

HANDLEIDING!

Door Shannon van Muijden – versie: april 2024. <https://www.linkedin.com/in/svmuijden/>

Ik heb een rommelige dataset. Ik zou al een heel stap verder zijn met de terugvindbaarheid van mijn documenten als ik alle wie, wat, waar lijsten uit mijn data heb opgeschoond. Hoe maak ik van deze rommelige termen-, personen of plaatsenlijst nette lijsten die ik vervolgens in OpenRefine kan reconcilen met een externe termenlijst zodat ik niet alleen mijn termen-, personen of plaatsenlijst in mijn systeem opschoon maak ook meteen voorzie van URI's?

In dit document vind je de uitleg van hoe je dit zou kunnen doen.

Wat heb je nodig?

- Een account bij ChatGPT: <https://chat.openai.com/>
- Excel: <https://www.microsoft.com/nl-nl/microsoft-365/excel?market=nl>
- Fuzzy Lookup extensie voor Excel: <https://www.microsoft.com/en-us/download/details.aspx?id=15011&>
- OpenRefine: <https://openrefine.org/>
- Optioneel: Notepad++: <https://notepad-plus.en.softonic.com/download>

In 4 stappen klaar

1. Laat je rommelige termen-, personen of plaatsenlijsten opschonen door ChatGPT
2. Zet het resultaat uit ChatGPT naast je oorspronkelijke lijst met behulp van Fuzzy Lookup in Excel
3. Reconcile (vergelijk) de opgeschoonde lijst in OpenRefine met die van een externe terminologiebron uit het termennetwerk van het Netwerk Digitaal Erfgoed. Haal de voorkeursspelling op en geef me daarbij de URI
4. Importeer de opgeschoonde persoonsnamen, termen en plaatsen + hun URI's in je eigen documentatiesysteem of collectiebeheersysteem

LETS BEGIN!

1. Gebruik ChatGPT voor eerste opschoonactie

1. Open de lijst met namen + hun unieke identificatienummer (recordnummers / priref of documentnummers). Kopieer de gehele lijst.
2. Ga naar ChatGPT. Geef een soortgelijk prompt als onderstaande op en plak daarachter de lijst met namen.

Ik heb hier een lijst met namen. Kun je de namen zo transformeren dat ze achternaam komma voornaam staan, zonder een enkele regel over te slaan ook al staan er doubletten in. Ik wil de gehele lijst getransformeerd hebben, met alle recordnummers erbij zodat ik zeker weet dat er geen regels worden overgeslagen. Kun je het in een tabel zetten? [plak hier de lijst met id-nummers + namen]

3. Als Chat klaar is met verwerken van de lijst, kopieer hem en plak hem in een apart werkblad in Excel.

2. Vergelijken van oorspronkelijke lijst met resultaat uit ChatGPT met Fuzzy Lookup in Excel

Na het verwerken van de gegevens in ChatGPT, moet je het resultaat kopiëren en plakken in een nieuw werkblad in hetzelfde Excel-document waarin je oorspronkelijke namenlijst staat. Op deze manier heb je nu twee werkbladen: één met de oorspronkelijke, vervuilde namenlijst en één met het resultaat van de ChatGPT-verwerking.

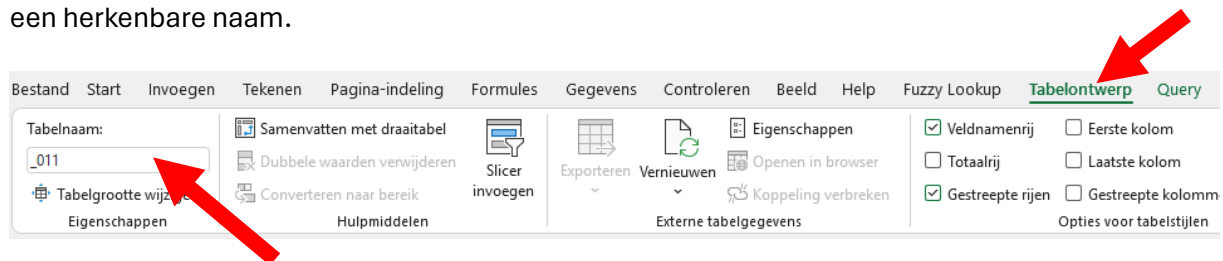
Download Fuzzy Lookup: <https://www.microsoft.com/en-us/download/details.aspx?id=15011> Installeer de extensie of laat deze door je systeembeheerder installeren.

Toetsenbordcombinaties:

- ctrl-A = alles selecteren
- ctrl-L = tabel maken in Excel

Tabel maken

Maak een tabel van de 'vervuilde' namenlijst: selecteer de eerste cel van het werkblad. Gebruik de toetsenbordcombinatie **ctrl-A** om alle cellen te selecteren waar waarden in staan (de lege cellen moeten niet geselecteerd zijn). Gebruik dan **ctrl-L** om een tabel te maken. Geef de tabel een herkenbare naam.



Dit zie je in het menu bovenin: **Tabelontwerp** en links **Tabelnaam**.

Doe hetzelfde met de lijst namen die ChatGPT heeft verbeterd.

Fuzzy Lookup gebruiken

Maak vervolgens een nieuw, leeg werkblad in hetzelfde Excel document aan. Zet de cursor in de eerste cel (A1).

Klink boven in de menubalk op Fuzzy Lookup (geheel rechts) en vervolgens op het icoontje Fuzzy Lookup.



Fuzzy
Lookup


Rechts van je Excel werkblad opent een venster.

Left table – Right table

Het begint met Left table en Right table. De Left table is het werkblad met de vervuilde namen. Right table is het werkblad waarmee je dit wilt gaan vergelijken (het resultaat uit ChatGPT).

Left Columns – Right Columns

Vervolgens kies je de kolommen die je wilt vergelijken. Dit zijn dus de kolommen waarin de uniek ID's staan, want door die twee rijen met elkaar te vergelijken weet je altijd dat je de juiste informatie bij elkaar krijgt. (in dit geval heet die kolom 'preref' en 'prerefChatGPT'). Klik op 'preref' om hem te selecteren. In Right Columns staan de kolommen uit de ChatGPT opschoon-actie. Daaruit kiezen we kolom 'prerefChatGPT'. Klik op de tekst om deze te selecteren. Druk vervolgens op het icoontje in het

midden . Je ziet nu dat bij Match Columns een regel is verschenen. Deze twee kolommen gaan met elkaar vergeleken worden.

Druk vervolgens op Go.

Hij gaat nu de twee werkbladen met elkaar vergelijken. Afhankelijk van de grootte van de lijst kan dit een tijdje duren.

Sorteren en filteren

- Als het matchen klaar is, selecteer de eerste rij en schakel een filter in (zie [filter aanbrengen](#)).

Fuzzy Lookup

Left Table: spak

Right Table: chat

Left Columns:

name

preref

Right Columns:

nameCHAT

prerefChatGPT

Match Columns:

Left Colu...	Right Col...	Configur...
preref	prerefChat...	Default

Ouput Columns:

☒ spak.preref

☒ spak.name

☒ chat.prerefChatGPT

☒ chat.nameCHAT

☒ FuzzyLookup.Similarity

Number of Matches: 1

Similarity Threshold:

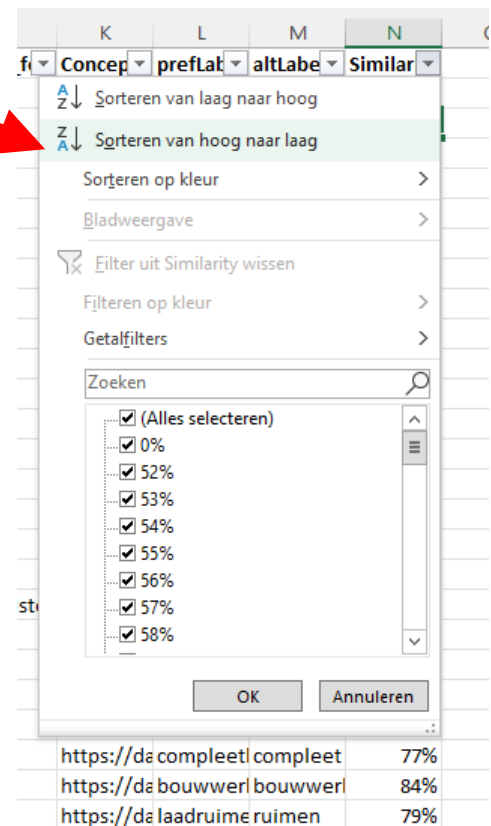
Undo

Configure...

Go

- Selecteer de kolom met Similarity (waar allee 00.001 / 00.000 in staan). En klik op het icoontje voor % om deze in procenten om te zetten.
- Sorteer deze kolom vervolgens van hoog naar laag.

Alle 100% match komen nu bovenaan te staan. Bij het vergelijken van unieke ID's zou je nu een hele lijst met 100% moeten hebben en dan heeft hij alle regels uit de oorspronkelijke, vervuilde lijst, naast het resultaat uit de ChatGPT actie gezet. Je kunt nu dus ook goed zien (en controleren) wat ChatGPT met je data heeft gedaan. Bij het matchen van termen, waar Fuzzy ook een hele handige tool voor is, krijg je naast 100% vaak ook lagere percentages, zoals het voorbeeld hiernaast. Fuzzy is ook heel handig om importdocumenten mee te maken voor het documentatiesysteem waarmee je werkt. Je kunt namelijk je importbestanden goed op één rij zetten met je unieke ID's zodat je altijd weet dat je imports op de juiste plaats terecht komen. Fuzzy Lookup is voor mij onmisbaar in mijn werk als datamanager. Ik gebruik het vaak in combinatie met OpenRefine en Notepad++.



	K	L	M	N
fi	Concept	preflat	altLabel	Similar
				Sorteren van laag naar hoog
				Sorteren van hoog naar laag
				Sorteren op kleur
				Bladweergave
				Filter uit Similarity wissen
				Filtreren op kleur
				Getalfilters
				Zoeken
				<input checked="" type="checkbox"/> (Alles selecteren)
				<input checked="" type="checkbox"/> 0%
				<input checked="" type="checkbox"/> 52%
				<input checked="" type="checkbox"/> 53%
				<input checked="" type="checkbox"/> 54%
				<input checked="" type="checkbox"/> 55%
				<input checked="" type="checkbox"/> 56%
				<input checked="" type="checkbox"/> 57%
				<input checked="" type="checkbox"/> 58%
				OK
				Annuleren
				https://da compleet compleet 77%
				https://da bouwwer bouwwer 84%
				https://da laadruime ruimen 79%

3. Reconcilen van mijn lijst met het termennetwerk in OpenRefine

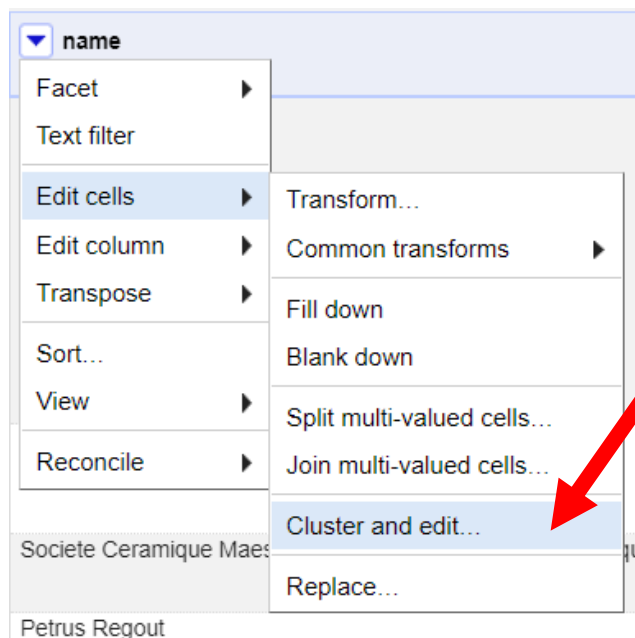
OpenRefine is een opensource programma, te downloaden via <https://openrefine.org/>. Installatie kan ook via ZIP-file met ingebouwde JAVA, die je enkel hoeft te uit te pakken om ermee te kunnen werken. Open OpenRefine en importeer je dataset waarin je de clustering en bewerking wilt uitvoeren. Klik op "Create Project" en selecteer je bestand of voer de URL in van het bestand dat je wilt bewerken.

Cluster & Edit

OpenRefine heeft een mooie functionaliteit die Cluster & Edit heet. Hiermee kun je snel en efficiënt groepen van vergelijkbare waarden identificeren en bewerken in je dataset. Door clustering en vervolgens bewerkingen toe te passen, kun je inconsistente of onjuiste gegevens corrigeren en je dataset schoonmaken.

Selecteer een kolom om te clusteren

Zodra je dataset is geladen, zoek je naar de kolom waarin je clustering wilt uitvoeren. Klik op de naam van de kolom om deze te selecteren. Klik op het pijltje naast de kolomnaam en selecteer "Cluster and Edit". OpenRefine zal automatisch vergelijkbare waarden in de geselecteerde kolom groeperen in clusters.



Bekijk en bewerk de clusters

OpenRefine zal nu een lijst met clusters tonen, waarbij elke cluster vergelijkbare waarden bevat. Bekijk elke cluster zorgvuldig en controleer of de waarden correct zijn gegroepeerd. Als je fouten opmerkt, kun je deze handmatig corrigeren door op de cluster te klikken en de bewerkingen uit te voeren in het bewerkenvenster.

Cluster size	Row count	Values in cluster	Merge?	New cell value
2	2	<ul style="list-style-type: none">J.H.B. koekkoekkoekkoek, J.H.B.	<input type="checkbox"/>	<input type="text" value="J.H.B. koekkoek"/>
2	2	<ul style="list-style-type: none">O.PrachthausenOprachthausen	<input type="checkbox"/>	<input type="text" value="O.Prachthausen"/>
2	2	<ul style="list-style-type: none">Gestel, LeoLeo Gestel	<input type="checkbox"/>	<input type="text" value="Gestel, Leo"/>
2	2	<ul style="list-style-type: none">Tegelfabriek Ravesteijn Westraven UtrechtTegelfabriek Ravesteijn Westraven Utrecht.	<input type="checkbox"/>	<input type="text" value="Tegelfabriek Ravesteijn Westrav"/>
2	2	<ul style="list-style-type: none">Koperslagerkoperslager.	<input type="checkbox"/>	<input type="text" value="Koperslager"/>

Pas bewerkingen toe op de clusters

Nadat je de clusters hebt beoordeeld en eventuele fouten hebt gecorrigeerd, kun je bewerkingen toepassen op de hele cluster om consistentie te garanderen. Klik op "Apply" nadat je tevreden bent met de bewerkingen.

Vergelijk je dataset met het termennetwerk

Wat is het termennetwerk?

Het Termennetwerk is een handige zoekmachine ontwikkeld binnen het Netwerk Digitaal Erfgoed (NDE). Het doel van het Termennetwerk is om collectiebeheerders van erfgoedorganisaties te helpen bij het vinden van eenduidige termen voor het beschrijven van erfgoedinformatie. Dit zijn de belangrijkste kenmerken van het Termennetwerk:

- **Realtime Zoeken:** Je kunt een zoekvraag invoeren, bijvoorbeeld “Rembrandt,” en het Termennetwerk zoekt rechtstreeks en realtime in verschillende terminologiebronnen naar overeenkomende termen. Zo kun je de juiste termen kiezen, zoals “Rembrandt” in RKDartists of “Rembrandt Harmensz. van Rijn” in de Nederlandse Thesaurus van Auteursnamen.
- **Onderdeel van Digitaal Erfgoed Bruikbaar:** Het Termennetwerk maakt deel uit van het programma Digitaal Erfgoed Bruikbaar van het NDE. Binnen dit programma werken erfgoedorganisaties samen aan het standaardiseren en verbinden van erfgoedinformatie.
- **Koppelbaar met Collectiebeheersysteem:** Het toekennen van termen aan erfgoed maakt deel uit van collectieregistratie. Het Termennetwerk is koppelbaar met collectiebeheersystemen, zodat je vanuit je systeem kunt zoeken in terminologiebronnen. De gevonden termen worden automatisch teruggegeven aan je collectiebeheersysteem.
- **Eigen Termenbronnen Toevoegen:** Organisaties kunnen ook hun eigen termenlijsten toevoegen aan het Termennetwerk. Neem hiervoor contact op met Ruben Schalk van de Rijksdienst voor het Cultureel Erfgoed.

Wil je meer weten of zelf aan de slag met het Termennetwerk? [Bezoek de officiële website van het Termennetwerk binnen het Netwerk Digitaal Erfgoed.](#)

Binnen OpenRefine is het termennetwerk ook te bevragen. Hierdoor kun je geautomatiseerd je gehele dataset vergelijken met die van een geautoriseerde externe termenbron en voorzien van URI's wat weer leidt naar beter terugvindbaarheid en koppelbaarheid van je collectie, zowel intern als online.

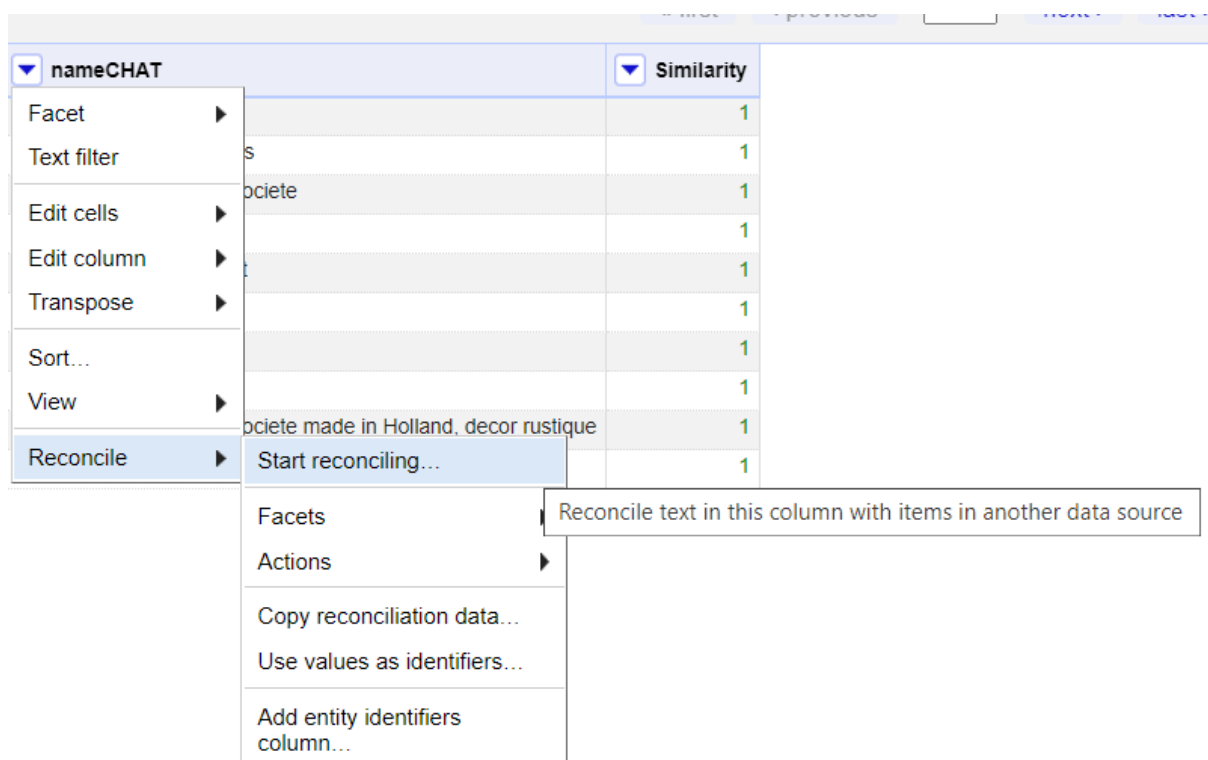
Vergelijk de namenlijst met een externe bron in OpenRefine

Als de data in OpenRefine is geladen ziet het er zo uit:

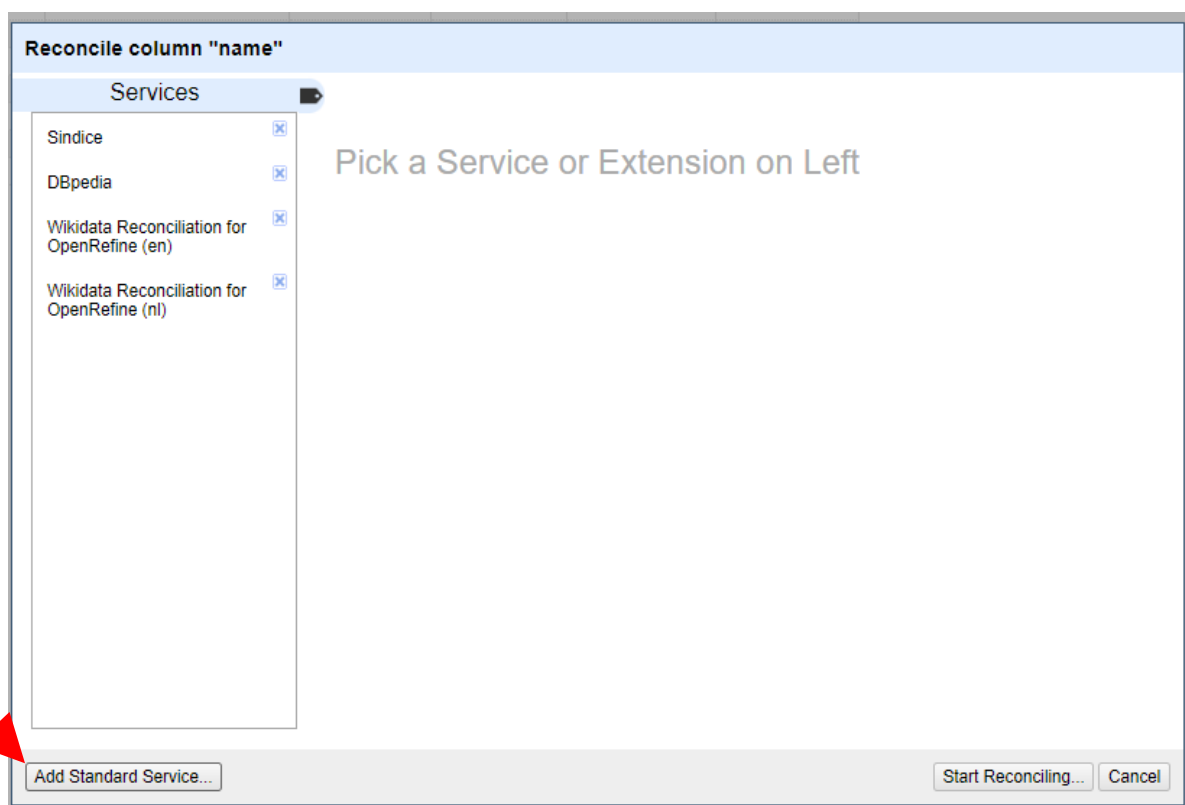
202 rows					
Show as: rows records		Show: 5 10 25 50 100 500 1000 rows			
« first < g					
All	preref	name	prerefChatGPT	nameCHAT	Similarity
1.	2097	Regout	2097	Regout	1
2.	2100	Regout, Petrus Laurentius	2100	Regout, Petrus Laurentius	1
3.	2112	Societe Ceramique Maestricht, made in Holland, decor rustique	2112	Ceramique Maestricht, Societe	1
4.	2113	Petrus Regout	2113	Regout, Petrus	1
5.	2114	PetrusRegout Maestricht	2114	Regout, PetrusMaestricht	1
6.	2115	Heju porselein	2115	Heju porselein	1
7.	2116	Gestel, Leo	2116	Gestel, Leo	1
8.	2117	Thuringer porzellan	2117	Thuringer porzellan	1
9.	2118	Societe Ceramique Maestricht, made in Holland,	2118	Ceramique Maestricht, Societe made in Holland, decor rustique	1
10.	2120	Sfinx Regout	2120	Regout, Sfinx	1

Kies een kolom die je wilt vergelijken met een externe terminologiebron. In dit geval is dat de kolom ‘nameCHAT’ met alle opgeschoonde namen erin. Klik op het pijltje (uitklapmenu).

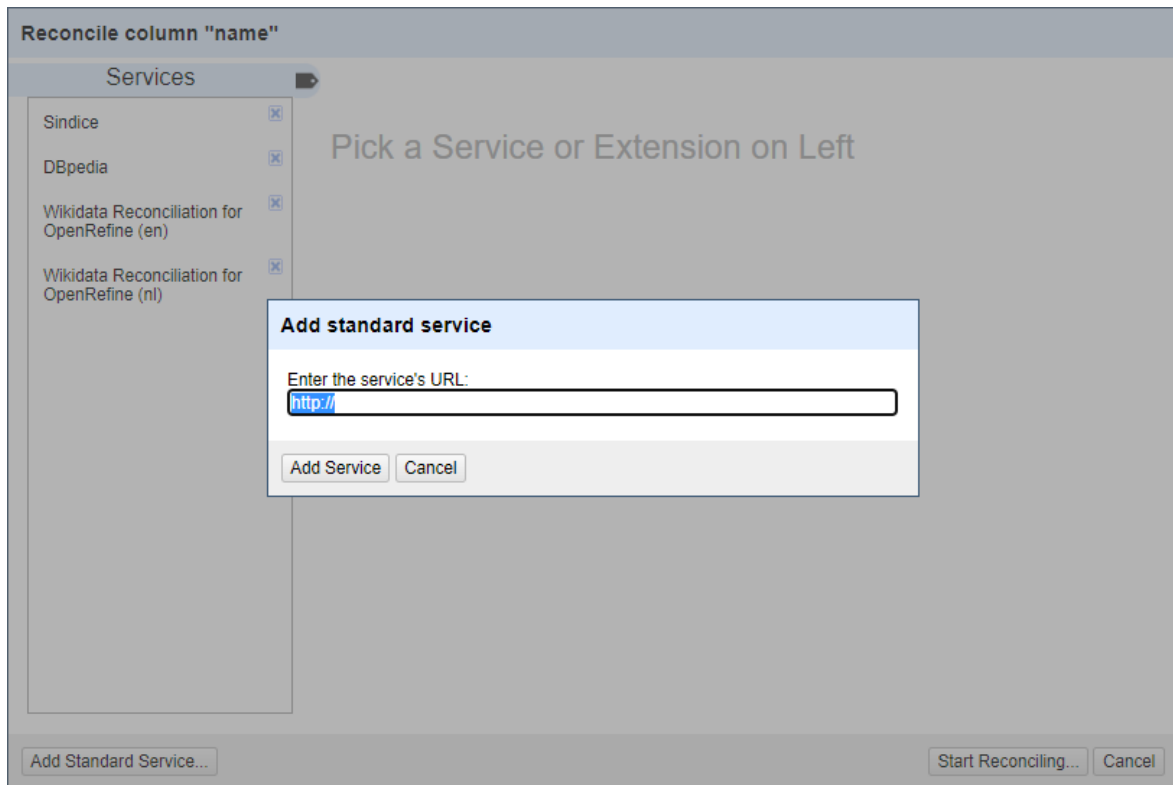
Kies voor **Reconcile** en vervolgens voor **Start reconciling**



Een venster opent. Klik op de knop Add Standard Service... linksonder.



Er wordt nu gevraagd om een URL. Hier moet je de URL opgeven die je toegang geeft tot de betreffende externe terminologiebron.



Ga nu naar:

<https://termennetwerk.netwerkdigitaalergoed.nl/reconciliation> en scroll naar beneden tot je bij de kop **Welke terminologiebronnen kan ik gebruiken voor reconciliation?** komt. Hieronder volgt een hele rij met allemaal URL's die je kunt aanspreken om toegang te krijgen tot die betreffende terminologiebron. De URL vul je in OpenRefine in bij **Enter the service's URL:**

In dit voorbeeld wordt gekozen voor de RKD artists database. Klik op de knop **Kopieer URL** en ga terug naar OpenRefine.

RKDartists RKD

Kopieer URL

Bekijk bij bron

Biografische gegevens van Nederlandse en buitenlandse kunstenaars van de middeleeuwen tot heden

<https://termennetwerk-api.netwerkdigitaalergoed.nl/reconcile/https://data.netwerkdigitaalergoed.nl/rkd/rkdartists/sparql>

Plak de zojuist gekopieerde URL in het venster **Enter the service's URL:** en klik op de knop **Add Service.**

Reconcile column "nameCHAT"

Services

Sindice

DBpedia

Wikidata Reconciliation for OpenRefine (en)

Wikidata Reconciliation for OpenRefine (nl)

VIAF

ORCID

Wikidata SPARQL

Wikidata (old) (en)

Getty Vocabularies Reconciliation Service

Wikidata (en)

Wikidata (nl)

Wikidata (en)

Pick a service or extension on left

Add standard service

Enter the service's URL

<https://termennetwerk-api.netwerkdigitaalergoed.nl/reconcile/https://data.netwerkdigitaalergoed.nl/rkd/rkdartists/sparql>

Add service

Cancel

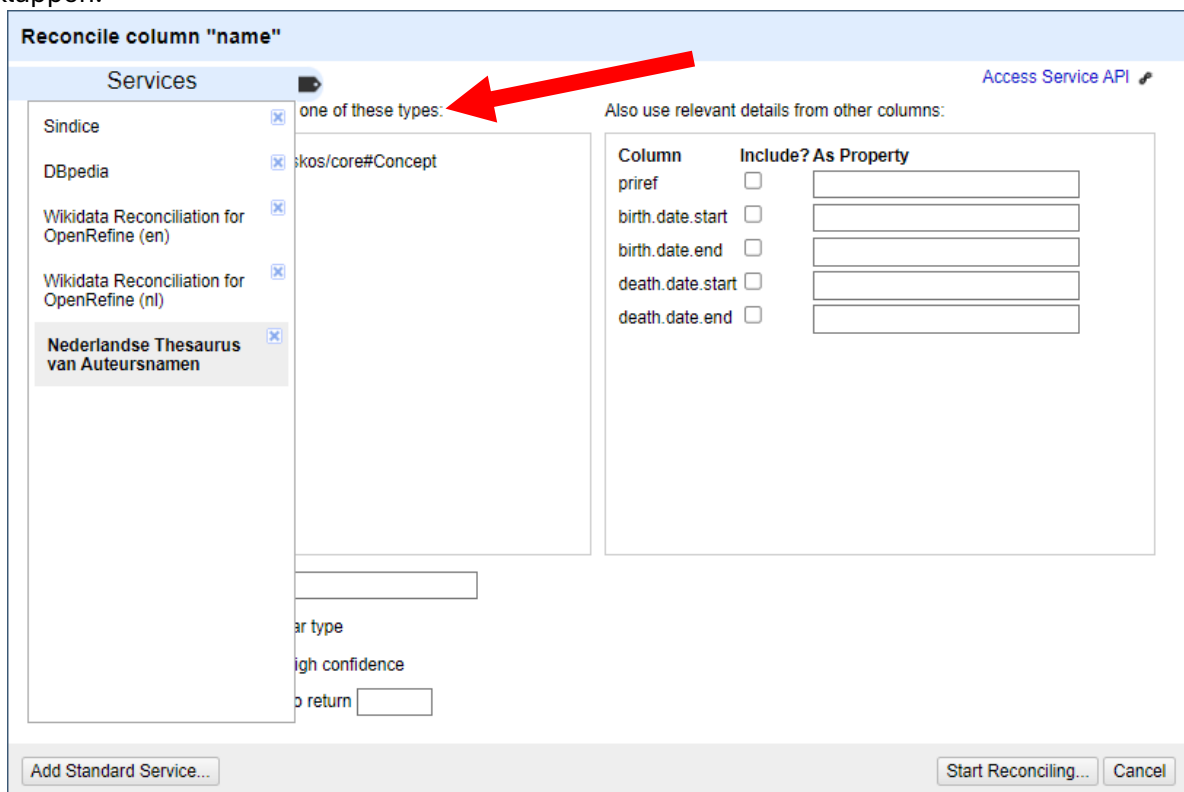
Add standard service...

Discover services...


Start reconciling...

Cancel

Het venster hieronder is het resultaat. Klik op het tab-icoontje (rode pijl) om dat tabblad in te klappen.



Reconcile column "name"

Services  [Access Service API](#)

one of these types:

- ☒ Sindice
- ☒ DBpedia
- ☒ Wikidata Reconciliation for OpenRefine (en)
- ☒ Wikidata Reconciliation for OpenRefine (nl)
- ☒ Nederlandse Thesaurus van Auteursnamen

Also use relevant details from other columns:

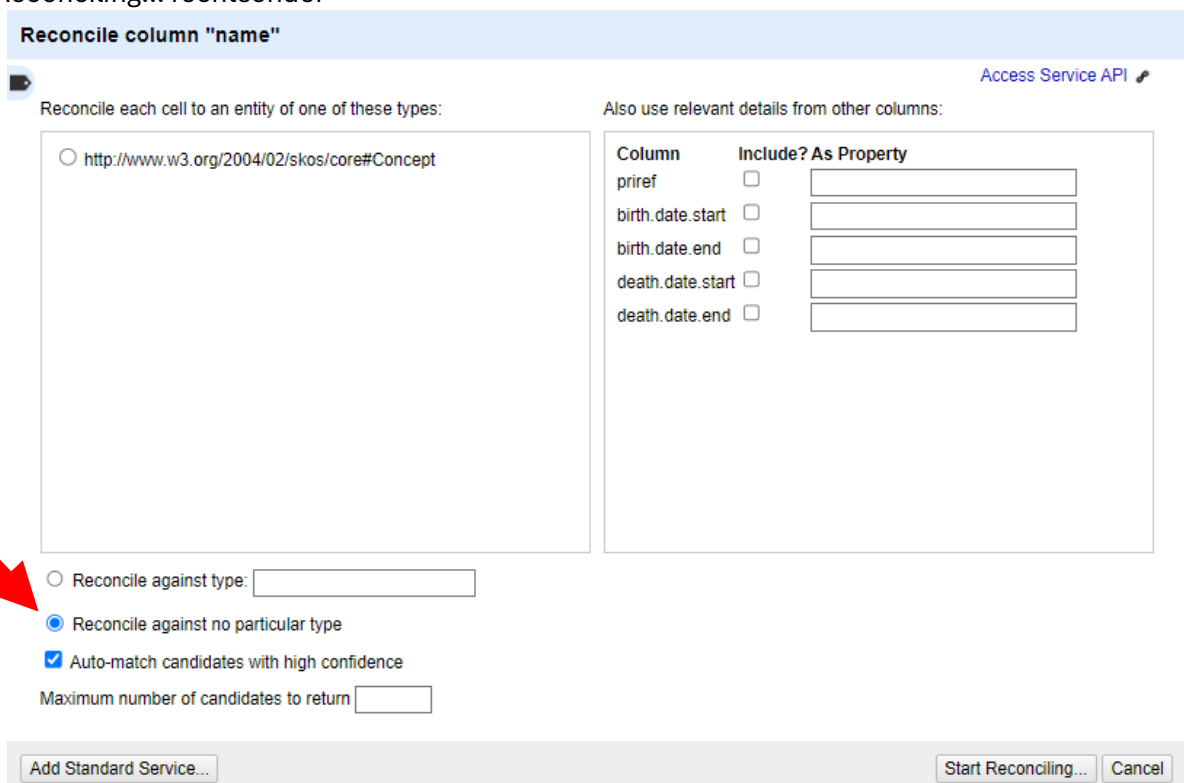
Column	Include?	As Property
preref	<input type="checkbox"/>	<input type="text"/>
birth.date.start	<input type="checkbox"/>	<input type="text"/>
birth.date.end	<input type="checkbox"/>	<input type="text"/>
death.date.start	<input type="checkbox"/>	<input type="text"/>
death.date.end	<input type="checkbox"/>	<input type="text"/>

ar type


☐ high confidence

o return

Klik dan op **Reconcile against no particular type** (rode pijl) en vervolgens op de knop Start Reconciling... rechtsonder



Reconcile column "name"

 [Access Service API](#)

Reconcile each cell to an entity of one of these types:

☐ <http://www.w3.org/2004/02/skos/core#Concept>

☐ Reconcile against type:

☒ Reconcile against no particular type

☒ Auto-match candidates with high confidence

Maximum number of candidates to return

Also use relevant details from other columns:

Column	Include?	As Property
preref	<input type="checkbox"/>	<input type="text"/>
birth.date.start	<input type="checkbox"/>	<input type="text"/>
birth.date.end	<input type="checkbox"/>	<input type="text"/>
death.date.start	<input type="checkbox"/>	<input type="text"/>
death.date.end	<input type="checkbox"/>	<input type="text"/>

Afhankelijk van de grootte van je lijst is dit een goed moment om koffie te gaan halen. De voortgang is bovenin in het gele kader te volgen.

uitbreidingMETCHAT.xlsx
[Permalink](#)

Reconcile cells in column nameCHAT to type null
 5% complete [Cancel](#)

202 rows					
Show as: rows records		Show: 5 10 25 50 100 500 1000 rows			
<input type="checkbox"/> All	<input type="checkbox"/> prireref	<input type="checkbox"/> name	<input type="checkbox"/> prirerefChatGPT	<input type="checkbox"/> nameCHAT	
	1.	2097	Regout	2097	Regout
	2.	2100	Regout, Petrus Laurentius	2100	Regout, Petrus l
	3.	2112	Societe Ceramique Maestricht, made in Holland, decor rustique	2112	Ceramique Mae
	4.	2113	Petrus Regout	2113	Regout, Petrus
	5.	2114	PetrusRegout Maestricht	2114	Regout, PetrusH
	6.	2115	Heju porselein	2115	Heju porselein
	7.	2116	Gestel, Leo	2116	Gestel, Leo
	8.	2117	Thuringer porzellan	2117	Thuringer porze
	9.	2118	Societe Ceramique Maestricht, made in Holland,	2118	Ceramique Mae
	10.	2120	Sfinx Regout	2120	Regout, Sfinx

Dit is het resultaat:

202 rows						
Show as: rows records		Show: 5 10 25 50 100 500 1000 rows				
▼ All	▼ prirref	▼ name	▼ prirrefChatGPT	▼ nameCHAT		
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item	
☆		75.	2186	P.Kuiten	2186	Kuiten, P. <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		76.	2187	Toon de Jong	2187	Jong, Toon de <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Toon de (100) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		77.	2188	Willem van de Berg	2188	van de Berg, Willem <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		78.	2189	Z. Pievowaroryk	2189	Pievowaroryk, Z. <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		79.	2190	Klaas Huijgen	2190	Huijgen, Klaas <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		80.	2191	Dirk de Jong edit	2191	de Jong, Dirk <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Dirk de (89) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Dirk W. de (84) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Dirk de (1962-) (67) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Coen de (54) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		81.	2192	Hendrikje Heinen	2192	Heinen, Hendrikje <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		82.	2193	Jaques Maris	2193	Maris, Jaques <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		83.	2195	Wouterus Beukers	2195	Beukers, Wouterus <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		84.	2196	Rie Cramer	2196	Cramer, Rie <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Cramer, Rie (100) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
☆		85.	2197	marie	2197	marie <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Marie, Zen (73) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Anne, Marie (67) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Lund, Marie (67)

Hij heeft gezocht op een exacte overeenkomst in naam tussen de aangeleverde dataset en de externe databron (in dit geval de RKDartists database). De naam in het zwart is de oorspronkelijke data, de naam in het blauw de match uit de RKDartists database. Wanneer er (100) in het grijs achter de naam staat betekent dat dat de namen exact overeenkomen dus de score is 100%.

	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
2189	Pievowaroryk, Z. <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
2190	Huijgen, Klaas <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
2191	de Jong, Dirk <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Dirk de (89) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Dirk W. de (8 <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Dirk de (1962 <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Jong, Coen de (54) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
2192	Heinen, Hendrikje <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
2193	Maris, Jaques <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
2195	Leukers, Wouterus <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item
2196	Cramer, Rie <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Cramer, Rie (100) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item

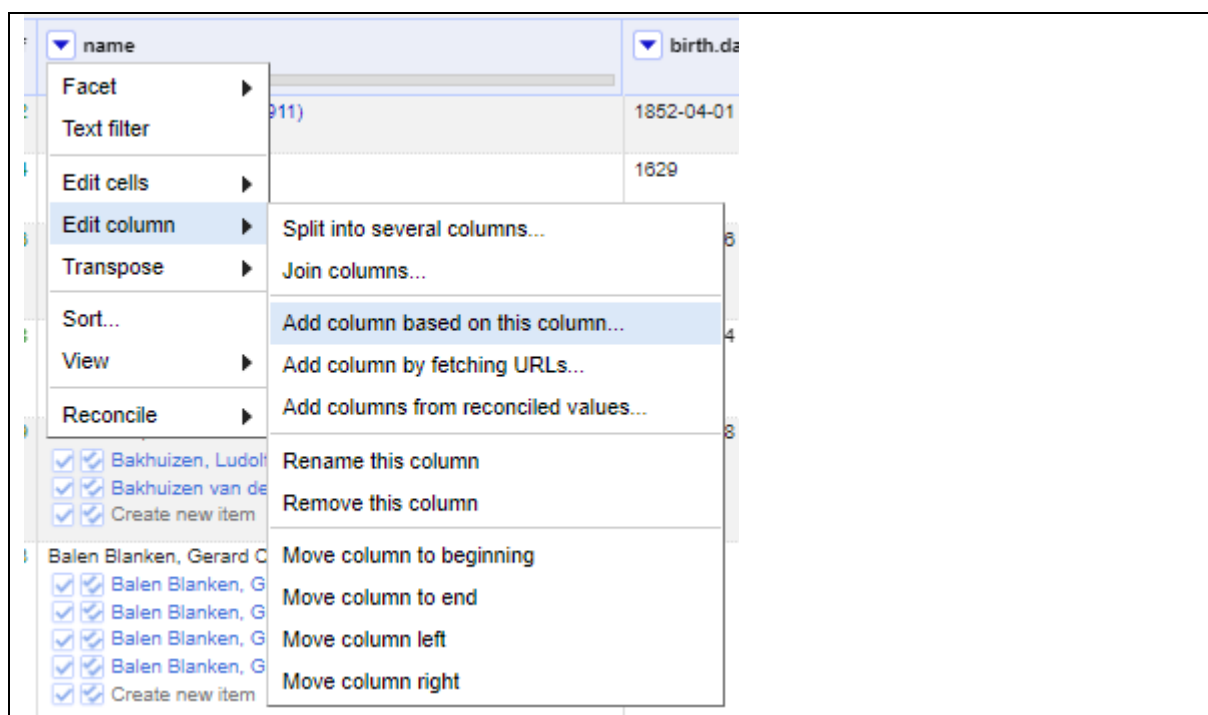
Cramer, Rie

remark names: also worked under pseudonyms Marc Holman and Annie Smit, monogrammed her work 'RC' or in full 'Rie Cramer' (Fichescollectie Monogrammen MHK) • remark family relation: Married from 1912 to 1914 to Peter Otten, clerk and poet (De Gids). Relationship with Albert Plasschaert ca. 1914-1919. Wife of actor and theater director Eduard Rutger Verkade between 1922 and 1933 • remark: Carbasius, Fransje

Wanneer je klikt op het vinkje voor de naam zal deze als match worden weergegeven. De oorspronkelijke data zal worden vervangen door de data uit de externe bron.

2196	Cramer, Rie
	Choose new match

Wil je de oorspronkelijke data behouden in een aparte kolom, zodat je de 2 met elkaar kunt vergelijken, klik dan op het pijltje naast name en kies voor **Edit column > Add column based on this column**



In het venster dat opent word je gevraagd een column name in te geven. Geef hier een naam op die duidelijk maakt wat de inhoud van de kolom is. Bij Expression staat standaard **value**. Laat dit zo staan; je ziet in het voorbeeld eronder al dat de gegevens in de nieuwe kolom een kopie zijn. Klik dan op OK.

Add column based on column nameCHAT

New column name

On error ☒ set to blank ☐ store error ☐ copy value from original column

Expression Language No syntax error.

Preview History Starred Help

row	value	value
1.	Regout	Regout
2.	Regout, Petrus Laurentius	Regout, Petrus Laurentius
3.	Ceramique Maestricht, Societe	Ceramique Maestricht, Societe
4.	Regout, Petrus	Regout, Petrus
5.	Regout, PetrusMaestricht	Regout, PetrusMaestricht
6.	Heju porselein	Heju porselein

OK Cancel

Omdat het fijner is om geautomatiseerd te werken (zeker met veel records) is het handig de match-score ook in een aparte kolom te hebben staan.

Klik op het pijltje bij de kolom Name en kies voor Edit column > Add column based on this column. Geef de kolom een herkenbare naam (bijvoorbeeld 'score'). En vul bij de Expression het volgende in: **cell.recon.best.score**

Add column based on column name

New column name

score

On error

☒ set to blank

☐ store error

☐ copy value from original column

Expression

Language

General Refine Expression Language (GREL) ▼

cell.recon.best.score

No syntax error.

Preview

History

Starred

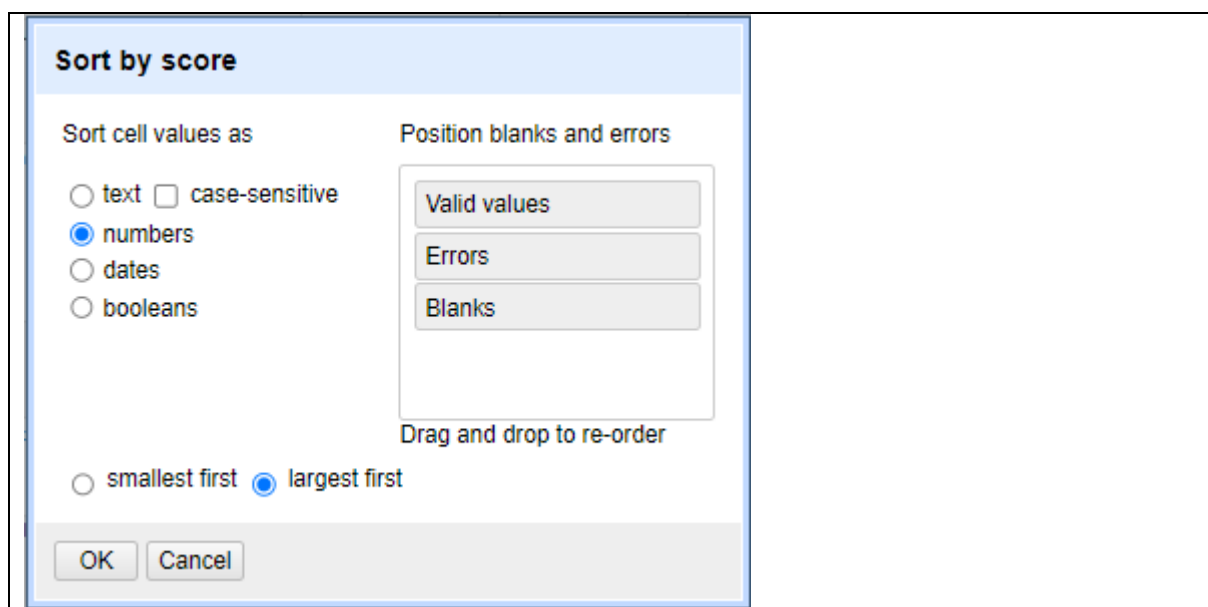
Help

row	value	cell.recon.best.score
1.	Abbey, Edwin Austin	100
2.	Adrichem, Dirk van	null
3.	Apol, Louis	87.5
4.	Bakels, Reinier Sybrand	100
5.	Bakhuizen, Ludolf	92.86
6.	Balen Blanken, Gerard Cornelis van	100

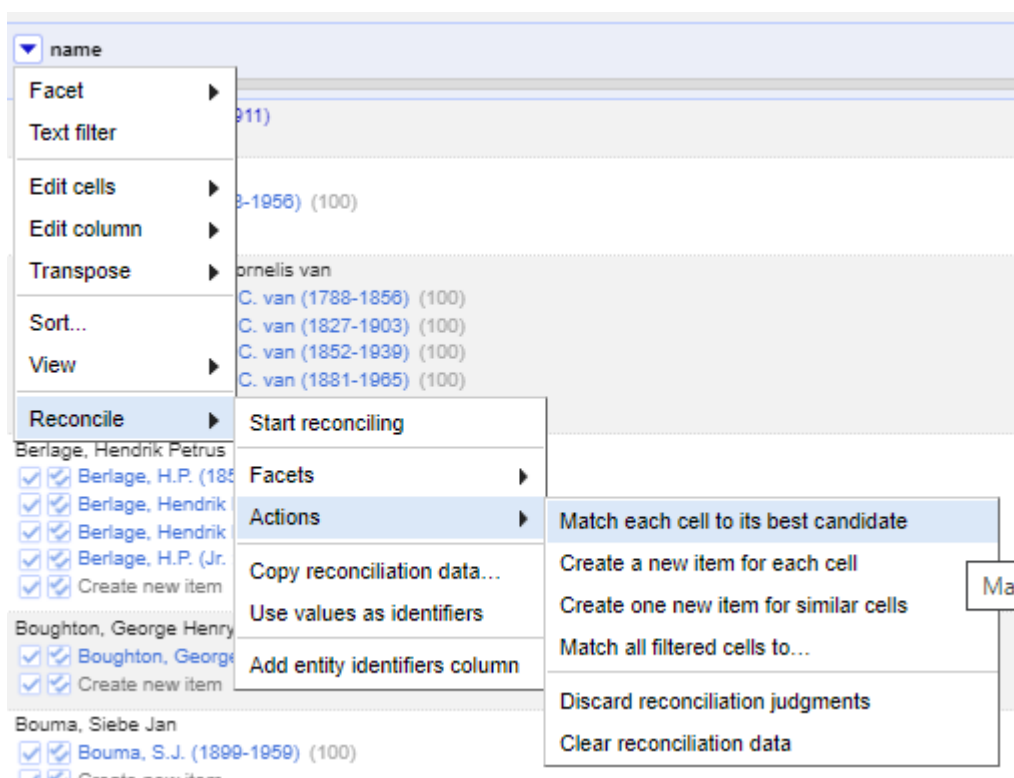
OK

Cancel

Vervolgens klik je op het pijltje bij de kolom 'score' en kies voor **Sort en selecter numbers en largest first**.



Nu kun je de data gaan controleren. Dat kan dus handmatig maar, als je behoorlijk zeker bent van je zaak, of je wilt de check achteraf uitvoeren (bijvoorbeeld door de namenrijen ChatGPTnaam en RKDartistnaam nog eens met Fuzzy Lookup te vergelijken), dan klik je op het pijltje bij kolom 'name'. Kies voor **Reconcile > Actions > Match each cell to its best candidate**. Hou er hierbij rekening mee dat hij ook de niet-100% matches nu gaat matchen met de externe terminologiebron en dat die match mogelijk niet klopt. Dit kun je echter achteraf ook nog vergelijken en rechtekken (je hebt altijd de kolom met de oorspronkelijke data nog).



Resultaat:

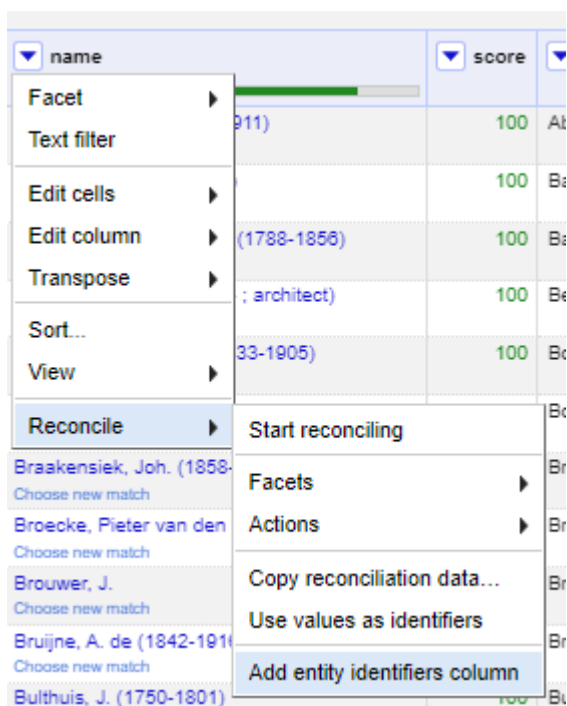
202 rows

Show as: **rows** records Show: 5 10 25 50 100 500 1000 rows Sort ▼

« first < previous 1

▼ All	▼ prirref	▼ name	▼ prirrefChatGPT	▼ nameCHAT	▼ score
☆		2. 2100 Regout, Petrus Laurentius	2100	Regout, Petrus Choose new match	100
☆		4. 2113 Petrus Regout	2113	Regout, Petrus Choose new match	100
☆		7. 2116 Gestel, Leo	2116	Gestel, Leo Choose new match	100
☆		12. 2122 Henri Cassiers	2122	Cassiers, Henry Choose new match	100
☆		14. 2124 Klaas Zwaan	2124	Zwaan, Klaas Choose new match	100
☆		16. 2126 Leo Gestel	2126	Gestel, Leo Choose new match	100
☆		17. 2127 Henk Bellaard	2127	Bellaard, Henk Choose new match	100
☆		23. 2133 David Schulman	2133	Schulman, David Choose new match	100
☆		31. 2142 Adriaan Lubbers	2142	Lubbers, Adriaan Choose new match	100
☆		33. 2144 Peter Dorleijn	2144	Dorleijn, Peter Choose new match	100
☆		35. 2146 Betsy Repelius	2146	Repelius, Betsy Choose new match	100
☆		36. 2147 Johan Traarbach	2147	Traarbach, Johan Choose new match	100
☆		37. 2148 Jo Bezaan	2148	Bezaan, Jo Choose new match	100
☆		44. 2155 Johan Conrad Greive	2155	Greive, Johan Conrad Choose new match	100
☆		50. 2161 Lussenburg Jos	2161	Lussenburg, Jos Choose new match	100
☆		60. 2171 Hans Hermann	2171	Herrmann, Hans (1858-1942) Choose new match	100
☆		61. 2172 Marie Henry Mackenzie	2172	Mackenzie, Marie Henry • Mackenzie, Marie Henrie (1878-1961) Choose new match	100
☆		72. 2183 Isaac Tinion	2183	Tinion, Isaac	100

Nu willen we het belangrijkste uit de externe bron ophalen; namelijk **de URI**, de unieke permalink die hoort bij elke naam. Klik op het pijltje bij kolom ‘name’ en kies voor **Reconcile > Add entity identifiers column**



Er wordt gevraagd om een kolomnaam. Kies voor een herkenbare naam en druk op OK.

Nu hebben alle namen die een match kregen in de RKDartists database een URI. Deze URI kun je vervolgens importeren in je documentatiesysteem. Het geeft een website developer de mogelijkheden om allerlei extra informatie op te halen uit de externe terminologiebron, denk aan een biografie, relaties met andere schrijvers, gepubliceerde werken, etc. Zolang als deze gegevens maar geregistreerd zijn in de externe bron. Hiermee kun je dus de gegevens op je collectiewebsite of bibliotheekportal aanzienlijk uitbreiden, en daarvoor heb je alleen maar 1 simpel linkje in je systeem hoeven importeren! Ook kun je veel gemakkelijker jouw collectie of documentatie koppelen aan andere collecties of documentatie, zowel intern als extern, en je data is ook nog eens opgeschoond! Hele grote winst met heel weinig moeite.

2100	Regout, Petrus Laurentius	2100	Regout, Petrus Choose new match	https://data.rkd.nl/artists/424226	100
2113	Petrus Regout	2113	Regout, Petrus Choose new match	https://data.rkd.nl/artists/424226	100
2116	Gestel, Leo	2116	Gestel, Leo Choose new match	https://data.rkd.nl/artists/31221	100
2122	Henri Cassiers	2122	Cassiers, Henry Choose new match	https://data.rkd.nl/artists/15786	100
2124	Klaas Zwaan	2124	Zwaan, Klaas Choose new match	https://data.rkd.nl/artists/86690	100
2126	Leo Gestel	2126	Gestel, Leo Choose new match	https://data.rkd.nl/artists/31221	100
2127	Henk Bellaard	2127	Bellaard, Henk Choose new match	https://data.rkd.nl/artists/6190	100
2133	David Schulman	2133	Schulman, David Choose new match	https://data.rkd.nl/artists/71321	100
2142	Adriaan Lubbers	2142	Lubbers, Adriaan Choose new match	https://data.rkd.nl/artists/51118	100
2144	Peter Dorleijn	2144	Dorleijn, Peter Choose new match	https://data.rkd.nl/artists/23923	100
2146	Betsy Repelius	2146	Repelius, Betsy Choose new match	https://data.rkd.nl/artists/89854	100
2147	Johan Traarbach	2147	Traarbach, Johan Choose new match	https://data.rkd.nl/artists/78067	100
2148	Jo Bezaan	2148	Bezaan, Jo Choose new match	https://data.rkd.nl/artists/8006	100
2155	Johan Conrad Greive	2155	Greive, Johan Conrad Choose new match	https://data.rkd.nl/artists/33671	100
2161	Lussenburg Jos	2161	Lussenburg, Jos Choose new match	https://data.rkd.nl/artists/51397	100
2171	Hans Hermann	2171	Herrmann, Hans (1858-1942) Choose new match	https://data.rkd.nl/artists/37930	100
2172	Marie Henry Mackenzie	2172	Mackenzie, Marie Henry • Mackenzie, Marie Henrie (1878-1961) Choose new match	https://data.rkd.nl/artists/51768	100

Je hebt je unieke ID binnen je documentatiesysteem, dus importeren van deze gegevens is nu heel gemakkelijk geworden. Elke naam is vervangen door de voorkeursspelting vanuit RKDartists. Je kunt deze voorkeursspelting overnemen, maar dat hoeft niet. Het schoont wel in één keer al je namen op want die stonden in de oorspronkelijke data nog als Voornaam Achternaam terwijl Achternaam, Voornaam wenselijker is. Daarnaast heb je meteen de URI waardoor je plotseling een hele stap dichterbij het 5-sterrenmodel van Linked Data bent gekomen!

4. Importeer je opgeschoonde lijst in je documentatie- of collectiebeheersysteem

De importroutine is afhankelijk van het soort systeem wat je in gebruik hebt. Het hebben van een uniek ID, priref, recordnummer of documentnummer horend bij de persoonsnaam, term of plaats zal echter in de meeste gevallen cruciaal zijn voor het goed kunnen importeren van de gegevens. Neem deze dus altijd mee als je een export maakt van je datasets om op te schonen in Excel of OpenRefine