

Workflow gestructureerde data toevoegen aan oude mediacollectie op Wikimedia Commons met OpenRefine

- [1. Categorie op Commons zoeken](#)
- [2. Lijst met bestandsnamen genereren](#)
- [3. Maak een nieuw project in OpenRefine](#)
- [4. Koppelen van je bestandsnamen aan Wikimedia Commons](#)
- [5. Gegevens ophalen uit Wikimedia Commons naar OpenRefine](#)
- [6. Tekst isoleren uit Wikitext](#)
- [7. Gegevens koppelen aan Wikidata](#)
- [8. Gekoppelde gegevens toevoegen aan Wikidata](#)
- [9. \(Zoek\)resultaten bekijken](#)

[Documentatie](#)

1. Categorie op Commons zoeken

Zoek categorie op Commons die je wilt gaan gebruiken, in dit voorbeeld:

[https://commons.wikimedia.org/wiki/Category:Photographs_by_Willem_van_de_Poll_in_Nick_erie,_Suriname_\(1947\)](https://commons.wikimedia.org/wiki/Category:Photographs_by_Willem_van_de_Poll_in_Nick_erie,_Suriname_(1947))

2. Lijst met bestandsnamen genereren

Maak een filelist van bestandsnamen in de categorie voor import in OpenRefine. Dat werkt als volgt:

- ga naar <https://petscan.wmflabs.org/>
- op pagina/tabblad 'Categories' vul in:
 - language = Commons
 - project = wikimedia
 - depth = 1 (of meer als je subcategorieën hebt)
- op pagina/tabblad 'Pages properties' vink de optie 'file' aan
- scroll naar beneden en klik 'do it'
- controleer: het totale aantal files moet hetzelfde zijn als het aantal files in de categorie
- bij het tabblad 'output', optie 'format' vink aan 'plain text'

- klik nogmaals op de knop 'do it'
- kopieer de filenamen uit het venster dat verschijnt

3. Maak een nieuw project in OpenRefine

- Open OpenRefine (downloaden: <https://openrefine.org/download>)
- Kies 'Create project' en 'clipboard'
- Plak de filenames die je in stap 2 gekopieerd hebt uit Petscan en klik 'next'
- Je ziet het veld om je data te importeren in OpenRefine.

Om de data in de goede vorm te krijgen, verwijder je de ';' in het 'custom' veld, en verander je de 1 in een 0 bij 'Parse next line(s) as column headers'.

Het zou eruit moeten zien zoals dit voorbeeld met de hele bestandsnaam in een kolom (Column 1), en beginnend op de eerste rij:

OpenRefine A power tool for working with messy data.

Create project « start over Configure parsing options Project name C

Open project
mport project
..anguage settings

Column 1
1. File:Het koninklijk gezelschap op de nieuwe zeedijk in het district Nickerie, Bestanddeelnr 252-4684.jpg
2. File:Het koninklijk gezelschap op de nieuwe zeedijk in het district Nickerie, Bestanddeelnr 252-4683.jpg
3. File:Ereboog in het dorp Paradise in het district Nickerie, Bestanddeelnr 252-4675.jpg
4. File:De koningin tekent het gouden boek van de Prins Bernhardpolder. Rechts gouverneu, Bestanddeelnr 252-4635.jpg
5. File:Het koninklijk gezelschap luistert naar een toespraak in een gebouw in de Prins , Bestanddeelnr 252-4634.jpg
6. File:Het koninklijk gezelschap komt aan bij een nederzetting in de Prins Bernhardpold, Bestanddeelnr 252-4632.jpg
7. File:Het koninklijk gezelschap op de nieuwe zeedijk van Nickerie, Bestanddeelnr 252-4631.jpg
8. File:Het koninklijk gezelschap op de nieuwe zeedijk van Nickerie, Bestanddeelnr 252-4630.jpg
9. File:Het koninklijk gezelschap verlaat Nickerie met Hr. Ms. Willem van der Zaan, Bestanddeelnr 252-4629.jpg
10. File:Het koninklijk gezelschap in Nickerie, Bestanddeelnr 252-4622.jpg
11. File:De koningin krijgt in Nickerie door vertegenwoordigers van de Creoolse bevolking, Bestanddeelnr 252-4616.jpg
12. File:De koningin krijgt in Nickerie door vertegenwoordigers van de Javaanse bevolking, Bestanddeelnr 252-4614.jpg
13. File:De koningin krijgt in Nickerie door vertegenwoordigers van de Javaanse bevolking, Bestanddeelnr 252-4613.jpg
14. File:De koningin en prins worden ontvangen in Nickerie, Bestanddeelnr 252-4601.jpg
15. File:Het koninklijk gezelschap verlaat Nickerie op de Beatrix, Bestanddeelnr 252-4593.jpg
16. File:Het koninklijk gezelschap komt met de Prinses Beatrix aan in Nickerie, Bestanddeelnr 252-4590.jpg
17. File:Het koninklijk gezelschap komt met de Prinses Beatrix aan in Nickerie, Bestanddeelnr 252-4589.jpg
18. File:De koningin aan boord van de Beatrix op weg naar Nickerie, Bestanddeelnr 252-4587.jpg
19. File:Een gebouw in de Bernhardpolder in Nickerie, Bestanddeelnr 252-4782.jpg
20. File:Hindostaanse vrouw in Nickerie, Bestanddeelnr 252-4688.jpg
21. File:Javaanse vrouw in Nickerie, Bestanddeelnr 252-4687.jpg
22. File:Hindostaanse vrouwe in Nickerie, Bestanddeelnr 252-4686.jpg
23. File:De kotomisi Pauline in Nickerie, Bestanddeelnr 252-4637.jpg
24. File:De kotomisi Pauline in Nickerie, Bestanddeelnr 252-4636.jpg
25. File:Het koninklijk gezelschap in Nickerie, Bestanddeelnr 252-4623.jpg
26. File:Julianadorp in Nickerie, Bestanddeelnr 252-4606.jpg
27. File:Julianadorp in Nickerie, Bestanddeelnr 252-4605.jpg
28. File:Weg bij Nickerie, Bestanddeelnr 252-4603.jpg
29. File:De Beatrix komt aan in Nickerie, Bestanddeelnr 252-4600.jpg
30. File:De Beatrix komt aan in Nickerie, Bestanddeelnr 252-4599.jpg
31. File:De sluis bij Nickerie, Bestanddeelnr 252-4596.jpg
32. File:Een sluis bij Nickerie, Bestanddeelnr 252-4595.jpg
33. File:De Beatrix legt aan in Nickerie, Bestanddeelnr 252-4594.jpg
34. File:Monding van de Corantijnrivier in Suriname vanuit een vliegtuig, Bestanddeelnr 252-2672.jpg
35. File:De koningin krijgt in Nickerie door vertegenwoordigers van de Creoolse bevolking, Bestanddeelnr 252-4615.jpg
36. File:Geïmproviseerde vlag in Nickerie, Bestanddeelnr 252-4685.jpg
37. File:Een gebouw in de Bernhardpolder in Nickerie, Bestanddeelnr 252-4781.jpg

- Geef je project een naam, en evt. een tag, en klik op 'create project'

4. Koppelen van je bestandsnamen aan Wikimedia Commons

Open Refine heeft nu je project gegenereerd.

De volgende stap is het koppelen van je bestandsnamen aan Wikimedia Commons, zodat OpenRefine weet waar het de informatie vandaan moet halen.

Hoe dit gaat, lees je in stap 5 van [de 'How to upload new files to Wikimedia Commons'](#) handleiding van Sandra Fauconnier, directeur van OpenRefine (in het Engels).

5. Gegevens ophalen uit Wikimedia Commons naar OpenRefine

Analyseer een aantal van de bestanden in je categorie om te bekijken waar de informatie zit die je als gestructureerde data wilt toevoegen aan Wikidata. Dit kan informatie zijn over wat er op de foto te zien is (P180 - depicts), waar een bestand gemaakt is (P107 - location of creation), het onderwerp van foto (P921 - main subject) of bijvoorbeeld de collectie waartoe het behoort (P195 - collection).

Nuttige informatie die kan worden toegevoegd als structured data zit vaak in:

- description \ beschrijving
- categorieën
- filename

In dit voorbeeld staat de informatie, de locatie waar de foto gemaakt is, in de wikitext. Die wikitext is toe te voegen aan de OpenRefine sheet met behulp van de koppeling met Commons. Dit werkt als volgt:

- Ga naar het drop down pijltje boven de kolom met de bestandsnaam, selecteer Edit column en kies 'Add columns from reconciled values'.

OpenRefine Photographs by Willem van de Poll in Nickerie, Suriname [Permalink](#) Open... Export Help

Facet / Filter Undo / Redo 2 / 2 **102 rows** Extensions Wikidata

Show as: **rows** records Show: 5 10 25 50 100 500 1000 rows « first « previous 1 next » last

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

Column 1 **Wikitext**

Facet
Text filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile

Split into several columns...
Join columns...
Add column based on this column...
Add column by fetching URLs...
Add columns from reconciled values...
Rename this column...
Remove this column
Move column to beginning
Move column to end
Move column left
Move column right

1.  leed in feestkleding in rood, wit, blauw anddeelnr 252-5462.jpg == {{int:filedesc}} ==
{{Information |Description={{nl|Collectie / Archief : Fotocollectie Van de Poll
 Reportage / Serie : Reis naar Suriname en de Nederlandse Antillen
 Beschrijving : Kotomisi Pauline gekleed in feestkleding in rood, wit, blauw en met portretten van koningin Wilhelmina, prinses Juliana en prins Bernhard
 Annotatie : Pauline was de gastvrouw van het gouvernementslogeerhuis in Nieuw-Nickerie, Suriname
 Trefwoorden : creolen, inheemse bevolking, kleding, vrouwen
 Fotograaf : Poll, Willem van de
 Auteursrechtthebbende : Nationaal Archief
 Materiaalsoort : Negatief (zwart/wit)
 Nummer archiefinventaris : bekijk toegang 2.24.14.02
 Bestanddeelnr : 252-5462}} |Source=http://proxy.handle.net/10648/aeb635e4-d0b4-102d-bcf8-003048976d84 |Date= 1947 |Author=Willem van de Poll |Permission= |other_versions= }} {{Nationaal Archief}} == {{int:license-header}} == {{CC-zero}} [[Category:Images from Nationaal Archief]] [[Category:Photographs by Willem van de Poll in Nickerie, Suriname (1947)]] [[Category:Kotomisi]]

2.  File:Kotomisi Pauline gekleed in feestkleding in rood, wit, blauw en met portretten v Bestanddeelnr 252-5461.jpg Choose new match == {{int:filedesc}} ==
{{Information |Description={{nl|Collectie / Archief : Fotocollectie Van de Poll
 Reportage / Serie : Reis naar Suriname en de Nederlandse Antillen
 Beschrijving : Kotomisi Pauline gekleed in feestkleding in rood, wit, blauw en met portretten van koningin Wilhelmina, prinses Juliana en prins Bernhard
 Annotatie : Pauline was de gastvrouw van het gouvernementslogeerhuis in Nieuw-Nickerie, Suriname
 Trefwoorden : creolen, inheemse bevolking, kleding, vrouwen
 Fotograaf : Poll, Willem van de
 Auteursrechtthebbende : Nationaal Archief
 Materiaalsoort : Negatief (zwart/wit)
 Nummer archiefinventaris : bekijk toegang 2.24.14.02
 Bestanddeelnr : 252-5462}} |Source=http://proxy.handle.net/10648/aeb635e4-d0b4-102d-bcf8-003048976d84 |Date= 1947 |Author=Willem van de Poll |Permission= |other_versions= }} {{Nationaal Archief}} == {{int:license-header}} == {{CC-zero}} [[Category:Images from Nationaal Archief]] [[Category:Photographs by Willem van de Poll in Nickerie, Suriname (1947)]] [[Category:Kotomisi]]

- Kies de kolommen die je wilt toevoegen aan je OR bestand, in dit geval 'Wikitext'.
Klik op OK.
Het programma genereert een preview.

Add columns from reconciled column Column 1

Add property Preview Reset

Suggested properties

- collection
- coordinates of the point of view
- copyright license
- copyright status
- creator
- depicts
- digital representation of
- inception
- location of creation
- main subject
- source of file
- Wikitext**

Column 1	Wikitext
File:Kotomisi Pauline gekleed in feestkleding in rood, wit, blauw en met portretten v Bestanddeelnr 252-5462.jpg	== {{int:filedesc}} == {{Information Description={{nl Collectie / Archief : Fotocollectie Van de Poll Reportage / Serie : Reis naar Suriname en de Nederlandse Antillen Beschrijving : Kotomisi Pauline gekleed in feestkleding in rood, wit, blauw en met portretten van koningin Wilhelmina, prinses Juliana en prins Bernhard Annotatie : Pauline was de gastvrouw van het gouvernementslogeerhuis in Nieuw-Nickerie, Suriname Trefwoorden : creolen, inheemse bevolking, kleding, vrouwen Fotograaf : Poll, Willem van de Auteursrechtthebbende : Nationaal Archief Materiaalsoort : Negatief (zwart/wit) Nummer archiefinventaris : bekijk toegang 2.24.14.02 Bestanddeelnr : 252-5462}} Source=http://proxy.handle.net/10648/aeb635e4-d0b4-102d-bcf8-003048976d84 Date= 1947 Author=Willem van de Poll Permission= other_versions= }} {{Nationaal Archief}} == {{int:license-header}} == {{CC-zero}} [[Category:Images from Nationaal Archief]] [[Category:Photographs by Willem van de Poll in Nickerie, Suriname (1947)]] [[Category:Kotomisi]]
File:Kotomisi Pauline	== {{int:filedesc}} == {{Information Description=

OK Cancel

De Wikitext is nu toegevoegd in een nieuwe kolom.

6. Tekst isoleren uit Wikitext

Om de gestructureerde data toe te kunnen voegen aan Wikidata, moet de tekst die je daarvoor wilt gebruiken geïsoleerd worden uit de wikitext.

In dit geval gebruiken we de locatie waar de foto gemaakt is, om het statement 'Location of creation' toe te voegen. Hiervoor maken we gebruik van een programma (Jython) dat deze tekst uit de wikitext kan halen.

Dat gaat als volgt:

- Klik op 'Edit column' in het drop down menu, en weer op 'Add column based on this column'.
 - Vul de new column name in, in dit geval 'Plaatsbepaling'
 - Bij het veld 'Expression' selecteer je bij language 'Python/Jython'.
- In het veld eronder vul je deze expressie in:

```
return cells["Wikitext"].value.split("Locatie : ")[-1].split("\n")[0].replace("<br />", "")
```

Deze expressie zoekt in de kolom Wikitext naar de tekst (string) "Locatie : " en geeft als resultaat de tekst die na de tekst op dezelfde regel volgt.

b.v.:

Locatie : **Nickerie, Suriname**
Maker: Willem van de Poll

Als je een ander wikitext element uit de tekst wilt halen, dan moet je "Locatie : " vervangen door dat element, b.v. "Maker : ".

Let op: dit werkt alleen als die informatie in een aparte regel staat, gescheiden door een enter.

Open Refine genereert een preview zodat je kunt kijken of het werkt.
Als het er goed uit ziet, klik je op oke.

Als het niet lukt, controleer dan of je de juiste gegevens op de juiste plek in de expressie hebt ingevuld. Lukt het dan nog steeds niet, vraag dan hulp via het [OpenRefine forum](#) of de [OpenRefine telegram groep](#).

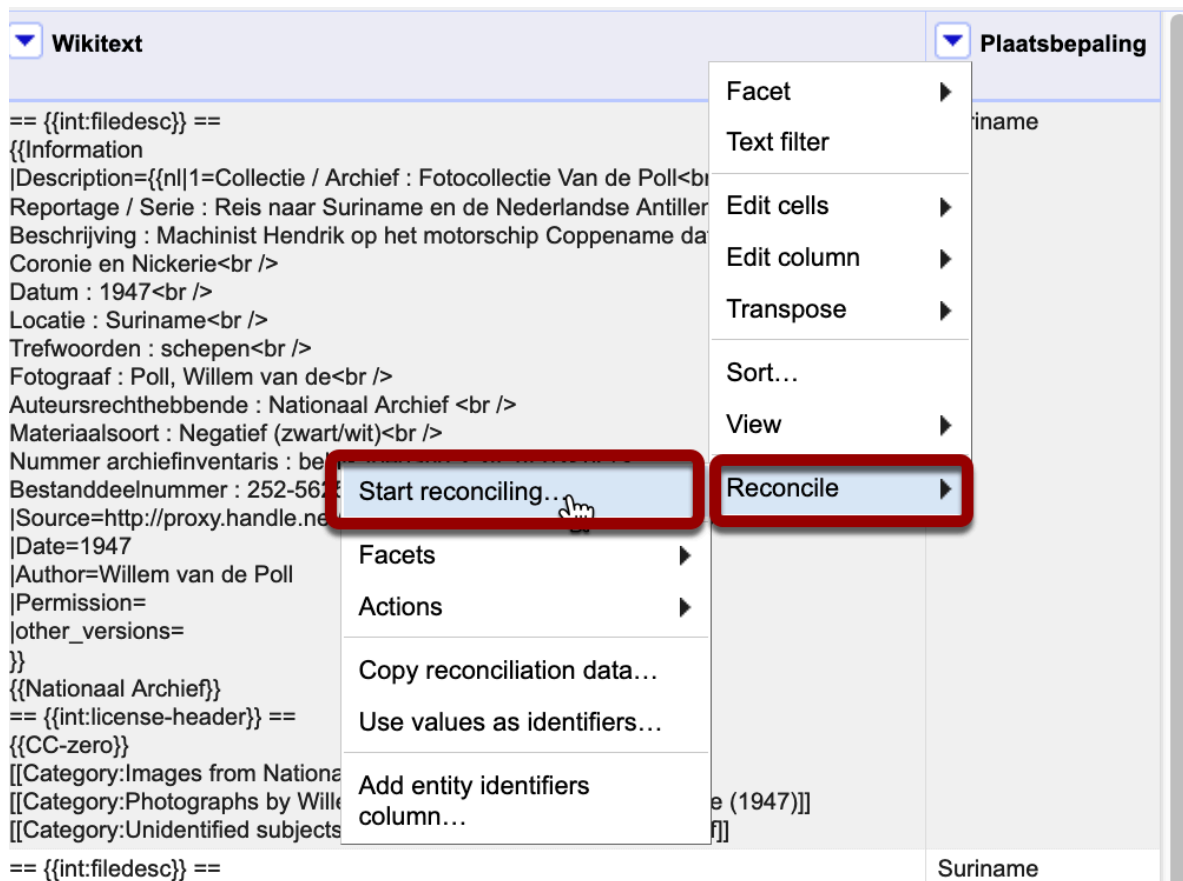
7. Gegevens koppelen aan Wikidata

In de nieuwe kolom staat de naam van de locatie.

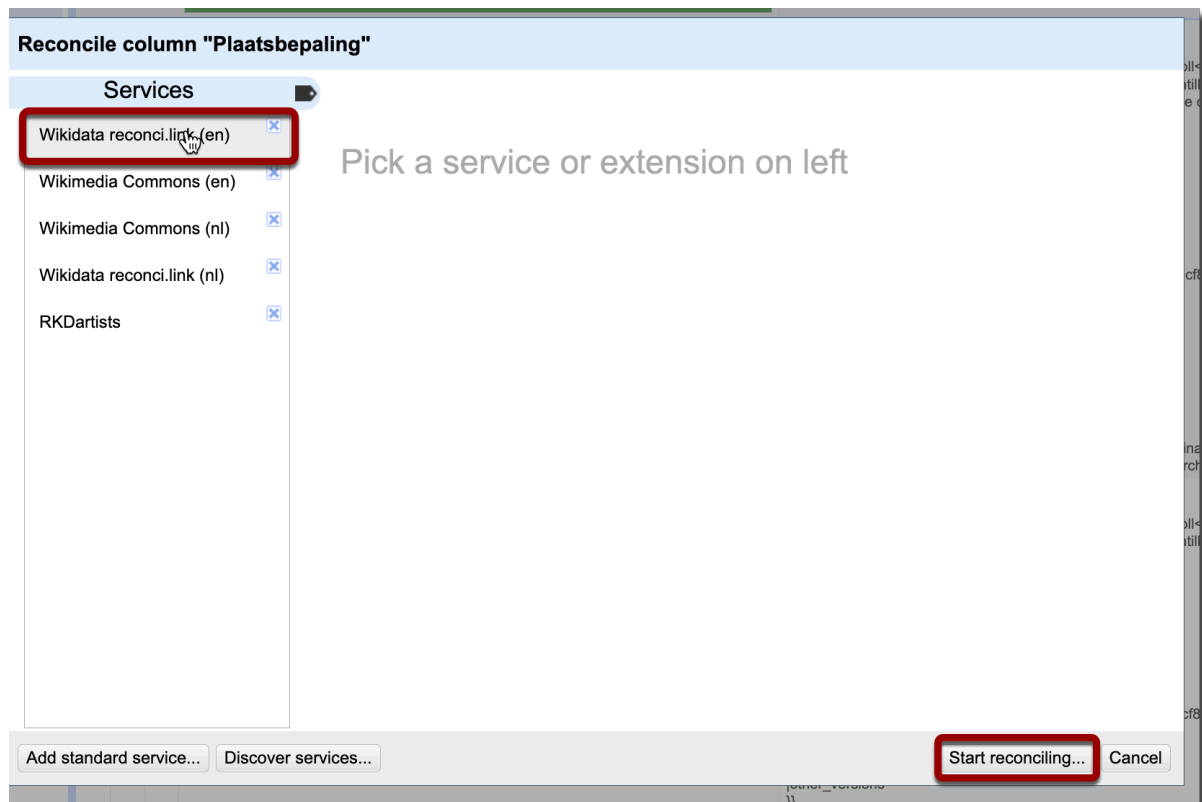
Deze kunnen we nu vervolgens gaan koppelen aan Wikidata.

Dat gaat als volgt.

- Ga weer naar het menu van de kolom 'Plaatsbepaling' en klik op het pijltje naar beneden. Kies de onderste optie 'reconcile'.



- OpenRefine vraagt je nu een database te kiezen waar je je data aan gaat koppelen. Kies Wikidata. De Engelse Wikidata is standaard toegevoegd, wil je een andere reconciliatie service toevoegen (of in een andere taal), bekijk dan de documentatie hierover (LINK NOG INVOEGEN) .



- In het pop-up venster dat verschijnt kun je de voorwaarden voor het koppelen van de data instellen.
OpenRefine geeft automatisch suggesties van soorten data waar je mee kunt matchen.
Kies een van deze suggesties, of vul je eigen soort in (de bullet 'Reconcile against type').
Als je veel verschillende soorten typen in je kolom hebt staan, kies dan 'Reconcile against no particular type'.
- Als je de koppeling van OpenRefine met Wikidata nog handmatig wilt goedkeuren, haal dan het vinkje voor 'auto-match candidates with high confidence' weg.
Anders koppelt OpenRefine automatisch gegevens waarvan het programma vrij zeker weet dat het om hetzelfde item gaat.
- Als je alles hebt ingesteld, klik je op de knop 'Start reconciling'.

OpenRefine gaat nu in Wikidata zoeken of het items kan vinden die overeenkomen met de waarden in je kolom. Het is afhankelijk van de hoeveelheid data hoe lang dit duurt. Het resultaat ziet er ongeveer zo uit:

Wikitext

```

= {{int:filedesc}} ==
[Information
Description={{nl|Collectie / Archief : Fotocollectie Van de Poll<br />
Reportage / Serie : Reis naar Suriname en de Nederlandse Antillen<br />
Beschrijving : Kotomisi Pauline gekleed in feestkleding in rood, wit, blauw en met
ortretten van koningin Wilhelmina, prinses Juliana en prins Bernhard<br />
Annotatie : Pauline was de gastvrouw van het gouvernementslogeerhuis in Nieuw-
lickerie<br />
Datum : 1947<br />
Locatie : Nickerie, Suriname<br />
Refwoorden : creolen, inheemse bevolking, kleding, vrouwen<br />
Fotoograaf : Poll, Willem van de<br />
auteursrechthebbende : Nationaal Archief <br />
Materiaalsoort : Negatief (zwart/wit)<br />
Lummer archiefinventaris : bekijk toegang 2.24.14.02<br />
Iestanddeelnnummer : 252-5462}}
Source=http://proxy.handle.net/10648/aeb635e4-d0b4-102d-bcf8-003048976d84
Date= 1947
Author=Willem van de Poll
Permission=
Other_versions=
}
[Nationaal Archief]]
= {{int:license-header}} ==
[CC-zero}}
Category:Images from Nationaal Archief]]
Category:Photographs by Willem van de Poll in Nickerie, Suriname (1947)]]
Category:Kotomisi]]

```

Plaatsbepaling

Nickerie, Suriname

☒ ☒ Nickerie River (77)

☒ ☒ Nickerie District (73)

☒ ☒ Nieuw Nickerie (71)

☒ ☒ NIEUW NICKERIE (71)

☒ ☒ PSV Nickerie (69)

☒ ☒ Asraf Peekhan Stadion (63)

☒ ☒ Lachmipersad Mungra Regional Hospital Nickerie (62)

☒ ☒ Stg. Drs. L. Mungra Streekziekenhuis Nickerie (62)

☒ ☒ Zeedijk (61)

☒ ☒ Glasgow (55)

☒ ☒ Nieuw Rotterdam (50)

☒ ☒ Praké Kreek (50)

☒ ☒ Kankantrie Kreek (48)

☒ ☒ Sipari Kreek (48)

☒ ☒ Warappe Kreek (47)

☒ ☒ Backtrack (46)

☒ ☒ Nickerie-expedition (46)

☒ ☒ Nickerie-expeditie (46)

☒ ☒ Tapoeripa Kreek (44)

☒ ☒ L'Espérance (Nickerie) (43)

☒ ☒ Boonacker Polder (42)

☒ ☒ Tiengivoloe Kreek (41)

☒ ☒ Henarbrug (41)

☒ ☒ Oostelijke Polders (38)

☒ ☒ Machirchinro Kreek (38)

☒ ☒ Create new item

Search for match

```

= {{int:filedesc}} ==
[Information
Description={{nl|Collectie / Archief : Fotocollectie Van de Poll<br />
Reportage / Serie : Reis naar Suriname en de Nederlandse Antillen<br />
Beschrijving : Kotomisi Pauline gekleed in feestkleding in rood, wit, blauw en met
ortretten van koningin Wilhelmina, prinses Juliana en prins Bernhard<br />
Annotatie : Pauline was de gastvrouw van het gouvernementslogeerhuis in Nieuw-
lickerie<br />
Datum : 1947<br />
Locatie : Nickerie, Suriname<br />
Refwoorden : creolen, inheemse bevolking, kleding, vrouwen<br />
Fotoograaf : Poll, Willem van de<br />
auteursrechthebbende : Nationaal Archief <br />

```

Nickerie, Suriname

☒ ☒ Nickerie River (77)

☒ ☒ Nickerie District (73)

☒ ☒ Nieuw Nickerie (71)

☒ ☒ NIEUW NICKERIE (71)

☒ ☒ PSV Nickerie (69)

☒ ☒ Asraf Peekhan Stadion (63)

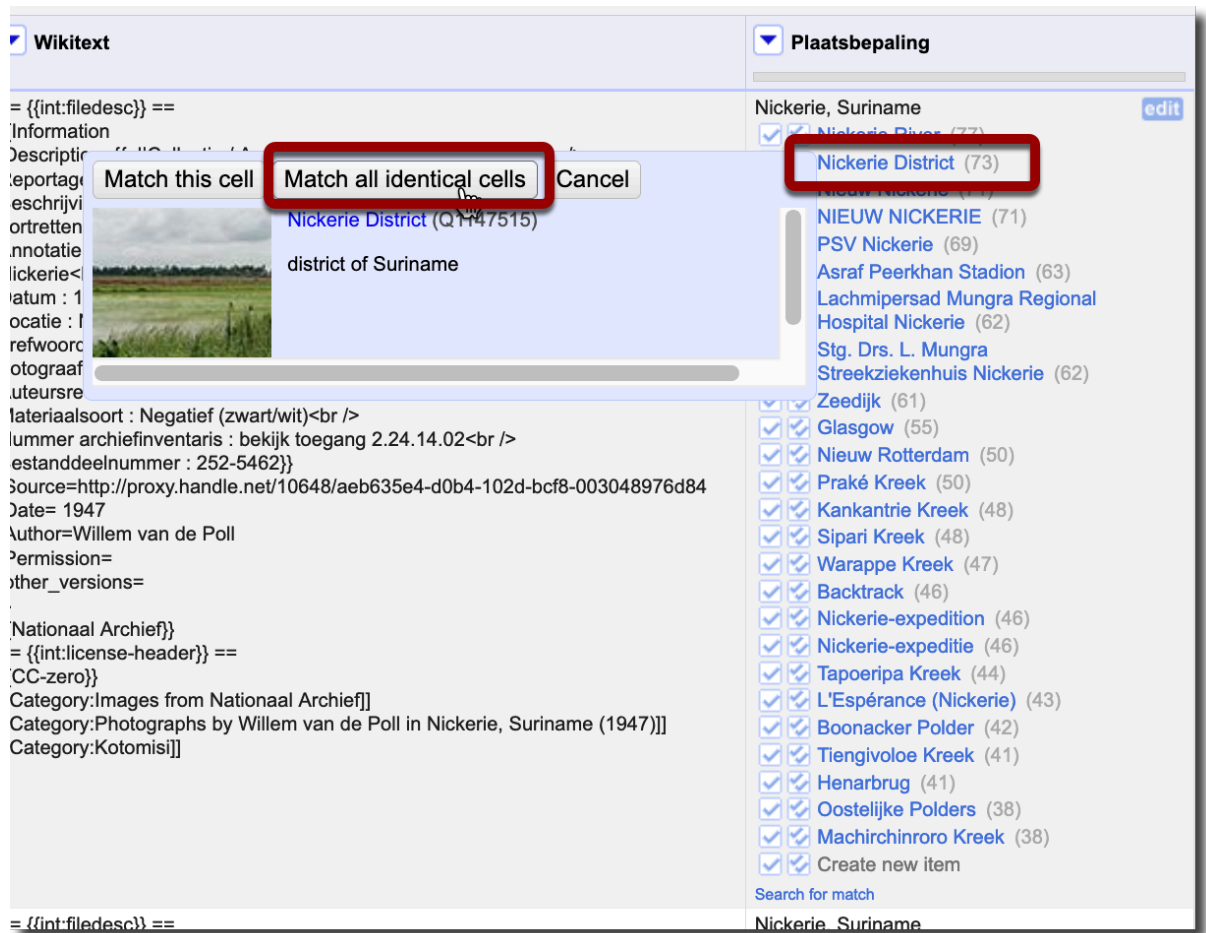
☒ ☒ Lachmipersad Mungra Regional Hospital Nickerie (62)

☒ ☒ Stg. Drs. L. Mungra Streekziekenhuis Nickerie (62)

☒ ☒ Zeedijk (61)

Nu is het zaak om je rijen handmatig na te gaan om ze aan de juiste items op Wikidata te koppelen. Dat gaat als volgt:

- Klik op het item in de rij met blauwe links, een pop-up verschijnt met meer informatie over het item op Wikidata en een foto, zodat je kunt controleren of dit het juiste item is.
- Als het het juiste item is, klik je 'Match this cell'.
Als je zeker weet dat alle cellen in je kolom met deze waarde naar hetzelfde item verwijzen, klik je op 'Match all identical cells'



OpenRefine koppelt dan het item uit Wikidata aan de gegevens in jouw kolom.
Ga hiermee door je alle items hebt gehad en tot de voortgangsbalk aan de bovenkant van de kolom helemaal groen is.

8. Gekoppelde gegevens toevoegen aan Wikidata

Nu gaan we deze aan Wikidata gekoppelde locatie toevoegen als gestructureerde data aan Wikimedia Commons. Volg hiervoor [stap 6 tot en met 8 uit de 'How to upload new files to Wikimedia Commons'](#) handleiding van Sandra Fauconnier, directeur van OpenRefine (in het Engels).

Klaar! Je hebt nu gestructureerde data toegevoegd aan je oude collectie afbeeldingen op Wikimedia Commons.

Op dezelfde manier is het uiteraard mogelijk om nog veel meer gestructureerde data toe te voegen, denk bijvoorbeeld aan de maker van de afbeeldingen (P170 - creator), het jaar waarin een object is gemaakt, de auteursrechtelijke status van een foto, de collectie waarin een item zich bevindt etc.

9. (Zoek)resultaten bekijken

Nieuwsgierig naar de doorzoekbaarheid van je data?

Ga naar [de structured search zoekmachine van Hay Kranen](#), en vul de gegevens in.

Voorbeeld:

<https://hay.toolforge.org/sdsearch/#q=haswbstatement:P1071=Q1147515>

Je kunt ook direct zoeken in Wikimedia Commons, gebruik daarvoor de zoekopdracht 'haswbstatement:P1071=Q1147515' (oftewel zoek items die wikibasestatement eigenschap (P-nummer) hebben met item X (Q nummer).

Documentatie

<https://docs.google.com/document/d/19eiMeq3XssiPrT9b04E-8XyE-desBEzYNgygLDYKP4o/edit>

<https://github.com/KBNLwikimedia/OpenRefine-WikimediaCommons-Workshop/>

tool van Hay Kranen om de Mnummers (unieke nummers) van Wikimedia Commons files te exporteren, voor als het matchen op bestandsnaam in OpenRefine niet lekker loopt:

<https://hay.toolforge.org/minefield/>