29 à 32 1	22	ALT (Jean François Joseph).
29	A	23	Né le 12 janvier 1762 à Dinant (Belgique).
30	AGUILLÉ (François, Julien).	24	Remarques : décédé le 16 août 1829, la
31	Né à Authon (Eure-et-Loir).	25	veuve née CHARLOT (Rolande Josèphe)
32	Remarques : Réclame le secours de 500 frs,	26	réclame un secours.
33	qui lui a été accordé. 2 pièces.	27	F/1dIII/29.
34	F/1dIII/29.		
35	ALEXANDRE (Jean Pierre).		
36	Né le 25 mars 1768 à Hestroff (Moselle).		

Copier le corps de l'inventaire

Petits nettoyages rapides

Dans le cas présent, le contenu de certains champs étant sur plusieurs lignes, on peut faire quelques petites transformations en masse rapides dans Notepad++

Faisable dans OpenRefine mais plus rapide dans Notepad++, en utilisant les expressions régulières dans le moteur Rechercher / Remplacer par (Ctrl + H)

rechercher : à\r (à en bout de ligne)

> remplacer par : à

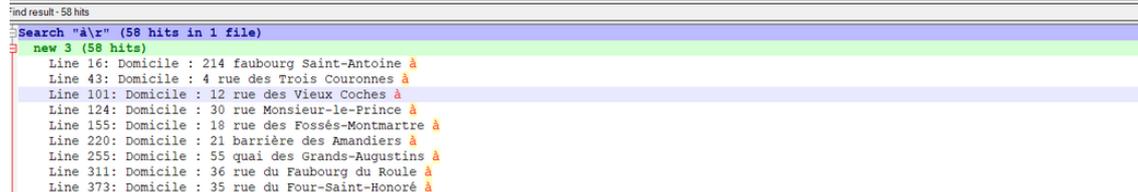
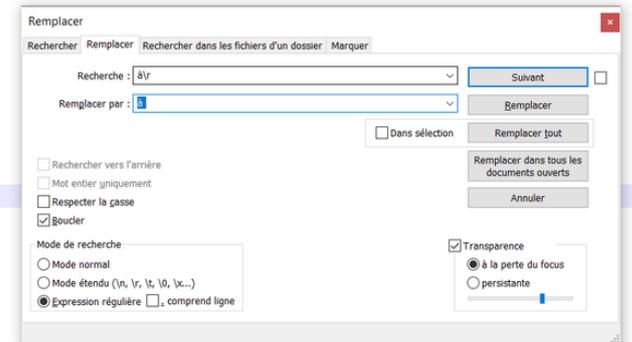
rechercher -\r (tiret en bout de ligne)

> remplacer par -

rechercher \n\ (saut de ligne suivi d'une parenthèse)

> remplacer par (

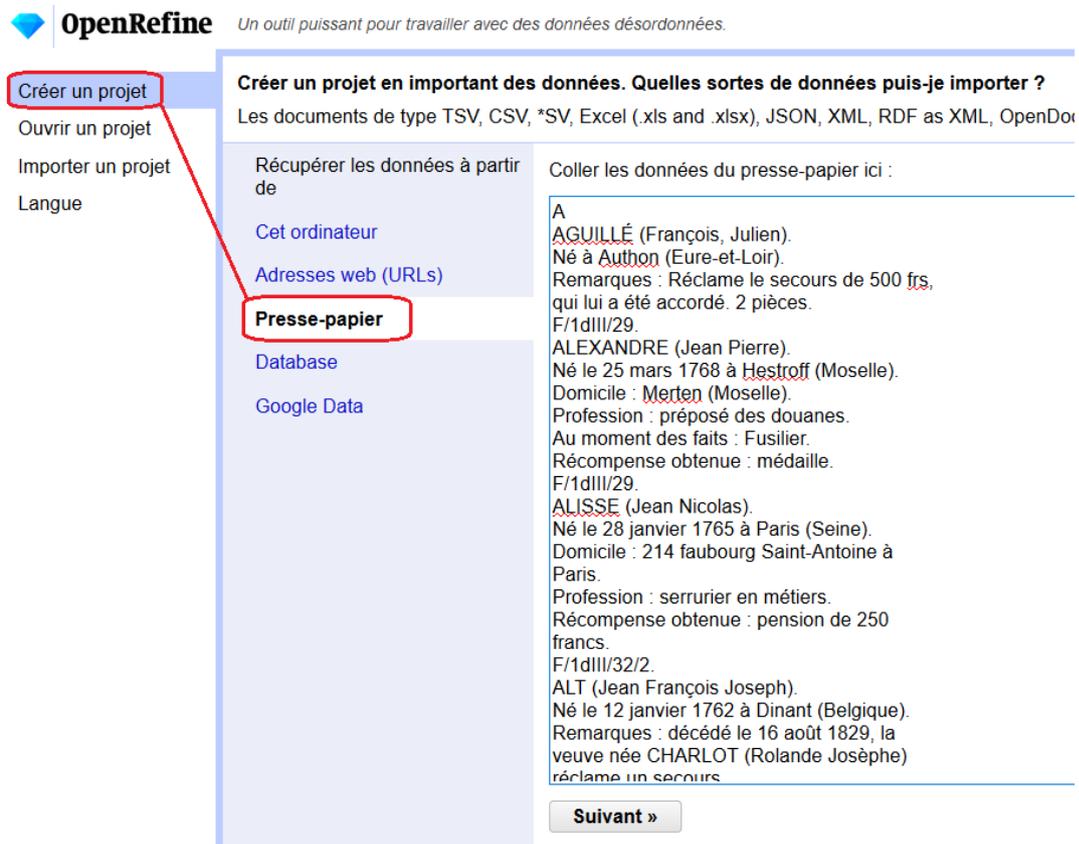
```
91 AUGÉAS (Nicolas).
92 Né le 2 septembre 1767 à Chatillon-sur-
93 Chalaronne (Ain).
94 Domicile : Oullins (Rhône).
95 Profession : tailleur d'habits.
96 Au moment des faits : Garde française.
97 F/1dIII/29.
98 AUMASSIP (Arnaud).
99 Né le 14 avril 1761 à Saint-Front-de-
100 Périgueux (Dordogne).
101 Domicile : 12 rue des Vieux Coches à
102 Versailles (Seine-et-Oise).
103 Profession : marchand de bestiaux.
104 F/1dIII/32/2.
105 B
106 BAETZ (François Etienne).
107 Au moment des faits : Garde française.
108 F/1dIII/29.
109 BAILLY (Louis Charles).
110 Né le 31 août 1764 à Aincourt (Seine-et-
```



1.2. Création du projet OpenRefine

Créer un projet OpenRefine en utilisant le mode Presse-Papier

Coller le contenu texte



OpenRefine *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 sortes de données puis-je importer ?
Les documents de type TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, OpenDoc

Récupérer les données à partir de
Cet ordinateur
Adresses web (URLs)
Presse-papier
Database
Google Data

Coller les données du presse-papier ici :

A
AGUILLÉ (François, Julien).
Né à Authon (Eure-et-Loir).
Remarques : Réclame le secours de 500 frs,
qui lui a été accordé. 2 pièces.
F/1dIII/29.
ALEXANDRE (Jean Pierre).
Né le 25 mars 1768 à Hestroff (Moselle).
Domicile : Merten (Moselle).
Profession : préposé des douanes.
Au moment des faits : Fusilier.
Récompense obtenue : médaille.
F/1dIII/29.
ALISSE (Jean Nicolas).
Né le 28 janvier 1765 à Paris (Seine).
Domicile : 214 faubourg Saint-Antoine à
Paris.
Profession : serrurier en métiers.
Récompense obtenue : pension de 250
francs.
F/1dIII/32/2.
ALT (Jean François Joseph).
Né le 12 janvier 1762 à Dinant (Belgique).
Remarques : décédé le 16 août 1829, la
veuve née CHARLOT (Rolande Josèphe)
réclame un secours.

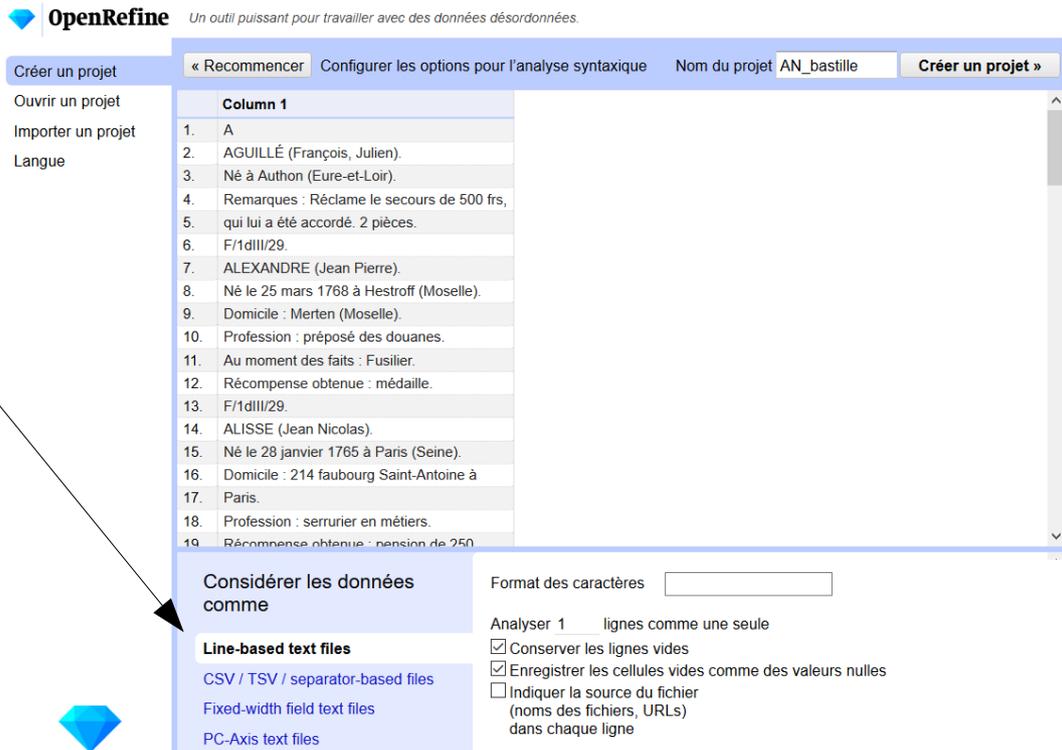
Suivant »

1.3. Paramétrage du projet

Vérifier l'aperçu des données dans OpenRefine (accents notamment).

Au besoin, forcer l'affichage en mode *Line-based text files* si ce n'est pas celui automatiquement détecté.

Nommer le projet et cliquer sur *Créer un projet*.



OpenRefine Un outil puissant pour travailler avec des données désordonnées.

Créer un projet « Recommencer Configurer les options pour l'analyse syntaxique Nom du projet AN_bastille Créer un projet »

Ouvrir un projet
Importer un projet
Langue

Column 1
1. A
2. AGUILLÉ (François, Julien).
3. Né à Authon (Eure-et-Loir).
4. Remarques : Réclame le secours de 500 frs,
5. qui lui a été accordé. 2 pièces.
6. F/1dIII/29.
7. ALEXANDRE (Jean Pierre).
8. Né le 25 mars 1768 à Hestroff (Moselle).
9. Domicile : Merten (Moselle).
10. Profession : préposé des douanes.
11. Au moment des faits : Fusilier.
12. Récompense obtenue : médaille.
13. F/1dIII/29.
14. ALISSE (Jean Nicolas).
15. Né le 28 janvier 1765 à Paris (Seine).
16. Domicile : 214 faubourg Saint-Antoine à
17. Paris.
18. Profession : serrurier en métiers.
19. Récompense obtenue : pension de 250

Considérer les données comme

Line-based text files

CSV / TSV / separator-based files

Fixed-width field text files

PC-Axis text files

Format des caractères

Analyser 1 lignes comme une seule

Conservier les lignes vides

Enregistrer les cellules vides comme des valeurs nulles

Indiquer la source du fichier (noms des fichiers, URLs) dans chaque ligne

1.4. Aperçu de l'interface de travail

On va maintenant travailler sur les 4315 lignes importées.

Au besoin, on peut modifier le nombre de lignes affichées pour un meilleur confort de lecture à l'écran.



OpenRefine AN_bastille [Permalien](#)

Facette / Filtre Défaire / Refaire 0 / 0

4315 lignes

Voir en: **lignes** entrées Afficher: **5 10 25 50** lignes

Toutes Column 1

<input type="checkbox"/>	<input type="checkbox"/>	1.	A
<input type="checkbox"/>	<input type="checkbox"/>	2.	AGUILLÉ (François, Julien).
<input type="checkbox"/>	<input type="checkbox"/>	3.	Né à Authon (Eure-et-Loir).
<input type="checkbox"/>	<input type="checkbox"/>	4.	Remarques : Réclame le secours de 500 frs,
<input type="checkbox"/>	<input type="checkbox"/>	5.	qui lui a été accordé. 2 pièces.
<input type="checkbox"/>	<input type="checkbox"/>	6.	F/1dIII/29.
<input type="checkbox"/>	<input type="checkbox"/>	7.	ALEXANDRE (Jean Pierre).
<input type="checkbox"/>	<input type="checkbox"/>	8.	Né le 25 mars 1768 à Hestroff (Moselle).
<input type="checkbox"/>	<input type="checkbox"/>	9.	Domicile : Merten (Moselle).
<input type="checkbox"/>	<input type="checkbox"/>	10.	Profession : préposé des douanes.

Utiliser les facettes et les filtres

Utiliser les facettes et les filtres pour sélectionner les sous-ensembles de données à traiter. Choisir les méthodes de facette et de filtre dans les menus situés dans les entêtes de colonne.

Vous ne savez pas par où commencer ? [Regarder ces tutoriels vidéos](#)

1.5. Restructurer les données

Cet inventaire décrit des dossiers créés au nom d'un individu, suivis de différents champs de description plus ou moins récurrents.

Il va falloir dégager les différents champs sur la base de caractéristiques communes.

A

AGUILLÉ (François, Julien).
Né à Authon (Eure-et-Loir).
Remarques : Réclame le secours de 500 frs,
qui lui a été accordé. 2 pièces.

F/1dIII/29.

ALEXANDRE (Jean Pierre).
Né le 25 mars 1768 à Hestroff (Moselle).
Domicile : Merten (Moselle).
Profession : préposé des douanes.
Au moment des faits : Fusilier.
Récompense obtenue : médaille.

F/1dIII/29.

☆	↩	1.	A
☆	↩	2.	AGUILLÉ (François, Julien).
☆	↩	3.	Né à Authon (Eure-et-Loir).
☆	↩	4.	Remarques : Réclame le secours de 500 frs,
☆	↩	5.	qui lui a été accordé. 2 pièces.
☆	↩	6.	F/1dIII/29.
☆	↩	7.	ALEXANDRE (Jean Pierre).
☆	↩	8.	Né le 25 mars 1768 à Hestroff (Moselle).
☆	↩	9.	Domicile : Merten (Moselle).
☆	↩	10.	Profession : préposé des douanes.

1.5.1. L'individu, entité-clé

Dans cet inventaire, l'individu (entité-clé des données) est facilement identifiable car décrit ainsi : NOM (Prénoms).

4315 lignes

Voir en: lignes entrées Afficher: 5 10 25 50 lignes

Toutes	Column 1
1.	Facette
2.	Filter le texte
3.	Éditer les cellules
4.	Éditer la colonne
5.	Transposer
6.	Trier...
7.	Aperçu
8.	Réconcilier
11.	Au moment des faits : Fusilier.
12.	Récompense obtenue : médaille.
13.	F/1dIII/29.
14.	ALISSE (Jean Nicolas).
15.	Né le 28 janvier 1765 à Paris (Seine).
16.	Domicile : 214 faubourg Saint-Antoine à
17.	Paris.

Un filtre de la colonne permet d'isoler les lignes dont les cellules commencent par 2 lettres majuscules qui se suivent :

$^[A-Z][A-Z]$

OpenRefine AN_bastille Permalien

Facette / Filtre Défaire / Refaire 0 / 6

Rafraîchir Tout réinitialiser Tout supprimer

Column 1 inverser réinitialiser

sensible à la casse expression rationnelle

723 matching lignes (4315 total)

Voir en: lignes entrées Afficher: 5 10 25 50 lignes

Toutes	Column 1
2.	AGUILLÉ (François, Julien).
7.	ALEXANDRE (Jean Pierre).
14.	ALISSE (Jean Nicolas).
22.	ALT (Jean François Joseph).
28.	ALTÊTE (Jean Baptiste Barthélemy).
33.	ALTMAYER (Guillaume).
41.	ANDRÉ (François).

Penser à cocher les cases !

Pour prendre en compte les éventuelles majuscules accentuées, on va complexifier un peu l'expression régulière (et éviter ainsi de reprendre 10 lignes à la main)

`^[A-Z|Â|Ê|Î|Ô|Û|Ä|Ë|Ï|Ö|Ü|À|Æ|æ|Ç|É|È|Œ|Ù][A-Z|Â|Ê|Î|Ô|Û|Ä|Ë|Ï|Ö|Ü|À|Æ|æ|Ç|É|È|Œ|Ù]`

The screenshot shows a search interface with a facet filter on the left and a list of results on the right.

Facette / Filtre Défaire / Refaire 0 / 3

Rafraîchir Tout réinitialiser Tout supprimer

Column 1 inverser réinitialiser

`^[A-Z|Â|Ê|Î|Ô|Û|Ä|Ë|Ï|Ö|Ü|À|Æ|æ|Ç|É|È|Œ|Ù][A-Z|Â|Ê|Î|Ô|Û|Ä|Ë|Ï|Ö|Ü|À|Æ|æ|Ç|É|È|Œ|Ù]`

sensible à la casse expression rationnelle

737 matching lignes (4315 total)

Voir en: **lignes** entrées Afficher: 5 10 25 50 lignes

Toutes		Column 1
☆	🗨	2. AGUILLÉ (François, Julien).
☆	🗨	7. ALEXANDRE (Jean Pierre).
☆	🗨	14. ALISSE (Jean Nicolas).
☆	🗨	22. ALT (Jean François Joseph).
☆	🗨	28. ALTÊTE (Jean Baptiste Barthélemy).
☆	🗨	33. ALTMAYER (Guillaume).
☆	🗨	41. ANDRÉ (François).
☆	🗨	47. ARNOLD (Joseph).
☆	🗨	52. ARNOI D (Jean)

Facette / Filtre Défaire / Refaire 0 / 3 **737 matching lignes** (4315 total)

Voir en: **lignes** entrées Afficher: 5 10 25 50 lignes

Rafraîchir Tout réinitialiser Tout supprimer

Column 1 inverser réinitialiser

^[A-Z|À|Ê|Ë|Ï|Ô|Û|Ä|Æ|Ï|Ö|Ü|À|Æ|æ|Ç|É|Ê|Ë|Ï|Ü][A-

sensible à la casse expression rationnelle

Toutes	Column 1
2.	Facette
7.	Filtrer le texte
14.	Éditer les cellules
22.	Éditer la colonne
28.	Transposer
33.	Trier...
41.	Aperçu
47.	Réconcilier
52.	ARRAS (Nicolas François)
59.	ASTIER (Claude Marin)
64.	AUBANEL (Jean Jacques)
71.	AUBUSSON (Jean Baptiste)

Facette
Filtrer le texte
Éditer les cellules
Éditer la colonne
Transposer
Trier...
Aperçu
Réconcilier
Ajouter une colonne en fonction de cette colonne...
Ajouter une colonne en moissonnant des URL...
Ajouter des colonnes à partir de valeurs réconciliées...
Renommer cette colonne
Supprimer cette colonne
Déplacer la colonne en premier

Ajouter une colonne en fonction de cette colonne pour dupliquer le contenu filtré.

value (par défaut) permet de dupliquer le contenu

Après vérification de l'aperçu, nommer la colonne et cliquer sur OK.

Ajouter une colonne en fonction d'une colonne Column 1

Nouveau nom de colonne

En cas d'erreur vider la cellule conserver l'erreur copier la valeur depuis la colonne originale

Expression Langue Pas d'erreur de syntaxe.

Aperçu Historique Étoilé Aide

row	value	value
2.	AGUILLÉ (François, Julien).	AGUILLÉ (François, Julien).
7.	ALEXANDRE (Jean Pierre).	ALEXANDRE (Jean Pierre).
14.	ALISSE (Jean Nicolas).	ALISSE (Jean Nicolas).
22.	ALT (Jean François Joseph).	ALT (Jean François Joseph).
28.	ALTÊTE (Jean Baptiste Barthélemy).	ALTÊTE (Jean Baptiste Barthélemy).
33.	ALTMAYER (Guillaume).	ALTMAYER (Guillaume).

OK Annuler

Facette / Filtre Défaire / Refaire 1 / 1

Rafraichir Tout réinitialiser Tout supprimer

Column 1 [inverser](#) [réinitialiser](#)

`^[A-Z|À|É|Ï|Ô|Ù|Â|Ê|Ï|Ó|Ü|Ä|Æ|œ|Ç|É|È|Œ|Û]A`

sensible à la casse expression rationnelle

737 matching lignes (4315 total)

Voir en: lignes entrées Afficher: 5 10 25 50 lignes

Toutes	Column 1	personne
2.	Facette	AGUILLÉ (François, Julien).
7.	Facette	ALEXANDRE (Jean Pierre).
14.	Facette	ALISSE (Jean Nicolas).
22.	Facette	Joseph).
28.	Facette	
33.	Facette	
41.	Facette	
47.	Facette	
52.	Facette	
59.	Facette	
64.	ARRAS (Nicolas François)	
71.	ASTIER (Claude Marin).	
76.	AUBANEL (Jean Jacques)	
87.	AUBUSSON (Jean Baptiste d')	AUBUSSON (Jean E
91.	AUGEAS (Nicolas).	AUGEAS (Nicolas).
98.	AUMASSIP (Arnaud).	AUMASSIP (Arnaud
106.	BAETZ (François Etienne).	BAETZ (François E
109.	BAILLY (Louis Charles).	BAILLY (Louis Char
116.	BAILLY (Simon Paul).	BAILLY (Simon Pau
123.	BAIZE (Jean Nicolas).	BAIZE (Jean Nicolas).
127.	BANQUE.	BANQUE.
133.	BANSE (Jean Joseph).	BANSE (Jean Joseph).
140.	BARASSE (Georges Pascal).	BARASSE (Georges Pascal).
145.	BARBAT (Claude).	BARBAT (Claude).
154.	BARBET (Joseph Charles).	BARBET (Joseph Charles).

Les Noms et prénoms des individus étant désormais dupliqués dans la colonne *personne*, on peut effacer ces données de la colonne initiale.

Éditer les cellules / Transformations courantes / Transformer en chaîne vide

4315 lignes

Voir en: [lignes entrées](#) Afficher: 5 10 25 50 lignes

Toutes	personne	Column 1
☆		1. A
☆		2. AGUILLÉ (François, Julien).
☆		3. Né à Authon (Eure-et-Loir).
☆		4. Remarques : Réclame le secours de 500 frs,
☆		5. qui lui a été accordé. 2 pièces.
☆		6. F/1dIII/29.
☆		7. ALEXANDRE (Jean Pierre).
☆		8. Né le 25 mars 1768 à Hestroff (Moselle).
☆		9. Domicile : Merten (Moselle).
☆		10. Profession : préposé des douanes.
☆		11. Au moment des faits : Fusilier.
☆		12. Récompense obtenue : médaille.
☆		13. F/1dIII/29.

4315 lignes

Voir en: [lignes entrées](#) Afficher: 5 10 25 50 lignes

Toutes	Column 1	personne
☆		1. A
☆		2. AGUILLÉ (François, Julien).
☆		3. Né à Authon (Eure-et-Loir).
☆		4. Remarques : Réclame le secours de 500 frs,
☆		5. qui lui a été accordé. 2 pièces.
☆		6. F/1dIII/29.
☆		7. ALEXANDRE (Jean Pierre).
☆		8. Né le 25 mars 1768 à Hestroff (Moselle).
☆		9. Domicile : Merten (Moselle).

Transformer

Facette

Éditer les lignes

Éditer les colonnes

Aperçu

Retrier / supprimer les colonnes...

Recopier les valeurs dans les cellules vides consécutives

Vider les valeurs répétées dans des cellules consécutives

Inversion de l'ordre des colonnes

Trier / Supprimer des colonnes

Glisser des colonnes pour les trier

Déposer de:

personne

Column 1

4315 lignes

Voir en: [lignes entrées](#) Afficher: 5 10 25 50 lignes

Toutes	personne	Column 1
☆		1. A
☆		2. AGUILLÉ (François, Julien).
☆		3. Né à Authon (Eure-et-Loir).
☆		4. Remarques : Réclame le secours de 500 frs,
☆		5. qui lui a été accordé. 2 pièces.
☆		6. F/1dIII/29.
☆		7. ALEXANDRE (Jean Pierre).
☆		8. Né le 25 mars 1768 à Hestroff (Moselle).
☆		9. Domicile : Merten (Moselle).
☆		10. Profession : préposé des douanes.
☆		11. Au moment des faits : Fusilier.
☆		12. Récompense obtenue : médaille.
☆		13. F/1dIII/29.

4315 lignes...

4315 lignes			
Voir en: lignes entrées Afficher: 5 10 25 50 lignes			
Toutes	personne	Column 1	
☆	🗨	1.	A
☆	🗨	2.	AGUILLÉ (François, Julien).
☆	🗨	3.	Né à Authon (Eure-et-Loir).
☆	🗨	4.	Remarques : Réclame le secours de 500 frs,
☆	🗨	5.	qui lui a été accordé. 2 pièces.
☆	🗨	6.	F/1dIII/29.
☆	🗨	7.	ALEXANDRE (Jean Pierre).
☆	🗨	8.	Né le 25 mars 1768 à Hestroff (Moselle).
☆	🗨	9.	Domicile : Merten (Moselle).
☆	🗨	10.	Profession : préposé des douanes.
☆	🗨	11.	Au moment des faits : Fusilier.
☆	🗨	12.	Récompense obtenue : médaille.
☆	🗨	13.	F/1dIII/29.

738 entrées (= enregistrements)

on retrouve l'équivalent de notre notice initiale

738 entrées			
Voir en: lignes entrées Afficher: 5 10 25 50 entrées			
Toutes	personne	Column 1	
☆	🗨	1.	A
☆	🗨	2.	AGUILLÉ (François, Julien).
☆	🗨		Né à Authon (Eure-et-Loir).
☆	🗨		Remarques : Réclame le secours de 500 frs,
☆	🗨		qui lui a été accordé. 2 pièces.
☆	🗨		F/1dIII/29.
☆	🗨	3.	ALEXANDRE (Jean Pierre).
☆	🗨		Né le 25 mars 1768 à Hestroff (Moselle).
☆	🗨		Domicile : Merten (Moselle).
☆	🗨		Profession : préposé des douanes.
☆	🗨		Au moment des faits : Fusilier.
☆	🗨		Récompense obtenue : médaille.
☆	🗨		F/1dIII/29.

Le mode *Entrées* permet de garder une notion de lien entre la première colonne (entité-clé) et les suivantes.

Toutes		personne	Column 1
☆	🗨	1.	A
☆	🗨	2. AGUILLÉ (François, Julien).	
☆	🗨		Né à Authon (Eure-et-Loir).
☆	🗨		Remarques : Réclame le secours de 500 frs,
☆	🗨		qui lui a été accordé. 2 pièces.
☆	🗨		F/1dIII/29.
☆	🗨	3. ALEXANDRE (Jean Pierre).	
☆	🗨		Né le 25 mars 1768 à Hestroff (Moselle).
☆	🗨		Domicile : Merten (Moselle).
☆	🗨		Profession : préposé des douanes.
☆	🗨		Au moment des faits : Fusilier.
☆	🗨		Récompense obtenue : médaille.
☆	🗨		F/1dIII/29.
☆	🗨	4. ALISSE (Jean Nicolas).	
☆	🗨		Né le 28 janvier 1765 à Paris (Seine).
☆	🗨		Domicile : 214 faubourg Saint-Antoine à
☆	🗨		Paris.
☆	🗨		Profession : serrurier en métiers.
☆	🗨		Récompense obtenue : pension de 250
☆	🗨		francs.
☆	🗨		F/1dIII/32/2.

Les étapes suivantes vont consister à restructurer les différents champs (domicile, donnée de naissance, profession, etc.) en tableur

Toutes		nom	id	prenom	cote	naissance_date	naissance_lieu	domicile
☆	🗨	1. AGUILLÉ	bastille_0	François Julien	F/1dIII/29		Authon (Eure-et-Loir)	
☆	🗨	2. ALEXANDRE	bastille_1	Jean Pierre	F/1dIII/29	25 mars 1768	Hestroff (Moselle)	Merten (Moselle)
☆	🗨	3. ALISSE	bastille_2	Jean Nicolas	F/1dIII/32/2	28 janvier 1765	Paris (Seine)	214 faubourg Saint-Antoine - Paris (Seine)
☆	🗨	4. ALT	bastille_3	Jean François Joseph	F/1dIII/29	12 janvier 1762	Dinant (Belgique)	
☆	🗨	5. ALTÈTE	bastille_4	Jean Baptiste Barthélemy	F/1dIII/29			3 rue de Sèvres - Paris (Seine)
☆	🗨	6. ALTMAYER	bastille_5	Guillaume	F/1dIII/29	6 novembre 1763	Berviller-en-Moselle (Moselle)	28 rue Mouffetard - Paris (Seine)

Continuons donc à extraire le contenu de *Column 1* (en mode *Lignes* !)

1.5.2. Données de naissance, etc.

Filtrer sur « tout ce qui commence par Né » : `^Né` (casse et expression rationnelles cochées)

Dupliquer le contenu : *Ajouter une colonne en fonction de cette colonne*

The screenshot shows a data table interface with a filter applied to the first column. The filter is '^Né', with checkboxes for 'sensible à la casse' and 'expression rationnelle' both checked. The table has 467 matching lines out of 4315 total. A context menu is open over the table, showing various actions. The option 'Ajouter une colonne en fonction de cette colonne...' is highlighted.

Facette / Filtre Défaire / Refaire 3 / 3

Rafraîchir Tout réinitialiser Tout supprimer

Column 1 inverser réinitialiser

^Né

sensible à la casse expression rationnelle

467 matching lignes (4315 total)

Voir en: lignes entrées Afficher: 5 10 25 50 lignes

Toutes	personne	Column 1
3.		
8.		
15.		
23.		
34.		
42.		
48.		
53.		
65.		
72.		

- Facette
- Filtrer le texte
- Éditer les cellules
- Éditer la colonne
- Transposer
- Trier...
- Aperçu
- Réconcilier

- Diviser en plusieurs colonnes...
- Joindre des colonnes...
- Ajouter une colonne en fonction de cette colonne...
- Ajouter une colonne en moissonnant des URL...
- Ajouter des colonnes à partir de valeurs réconciliées

Création de la colonne *naissance*,
 puis *cote* (après un filtre sur : F/1dIII)
 et ainsi de suite...

467 matching lignes (4315 total)

Voir en: **lignes entrées** Afficher: 5 10 25 50 lignes

Toutes	personne	Column 1	naissance
3.		Facette	Né à Authon (Eure-et-Loir).
8.		Filter le texte	Né le 25 mars 1768 à Hestroff (Moselle).
15.		Éditer les cellules	Né le 28 janvier 1765 à Paris (Seine).
23.		Éditer la colonne	Né le 12 janvier 1762 à Dinant (Belgique).
34.		Transposer	Né le 6 novembre 1763 à Berviller-en-
42.		Trier...	Né le 3 février 1763 à Martinville (Vosges).
48.		Aperçu	Né à Millincourt (Somme).
53.		Réconcilier	Né le 26 décembre 1765 à Aincourt (Seine-et-
55.			Né le 5 décembre 1763 à Aincourt (Seine-et-
65.			Né le 2 septembre 1767 à Vitry-le-François
72.			Né le 14 avril 1761 à Saint-Front-de-
77.			Né le 31 août 1764 à Aincourt (Seine-et-
88.			Né le 28 octobre 1766 à Aincourt (Seine-et-
92.			Né le 8 février 1772 à Paris (Seine).
99.			Né le 8 avril 1768 à Avesnes (Nord).
110.			Né le 9 juin 1766 à Censy-Gailly (Somme).
117.			Né le 6 juillet 1759 à Vitry-le-François
128.			
134.			
141.			
146.			

Ajouter une colonne en fonction d'une colonne Column 1

Nouveau nom de colonne:

En cas d'erreur: vider la cellule conserver l'erreur copier la valeur depuis la colonne originale

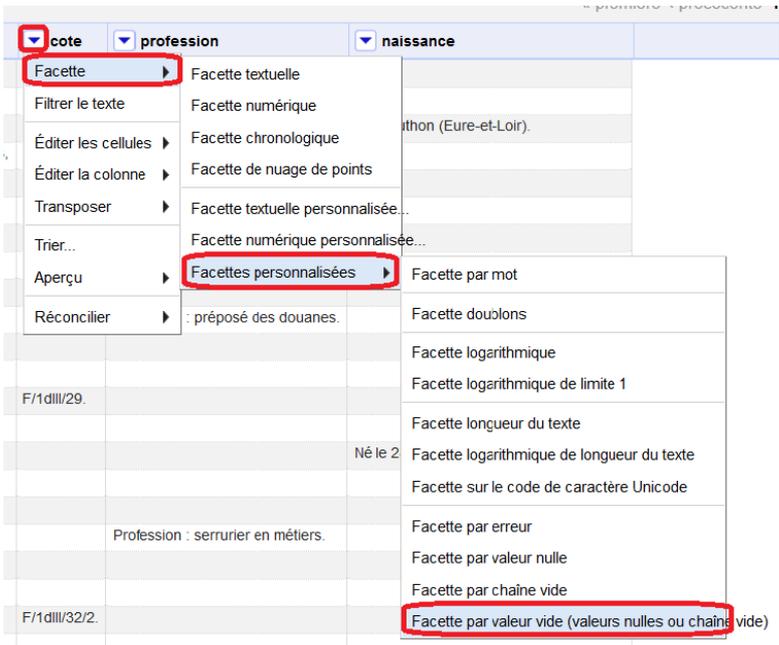
Expression: Langue: Pas d'erreur de syntaxe.

Aperçu Historique Étoilée Aide

row	value	value
3.	Né à Authon (Eure-et-Loir).	Né à Authon (Eure-et-Loir).
8.	Né le 25 mars 1768 à Hestroff (Moselle).	Né le 25 mars 1768 à Hestroff (Moselle).
15.	Né le 28 janvier 1765 à Paris (Seine).	Né le 28 janvier 1765 à Paris (Seine).
23.	Né le 12 janvier 1762 à Dinant (Belgique).	Né le 12 janvier 1762 à Dinant (Belgique).
34.	Né le 6 novembre 1763 à Berviller-en-	Né le 6 novembre 1763 à Berviller-en-
42.	Né le 3 février 1763 à Martinville (Vosges).	Né le 3 février 1763 à Martinville (Vosges).
48.	Né à Millincourt (Somme).	Né à Millincourt (Somme).

OK Annuler

Et surtout on n'oublie pas de *transformer en chaîne vide* le contenu dupliqué !



Si on a oublié d'effacer un contenu dupliqué, utiliser le système de facettes par valeur vide :

Facette / Facettes personnalisées / Facette par valeur vide

Facette / Filtre Défaire / Refaire 8 / 8

Rafraîchir Tout réinitialiser Tout supprimer

698 matching lignes (4315 total)

Voir en: **lignes entrées** Afficher: 5 10 25 50 lignes

Toutes
 personne
 Column 1
 cote
 profession

<input type="checkbox"/>	<input type="checkbox"/>	6.	F/1dIII/29.	F/1dIII/29.
<input type="checkbox"/>	<input type="checkbox"/>	13.	F/1dIII/29.	F/1dIII/29.
<input type="checkbox"/>	<input type="checkbox"/>	21.	F/1dIII/32/2.	F/1dIII/32/2.
<input type="checkbox"/>	<input type="checkbox"/>	27.	F/1dIII/29.	F/1dIII/29.
<input type="checkbox"/>	<input type="checkbox"/>	32.	F/1dIII/29.	F/1dIII/29.
<input type="checkbox"/>	<input type="checkbox"/>	40.	F/1dIII/29.	F/1dIII/29.
<input type="checkbox"/>	<input type="checkbox"/>	46.	F/1dIII/32/2.	F/1dIII/32/2.
<input type="checkbox"/>	<input type="checkbox"/>	51.	F/1dIII/29.	F/1dIII/29.
<input type="checkbox"/>	<input type="checkbox"/>	58.	F/1dIII/32/2.	F/1dIII/32/2.
<input type="checkbox"/>	<input type="checkbox"/>	63.	F/1dIII/29.	F/1dIII/29.

cote changer inverser réinitialiser
 2 choix Trier par: **nom** compte
 false 698 exclude
 true 3617
 Facette par nombre de choix

False / true permet de sélectionner les lignes vides / non vides (pour les transformer en chaîne vide au besoin)

1.5.3. Éléments de mise en page

Les données ayant été copiées d'un PDF, il y a des éléments de mise en page à supprimer en masse.

- Numéro de page : `^[0-9]{1,2}$` (cellules avec un ou deux chiffres uniquement)
- Entête : `^Vainqueurs de la Bastille F`
- Lettres du classement alphabétique

Facette / Filtre Défaire / Refaire 12 / 12

Rafraîchir Tout réinitialiser Tout supprimer

Column 1 inverser réinitialiser

sensible à la casse expression rationnelle

52 matching lignes (4210 total)

Voir en: lignes entrées Afficher: 5 10 25 50 lignes

Toutes	personne	Column 1
☆	🗨	2080. Vainqueurs de la Bastille F 1d III 29 à 32 1
☆	🗨	2163. Vainqueurs de la Bastille F 1d III 29 à 32 1
☆	🗨	2244. Vainqueurs de la Bastille F 1d III 29 à 32 1
☆	🗨	2323. Vainqueurs de la Bastille F 1d III 29 à 32 1

Facette / Filtre Défaire / Refaire 11 / 15

Rafraîchir Tout réinitialiser Tout supprimer

Column 1 inverser réinitialiser

sensible à la casse expression rationnelle

105 matching lignes (4315 total)

Voir en: lignes entrées Afficher: 5 10 25 50

Toutes	personne	Column 1
☆	🗨	84. 3
☆	🗨	85. 4
☆	🗨	168. 4
☆	🗨	169. 5
☆	🗨	250. 5
☆	🗨	251. 6

1.5.4. Le champ Remarques

L'essentiel du contenu restant dans la colonne initiale relève désormais du champ *Remarques*. Mais le contenu a souvent été tronqué sur plusieurs lignes...

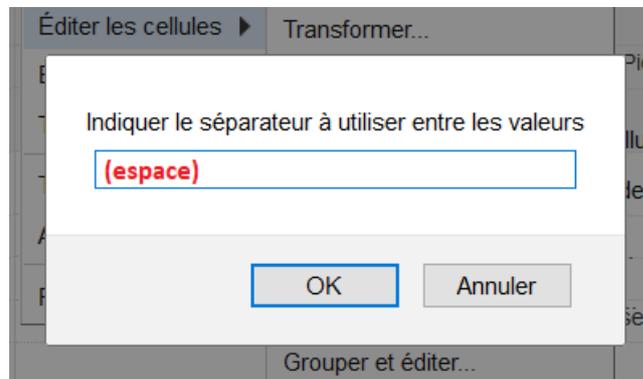
Solution proposée :

- Faire un filtre : *^Remarques*
- Ajouter une colonne en fonction de cette colonne (ici *remarques1*)
- Mettre la colonne *remarques1* en première colonne et basculer en mode *entrées*
- On va pouvoir ainsi agir sur *Column1* en la liant à la colonne clé *remarques1*

Toutes	remarques1	Column 1
3.	Remarques : Réclame le secours de 500 frs,	qui lui a été accordé. 2 pièces.
96.	Remarques : décédé le 2 mai 1832, la veuve	réclame un secours, 2 pièces.
154.	Remarques : décédé le 24 novembre 1834, 2	pièces.
191.	Remarques : n'a pas pris les armes, a assisté	à la prise de la Bastille, a secouru les assiégeants, a exposé sa vie par son intrépidité et espère être récompensée et reconnue comme Vainqueur de la Bastille, 1 pièce.

- Fusion des contenus de *Column1* divisés sur plusieurs lignes qui relèvent de la même entrée

remarques1	Column 1	personne
Remarques : Réclame le secours de 500 frs,		
	2 pièces.	
Remarques : décédé le 2 mai 1832, la veuve		
Remarques : décédé le 24 novembre 1834, 2		
	pièces.	
		BOYER (Gabriel Et
Remarques : n'a pas pris les armes, a assisté		
	à la prise de la Bastille, a secouru les	
	assiégeants, a exposé sa vie par son	
	intrépidité et espère être récompensée et	
	reconnue comme Vainqueur de la Bastille,1	



Remarques : décédé le 2 mai 1832, la veuve	réclame un secours, 2 pièces.
Remarques : décédé le 24 novembre 1834, 2	pièces.
Remarques : n'a pas pris les armes, a assisté	à la prise de la Bastille, a secouru les assiégeants, a exposé sa vie par son intrépidité et espère être récompensée et reconnue comme Vainqueur de la Bastille,1 pièce.

- Fusion des colonnes contenant les éléments du champ *Remarques* :
éditer les cellules / value + " " + cells["Column 1"].value

Transformation textuelle personnalisée sur la colonne remarques1

Expression Langue General Refine Expression Language (GREL) ▾

`value + " " + cells["Column 1"].value`

Pa:

Aperçu Historique Étoilée Aide

row	value	value + " " + cells["Column 1" ...
3.	Remarques : Réclame le secours de 500 frs,	Remarques : Réclame le secours de 500 frs, qui lui a été accordé. 2 pièces.
4.	null	null
5.	null	null
6.	null	null
301.	Remarques : décédé le 2 mai 1832, la veuve	Remarques : décédé le 2 mai 1832, la veuve réclame un secours, 2 pièces.
302.	null	null
...

En cas d'erreur conserver l'original vider la cellule conserver l'erreur

Retransformer fois maximum, tant que les données changent

On obtient finalement, après quelques résidus à reprendre à la main, quelque chose qui ressemble à ça :

- une colonne clé *peersonne*
- un contenu structuré dans les colonnes suivantes

Il ne reste plus qu'à dupliquer le contenu de chaque colonne afin de ne conserver qu'une seule ligne avec tous les champs renseignés.

737 entrées

Voir en: lignes entrées Afficher: 5 10 25 50 entrées

Toutes	personne	recompense	domicile	remarques	profession_ba
1.	AGUILLÉ (François, Julien).				
				Remarques : Réclame le secours de 500 frs, qui lui a été accordé. 2 pièces.	
2.	ALEXANDRE (Jean Pierre).				
			Domicile : Merten (Moselle).		
		Récompense obtenue : médaille.			Au m Fusil
3.	ALISSE (Jean Nicolas).				

1.5.5. Duplication du contenu

En mode ***Entrées***, pour chaque colonne sauf la colonne clé (la première), en changeant à chaque fois le nom de la "colonne" :

éditer les cellules / row.record.cells["domicile"].value

Transformation textuelle personnalisée sur la colonne domicile

Expression Langue General Refine Expression Language (GREL)

```
row.record.cells["domicile"].value.slice(0).join(" ")
```

Aperçu Historique Étoilée Aide

3.	null	
4.	null	
5.	null	Domicile : Merten (Moselle).
6.	null	Domicile : Merten (Moselle).
7.	Domicile : Merten (Moselle).	Domicile : Merten (Moselle).
8.	null	Domicile : Merten (Moselle).
9.	null	Domicile : Merten (Moselle).
10.	null	Domicile : Merten (Moselle).

En cas d'erreur conserver l'original Retransformer 10 fois maximum, tant que les données changent
 vider la cellule
 conserver l'erreur

OK Annuler

1.5.6. Vers un tableur structuré...

En mode **Lignes**, sélection des lignes dont la colonne clé est vide :

*Facette / Facettes personnalisées
/ Facette par valeur vide / True*

Facette / Filtre Défaire / Refaire 362 / 362

2988 matching lignes (3721 total)

Voir en: lignes entrées Afficher: 5 10 25 50 lignes

Toutes personne récompense

Éditer les lignes

Supprimer les lignes correspondantes

Voir en: lignes entrées Afficher: 5 10 25 50 entrées

Toutes personne récompense domicile remarques profession_bas cote

Facette

Facettes personnalisées

Facette par valeur vide (valeurs nulles ou chaînes vides)

Suppression : *Éditer les lignes / Supprimer les lignes correspondantes*

2. Harmonisation et enrichissement

6. Harmonisation et enrichissement

En deux grosses heures, le PDF textuel est devenu un beau tableur structuré (qui demande encore un peu de toilettage, mais le plus gros est fait).

733 lignes										Extensions: Wikidata ▾									
Voir en: lignes entrées Afficher: 5 10 25 50 lignes										« première < précédente 1 - 25 suivante > dernière »									
<input type="checkbox"/>	<input type="checkbox"/>	Toutes	<input type="checkbox"/>	personne	<input type="checkbox"/>	recompense	<input type="checkbox"/>	domicile	<input type="checkbox"/>	remarques	<input type="checkbox"/>	profession_bas	<input type="checkbox"/>	cote	<input type="checkbox"/>	profession	<input type="checkbox"/>	naissance	
<input type="checkbox"/>	<input type="checkbox"/>	1.		AGUILLÉ (François, Julien).						Remarques : Réclame le secours de 500 frs, qui lui a été accordé. 2 pièces.					F/1dIII/29.			Né à Authon (Eure-et-Loir).	
<input type="checkbox"/>	<input type="checkbox"/>	2.		ALEXANDRE (Jean Pierre).	Récompense obtenue : médaille.			Domicile : Merten (Moselle).				Au moment des faits : Fusilier.			F/1dIII/29.			Né le 25 mars 1768 à Hestroff (Moselle).	
<input type="checkbox"/>	<input type="checkbox"/>	3.		ALISSE (Jean Nicolas).	Récompense obtenue : pension de 250 francs.			Domicile : 214 faubourg Saint-Antoine à Paris							F/1dIII/32/2.			Né le 28 janvier 1765 à Paris (Seine).	
<input type="checkbox"/>	<input type="checkbox"/>	4.		ALT (Jean François Joseph).						Remarques : décédé le 16 août 1829, la veuve née CHARLOT (Rolande Josèphe) réclame un secours.					F/1dIII/29.			Né le 12 janvier 1762 à Dinant (Belgique).	
<input type="checkbox"/>	<input type="checkbox"/>	5.		ALTÈTE (Jean Baptiste Barthélemy).	Récompense obtenue : secours.			Domicile : 3 rue de Sèvres à Paris.							F/1dIII/29.				
<input type="checkbox"/>	<input type="checkbox"/>	6.		ALTMAYER (Guillaume).	Récompense obtenue : médaille.			Domicile : 28 rue Mouffetard à Paris.				Au moment des faits : Grenadier.			F/1dIII/29.			Né le 6 novembre 1763 à Berviller-en-Moselle (Moselle).	
<input type="checkbox"/>	<input type="checkbox"/>	7.		ANDRÉ (François).				Domicile : 4 rue des Trois Couronnes à Paris							F/1dIII/32/2.			Né le 3 février 1763 à Martinville (Vosges).	
<input type="checkbox"/>	<input type="checkbox"/>	8.		ARNOLD (Joseph).	Récompense obtenue : médaille.			Domicile : idem.							F/1dIII/29.			Né à Mollkirch (Bas-Rhin).	
<input type="checkbox"/>	<input type="checkbox"/>	9.		ARNOLD (Jean).				Domicile : 25 rue de la Croix à Paris.				Au moment des faits : Garde française.			F/1dIII/32/2.			Né le 17 novembre 1747 à Grosbliederstroff (Moselle).	

2.1. Colonnes Nom et prénoms

Division de la colonne *personne* en une colonne *nom* et une autre *prenoms*

Toutes	personne	cote	naissance	domi
1.	Facette	F/1dIII/29.	Né à Authon (Eure-et-Loir).	
2.	ALTÈTE (Jean Baptiste Barthélemy).			
3.	ALTMAYER (Guillaume).			
4.	ANDRÉ (François).			
5.	ARNOLD (Joseph).			
6.	ARNOLD (Jean).			
7.	ARNOLD (Nicolas).			
8.	ARNOLD (Nicolas).			
9.	ARNOLD (Nicolas).			
10.	ARNOLD (Nicolas).			

Diviser la colonne personne en plusieurs colonnes

Méthode de division de la colonne

par séparateur
Séparateur () expression rationnelle
Diviser en 2 colonnes au plus (laisser vide pour ne pas limiter)

par les longueurs de champs

Liste les longueurs en les séparant par des virgules, par exemple 5, 7, 15

Après la division

Deviner le type de cellule
 Supprimer cette colonne

OK Annuler

Suppression des renvois fréquemment utilisés dans les index papier

Facette / Filtre Défaire / Refaire 398 / 398

Rafraîchir Tout réinitialiser Tout supprimer

nom inverser réinitialiser

voir

sensible à la casse expression rationnelle

25 matching entrées (733 total)

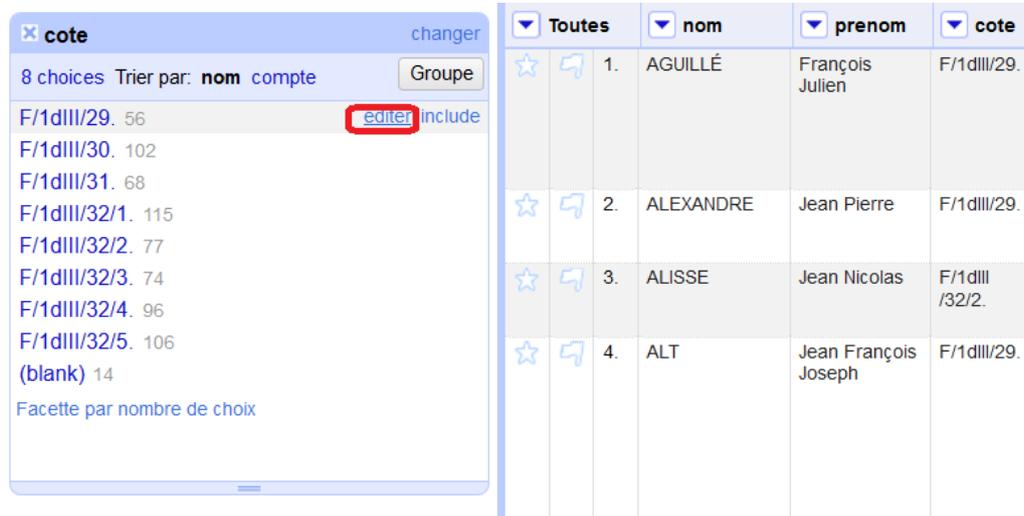
Voir en: lignes **entrées** Afficher: 5 10 **25** 50 entrées

Toutes		nom	prenom
☆	↩	91. BOULARD, voir BARTHELIEZ.	
☆	↩	97. BOURGUIGNON, voir BOURSIN de MÉRY dit BOURGUIGNON	Georges François Joseph
☆	↩	107. BRANCHON, voir PASQUIER dit	
☆	↩	109. BREITEL, voir RICHERT.	
☆	↩	119. CAQUET, voir BAUER	Jean Louis
☆	↩	127. CHARPENTIER, voir HAUCERNE née	
☆	↩	178. DESMET, voir DEMAY.	

Entête de tableau / éditer les lignes / supprimer les lignes

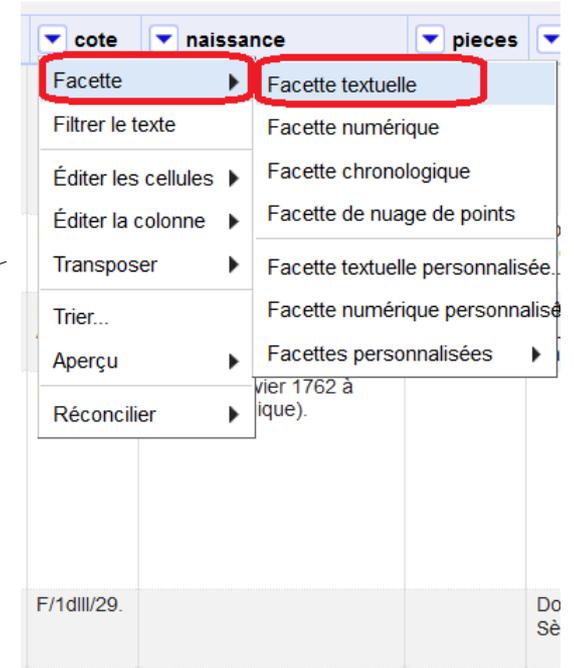
2.2. Colonne cote

L'inventaire concerne 8 cotes. Le plus rapide pour normaliser leur écriture est d'utiliser la facette textuelle.



The image shows a software interface with a facet for 'cote' on the left and a table of records on the right. The facet is titled 'cote' and has a 'changer' button. It shows 8 choices, sorted by 'nom' and 'compte'. The choices are: F/1dIII/29. 56, F/1dIII/30. 102, F/1dIII/31. 68, F/1dIII/32/1. 115, F/1dIII/32/2. 77, F/1dIII/32/3. 74, F/1dIII/32/4. 96, F/1dIII/32/5. 106, and (blank) 14. There is an 'éditer' button next to the first choice. The table has columns for 'Toutes', 'nom', 'prenom', and 'cote'. It contains 4 rows of data:

Toutes	nom	prenom	cote
1.	AGUILLÉ	François Julien	F/1dIII/29.
2.	ALEXANDRE	Jean Pierre	F/1dIII/29.
3.	ALISSE	Jean Nicolas	F/1dIII/32/2.
4.	ALT	Jean François Joseph	F/1dIII/29.



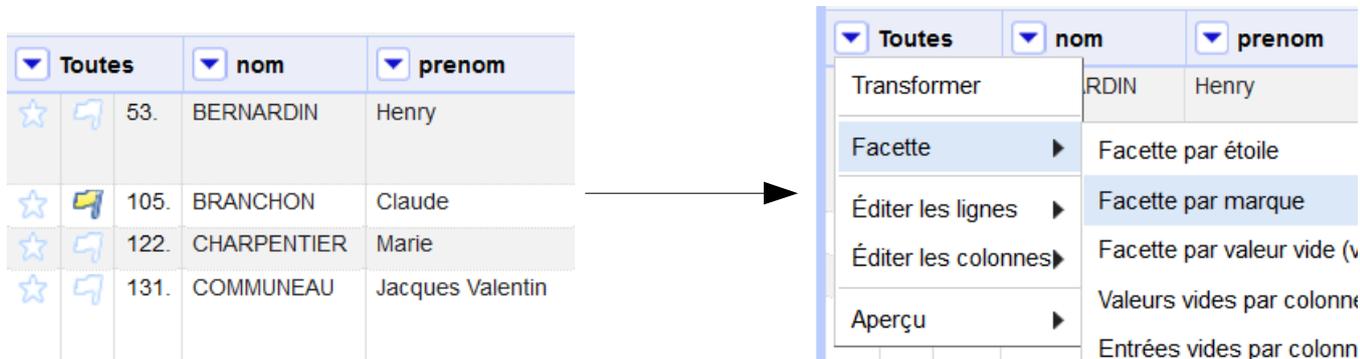
The image shows a context menu for the 'cote' facet. The menu is open, and the 'Facette' option is selected, showing a sub-menu with the following options: 'Facette textuelle', 'Facette numérique', 'Facette chronologique', 'Facette de nuage de points', 'Facette textuelle personnalisée...', 'Facette numérique personnalisée...', and 'Facettes personnalisées'. The 'Facette textuelle' option is highlighted with a red box. An arrow points from the 'Facette textuelle' option in the sub-menu to the 'Facette' option in the main menu.

On peut éditer directement les facettes pour faire les modifications nécessaires (ici suppression du point final).

Voire repérer 14 cellules vides (blank) à recompléter d'après l'original éventuellement

2.3. Étoiles / drapeaux

Certaines sont en fait à éliminer car ce sont des renvois. Les étoiles et drapeaux permettent de marquer des ensembles de lignes (manuellement ou en masse) pour ensuite leur appliquer un traitement (suppression pour exemple).



The image shows a data table with four columns: 'Toutes', 'nom', and 'prenom'. The 'Toutes' column contains icons for marking rows (star, flag, speech bubble). A context menu is open over the 'Toutes' column, listing actions like 'Facette', 'Éditer les lignes', and 'Aperçu'. The 'Facette' option is expanded, showing sub-options like 'Facette par étoile' and 'Facette par marque'.

▼ Toutes	▼ nom	▼ prenom
☆ 🗨️	53.	BERNARDIN Henry
☆ 🚩	105.	BRANCHON Claude
☆ 🗨️	122.	CHARPENTIER Marie
☆ 🗨️	131.	COMMUNEAU Jacques Valentin

▼ Toutes	▼ nom	▼ prenom
Transformer	RDIN	Henry
Facette	▶ Facette par étoile	
Éditer les lignes	▶ Facette par marque	
Éditer les colonnes	▶ Facette par valeur vide (v	
Aperçu	▶ Valeurs vides par colonn	
	▶ Entrées vides par colonn	

2.4. Colonne naissance

On veut séparer lieu et date de naissance :

- Facette par valeur vide de la colonne naissance
- Division en 2 colonnes sur la base du séparateur « à »
- Couper / coller à la main des 6 lignes non traitées qui contenaient « au »
- Renommage des colonnes

Voire, on peut ajouter une colonne *annee* à partir de la colonne *naissance_date* :

ajouter une colonne en fonction de cette colonne / `value.match(/.(\d{4}).*)[0]`*

Toutes	nom	id	prenom	cote	naissance_date	naissance_lieu	annee
☆ ↻	1. AGUILLÉ	bastille_0	François Julien	F/1dIII/29		Authon (Eure-et-Loir)	
☆ ↻	2. ALEXANDRE	bastille_1	Jean Pierre	F/1dIII/29	25 mars 1768	Hestroff (Moselle)	1768
☆ ↻	3. ALISSE	bastille_2	Jean Nicolas	F/1dIII/32/2	28 janvier 1765	Paris (Seine)	1765
☆ ↻	4. ALT	bastille_3	Jean François Joseph	F/1dIII/29	12 janvier 1762	Dinant (Belgique)	1762

(l'expression rationnelle permet d'extraire 4 chiffres qui se suivent)

2.5. Colonne domicile

Le libellé « Domicile : » n'étant pas du contenu, on va l'enlever en masse :

éditer les cellules / value.replace("Domicile :", "")

Les *idem*, d'après le PDF source, correspondent au lieu de naissance. On va donc les remplacer par le contenu de la colonne *naissance_lieu* :

The screenshot shows a data management interface with a table. The table has columns for 'domicile', 'naissance_date', 'naissance_lieu', 'cote', 'prenom', and 'nom'. The 'domicile' column contains values like 'idem.' and 'Mollkirch (Bas-Rhin)'. A red box highlights the 'idem.' values. A dialog box titled 'Transformation textuelle personnalisée sur la colonne domicile' is open, showing the expression 'cells[\"naissance_lieu\"].value' in the 'Expression' field. The 'Aperçu' tab shows a preview of the transformation results.

row	value	cells[\"naissance_lieu\"].value
8.	idem.	Mollkirch (Bas-Rhin)
19.	idem.	Aincourt (Seine-et-Oise)

- facette sur la colonne *domicile*

- sélection des *idem*

- *éditer les cellules / cells[\"naissance\"].value*

Harmonisation de la colonne pour que la commune soit suivie du département (ce qui n'est pas le cas de Paris)

Facette / Filtre Défaire / Refaire 473 / 473

Rafraîchir Tout réinitialiser Tout supprimer

domicile changer

204 choices Trier par: nom compte Groupe

- hôtel des Invalides à Paris 14
- 10 rue Saint-Bernard à Paris 2
- 11 passage Saint-Pierre à Paris 2
- 13 rue d'Anjou-Saint-Honoré à Paris 2
- 19 rue du Faubourg du Temple à Paris 2
- 25 rue de Reuilly à Paris 2

domicile inverser réinitialiser

paris

sensible à la casse expression rationnelle

domicile inverser réinitialiser

seine

sensible à la casse expression rationnelle

223 matching entrées (705 total)

Voir en: lignes entrées Afficher: 5 10 25 50 entrées

Toutes	nom	prenom	cote	domicile	
★	3.	ALISSE	Jean Nicolas	F/1dIII/32/2	214 faubourg Saint-Antoine à Paris
★	5.	ALTÊTE	Jean Baptiste Barthélemy	F/1dIII/29	3 rue de Sèvres à Paris
★	6.	ALTMAYER	Guillaume	F/1dIII/29	28 rue Mouffetard à Paris
★	7.	ANDRÉ	François	F/1dIII/32/2	4 rue des Trois Couronnes à Paris
★	9.	ARNOLD	Jean	F/1dIII/32/2	25 rue de la Croix à Paris
★	12.	ASTIER	Claude Marin	F/1dIII/32/2	13 rue de Charonne Impasse Delaunay à Paris
★	20.	BAIZE	Jean Nicolas	F/1dIII/29	30 rue Monsieur-le-Prince à Paris
★	25.	BARBET	Joseph Charles	F/1dIII/29	18 rue des Fossés-Montmartre à Paris
★	28.	BARON	Olivier	F/1dIII/32/2	hôtel des Invalides à Paris
★	36.	BAUDOT	Jean Baptiste	F/1dIII/32/2	21 Place Dauphine à Paris
★	40.	BEAUMONT	Dominique François	F/1dIII/29	199 rue Saint-Martin à Paris
★	42.	BEAUALET	Jacques Marie	F/1dIII/29	55 quai des Grands-Augustins à Paris

- filtre « paris » sur la colonne *domicile*
- filtre inversé « seine »
- éditer les cellules / `value.replace("Paris", "Paris (Seine)")`

Transformation textuelle personnalisée sur la colonne domicile

Expression Langue: General Refine Expression Language (GREL) ▼

`value.replace("Paris", "Paris (Seine)")`

Aperçu Historique Étoilée Aide

row	value	value.replace("Paris", "Paris (...)
3.	214 faubourg Saint-Antoine à Paris	214 faubourg Saint-Antoine à Paris (Seine)
5.	3 rue de Sèvres à Paris	3 rue de Sèvres à Paris (Seine)

2.6. Colonne profession_bastille

On enlève le libellé « Au moment des faits » comme précédemment, ainsi que les points finaux.

Pour homogénéiser le contenu, on va utiliser les facettes textuelles, et plus particulièrement la possibilité de faire des fusions automatisées d'occurrences proches...

The screenshot shows a search interface with a table of 461 matching entries (705 total). The table has columns for 'nom', 'cote', 'naissance_date', 'naissance_lieu', 'domicile', and 'profession_bas'. Two facet panels are visible on the left, both for the 'profession_bastille' field. The top facet shows 2 choices: 'false' (461) and 'true' (244). The bottom facet shows 41 choices, including 'Aide de camp du commandant de Paris son père' (1), 'Brosseur' (1), 'Canonnier' (1), 'Caporal' (3), 'Charpentier' (1), 'Chasseur à cheval' (1), 'Chaudronnier' (1), 'Compagnon' (1), 'Elève de l'Ecole Royale des orphelins militaires' (1), and 'Epicier' (1).

nom	cote	naissance_date	naissance_lieu	domicile	profession_bas
e	F/1dIII/29	25 mars 1768	Hestroff (Moselle)	Merten (Moselle)	Fusilier
	F/1dIII/29	6 novembre 1763	Berviller-en-Moselle (Moselle)	28 rue Mouffetard - Paris (Seine)	Grenadier
	F/1dIII/32/2	17 novembre 1747	Grosbliederstroff (Moselle)	25 rue de la Croix - Paris (Seine)	Garde française
	F/1dIII/29	12 janvier 1764	Paris (Seine)	6 rue Saint-Claude - Tours (Indre-et-Loire)	Volontaire de la Bazoche
	F/1dIII/29	26 décembre 1765	Montpellier (Hérault)	37 Grande-Rue - Montpellier (Hérault)	Fusilier
	F/1dIII/29	2 septembre 1767	Chatillon-sur-Chalaronne (Ain)	Oullins (Rhône)	Garde française
	F/1dIII/29				Garde française
	F/1dIII/29	31 août 1764	Aincourt (Seine-et-Oise)	Jambville (Seine-et-Oise)	Garde française
ul	F/1dIII/29	28 octobre 1766	Aincourt (Seine-et-Oise)	Aincourt (Seine-et-Oise)	Garde française
ph	F/1dIII/29	8 avril 1768	Avesnes (Nord)	La Flèche (Sarthe)	Garde française

A partir des facettes textuelles, le bouton Groupe permet de faire des regroupements d'occurrences en utilisant différents algorithmes de (listes déroulantes).

- Définir la graphie souhaitée
- Cocher les cases des contenus à fusionner
- Fusionner la sélection

The screenshot shows a software interface for grouping data. The main window is titled "Regrouper & éditer une colonne 'profession_bastille'". Below the title, there is a text box explaining the tool's purpose: "Cet outil vous aide à identifier des groupes de cellules ayant des valeurs différentes mais qui peuvent correspondre à des représentations de la même valeur. Par exemple, les deux chaînes 'New York' et 'new york' n'ont qu'une différence de casse et font très certainement référence à la même personne. En savoir plus...".

At the top of the main window, there are two dropdown menus: "Méthode" set to "collision de clés" and "Fonction de codage" set to "metaphone3".

The central part of the interface is a table with the following columns: "Taille du groupe", "Nombre de lignes", "Valeurs dans le groupe", "Fusionner ?", "Nouvelle valeur pour la cellule", and "# Choix dans".

Taille du groupe	Nombre de lignes	Valeurs dans le groupe	Fusionner ?	Nouvelle valeur pour la cellule	# Choix dans
3	324	<ul style="list-style-type: none">• Garde française (321 rows)• Garde-française (2 rows)• Garde françaises (1 rows)	<input checked="" type="checkbox"/>	Garde française	
2	37	<ul style="list-style-type: none">• Garde nationale (36 rows)• Garde-nationale (1 rows)	<input checked="" type="checkbox"/>	Garde nationale	
2	2	<ul style="list-style-type: none">• Instructeur (1 rows)• Instructeur des gardes françaises (1 rows)	<input type="checkbox"/>	Instructeur	

At the bottom of the interface, there are several buttons: "Tout sélectionner", "Tout désélectionner", "Exporter les groupes", "Fusionner la sélection & regrouper", and "Fusionner la".

Red annotations in the image include: a box around the "Groupe" button on the left sidebar; a box around the "collision de clés" dropdown; a box around the "metaphone3" dropdown; a box around the "Fusionner ?" checkbox for the first row; a box around the "Nouvelle valeur pour la cellule" input for the first row; and a box around the "Fusionner la sélection & regrouper" button.

Et recommencer en jouant sur les listes déroulantes pour changer d'algorithme...

Etc. sur les colonnes suivantes

Une heure après, le contenu est propre, enrichi et harmonisé...

Et prêt à être encodé en XML EAD !

Ou presque : un numéro d'identifiant unique peut être requis pour chaque ligne.

À partir de n'importe quelle colonne :

ajouter une colonne en fonction de cette colonne / "bastille_" + rowIndex

Le contenu "bastille_" devant être adapté en fonction des prérequis de caractères (chiffre, lettre, etc.) requis par le système d'information archivistique final.

Ajouter une colonne en fonction d'une colonne nom

Nouveau nom de colonne:

En cas d'erreur: vider la cellule conserver l'erreur copier la valeur de originale

Expression: Langue:

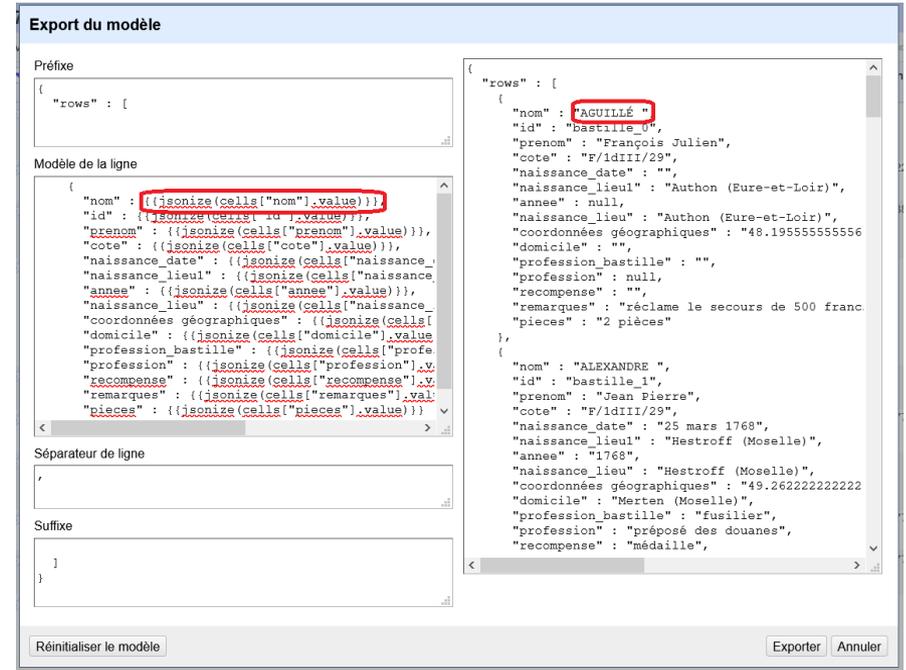
Aperçu Historique Étoilée Aide

row	value	"bastille_" + rowIndex
1.	AGUILLÉ	bastille_0
2.	ALEXANDRE	bastille_1
3.	ALISSE	bastille_2
4.	ALT	bastille_3
5.	ALTÊTE	bastille_4
6.	ALTMAYER	bastille_5
7.	ANDRÉ	bastille_6

OK Annuler

3. Exporter en XML EAD

3. Exporter en XML EAD



Et dans ce code en JSON qui peut faire peur, une seule chose à retenir :

pour chaque ligne, le contenu d'une colonne est encodé `{{jsonize(cells["nom"].value)}}`

["nom"] étant l'intitulé de la colonne

avant et après ce code, on met ce qu'on veut, y compris des balises XML EAD.

L'export est plus facile à modéliser sur Notepad++

```
1 <c id={{jsonize(cells["id"].value)}}>
2 <did>
3 <unitid type="cote-de-consultation"{{jsonize(cells["cote"].value).replace(' ','')}}</unitid>
4 <unitid>Dossier</unitid>
5 <unittitle>
6 <persname>{{jsonize(cells["nom"].value).replace(' ','')},
  {{jsonize(cells["prenom"].value).replace(' ','')}}</persname>
7 </unittitle>
8 <unitdate calendar="gregorian" era="ce" normal="1830-01-01/1833-12-31">[1830-1833]</unitdate>
9 <physdesc><extent>{{jsonize(cells["pieces"].value).replace(' ','')}}</extent></physdesc>
10 </did>
11 <scopecontent>
12 <p>Date de naissance : {{jsonize(cells["naissance_date"].value).replace(' ','')}}</p>
13 <p>Lieu de naissance : {{jsonize(cells["naissance_lieu"].value).replace(' ','')}}</p>
14 <p>Domicile : {{jsonize(cells["domicile"].value).replace(' ','')}}</p>
15 <p>Profession : {{jsonize(cells["profession"].value).replace(' ','')}}</p>
16 <p>Profession au moment des faits : {{jsonize(cells["profession_bastille"].value).replace(' ','')}}</p>
17 <p>Récompense : {{jsonize(cells["recompense"].value).replace(' ','')}}</p>
18 <p>Remarques : {{jsonize(cells["remarques"].value).replace(' ','')}}</p>
19 </scopecontent>
20 </c>
```

Le contenu de type texte étant exporté entre guillemets, on va supprimer ces derniers dès l'export :

```
{{jsonize(cells["nom"].value).replace(' ','')}}
```

Vider les blocs Préfixe, Séparateur de ligne et Suffixe

Coller le patron modélisé en XML EAD dans Modèle de la ligne

Penser à enregistrer quelque part le modèle qui n'est pas sauvegardé dans OpenRefine!

Jeter un œil à l'aperçu

Exporter

Copier / coller le résultat dans les balises <dsc></dsc> de l'inventaire

Export du modèle

Préfixe

Modèle de la ligne

```
<c id={{jsonize(cells["id"].value)}}>
<did>
<unitid type="cote-de-consultation"{{jsonize(cells["co
<unitid>Dossier</unitid>
<unittitle>
<persname>{{jsonize(cells["nom"].value).replace("'",'
</unittitle>
<unitdate calendar="gregorian" era="ce" normal="1830-01
<physdesc><extent>{{jsonize(cells["pieces"].value).repla
</did>
<scopecontent>
<p>Date de naissance : {{jsonize(cells["naissance_date"]
<p>Lieu de naissance : {{jsonize(cells["naissance_lieu"]
<p>Domicile : {{jsonize(cells["domicile"].value).repla
<p>Profession : {{jsonize(cells["profession"].value).rep
<p>Profession au moment des faits : {{jsonize(cells["pro
<p>Récompense : {{jsonize(cells["recompense"].value).rep
<p>Remarques : {{jsonize(cells["remarques"].value).repla
</scopecontent>
</c>
```

Séparateur de ligne

Suffixe

Réinitialiser le modèle

Exporter Annuler

```
<c id="bastille_0">
<did>
<unitid type="cote-de-consultation">F/1dIII/2
<unitid>Dossier</unitid>
<unittitle>
<persname>AGUILLÉ , François Julien</persname>
</unittitle>
<unitdate calendar="gregorian" era="ce" norma
<physdesc><extent>2 pièces</extent></physdesc>
</did>
<scopecontent>
<p>Date de naissance : </p>
<p>Lieu de naissance : Authon (Eure-et-Loir)</p>
<p>Domicile : </p>
<p>Profession : null</p>
<p>Profession au moment des faits : </p>
<p>Récompense : </p>
<p>Remarques : réclame le secours de 500 franc
</scopecontent>
</c><c id="bastille_1">
<did>
<unitid type="cote-de-consultation">F/1dIII/2
<unitid>Dossier</unitid>
<unittitle>
<persname>ALEXANDRE , Jean Pierre</persname>
</unittitle>
<unitdate calendar="gregorian" era="ce" norma
<physdesc><extent>>null</extent></physdesc>
</did>
<scopecontent>
<p>Date de naissance : 25 mars 1768</p>
<p>Lieu de naissance : Hestroff (Moselle)</p>
<p>Domicile : Merten (Moselle)</p>
```

Il ne reste plus qu'à faire l'introduction...

Dernière remarque : lorsque la cellule dans OpenRefine est vide, le contenu exporté indique *null*.

Le XML n'étant pas forcément le format idéal ce type de données (mais parfois le seul qu'on peut utiliser), mieux vaut ne pas supprimer cette mention *null*, mais la remplacer par « non renseigné » par exemple, afin de conserver une structure identique pour toutes les notices et faciliter l'interopérabilité et la conversion vers d'autres formats.

The image shows a side-by-side comparison between the OpenRefine interface and its XML export. On the left, a tree view shows a record with ID 'c "bastille_652"' and a sub-record 'did F/1dIII/32/1'. The 'did' record has several fields: 'unitid "cote-de-consultation" F/1dIII/32/1', 'unitid Dossier', 'unittitle TABOUREUX, François Galien', 'persname TABOUREUX, François Galie', and 'unitdate "gregorian" [1830-1833]'. Below this, another record 'c "bastille_653"' is shown with a 'did F/1dIII/32/1' sub-record. The 'scopecontent' field for this 'did' is empty, which is highlighted in the interface. On the right, the corresponding XML export is shown. The XML structure mirrors the tree view, with elements like <id>, <unitid type=>, <unittitle>, <persname>, <unitdate calendar=>, <physdesc>, <scopecontent>, and <p>. The empty 'scopecontent' field is exported as <p>Date de naissance : non renseigné</p>.

```
13795
13796 <c id="bastille_653">
13797   <did>
13798     <unitid type="cote-de-consultation">F/1dIII/32/1</unit
13799     <unitid>Dossier</unitid>
13800     <unittitle>
13801     <persname>TERIÈRE, Marc</persname>
13802     </unittitle>
13803     <unitdate calendar="gregorian" era="ce" normal="1830-(
13804     <physdesc><extent>1 pièce.</extent></physdesc>
13805   </did>
13806 <scopecontent>
13807   <p>Date de naissance : non renseigné</p>
13808   <p>Lieu de naissance : non renseigné</p>
13809   <p>Domicile : non renseigné</p>
13810   <p>Profession : scieur de long</p>
13811   <p>Profession au moment des faits : non renseigné</p>
13812   <p>Récompense : non renseigné</p>
13813   <p>Remarques : non renseigné</p>
13814 </scopecontent>
```

4. Des pistes pour aller plus loin...

4. Des pistes pour aller plus loin...

Les facettes textuelles sont exportables facilement en CSV (datavisualisation)...

Facette / Filtre Défaire / Refaire 914 / 914 705 entrées

Rafraîchir Tout réinitialiser Tout supprimer Voir en: lignes entrées Afficher: 5 10 25 50 entrées

profession changer

139 choices Trier par: nom compte Groupe

profession	compte
gendarme	25
capitaine	20
militaire	16
cultivateur	15
propriétaire	14
cordonnier	11
journalier	9
officier	9
menuisier	8
garde forestier	7
lieutenant	5
caporal	4
concierge	4
ébéniste	4
employé	4
maçon	4
sergent	4
boulangier	3
brigadier de gendarmerie	3
charpentier	3
charron	3
clerc de notaire	3
cloutier	3
colonel	3
perruquier	3
serrurier	3

Choix des facettes en tant que valeurs séparées par des tabulations

```
gendarme 25
capitaine 20
militaire 16
cultivateur 15
propriétaire 14
cordonnier 11
journalier 9
officier 9
menuisier 8
garde forestier 7
lieutenant 5
caporal 4
concierge 4
ébéniste 4
employé 4
maçon 4
sergent 4
boulangier 3
brigadier de gendarmerie 3
charpentier 3
charron 3
clerc de notaire 3
cloutier 3
colonel 3
perruquier 3
serrurier 3
```

Fermer



Visualisation des principales professions

(outil : rawgraphs)

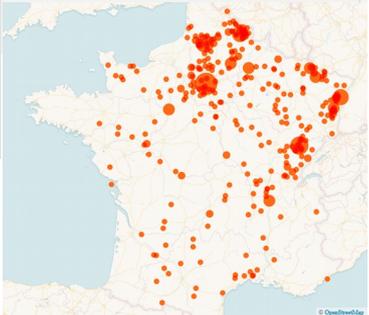
4. Des pistes pour aller plus loin...

Réconciliation Wikidata pour enrichir encore plus...

- **naissance_lieu** pour obtenir les coordonnées géographiques (cartographie)

individu	wp	date de naissance	date de mort	identifiant Bibli
Joseph-David de Barquier <small>Choisir une nouvelle correspondance</small>	https://fr.wikipedia.org/wiki/Joseph-David_de_Barquier	1757-06-07T00:00:00Z	1844-10-31T00:00:00Z	
Pierre-François Palloy <small>Choisir une nouvelle correspondance</small>	https://fr.wikipedia.org/wiki/Pierre-Fran%C3%A7ois_Palloy	1755-01-23T00:00:00Z	1835-01-19T00:00:00Z	11997800m
Joseph Souberbielle <small>Choisir une nouvelle correspondance</small>	https://fr.wikipedia.org/wiki/Joseph_Souberbielle	1754-03-18T00:00:00Z	1846-07-10T00:00:00Z	165537414

naissance_lieu	coordonnées géographiques
Authon-du-Perche <small>Choisir une nouvelle correspondance</small>	48.1955555555556,0.8925
Hestroff <small>Choisir une nouvelle correspondance</small>	49.2622222222222,6.4347222222222
Paris <small>Choisir une nouvelle correspondance</small>	48.8569444444444,2.3513888888889
Dinant <small>Choisir une nouvelle correspondance</small>	



- **individu (prenoms, nom)** pour obtenir les dates de décès, un identifiant BNF, une profession, un lien vers l'article Wikipédia...