

# VERBIND JE TERMEN

*Link je Adlib-thesaurus in Excel aan de AAT*

*& Introductie CultuurLINK, Importeren met Adlib Designer, Eenvoudig opschonen*





*Met dank aan:*



# INHOUD

<b>1</b>	<b>Voorwoord Kaat Debo</b>	<b>5</b>
	<b>Voorwoord Annette Gaalman en Wilbert Helmus</b>	<b>6</b>
<b>2</b>	<b>Introductie</b>	<b>7</b>
<b>3</b>	<b>Linken: waarom en hoe?</b>	<b>8</b>
3.1	Omgaan met terminologieën in (inter)nationaal verband	9
3.2	Welke bronnen?	10
3.3	Manieren van aanpak	12
<b>4</b>	<b>Casus: AAT-links documenteren in Adlib</b>	<b>14</b>
4.1	<b>Maak een kopie van de AAT in een CSV-bestand</b>	<b>14</b>
STAP 1:	<i>Zoek de overkoepelende term voor de concepten</i>	14
STAP 2:	<i>Zoek de concepten op via de AAT SPARQL API</i>	15
STAP 3:	<i>Download de concepten in een CSV-bestand</i>	16
4.2	<b>Maak een kopie van Adlib-termen in een CSV-bestand</b>	<b>18</b>
STAP 1:	<i>Zoek termen in de Adlib Thesaurus</i>	18
STAP 2:	<i>Plaats Adlib-termen in een Excel-bestand</i>	18
4.3	<b>Koppel Adlib-termen aan AAT-termen</b>	<b>19</b>
	<b>Vorbereiding</b>	<b>19</b>
STAP 1:	<i>Adlib- en AAT-termen samenbrengen in een Excel-bestand</i>	19
STAP 2:	<i>Adlib- en AAT-termen vergelijken met Fuzzy Lookup</i>	21
STAP 3:	<i>Matches evalueren (en opschoning voorbereiden)</i>	24
4.4	<b>Termen en URI's importeren in Adlib</b>	<b>26</b>
	<b>URI's importeren met een Adlib-Tagged</b>	<b>26</b>
<b>5</b>	<b>Literatuur</b>	<b>33</b>
<b>6</b>	<b>Colofon</b>	<b>34</b>
	<b>Over de projecten</b>	<b>35</b>
<b>ADDENDUM 1</b>	<b>KOPPELEN MET CULTUURLINK</b>	<b>36</b>
	<b>Verskossen van je thesaurus</b>	<b>36</b>
STAP 1:	<i>Installatie</i>	37
STAP 2:	<i>RDF/SKOS maken</i>	37
STAP 3:	<i>Export</i>	37
	<b>Op weg met CultuurLINK</b>	<b>38</b>
STAP 1:	<i>De externe thesaurus kiezen</i>	38
STAP 2:	<i>Linken: string match</i>	38
STAP 3:	<i>Eerste resultaten bekijken</i>	39
STAP 4:	<i>Je strategie verder uitbouwen</i>	39
STAP 5:	<i>Exporteren en gebruiken</i>	42

<b>ADDENDUM 2</b>	<b>URI'S IMPORTEREN IN ADLIB DESIGNER</b>	<b>44</b>
	URI's importeren met een Adlib-Tagged in Adlib Designer	44
	URI's importeren met een CSV in Adlib Designer	46
STAP 1:	<i>Vorbereiding</i>	46
STAP 2:	<i>De import configureren</i>	47
STAP 3:	<i>Import &amp; nazorg</i>	49
<b>ADDENDUM 3</b>	<b>EENVOUDIG OPSCHONEN</b>	<b>50</b>
	Opschonen voorbereiden	50
	Veel voorkomende 'vervuiling'	51
	Stappenplan opschonen	53
STAP 1:	<i>Zorg voor een referentielijst</i>	53
STAP 2:	<i>Analyseer de vervuiling</i>	53
STAP 3:	<i>Maak je eigen stappenplan en bedenk nieuwe regels</i>	54
STAP 4:	<i>Vul de opgeschoonde termen in</i>	54
STAP 5:	<i>Voer de opschoning door in je collectiebeheersysteem</i>	54
<b>ADDENDUM 4</b>	<b>OPSCHONING VERWERKEN IN ADLIB</b>	<b>55</b>
	Handmatig opschonen: Voorkeurstermen en Niet-voorkeurstermen	55
	Hoe het werkt achter de schermen van Adlib: een voorbeeld	56
	Use en Used for	56
	Het resultaat van deze handeling: Van Gogh	56
	Stappenplan handmatig opschonen	57

# 1 VOORWOORD

KAAT KEBO

Het ModeMuseum Antwerpen (MoMu) is initiatiefnemer van het erfgoedproject 'Visuele Thesaurus voor Mode & Kostuums', een project dat ondersteund wordt door de Vlaamse Gemeenschap. Zoals de meeste musea, investeert MoMu een substantieel deel van haar tijd en middelen aan de registratie en ontsluiting van de museumcollectie. In een eerste fase gebeurde dit op papieren fiches en sinds het einde van de jaren negentig worden deze gegevens ontsloten in een databank. Het gebruik van een gecontroleerde terminologie is een eerste en belangrijk instrument om deze data consistent te houden.

De laatste jaren komt er als maar meer nadruk op het ontsluiten van collecties in een netwerkband. Initiatieven als Europeana, Europeana Fashion en Modemuze zijn hier niet vreemd aan. Door deze gezamenlijke ontsluiting groeit de noodzaak om gegevensinvoer tussen musea onderling af te stemmen en volstaan geïsoleerde termenlijsten niet langer om je collectie adequaat te beschrijven en ontsluiten. Een project als 'Visuele Thesaurus voor Mode & Kostuums' is exemplarisch voor deze trend. Het initiatief poogt op pragmatische wijze de gecontroleerde termenlijsten van een tiental Nederlandse en Vlaamse musea af te stemmen aan de hand van een gedeelde thesaurus. Het project liep van het najaar 2015 tot de zomer van 2017 en met het aflopen ervan is het belangrijk te focussen op de praktische implementatie van een dergelijk instrument. Het MoMu is dan ook erg blij met deze handleiding, omdat dit document concreet ingaat op de vraag hoe je een thesaurus inzet om meer consistente en gelinkte data te krijgen.

Ik zou graag enkele betrokken instellingen en personen bedanken die dit project mogelijk maakten. In de eerste plaats dank ik de betrokken partnermusea om mee in dit traject te stappen en hun terminologieën ter beschikking te stellen. Verder zou ik graag de leden van de projectstuurgroep willen danken voor hun engagement, ondersteuning en expertise, met name Wilbert Helmus (Netwerk Digitaal Erfgoed), Joop Vanderheiden (Rijksdienst Cultureel Erfgoed), Eva Verdoodt en Matthias Priem (Vlaams Instituut voor Archivering) en Mila Ernst (Modemuze). Tot slot zou ik graag mijn eigen medewerkers Dieter Suls en vooral Ykje Wildenborg bedanken die instonden voor het bedenken en uitvoeren van dit mooie project.

*Kaat Debo, directeur ModeMuseum Antwerpen (MoMu)*

## VOORWOORD

ANNETTE GAALMAN EN WILBERT HELMUS

Een van de grootste uitdagingen van digitalisering ligt volgens het Netwerk Digitaal Erfgoed in het beter verbinden van collecties voor een efficiënt gebruik. Digitaal Erfgoed Bruikbaar richt zich specifiek op de mogelijkheden om collecties gezamenlijk online beschikbaar te stellen, door data te verbinden en te verrijken met behulp van terminologiebronnen, thematische ontsluiting en gedeelde tools daarvoor te ontwikkelen.

Het is duidelijk dat dit alleen kan door meer samenwerking tussen erfgoedinstellingen, ook over landsgrenzen heen. En dit is precies de reden waarom het project Visuele Thesaurus voor Mode en Kostuums zo goed past in de Nationale Strategie Digitaal Erfgoed.

Het Netwerk Digitaal Erfgoed complimenteert de initiatiefnemers en partners die met elkaar hebben samengewerkt en die elkaars expertise hebben samengevoegd in een hele praktische en heldere handleiding. We hopen dat dit handboek de drempel kan verlagen en dat veel meer erfgoedinstellingen de stap zetten om hun collecties te verbinden aan gestandaardiseerde (inter)nationale terminologiebronnen.

De gebruiker staat centraal in de Nationale Strategie Digitaal Erfgoed. Wij zijn daarom verheugd dat erfgoedprofessionals met hulp van deze handleiding de juiste 'tools' en technieken aangereikt krijgen om (data)collecties zo rijk mogelijk open te stellen voor eindgebruikers.

Het Netwerk Digitaal Erfgoed spreekt grote dank uit aan de initiatiefnemers, de projectpartners, de Vlaamse Overheid en alle betrokkenen die een aandeel hebben gehad in de realisering van het project Visuele Thesaurus Mode en Kostuums. Het vormt een mooie illustratie van de ambitie om slimme verbindingen tussen collecties te maken. Het publiek kan ons prachtige erfgoed, in dit geval mode en kostuums, in een veel rijkere context raadplegen, beleven en hergebruiken. En daar ligt de maatschappelijke winst.

*Annette Gaalman, voorzitter stuurgroep Digitaal Erfgoed Bruikbaar*  
*Wilbert Helmus, domeinmanager Digitaal Erfgoed Bruikbaar*

## 2 INTRODUCTIE

Dit rapport is geschreven voor erfgoedbeherende instellingen in Nederland en Vlaanderen. Met oog op de toekomst streven wij ernaar ondersteuning te bieden in het **gebruik van terminologieën en andere vrij bruikbare bronnen**, zoals namenlijsten en locaties. Het is tegelijk een document voor zogenaamde ‘non-techy people’ én technisch onderlegde medewerkers.<sup>1</sup>

Medewerkers van instellingen hebben tijdens onze projecten verzocht om heldere richtlijnen voor het **opschonen en koppelen van metadata**. Netwerk Oorlogsbronnen, Netwerk Zuiderzeecollectie en Visuele Thesaurus voor Mode & Kostuums hebben alle tot doel **bewustzijn te creëren over de effecten van gedeelde terminologieën**.<sup>2</sup> Doelstelling: betere zoekresultaten binnen metadata op het web. Op verschillende manieren stimuleren deze projecten het gebruiken van links naar deze terminologieën in de objectregistratie van de partnermusea. Deze handleiding beoogt meer instellingen die kans te bieden.

Het was **geenszins eenvoudig** om tot een helder gestructureerd document te komen dat elke hand dient. Met het **diverse palet aan behoeften** bij de (partner)musea, collectiebeheersystemen en aanwezige technische kennis is daarom voor een specifieke opzet gekozen.

Dit rapport beschrijft één voorbeeld-linkproces uitgebreid en is verder opgebouwd uit **addenda**, die alternatieve werkwijzen het toneel opschuiven. Met deze **kleine en grote bouwstenen** kunnen instellingen zelf bepalen hoe ze te werk gaan. Collega's kunnen in **Hoofdstuk 4** direct aan de slag met het linken van hun Adlib-termen aan de Art & Architecture Thesaurus (AAT), geholpen door Fuzzy Lookup. In de **addenda** kan men informatie terugvinden over het ideale opschoonproces, veel geziene ‘datavervuiling’ en het gebruik van meer geavanceerde tools als CultuurLINK en Adlib Designer. De meeste methodes in deze gids zijn **voor elk systeem van toepassing**, zo lang er een data-export mogelijk is in CSV of Excel. De algemene introductie beschrijft het waarom van koppelen en plaatst dit in een Vlaams-Nederlandse context.

Een laatste woord van dank voor de gulle ondersteuning van de **Vlaamse Overheid**, zonder welke dit rapport niet zou zijn verschenen, en van **Netwerk Digitaal Erfgoed**, dat het voorzag van vormgeving.

Bert Lemmens van PACKED vzw vormde een onmisbare schakel in het redactieproces.

*Shannon van Muijden, data manager collectie, Zuiderzeemuseum*

*shannon.van.muijden@zuiderzeemuseum.nl*

*Inge van Stokkom, data curator, Netwerk Oorlogsbronnen*

*i.van.stokkom@niod.knaw.nl*

*Ykje Wildenborg, projectmedewerker VTMK, ModeMuseum Antwerpen*

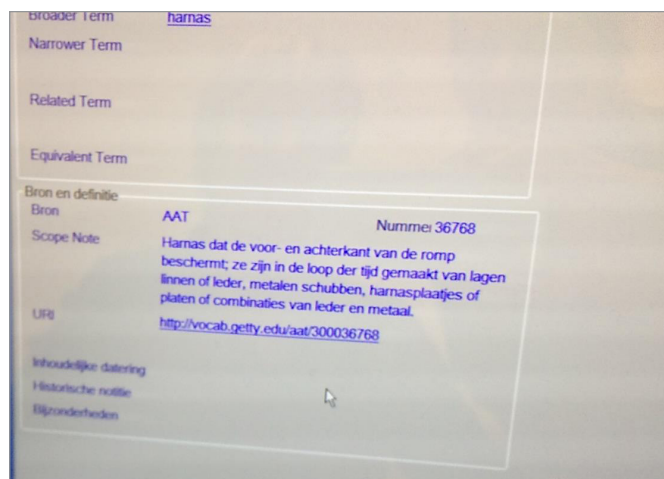
*ykje.wildenborg@momu.be*

### 3 LINKEN: WAAROM EN HOE?

Collectiebeheerders staan voor de uitdaging om hun data *machine readable* te maken: data zó indelen, dat zoeksoftware ze beter begrijpt. Hierdoor krijgt de scholier, verzamelaar of ontwerper van de toekomst op zoek naar **zonnebloemen** met één klik toegang tot alle relevante objecten uit het Van Gogh, Museum Plantin Moretus én Naturalis – zonder speurwerk op al hun afzonderlijke webpagina's. De meest efficiënte methode om dit te bereiken is **het linken van informatie in gecontroleerde velden aan de concepten van gedeelde terminologiebronnen**.<sup>3</sup> Bijvoorbeeld objectnamen, materialen of personen.

De links tussen deze termen en de terminologiebronnen documenteer je in je collectiebeheersysteem. Wat betekent dit voor de praktijk? Naast een term of naam (een set letters ofwel *text string*), registreer je in een ander veld van je thesaurus een **Uniform Resource Identifier** (URI). Dit is een **identifier** in de vorm van een webadres dat het concept dat je met dat woord aanduidt uniek identificeert op het web<sup>4</sup>. Via dat webadres kun je meteen ook meer info vinden over het betreffende concept. In de terminologiebron staan een omschrijving, alternatieve namen, spellingsvarianten, hoe je het concept benoemt in een andere taal. Dus je lost in één keer het probleem op van meertaligheid én van verschil in schrijfwijzen binnen een taal. Zo staat <https://rkd.nl/en/explore/artists/66219> bijvoorbeeld voor de schilder **Rembrandt**. Zoekmachines beschikken via dit webadres over een hoop informatie waarmee ze een zoekopdracht ze goed mogelijk kunnen beantwoorden. **Projecten** zoals Modemuze, Netwerk Oorlogsbronnen en Netwerk Zuiderzeecollectie of zelfs Europeana, die **verschillende collecties samen doorzoekbaar** maken, varen er wel bij als deze webadressen beschikbaar komen in aangeleverde metadata.

Figuur 1:  
Een binnengehaalde  
URI in de Adlib-  
database van het  
Amsterdam Museum.



Musea hebben er belang bij om **zelf controle te houden** over die links. Zo kunnen ze de juistheid van informatie zo goed mogelijk waarborgen. Tot nu toe gebeurt het linken tijdens het aggregatieproces: net voordat ze samenkomen in een portaal, vaak door medewerkers van dat portaal, en *niet* van de instelling zelf. Dit is een **onduurzame werkwijze** waarbij de instelling weinig zeggenschap heeft.

In een **ideale situatie** kunnen registratoren concepten (en de bijhorende webadressen) selecteren en invoeren tijdens de registratie en zoekt het collectiebeheersysteem die concepten via het web in publiek toegankelijke terminologiebronnen zoals AAT, RKDartists of Geonames. In de praktijk bestaan er nog weinig collectiebeheersystemen die dit kunnen; **in enkele gevallen is dit gebeurd na specifieke vraag van de instelling**.<sup>5</sup> Met deze nieuwe functie is het ook voor die



instellingen echter niet mogelijk om bestaande collectieregistratie te verrijken met links – dit vergemakkelijkt alleen het registreren van **nieuwe objecten**. Deze handleiding biedt daarom een werkwijze om toch URI's te documenteren in je collectiebeheersysteem.

★

*In het Zuiderzeemuseum bevindt zich een wildschietter (zie afbeelding cover). Een museum in Friesland heeft ook een wildschietter. Die term heeft een beschrijving en synoniemen in de erfgoedthesaurus en heeft als URI <http://data.cultureelerfgoed.nl/semnet/581cee4b-9ee1-476e-b8e7-15e4f43ded44>. Maar zij noemen dit een wyldsjitte, en hebben deze term ook als zodanig in hun collectiebeheersysteem opgenomen. Op het moment dat deze instelling de bovengenoemde URI naar wildschietter bij hun term wyldsjitte toevoegt, is deze online terug te vinden onder de noemer wildschietter. Als er in een gezamenlijke online presentatie op de term wildschietter wordt gezocht zal zowel het object uit het Zuiderzeemuseum als het object uit het Friese museum worden getoond.*

### 3.1 Omgaan met terminologieën in (inter)nationaal verband

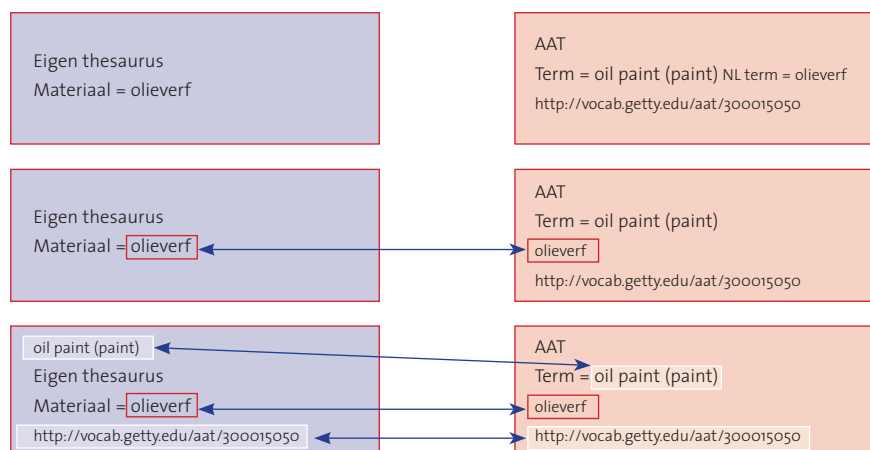
Nederlandse en Vlaamse overheden vinden het belangrijk dat erfgoedcollecties gebruik gaan maken van **gedeelde terminologiebronnen**. In Vlaanderen werken organisaties als **PACKED vzw**, expertisecentrum voor digitaal erfgoed en **VIAA**, Vlaams Instituut voor Archivering samen met het werkveld om een beweging in gang te brengen. De platformen **depotwijzer.be** en **CEST** zijn daarbij dé referentie voor erfgoedmedewerkers op zoek naar richtlijnen en standaarden. In Nederland vind je standaarden en richtlijnen terug op de website van DEN (Digitaal Erfgoed Nederland).<sup>6</sup> Daarnaast werkt het **Netwerk Digitaal Erfgoed** (NDE) aan een nationale strategie. NDE is een samenwerkingsverband dat een stelsel van **landelijke voorzieningen en diensten** ontwikkelt voor het verbeteren van de zichtbaarheid, bruikbaarheid en houdbaarheid van digitaal erfgoed. Het hoofddoel van deelproject **Digitaal Erfgoed Bruikbaar** richt zich op het **beter vindbaar maken** van de (versnipperde) digitale erfgoedinformatie.<sup>7</sup> Hierbij wordt ook de koppeling gemaakt naar het Vlaamse erfgoedwerkveld.

#### *Het beheer en gebruik van de bronnen*

De werkwijze die deze initiatieven propageren past binnen de gedeelde ideeën over het 'internet van de toekomst'. Als we in de toekomst van het web blikken lijken de begrippen 'Linked data' en 'semantisch web' érg relevant; de glazen bol toont dat informatie op het internet beter is gestructureerd en aan elkaar verbonden. Dat scenario omvat – naast het verbinden van onze data – ook het op hoger niveau met *elkaar* koppelen van (inter)nationale terminologiebronnen. In de toekomst hoeven instellingen daarom maar één webadres op te nemen om te linken met alle andere terminologiebronnen. Terminologiebronnen worden centraal beheerd door externe organisaties, zoals de Erfgoedthesaurus van de Rijksdienst voor het Cultureel Erfgoed, de AAT van het Getty Research Institute of de namenlijst Virtual International Authorities File (VIAF) van OCLC, het Online Computer Library Center.<sup>8</sup> En nee, **dan hoeven musea zich niet meer te bezighouden** met hiërarchieën binnen thesauri, scope notes bij termen, of aanvullende informatie bij kunstenaars, zoals geboorte- en sterfdata. Die informatie wordt **op een centrale plaats beheerd** door een organisatie die daar de nodige mensen en middelen voor ter beschikking heeft. Het beheer wordt hierdoor veel efficiënter: wijzigen hoeft slechts op die plek. In het collectiebeheersysteem documenteert de registrator in principe enkel het webadres van één concept, kunstenaar of geografische locatie.

Maar het kan in de actuele situatie nog nuttig zijn om **extra informatie** (scope notes, geboortedata) over te nemen in het eigen collectiebeheersysteem. Het Amsterdamse Rijksmuseum heeft dit enkele jaren geleden gedaan met de Engelse vertalingen van hun materiaaltermen, die binnengehaald werden uit de AAT door middel van de term-URI's. Via het collectiebeheersysteem worden die termen nu gebruikt in de Engelse versie van de eigen website. Maar dit is een scenario dat in het nu thuis hoort, omdat we verwachten dat die informatie in de toekomst door een zoekstelsel opgehaald zal worden uit de externe terminologiebron.

*Figuur 2:*  
In de thesaurus van het Rijksmuseum staan Engelse materiaaltermen die zijn binnengehaald vanuit de Art & Architecture Thesaurus (AAT).



### 3.2 Welke bronnen?

De bedoeling is dat je de termen uit je collectiebeheersysteem bij voorkeur matcht met **exact hetzelfde begrip** uit de terminologiebron van jouw keuze. Als er geen geschikte term in de terminologiebron beschikbaar is, wordt soms het webadres van een algemenere term toegevoegd.<sup>9</sup>

Om te **verifiëren** of de term uit je collectiebeheersysteem **exact overeenkomt** met die uit de terminologiebron, controleer je best ook de scope note bij die term. **Scope notes** zijn **geen definities**, maar beschrijven een begrip binnen de context van de thesaurus ('op welke manier wordt de term gebruikt in de thesaurus'). 'Transporten' is in de context van de Tweede Wereldoorlog bijvoorbeeld iets heel anders dan in normaal taalgebruik. Daarom is de bedoelde betekenis in de thesaurus van belang, om (pijnlijke) verkeerde toekenningen te voorkomen. En tijdens het link-proces is een controle van scope notes bij mogelijk dubbelzinnige begrippen dus aan te raden. Bedenk aan de hand van de aard van de collectie en het veld dat je wilt linken welke thesaurus erbij beschikbaar is.<sup>10</sup> **Aarzel niet** om musea met vergelijkbare collecties te vragen welke bronnen zij aanwenden.

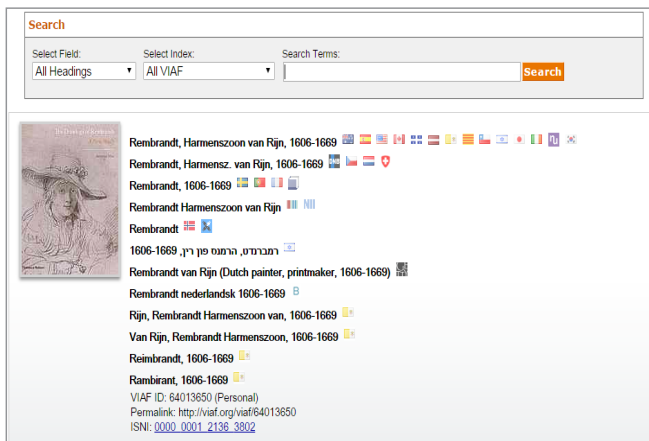
Het is niet uitzonderlijk dat men **meerdere thesauri** gebruikt, ook binnen één veld. Sommige concepten zijn simpelweg nog niet aanwezig in een bepaalde thesaurus. Soms blijft het beter zo **dicht mogelijk bij de bron** te blijven. De term 'pijlstaartrog' kan bijvoorbeeld beter naar het Soortenregister linken, terwijl de term ook aanwezig is in de Erfgoedthesaurus, met een link naar het Soortenregister.

Uiteraard **blijft het mogelijk om termen niet te linken**, wanneer de term of scope note van een term in een terminologiebron niet inhoudt wat er intern bedoeld wordt. In dat geval wordt ook

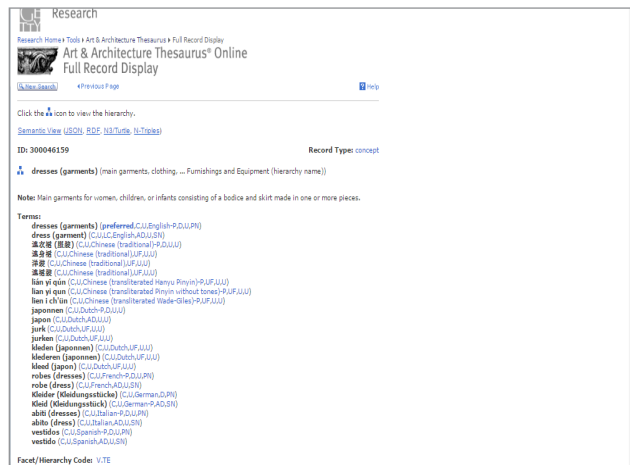
geen link gelegd met die terminologiebron en is het handig om jouw term eventueel daarin te laten opnemen.<sup>11</sup> **Zorg er in ieder geval voor dat je die niet-linkbare termen steeds opnieuw gebruikt, en geen spellingsvarianten introduceert in je collectiebeheersysteem.**

Hieronder vind je alvast een overzicht van de **meest gebruikte terminologiebronnen** in de museumsector.<sup>12</sup>

TITEL	ONDERWERP	HUMAN READABLE VERSIE	MACHINE READABLE VERSIE
RKD Artists	Personen (kunstenaars)	<a href="https://rkd.nl/nl/explore/artists/">https://rkd.nl/nl/explore/artists/</a>	<a href="http://opendata.rkd.nl/opensearch/artists/eac-cpf?q">http://opendata.rkd.nl/opensearch/artists/eac-cpf?q</a> (voor meer informatie neem contact op met het RKD)
Gemeenschappelijke Thesaurus voor Audiovisuele Archieven (GTAA)	Audiovisueel cultureel erfgoed	<a href="http://gtaa.beeldengeluid.nl/">http://gtaa.beeldengeluid.nl/</a>	<a href="http://gtaa.beeldengeluid.nl/">http://gtaa.beeldengeluid.nl/</a> zie 'Waar kan ik de GTAA downloaden?'
Erfgoedthesaurus	Breed referentienetwerk voor erfgoed	<a href="https://data.culturelerfgoed.nl/PoolParty/wiki/term/id/cht">https://data.culturelerfgoed.nl/PoolParty/wiki/term/id/cht</a>	<a href="https://digitaal erfgoed.poolparty.biz/erfgoedthesaurus.html">https://digitaal erfgoed.poolparty.biz/erfgoedthesaurus.html</a>
Soortenregister	Wetenschappelijke soortnamen van Nederlandse dieren en planten	<a href="http://www.nederlandsesoorten.nl/">http://www.nederlandsesoorten.nl/</a>	<a href="http://docs.biodiversitydata.nl/en/latest">http://docs.biodiversitydata.nl/en/latest</a> of <a href="http://www.gbif.org/dataset/4dd32523-a3a3-43b7-84df-4cdao2f15cf7">http://www.gbif.org/dataset/4dd32523-a3a3-43b7-84df-4cdao2f15cf7</a> (downloadbare XML)
Iconclass	Onderwerpsclassificatie	<a href="http://www.iconclass.nl/home">http://www.iconclass.nl/home</a>	<a href="http://www.iconclass.org/help/lod">http://www.iconclass.org/help/lod</a>
WO2-thesaurus	Tweede Wereldoorlog in Nederland	<a href="https://data.niod.nl/WO2_Thesaurus.html">https://data.niod.nl/WO2_Thesaurus.html</a>	<a href="https://data.niod.nl/WO2_Thesaurus.html">https://data.niod.nl/WO2_Thesaurus.html</a>
Erfgeo	Verzameling terminologiebronnen voor geografische namen, inclusief historische	<a href="http://erfgeo.nl">http://erfgeo.nl</a>	<a href="http://erfgeo.nl/tools/api.html">http://erfgeo.nl/tools/api.html</a>
Biografisch portaal	Persoonsnamen	<a href="http://www.biografischportaal.nl">www.biografischportaal.nl</a>	<a href="http://www.biografischportaal.nl/personen/biodes">http://www.biografischportaal.nl/personen/biodes</a> en <a href="http://www.biografischportaal.nl/personen/json">http://www.biografischportaal.nl/personen/json</a>
Visuele Thesaurus voor Mode & Kostuums (VTMK)	Objectnamen voor mode en kostuums	<a href="https://vtmk.data.momu.be/id.html">https://vtmk.data.momu.be/id.html</a>	<a href="https://vtmk.data.momu.be/PoolParty/sparql/id">https://vtmk.data.momu.be/PoolParty/sparql/id</a>
Art & Architecture Thesaurus	Kunst, architectuur, toegepaste kunst (o.a. objectnamen, materialen en technieken)	<a href="http://www.getty.edu/research/tools/vocabularies/aat/">http://www.getty.edu/research/tools/vocabularies/aat/</a>	<a href="http://vocab.getty.edu/">http://vocab.getty.edu/</a> en <a href="http://vocab.getty.edu/sparql">http://vocab.getty.edu/sparql</a>
Musical Instruments Museum Online (MIMO)	Muziekinstrumenten	<a href="http://www.mimo-international.com/MIMO/">http://www.mimo-international.com/MIMO/</a>	<a href="http://data.mimo-db.eu/sparql">http://data.mimo-db.eu/sparql</a>
Virtual International Authorities File (VIAF)	Persoonsnamen (m.n. auteurs)	<a href="https://viaf.org/">https://viaf.org/</a>	<a href="http://viaf.org/viaf/data/">http://viaf.org/viaf/data/</a>
Union List of Artist Names (ULAN)	Persoonsnamen	<a href="http://www.getty.edu/research/tools/vocabularies/ulan/index.html">http://www.getty.edu/research/tools/vocabularies/ulan/index.html</a>	<a href="http://vocab.getty.edu/">http://vocab.getty.edu/</a> en <a href="http://vocab.getty.edu/sparql">http://vocab.getty.edu/sparql</a>
Geonames	Geografische namen	<a href="http://www.geonames.org/">http://www.geonames.org/</a>	<a href="http://www.geonames.org/ontology/documentation.html">http://www.geonames.org/ontology/documentation.html</a>
Thesaurus of Geographic Names (TGN)	Geografische namen	<a href="http://www.getty.edu/research/tools/vocabularies/tgn/index.html">http://www.getty.edu/research/tools/vocabularies/tgn/index.html</a>	<a href="http://vocab.getty.edu/">http://vocab.getty.edu/</a> en <a href="http://vocab.getty.edu/sparql">http://vocab.getty.edu/sparql</a>
Wikidata	Encyclopedie	<a href="https://www.wikidata.org/wiki/Wikidata:Main_Page">https://www.wikidata.org/wiki/Wikidata:Main_Page</a>	<a href="https://query.wikidata.org/">https://query.wikidata.org/</a>



Figuur 3:  
Een record uit VIAF (personen)



Figuur 4:  
Een record uit de AAT  
(cultuurgerelateerde concepten)

### 3.3 Manieren van aanpak

Om de handleiding **helder en beknopt** te houden, gebruiken we in de kerncasus een **concreet collectiebeheersysteem**, Adlib, en een **concrete terminologiebron**, de AAT. Hierdoor kunnen we het basisproces zo concreet en eenvoudig mogelijk uitleggen. Maar vele wegen leiden naar Rome - dat geldt niet in de laatste plaats voor het link- en opschonenproces. **Probeer voordat je aan het werk gaat, de handleiding en de addenda door te kijken om voor jezelf een methode te bepalen.** Hoe je het best te werk gaat hangt onder meer af van met welke software je bekend bent en of je data niet of juist erg vervuild is. In het keuzeprocess hierbij enkele **handreikingen**. Wij adviseren de instellingen die nu **nog niet willen beginnen met linken** om in elk geval te starten met het opschonen.

#### *Instellings-afhankelijke werkwijze*

- **Opgeschoonde data?** Is je data (behoorlijk) opgeschoond, ga dan linken met Fuzzy Lookup of CultuurLINK. Voor beginners leggen we in de **Casus in Hoofdstuk 4** stap-voor-stap uit hoe je de URI's, horend bij AAT-concepten, kunnen binnenhalen in je Adlib-collectiebeheersysteem. Wie liever een stap verder neemt, kan het linkproces aangaan in de tool CultuurLINK, in **Addendum 1**.
- **Vervuilde data?** Het linken kan een eenvoudige eerste stap voor het opschonen van je data. Je gaat dan aan de slag met het opschonen van de termen, die niet gelinkt konden worden. Dit kan extra handig zijn als je bijvoorbeeld wel een **referentielijst** hebt, maar je data zelf niet meer daaraan voldoen. Wil je zo werken, ga dan ook naar de **Casus in Hoofdstuk 4**.
- Wil je echter liever beginnen met opschonen kijk dan eerst naar **Addenda 3 en 4**.

#### *Voorwaarden om aan de slag te kunnen*

- Een **export- en bij voorkeur import-functie** in je collectiebeheersysteem
- **Software om te linken**; deze is vrij beschikbaar
- Een **veld** om de URI in te documenteren

#### *Introductie van enkele tools*

- In de **Casus in Hoofdstuk 4** maken we gebruik van Fuzzy Lookup: een add-in voor Excel. Deze add-in vergelijkt termen **tussen** twee kolommen. Een meer geavanceerde tool voor



gevorderden is **CultuurLINK**, zie **daarvoor Addendum 1**. Deze biedt de mogelijkheid om de vergelijking van termenlijsten zelf vorm te geven, maar het is daarbij handig als je niet terugdeinst voor het maken van een versie van jouw termenlijst in SKOS.

- Voor uitgebreid opschonen is **OpenRefine** een andere nuttige tool. OpenRefine lijkt qua uiterlijk op Excel, maar heeft uitgebreide functies om data te manipuleren en op te schonen.<sup>12</sup> De **'Cluster & edit'-functie** is buitengewoon nuttig om vergelijkbare termen en namen binnen erg vervuilde data op te sporen. Het programma biedt voor gevorderden ook de mogelijkheid tot linken. Wij verwijzen voor informatie in de literatuurlijst naar reeds bestaande, uitgebreide documentatie.

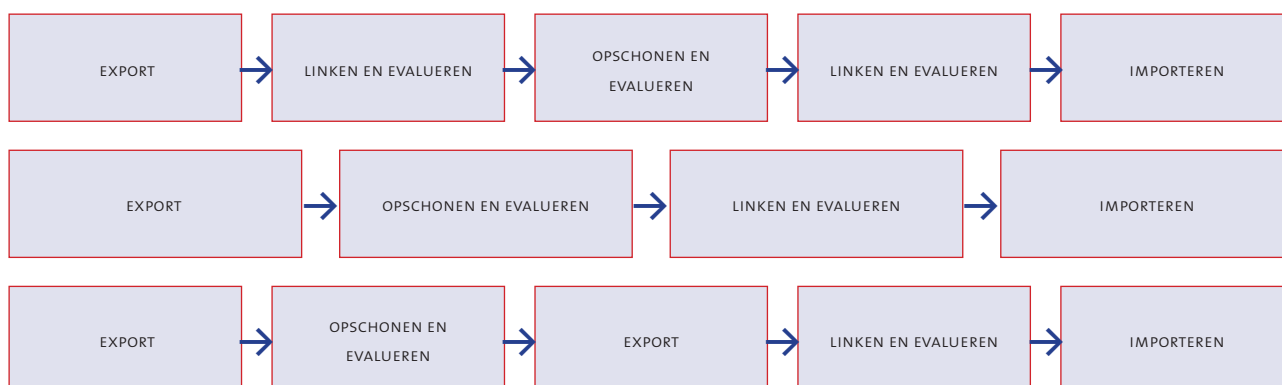
#### Importeren en velden aanmaken

- **Velden voor URI's** kun je in **Adlib Designer** aanmaken of je kunt hiervoor een bestaand Adlib-veld gebruiken. Een **nieuw veld** aanmaken is de meest zuivere oplossing. Ben je hiermee niet bekend, gebruik dan een **bestaand veld**. Zorg daarbij dat je zeker weet dat dit veld leeg is; dat er bijvoorbeeld geen collega's zijn die dit veld gebruiken voor andere informatie. In dit geval is het handigst om de velden 'Bron' en 'Nummer' samen te gebruiken, zoals in onderstaand voorbeeld is gedaan met een Wikidata-URI. Het deel voor de laatste slash staat dus bij 'Bron' en de rest, de *identificer* in het veld 'Nummer'.

Bron en definitie		
Bron	<a href="https://www.wikidata.org/wiki/">https://www.wikidata.org/wiki/</a>	Nummer Q3674992
URL	<a href="https://www.wikidata.org/wiki/Q3674992">https://www.wikidata.org/wiki/Q3674992</a>	
Scope Note		

- Als je niet, zoals in het basisproces met een Adlib-tagged wilt importeren, kun je met Adlib Designer werken. Daarvoor gebruik je ook een **Adlib-Tagged** of een **CSV**. **Als je dit wilt doen, zie dan Addendum 2.**
- Als importeren van data niet lukt, omdat het proces te ingewikkeld blijkt voor jouw situatie, kun je natuurlijk gewoon **knippen en plakken**, bijvoorbeeld vanuit de website van de AAT.<sup>14</sup> Deze laatste werkwijze zal echter niet meer worden genoemd in de handleiding.

	FUZZY LOOKUP	CULTUURLINK	OPENREFINE
doelpubliek	beginners	gevorderden	beginners (opschonen), gevorderden
input	Termen in CSV	Termen in RDF/Turtle (SKOS)	Termen in CSV, XML, JSON, etc
tool	Excel-plugin	Webtool	Java-applicatie
terminologiebronnen	Via export	Aanwezig, via web	Via web en export



Figuur 5: Het ideale scenario voor het linkproces kan per instelling verschillen.

## 4 CASUS: AAT-LINKS DOCUMENTEREN IN ADLIB

Deze casus bestaat uit **vier onderdelen**: een **kopie van de AAT** maken, een kopie maken van je **Adlib-termen**, het **koppelen** van de termen en het **importeren van de URI's**.

### 4.1 Maak een kopie van de AAT in een CSV-bestand

De eerste stap is **een lijst maken van de AAT-concepten** waarmee je de termen in je collectiebeheersysteem wilt linken. Dat doe je door een **kopie** te maken van de AAT, of althans van dat deeltje van de AAT dat je nodig hebt. Eventueel roep je bij dit proces **hulp** in van een **technische collega**.




Er bestaan ook groepen waar je met vragen terecht kan, zoals in het geval van de AAT bijvoorbeeld <https://groups.google.com/forum/#!forum/gettyvocablod>. AAT-Nederlandstalig is bereikbaar via <http://www.aat-ned.nl> en [aat@rkd.nl](mailto:aat@rkd.nl).

#### STAP 1: Zoek de overkoepelende term voor de concepten

Hier geven we een voorbeeld van een query, die URI's, Nederlandse Pref en AltLabels en de scopenote teruggeeft van een deel van de AAT.

- Ga naar <http://www.getty.edu/research/tools/vocabularies/aat/>
- Zoek de **broader term** van de objectnamen die jij nodig hebt: **zoek hiervoor een term die belangrijk is in de te linken deelcollectie**. In ons voorbeeld, waarbij we zo veel mogelijk soorten kleding willen hebben, is dat 'broek'.
- Klik op het  icoontje voor de term om te kijken waar het concept past in de structuur van de AAT. Je ziet dan de hiërarchie van de term.
- **Zoek naar het concept waar de termen die jij nodig hebt grotendeels onder vallen**. In de hiërarchie van AAT-termen zie je dat 'trousers' een nauwer concept is van 'costume (hierarchy name)'. Onder dat concept vind je ook 'costume accessories', 'costume components', 'costume by function' of 'costume by wearer' terug. Costume (hierarchy name) lijkt dus een goeie optie voor het concept dat alle AAT-termen over kleding omvat.



The screenshot shows a record for 'trousers' with the following details:

- ID:** 300209935
- Record Type:** concept
- Definition:** trousers (<main garments for the lower body>, main garments, ... Furnishings and Equipment (hierarchy name))
- Note:** Exposed bifurcated garments which extend from waist or hips to the ankle or sometimes to the knee or just below.
- Terms:**
  - trousers (preferred, C,U,LC,English-P,D,U,PN)
  - pants (trousers) (C,U,English,UF,U,N)
  - broeken (C,U,Dutch-P,D,U,U)
  - broek (C,U,Dutch,AD,U,U)
  - Hosen (C,U,German,D,U,PN)
  - Hose (C,U,German-P,AD,SN)
  - pantalón (C,U,Spanish-P,D,U,SN)
- Facet/Hierarchy Code:** V.TE
- Hierarchical Position:**
  - Objects Facet
  - .... Furnishings and Equipment (hierarchy name) (G)
  - ..... Costume (hierarchy name) (G)
  - ..... costume (mode of fashion) (G)
  - ..... clothing (G)
  - ..... main garments (G)
  - ..... <main garments for the lower body> (G)
  - ..... trousers (G)

A red arrow points to the 'trousers (G)' entry in the Hierarchical Position list.

- Klik op deze term en noteer het nummer van dat concept. Dit staat achter ID. Hieronder is dat 300209261.

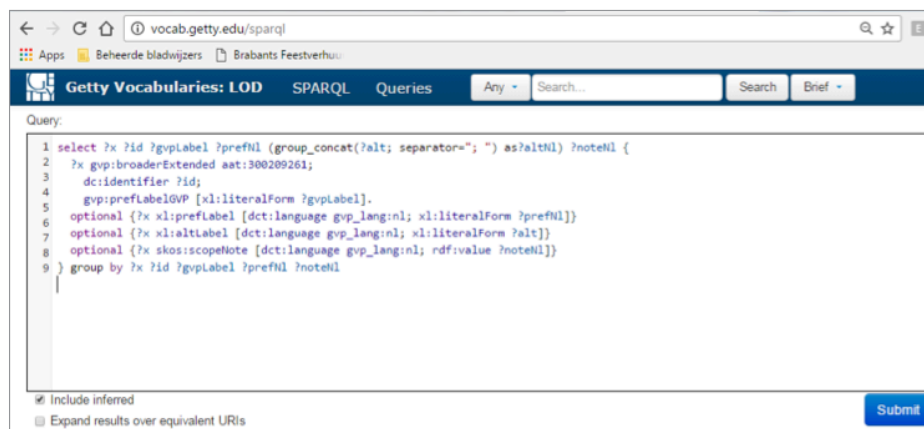


## STAP 2: *Zoek de concepten op via de AAT SPARQL API*

- Ga nu naar de API van AAT waar je alle concepten kan downloaden. **Type het webadres van de AAT-API in je webbrowser: <http://vocab.getty.edu>**
- Klik op **SPARQL** (linksboven) en vul onderstaande query in.

```
select ?x ?id ?gvpLabel ?prefNI (group_concat(?alt; separator="; ") as?altNI) ?noteNI {
  ?x gvp:broaderExtended aat:300209261;
  dc:identifier ?id;
  gvp:prefLabelGVP [xl:literalForm ?gvpLabel].
  optional {?x xl:prefLabel [dct:language gvp_lang:nl; xl:literalForm ?prefNI]}
  optional {?x xl:altLabel [dct:language gvp_lang:nl; xl:literalForm ?alt]}
  optional {?x skos:scopeNote [dct:language gvp_lang:nl; rdf:value ?noteNI]}
} group by ?x ?id ?gvpLabel ?prefNI ?noteNI
```

- Je ziet dan dit scherm:



Figuur 6:  
SPARQL-query invoer  
voor de AAT

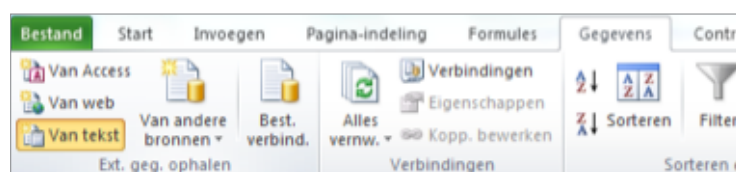
- Het **cijfer achter 'aat:'** in de query (tweede regel) vervang je door het nummer van de **broader term** waar de termen onder vallen die jij nodig hebt. Let goed op, dat je **alleen het nummer** aanpast en geen extra spaties maakt!
- Klik op **Submit**. Een lijst, zoals in Figuur 7, komt tevoorschijn.

x	id	gvpLabel	prefNI	altNI	noteNI
aat:300410869	300410869	fajas trenzadas@en	-		-
aat:300404137	300404137	academic robes@en	academische toga's@nl	academische toga	Tot de grond reikende gewaden met wijde mouwen, gemaakt van zijde of soortgelijk materiaal. Ze worden gedragen tijdens rechtspraak in rechtszalen door juristen, zoals rechters en advocaten. De kleur van de toga's is meestal zwart. In Groot-Brittannië bestaan er kleurvariaties op basis van het soort rechtbank. @nl
aat:300400705	300400705	laurel wreaths@en	lauwerkransen@nl	lauwerkrans	Hoef- of cirkelvormige kransen gevlochten uit de takken en bladeren van de laurier. Werd door Romeinse officiers gedragen tijdens triomftochten als teken van de overwinning. De laurierkrans is nog steeds een symbool voor overwinning en succes. @nl
aat:300209945	300209945	bathrobes@en	badjassen@nl	badjas	Ruimvallende knie- of enkellange kledingstukken, vaak dichtgehouden met een ceintuur, die meestal van warm, absorberend materiaal zijn gemaakt en worden gedragen voor of na het baden of als gemakkelijke kleding binnenshuis. @nl

Figuur 7: Resultaat SPARQL-query. Rechtsboven: mogelijkheid CSV-export.

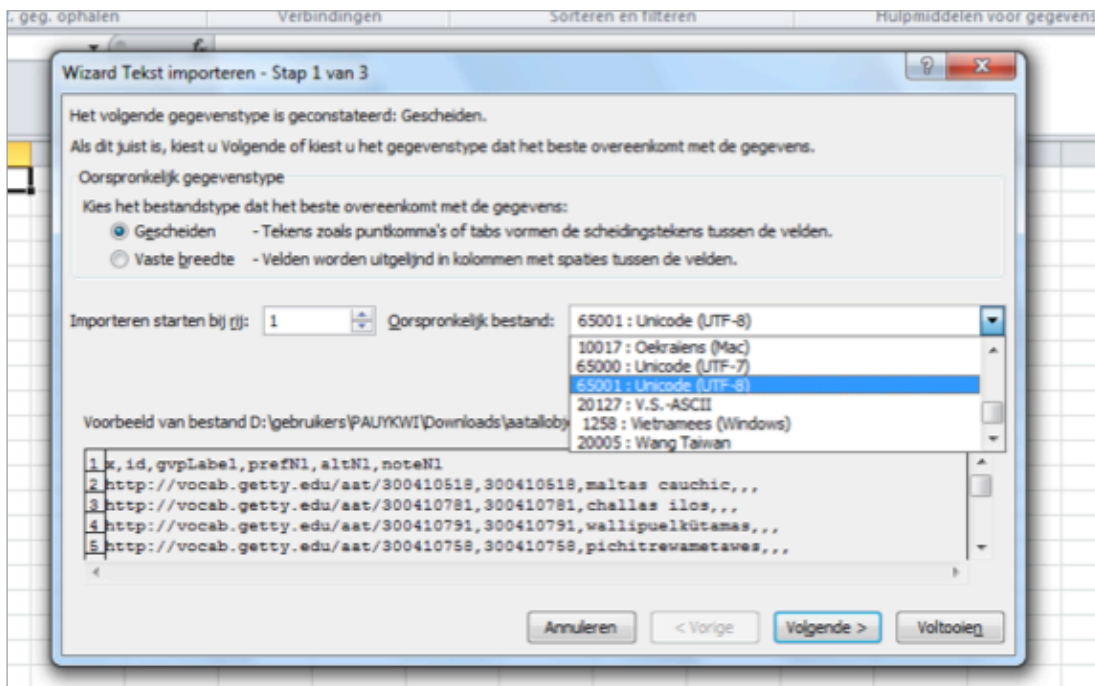
### STAP 3: *Download de concepten in een CSV-bestand*

- Klik vervolgens op **CSV**, rechtsboven in het scherm. Het bestand dat je dan automatisch downloadt, met de termen die je hebt gevraagd, kun je openen in **Excel**.
- Open eerst **Excel**.
- Klik op **Tabblad Gegevens**, dan op **Van tekst**.



- In de wizard kies je bij **Oorspronkelijk bestand 65001: Unicode (UTF-8)**. Hiermee zorg je dat alle diakritische tekens op de juiste manier overkomen.





- Klik **volgende** en kies 'komma' als scheidingsteken.
- Kies **volgende** en bij 'Gevenstype': **Tekst**.
- **OK**
- Nu zie je verschillende gegevens uit de AAT: URI (x), ID (id), Engelse voorkeursterm (gvplabel), Nederlandse voorkeurs- en alternatieve term (prefNI, altNI) en de scope note (noteNI).
- Met de kolom prefNI kun je gaan linken in de volgende stap.

	A	B	C	D	E	F	G	H
1	x	id	gvplabel	prefNI	altNI	noteNI		
1125	http://vocab.getty.edu/aat/300232754	300232754	bushwack boats	bushwack boats	bushwack boat	Te gebruiken voor sloepachtige boten met een brede a		
1126	http://vocab.getty.edu/aat/300232759	300232759	factory ships	fabrieksschepen	fabrieksschip	Vaartuigen die speciaal zijn ontworpen om gevangen w		
1127	http://vocab.getty.edu/aat/300232776	300232776	competition craft	wedstrijdvaartuigen	wedstrijdvaartuig			
1128	http://vocab.getty.edu/aat/300232770	300232770	bassboats	bassboats	bassboot	Te gebruiken voor 4,5 tot 5,5 m lange boten met buiter		
1129	http://vocab.getty.edu/aat/300232772	300232772	Adirondack guidebo	Adirondack guidebo	Adirondack guidebo	Te gebruiken voor lichte, schuin afgekante, voor en ach		
1130	http://vocab.getty.edu/aat/300232777	300232777	class boats	klasseboten	klasseboot	Zeilboten, die voldoende op elkaar lijken, of waarvan de		
1131	http://vocab.getty.edu/aat/300232779	300232779	development class	development class	development class	Te gebruiken voor wedstrijdzeilboten die conform bepa.		
1132	http://vocab.getty.edu/aat/300232786	300232786	<merchant vessels	<koopvaardischepe	<koopvaardischip naar lading>			
1133	http://vocab.getty.edu/aat/300212694	300212694	cargo vessels	vrachtschepen	vrachtschip, vrachtb	Te gebruiken voor vaartuigen die hoofdzakelijk zijn ontv		

- Als je je weg met Excel niet zo goed kent: ga door naar §4.2.

★  
**Als je bekend bent met Excel:** je kunt nu enkel vergelijken op het niveau van de prefLabels (hetgeen je ook een acceptabel resultaat zal geven, ondanks dat de AAT-voorkeurstermen meervoudsvormen zijn). Wanneer er meerdere waarden in één kolom staan, zoals bij de altLabels, zal een vergelijking namelijk niet werken. Als je uitgebreid te werk wilt gaan, is het daarom handig om de AltLabels 'uit elkaar te trekken' en vervolgens onder elkaar onderaan in de kolom prefNI te plakken. Dit kan met de functie **Tekst naar Kolommen** op het Tabblad **Gegevens** (ze staan door komma's gescheiden in één kolom). Dit is een minutieus werkje waarbij je de altLabels verspreidt over kolommen. Zorg dat de bijbehorende URI's meekomen. Bij onderstaand voorbeeld is dit gebeurd, en is er een kolom toegevoegd (de tweede); waarin de tekst 'AAT' of 'AAT-Alt' is toegevoegd, afhankelijk van de status van de term.

1210	zakken (kl AAT	http://voca	300210517	pockets (c zak (kledir	Stukken stof bevestigd aan de binnen- of buitenkant van kledingstukken gebruikt voor d
1211	zandloperf AAT	http://voca	300224852	hourglass zandloperf	Pantserhandschoenen in de vorm van een zandloper die populair waren in de 14e en be
1212	zeemanstr AAT	http://voca	300255848	guemseys zeemanstr	Dikke nauwsluitende truien meestal rond- of op vier pennen gebreid in ingewikkelde pat
1213	zegelringe AAT	http://voca	300046014	signet ring zegelring	Vingerringen die zijn voorzien van een signet of zegel en al heel vroeg werden gebruikt
1214	zeilpetten AAT	http://voca	300254618	yachting c zeilpet; ka	Hoofddeksels van witte of marineblauwe stof met een platte hoedenbol en een onbuigz
1215	zelf opwinc AAT	http://voca	300201685	self-windin zelf opwinc	Wordt gebruikt voor horloges die zichzelf opwinden bijvoorbeeld doordat bewegingen va
1216	zibellini AAT	http://voca	300261014	flohpelze zibellino; v	Te gebruiken voor pelswerk gewoonlijk sabel- of marterbont meestal uitgerust met versi
1217	Zischäggen AAT	http://voca	300201796	Zischäggen	Wapenhelmen met een halfronde of puntige kroon verschuifbare neusdekking of een viz
1218	zonnehoec AAT	http://voca	300210803	sunbonnet zonnehoec	Bonnetten van stof of stro met een brede rand om het gezicht tegen de zon te bescher
1219	zonnejurke AAT	http://voca	300221290	sundresse zonnejurk;	Jurken met een verkort lijfje dat de schouders armen en rug vrijhoudt.
1220	zuidwester AAT	http://voca	300248710	sou'wester zuidwester	Waterafstotende hoeden die als regenkleiding worden gedragen of door matrozen bij sle
1221	zware heln AAT	http://voca	300036811	helms (boc zware heln	Zware helmen die het hoofd en gezicht volledig omsluiten oorspronkelijk met een platte
1222	zware rouw AAT	http://voca	300210394	deep mourning	Rouwdrachten die de hoogste graad van rouwbetoon tot uitdrukking brengen; verwijst in
1223	zweetband AAT	http://voca	300046116	sweatband zweetband	Banden van absorberend materiaal die om het hoofd of de polsen worden gedragen om
1224	zweedoek AAT	http://voca	300263206	sudaria zweedoek	Verwijst naar vierkante doeken waarmee het gezicht werd schoongeveegd gedragen do
1225	zwembroel AAT	http://voca	300209936	trunks (costume)	Korte broeken gedragen door mannen voornamelijk bij het boxen of zwemmen.
1226	sokken en AAT-Alt	http://voca	300255776	hosiery su ophouders van kousen	
1227	<accessoi AAT-Alt	http://voca	300211602	<accessor <accessoires die boven het middel gedragen worden>	
1228	<accessoi AAT-Alt	http://voca	300209283	<accessor <accessoires die om het middel of daaronder gedragen worden>	
1229	<accessoi AAT-Alt	http://voca	300211601	<accessor <accessoires die op het hoofd worden gedragen>	
1230	<hamas n; AAT-Alt	http://voca	300224810	<armors b; <hamassen naar functie>	
1231	<japon na; AAT-Alt	http://voca	300221276	<dresses l <japonnen naar functie>	
1232	<japon na; AAT-Alt	http://voca	300221275	<dresses l <japonnen naar vorm>	
1233	<schoen n AAT-Alt	http://voca	300213744	<shoes by <schoenen naar functie>	
1234	<schoen n AAT-Alt	http://voca	300252214	<shoes by <schoenen naar locatie of context>	
1235	<schoen n AAT-Alt	http://voca	300252215	<shoes by <schoenen naar maaksel>	
1236	<schoen n AAT-Alt	http://voca	300213743	<shoes by <schoenen naar vorm>	

## 4.2 Maak een kopie van Adlib-termen in een CSV-bestand

### STAP 1: Zoek termen in de Adlib Thesaurus

- Ga in Adlib naar de database **Thesaurus** en ga naar zoekingang **Recordnummer**.

★

Als je slechts een deel van de thesaurus wilt exporteren, zoek je eerst die selectie op en ga je daarna verder met het volgende punt.

★

Zie ook het hoofdstuk **Gemarkeerde records exporteren** in de Adlib gebruikersgids:

<http://www.adlibsoft.nl/Content/Documentation/Dutch/Adlib%20Gebruikersgids%20A5.pdf>

- Druk op **Volgende** en alle termen in de thesaurus worden gepresenteerd.
- Gebruik de knop **Markeringen wisselen** om alle termen te selecteren.

### STAP 2: Plaats Adlib-termen in een Excel-bestand

- Ga naar **Bestand** en dan **Exporteren**.
- Kies voor **CSV-bestand**. Als je in OpenRefine wil werken, is XML ook een prima format om te exporteren.
- Exporteer de volgende velden:
  - Preref:** recordnummer van de term in Adlib. Deze is later weer nodig om de URI's te importeren
  - Term:** de term zelf
  - (eventueel) Broader\_term:** als deze is ingevuld geeft hij wat context aan de term, mocht de term een homoniem zijn.
- Je opent **Excel**. Dan kies je het bestand via **Tabblad Gegevens, Van Tekst, scheiden op komma, zoals ook werd uitgelegd in stap 3 van §4.1**. Je kunt het bestand ook leesbaar maken door het

direct te openen in Excel, de eerste kolom te selecteren en de informatie dan via **Gegevens, Tekst naar Kolommen** te verspreiden, zoals in het kader aan het einde van stap 3 van §4.1.

- Alle gegevens staan dan in losse kolommen. Sla het bestand dan op als een Excel-werkmap (.xlsx).

### 4.3 Koppel Adlib-termen aan AAT-termen

#### Vorbereiding

Er zijn grofweg drie stappen te onderscheiden: de twee lijsten samenbrengen in Excel, Vergelijken met Fuzzy Lookup, Matches evalueren (en opschoning voorbereiden).

- Download de Fuzzy Lookup add-in: <https://www.microsoft.com/en-us/download/details.aspx?id=15011>.

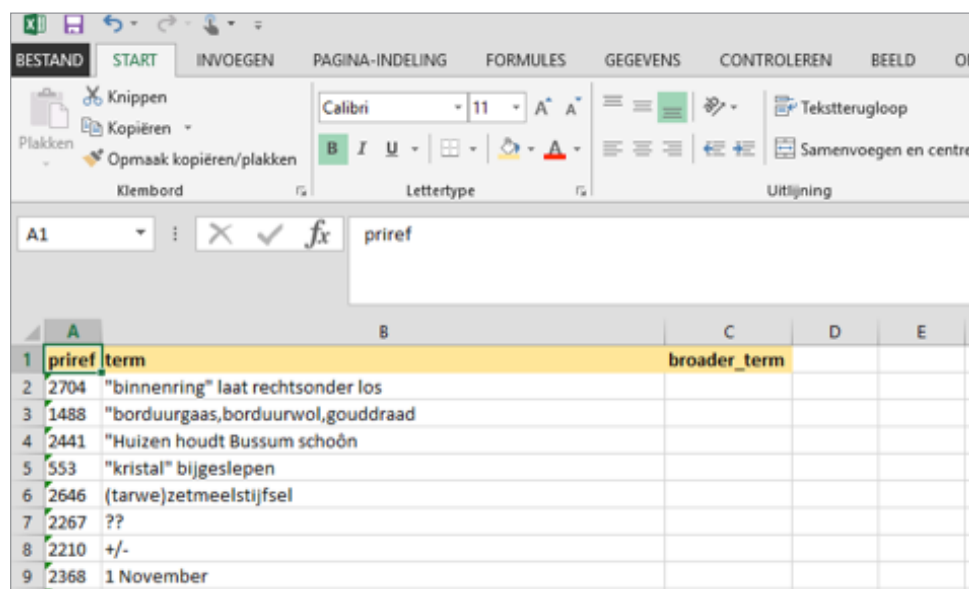
★

Dit is een goed filmpje om je bekend te maken met Fuzzy Lookup, van 'Excel Bytes': **How to use Fuzzy**

**Lookup in Excel** [https://www.youtube.com/watch?v=aJpL\\_wNzeMs](https://www.youtube.com/watch?v=aJpL_wNzeMs)

#### STAP 1: *Adlib- en AAT-termen samenbrengen in een Excel-bestand*

- Open het CSV-bestand met de Adlibtermen. Dit ziet er nu bijvoorbeeld zo uit:

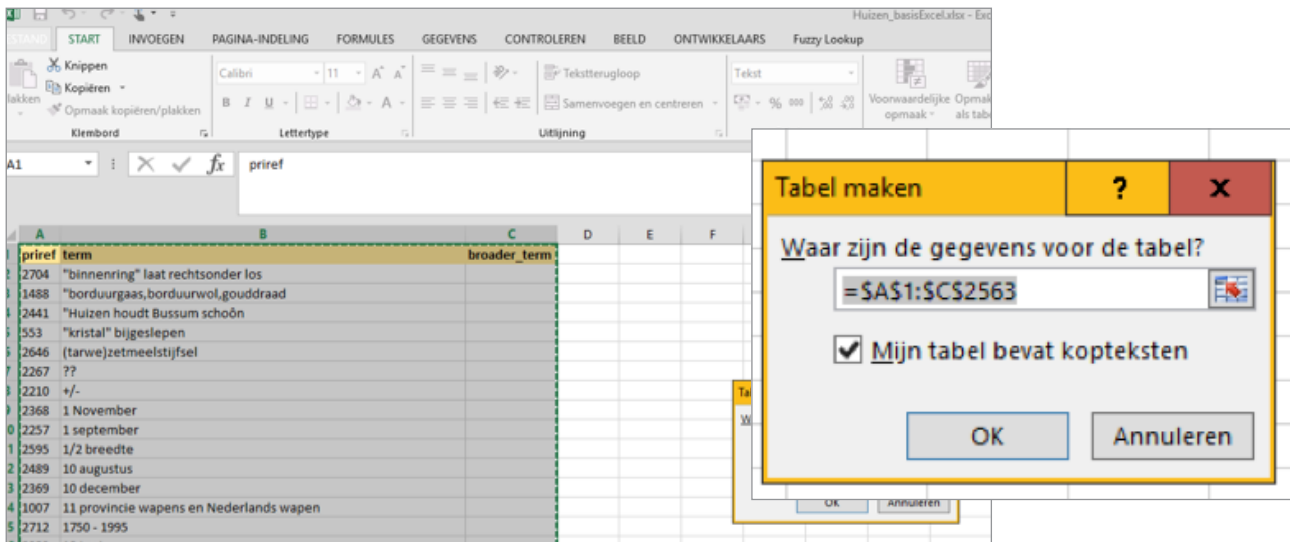


	A	B	C	D	E
1	preref	term	broader_term		
2	2704	"binnenring" laat rechtsonder los			
3	1488	"borduurgaas,borduurnwol,gouddraad			
4	2441	"Huizen houdt Bussum schoon			
5	553	"kristal" bijgeslepen			
6	2646	(tarwe)zetmeelstijfjel			
7	2267	??			
8	2210	+/-			
9	2368	1 November			

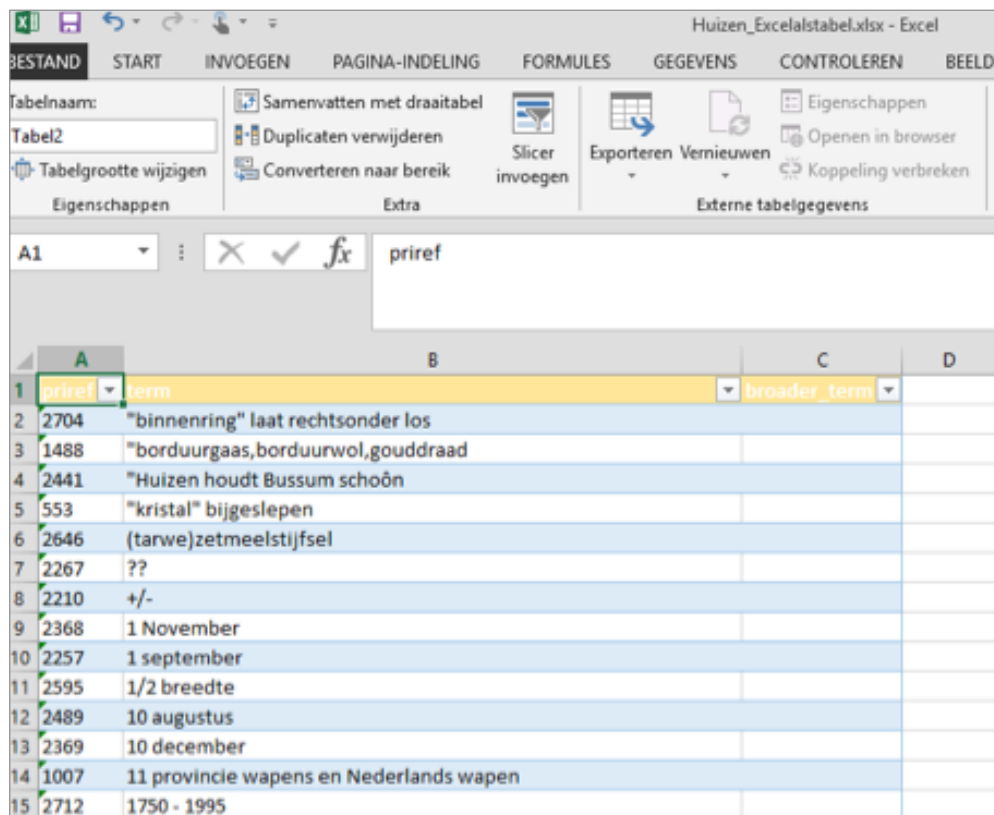
- Je gaat nu in één bestand, op twee verschillende tabbladen, de terminologiebronnen plaatsen (Adlib en AAT) en omzetten tot tabellen. Je begint in het geopende bestand: dat van je eigen instelling.
- Zet de cursor in kolom A, rij 1 en druk op de toetsenbordcombinatie **Ctrl+A**. Alle kolommen en rijen waar tekst in staat worden nu geselecteerd. Zorg ervoor dat niet alle rijen en alle kolommen

uit de Excel zijn geselecteerd, alleen de rijen en kolommen waar gegevens in staan.

- Druk vervolgens op de toetsenbordcombinatie **Ctrl+L**. Het volgende schermje verschijnt:

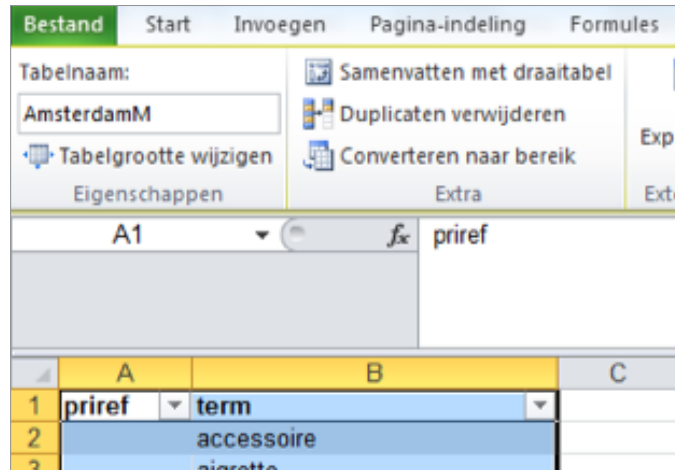


- Zorg dat 'Mijn tabel bevat kopteksten' is aangevinkt. Druk op **OK**. De tabel ziet er dan zo uit:





- Verander de tabelnaam links bovenin; geef die **de naam van je eigen museum**.



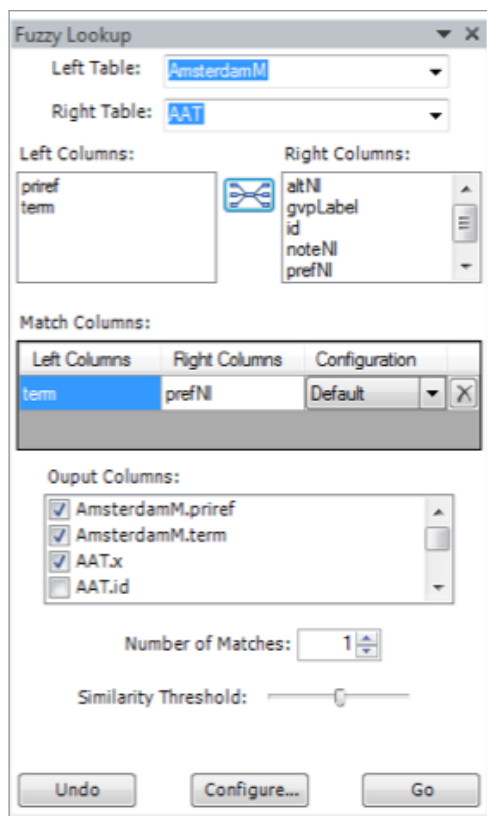
- Open het **Excel-document met de AAT**, dat je in §4.1 hebt gemaakt.
- Gebruik **Ctrl+A** om alle met gegevens gevulde kolommen en rijen te selecteren. Druk vervolgens op **Ctrl+C** om deze te kopiëren.
- Ga naar het oorspronkelijke Excel-document waar je mee begon en zet de cursor bovenaan in een **nieuw tabblad**. Druk op **Ctrl+V** om de gegevens te plakken.
- Hier bestaan, naast de voorkeurs- en niet-voorkeurstermen, de kolommen voor de **URI**, de **ID**, de **Engelse term**, en de **scope note**. Om het jezelf gemakkelijk te maken, kun je kolommen die je nu niet nodig hebt weghalen (zoals ID, Engelse term).
- Druk dan weer met de cursor in **Cell A1** op **Ctrl+A** en op **Ctrl+L** om hier een nieuwe tabel van te maken. Geef deze tabel de naam **AAT**.
- **Je hebt nu op twee tabbladen twee tabellen, waarbinnen je waarden kunt gaan vergelijken.**

## STAP 2: *Adlib- en AAT-termen vergelijken met Fuzzy Lookup*


Ga vervolgens naar de **extensie Fuzzy Lookup** (tabje rechts bovenin, naast 'Beeld' of 'Ontwikkelaars') en klik op deze knop:



- Aan de rechterkant van het scherm verschijnt een venster:



Figuur 8  
Fuzzy Lookup-paneel

- Bij **Left table** kies je je eigen thesaurus, bij **Right table** je doelthesaurus (AAT).
- Als er iets staat in het veld **Match Columns**, verwijder dat dan door op het kruisje rechts te klikken.
- Vervolgens kies je daarboven bij **Left en Right Columns**, voor velden die je wilt matchen met elkaar: links de kolomnaam van jouw eigen termen en rechts de kolomnaam van de AAT-termen waarmee je wilt matchen (vaak prefNI).
- Klik op het **icoontje**  om een nieuwe regel te creëren in **Match Columns**.
- Je ziet nu in **Match Columns** de namen van de kolommen die je met elkaar vergelijkt, zoals Figuur 8 illustreert.
- **Je vinkt vervolgens de informatie aan die je wilt terugzien in je resultaat.** Zorg dat je in ieder geval kiest: je priref, eigen term, de doelterm, de URI (als beschikbaar) en de Similarity. Informatie die je niet nodig hebt vink je beter uit.
- Open een derde **nieuw tabblad** in het Excel-document en zet je cursor in de eerste cel. Hier komen zo je resultaten terecht.
- Druk vervolgens in het Fuzzy Lookup venster op **Go**.

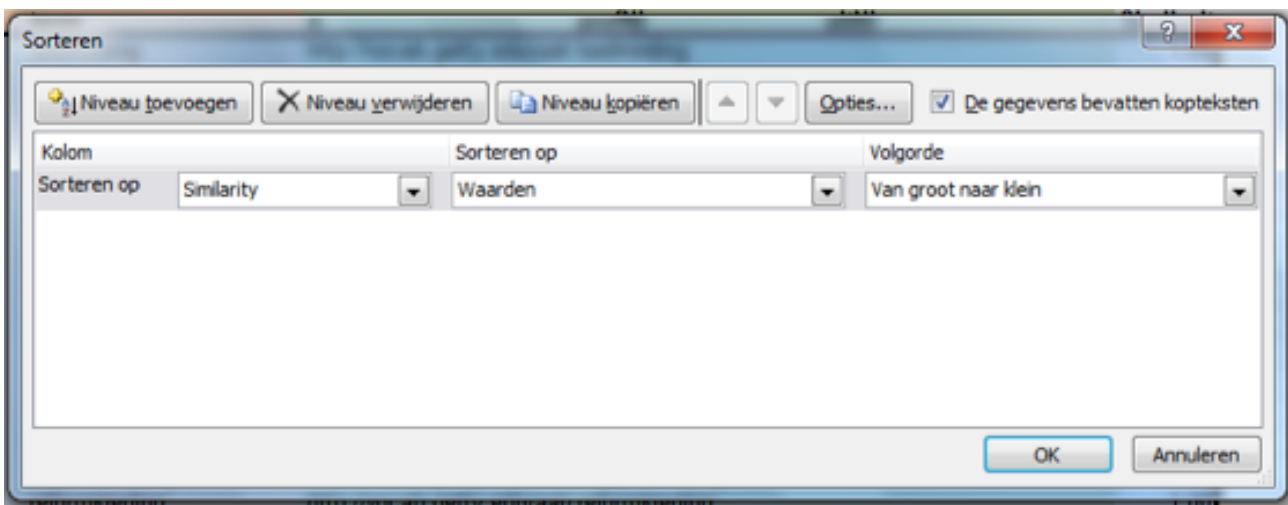
priref	term	broader_t	priref	term	erfgoedth	erfgeo.uri	latitude	longitude	Similarity
2704	"binnenring" laat rec		11830	rechts onc	http://data.cultureelerfgoed.nl/semnet/				00.001
1488	"borduurgaas, bordu		9919	borduurgi	http://data.cultureelerfgoed.nl/semnet/				00.001
2441	"Huizen houdt Bussu		4645	Bussum	http://sw: 52,27333				
553	"kristal" bijgeslepen								
2646	(tarwe)zetmeelstijfjel								
2267	??								
2210	+/-								
2368	1-nov								
2257	1-sep		7966	Sappemeer	http://sw: 53,16417				
2595	1/2 breedte		4180	breedte	http://data.cultureelerfgoed.nl/				
2489	10 augustus								00.000
2369	10-dec								00.000
1007	11 provincie wapens		9149	Nederlani	http://data.cultureelerfgoed.nl/semnet/				00.001
2712	1750 - 1995		48	1700-1750					00.001
2253	18 juni								00.000
2321	18-sep		7966	Sappemeer	http://sw: 53,16417		6,79028		00.001
2407	18-06								00.000
765	1860-1950								00.000
2343	19 augustus								00.000
526	1900 tot 1910		3036	tol	http://data.cultureelerfgoed.nl/semnet/				00.001
34	1900 tot 1920		3036	tol	http://data.cultureelerfgoed.nl/semnet/				00.001
1283	1900 tot1940		55	1900-1945					00.001
2692	1924								00.000
643	1925								00.000
43	1930								00.000
44	1950								00.000

- Als hij klaar is met het vergelijken is het soms niet overduidelijk welke rijen bij welke thesaurus horen. Geef ze daarom even een andere kleur. Je ziet dan links jouw eigen priref en term, terwijl rechts de daarop lijkende termen uit de doelthesaurus staan, inclusief de URI en andere informatie die je had aangevinkt. Helemaal rechts is een kolom met 'Similarity'.

priref	term	x	prefNI	altNI	Similarity
15034	accessoire		http://vocab.getty.edu/aat/accessoires	accessoire	00.001
18304	aigrette		http://vocab.getty.edu/aat/aigrettes	aigrette	00.001
126	aktetas				00.000
384	albe		http://vocab.getty.edu/aat/alben	albe; albes	00.001
348	ambtskleding		http://vocab.getty.edu/aat/ambtsketens	ambtsketen	00.001
239	ambtssymbool		http://vocab.getty.edu/aat/ambtssymbolen		00.001
2455	anorak		http://vocab.getty.edu/aat/andra	andron	00.001
432	armband		http://vocab.getty.edu/aat/armbanden (bovenarm)	armband (bovenarm)	00.001
3847	armstuk				00.000
2874	avondjapon		http://vocab.getty.edu/aat/avondjaponnen	avondjapon; avondjurk; avont	00.001
2984	avondjas				00.000
283	avondkleding		http://vocab.getty.edu/aat/strandkleding		00.001
948	avondschoen				00.000
478	avonttas				00.000

**STAP 3:** *Matches evalueren (en opschoning voorbereiden)*

- Selecteer de kolom **Similarity**.
- Klik de rechtermuisknop in de kolom, kies **Celeigenschappen**: geef deze als celeigenschap 'getal'. Dan zie je meer decimalen.
- Sorteert dan de rij van groot naar klein, door weer via **Cell A1** alles te selecteren met **Ctrl+A**; Klik rechtsboven op **Sorteren en Filteren; Aangepast sorteren**. Daar kies je voor de Kolom **Similarity**, in de volgorde **Van groot naar klein**.



- Nu komen de **exact matches** bovenaan. Dit is het begin van evaluatie en mogelijk ook van je opschoning.

B	C	D	E	F
term	x	prefNI	altNI	Similarity
badkleding	http://vocab.getty.edu/aat/badkleding			1,00
beddengoed	http://vocab.getty.edu/aat/beddengoed			1,00
breiwerk	http://vocab.getty.edu/aat/breiwerk			1,00
bretels	http://vocab.getty.edu/aat/bretels			1,00
ceremoniële kleding	http://vocab.getty.edu/aat/ceremoniële kleding			1,00
hot pants	http://vocab.getty.edu/aat/hot pants		hot pant	1,00
kinderkleding	http://vocab.getty.edu/aat/kinderkleding			1,00
kleding	http://vocab.getty.edu/aat/kleding		kleren; kledingstukken; kled	1,00
liturgische kleding	http://vocab.getty.edu/aat/liturgische kleding			1,00
mannenkleding	http://vocab.getty.edu/aat/mannenkleding		herenkleding	1,00
memorabilia	http://vocab.getty.edu/aat/memorabilia		memorabilium	1,00
miniaturkleding	http://vocab.getty.edu/aat/miniaturkleding			1,00
naaigerei	http://vocab.getty.edu/aat/naaigerei		naaigereedschappen en ben	1,00
ondergoed	http://vocab.getty.edu/aat/ondergoed		onderkleding	1,00
poppenkleding	http://vocab.getty.edu/aat/poppenkleding			1,00
reformkleding	http://vocab.getty.edu/aat/reformkleding			1,00
schoeisel	http://vocab.getty.edu/aat/schoeisel			1,00
sieraden	http://vocab.getty.edu/aat/sieraden		sieraad; juwelen (sieraden)	1,00



- Kijk eerst naar de **exact matches**, met als getal 1,00. Dit zijn termen uit de eigen terminologie en de AAT-prefLabels, die volledig overeenkomen.
- **Maak een nieuw tabblad en plak daar al deze exact matches in.** Ga er steekproefsgewijs doorheen, want er is een kleine kans op fouten. **Homoniemen**, zoals vos (accessoire) en vos (dier) zullen relatief weinig voorkomen, afhankelijk van de terminologiebron die je gebruikt. Denk erover na of dit mogelijk is binnen jouw eigen lijst of de terminologiebron: zijn er termen die hetzelfde zijn, maar iets anders betekenen?
- Scroll naar beneden. De matches die niet exact zijn (lager dan 1) zal je moeten nakijken, om er zoveel mogelijk correcte links tussenuit te halen. Bij de correcte laat je de URI staan; knip en plak ze naar tabblad **Exact matches** dat je zojuist hebt gemaakt.

94	naarborster	naarborsters	http://vocab.getty.edu/aat/300230713	0,99
95	portemonnee	portemonnees	http://vocab.getty.edu/aat/300215521	0,98
96	espadrille	espadrilles	http://vocab.getty.edu/aat/300046068	0,97
97	ski pak	skipak	http://vocab.getty.edu/aat/300210460	0,97
98	vrijetijdsleding	vrije-tijdsleding	http://vocab.getty.edu/aat/300221546	0,97
99	hotpants	hot pants	http://vocab.getty.edu/aat/300243776	0,97
100	salopette	salopettes	http://vocab.getty.edu/aat/300046178	0,97
101	ensemble	ensembles	http://vocab.getty.edu/aat/300209844	0,96
102	draagmand	draagband	http://vocab.getty.edu/aat/300258894	0,96
103	damesdas	damestas	http://vocab.getty.edu/aat/300046219	0,96
104	damesjas	damestas	http://vocab.getty.edu/aat/300046219	0,96
105	avondtas	avondjas	http://vocab.getty.edu/aat/300257524	0,96
106	smoking	smokings	http://vocab.getty.edu/aat/300216052	0,96
107	mantille	mantilla	http://vocab.getty.edu/aat/300046124	0,96
108	heren-T-shirt	T-shirt	http://vocab.getty.edu/aat/300209903	0,95
109	dames-T-shirt	T-shirt	http://vocab.getty.edu/aat/300209903	0,95
110	blouse	blouses	http://vocab.getty.edu/aat/300046133	0,95
111	koffer	koffers	http://vocab.getty.edu/aat/300046224	0,95
112	kinderschoen	kinderschorten	http://vocab.getty.edu/aat/300046186	0,95
113	parure	parures	http://vocab.getty.edu/aat/300261068	0,95
114	bontjas	bontjes	http://vocab.getty.edu/aat/300210071	0,95
115	panier	pannier	http://vocab.getty.edu/aat/300216740	0,95
116	handbeschermer	halsbeschermer	http://vocab.getty.edu/aat/300036921	0,95
117	bretel	bretels	http://vocab.getty.edu/aat/300209996	0,95
118	slak	slakken	http://vocab.getty.edu/aat/300209957	0,95

- Bij niet kloppende matches verwijder je de URI, je plaatst ze in een ander tabblad, **Geen match**. Daar plaats je ook alle termen met '0' **Similarity**. Het kan zijn dat deze termen gewoon niet in de doelthesaurus zitten, omdat je werkt met een klein deel van een grotere thesaurus, zoals hier:

521	guimpe haakinstr	guimpe (halsboord)		0,57
522	aansteker	hoop fastener nails	http://vocab.getty.edu/aat/300048941	0,57
523	decoratie	cement gecoate spijke	http://vocab.getty.edu/aat/300065935	0,57
524	deux pièces	double pieces	http://vocab.getty.edu/aat/300224812	0,54
525	deux-pièces	double pieces	http://vocab.getty.edu/aat/300224812	0,54
526	aangeklede haan	haak en oog	http://vocab.getty.edu/aat/300051508	0,52
527	manicure-set	formed seat hangers	http://vocab.getty.edu/aat/300081606	0,50
528	interieurtextiel			0,00
529	damast			0,00
530	speelgoed			0,00
531	doos			0,00
532	prent			0,00
533	affiche			0,00
534	muil			0,00
535	damesjabot			0,00
536	album			0,00
537	degen			0,00
538	gordijn			0,00
539	uniseks			0,00
540	bruidsaccessoire			0,00
541	lint			0,00
542	dameskraag			0,00

*Figuur 9*  
Bij deze thesaurusterminen vindt Fuzzy Lookup geen gelijke. Sommige zijn correct maar horen niet bij het domein waar de thesaurus in voorziet (hier: kostuum).

- Misschien ben je over sommige exact matches toch niet zeker. Uiteindelijk eindig je bijvoorbeeld met deze tabbladen: **Exact Match**, **Exact Match controle**, **Geen match**.
- Bij de termen in tabblad **Exact match** kunnen de URI's direct worden geïmporteerd in Adlib. Dat wordt uitgelegd in de volgende paragraaf, §4.4. Bij **Exact match controle** kan dit, na de controle door bijvoorbeeld een conservator, ook gebeuren. Geen match kan een **basis** zijn voor een **opschoonproces**. **Daarover lees je meer in Addenda 3 en 4.**

#### 4.4 Termen en URI's importeren in Adlib

Als je de termen hebt gematched en ze zijn voorzien van URI's kun je ze gaan importeren in Adlib. Hiervoor wordt wederom gebruik gemaakt van een bestand dat op een bepaalde manier moet zijn ingedeeld. In deze Casus gebruiken we een **Adlib Tagged** bestand voor een directe import.

Wat je nodig hebt voor de import zijn de velden:

- 1 **Priref**<sup>5</sup> (en de **tag ofwel veldcode**, in dit geval '%o')
- 2 **URI** uit de AAT (en de **tag** van het veld waar hij in moet komen, hier 'ur').

Het moet een Adlib Tagged worden, dus je doelstelling moet zijn:

```
%o 12347
ur http://vocab.getty.edu/aat/300209273
**
```

ONDERDELEN ADLIB-TAGGED	BETEKENIS
%o 12347	Tag/veldcode en inhoud veld 'priref'
ur http://vocab.getty.edu/aat/300209273	Tag/veldcode en inhoud veld 'URI'
**	Nieuw record/separator

#### URI's importeren met een Adlib Tagged

Dit kan met bijvoorbeeld **Word**, **OpenRefine**, of tekstprogramma's als **Ultraedit** of **Notepad++**. Deze uitleg is voor **Word**.

##### STAP 1: *Bewerk je Excel-bestand*

- **Knip** uit het matchingsbestand (de exact matches) de kolommen met de priref en de URI.
- **Plak** deze in een nieuw Excel-bestand en sla dit op.
- **Voeg een kolom toe** links van elke gevulde kolom.
- **Selecteer** het geheel (**Ctrl+A**). Klik rechts op **Celeigenschappen** en kies voor **Tekst**, zodat de tag '%o' zo zal blijven staan.
- Plaats dan in de lege kolommen: **%o** voor de priref in Kolom A en de tag voor je URI-veld in Kolom C, bijvoorbeeld **ur**.



Dit gaat **extra snel** door de tag een keer bovenin in te vullen en vervolgens ook alle cellen te selecteren waarin hij moet terechtkomen. Dat doe je bijvoorbeeld door in die bovenste cel te klikken en dan, met Shift ingedrukt, in de onderste cel van die Kolom te klikken waarin je hem wilt plaatsen (alle cellen waarin hij moet komen zijn dan geselecteerd). Dan: **Ctrl+D**.

	A	B	C	D
1	Tag-Preref	Preref	Tag-URI	URI
2	%0	2015	ur	http://vocab.getty.edu/aat/300129420
3	%0	4565	ur	http://vocab.getty.edu/aat/300204952
4	%0	465	ur	http://vocab.getty.edu/aat/300264039
5	%0	365	ur	http://vocab.getty.edu/aat/300209996
6	%0	52	ur	http://vocab.getty.edu/aat/300210387
7	%0	51331	ur	http://vocab.getty.edu/aat/300243776
8	%0	8621	ur	http://vocab.getty.edu/aat/300379348
9	%0	541	ur	http://vocab.getty.edu/aat/300266639

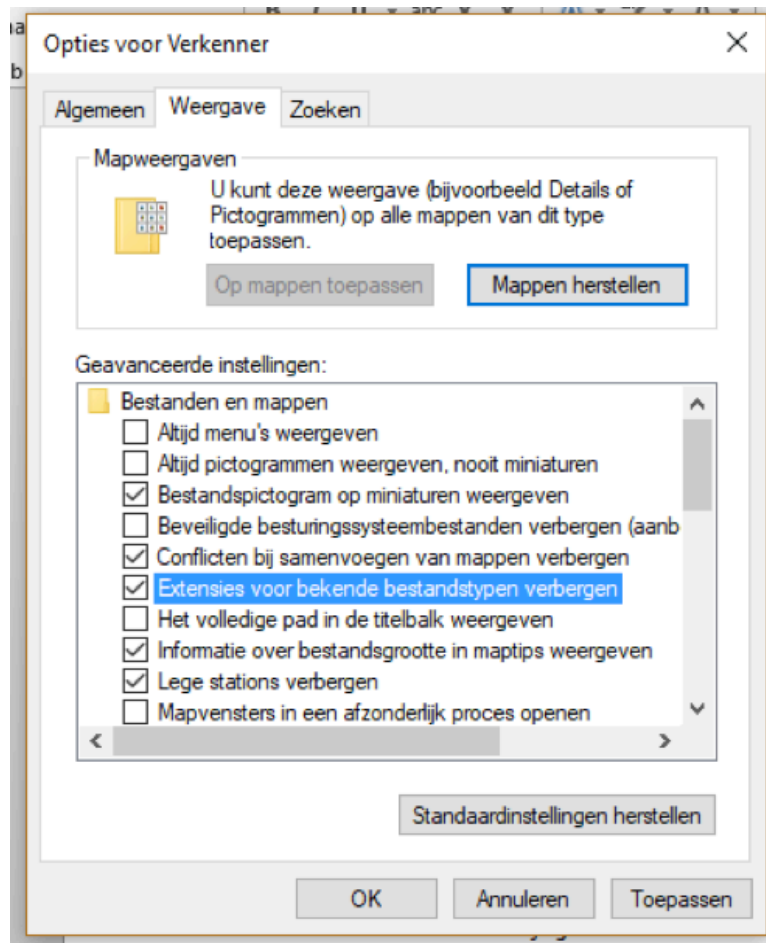
- Sla het Excel-tabblad op.
- Sla het Excel-tabblad onder een andere naam op als tekst-bestand (KIES: 'txt, tab is scheidingsteken'). Klik op OK.



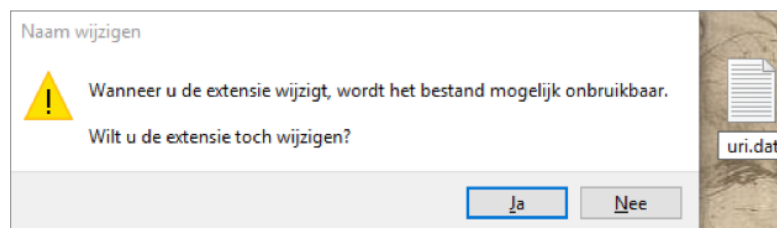
- Verlaat Excel.
- Open de **Verkenner** en zoek het bestand. Is de extensie '**txt**' zichtbaar achter de bestandsnaam van je tekst-bestand? Dan kun je door naar de volgende stap.



Is de extensie van je bestand niet zichtbaar in de Verkenner? Zoek dan bij Start op 'extensie'. Klik op 'Bestandsextensies weergeven of verbergen'. Zorg dat je dan dit uitvinkt: 'Extensies voor bekende bestandstypen verbergen'. Zie onderstaande afbeelding.



- Pas de extensie van het tekst-bestand aan van '.txt' naar '.dat'. Klik op Ja.

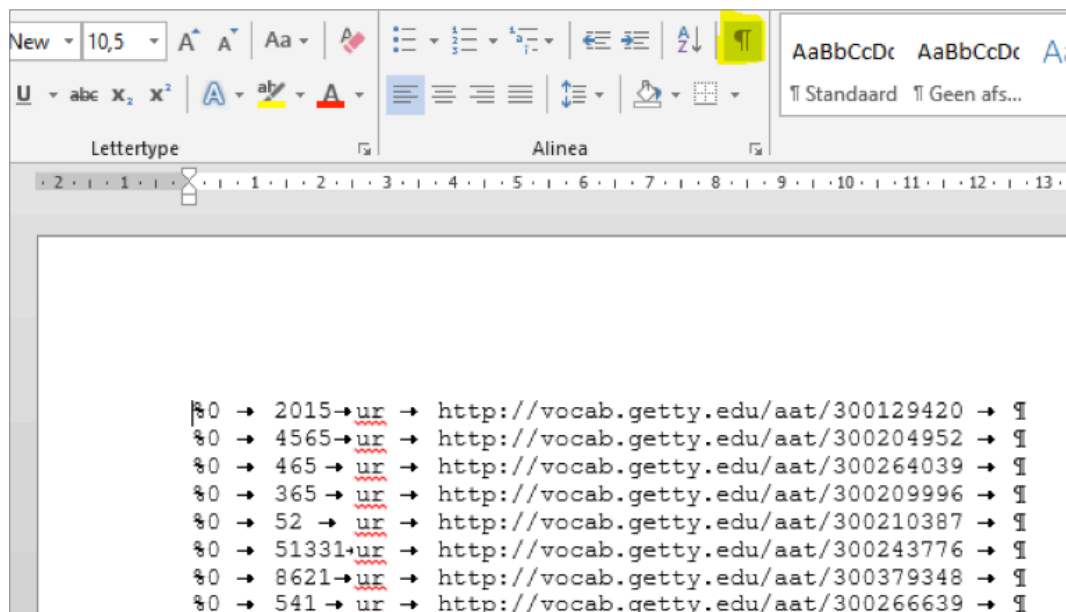


## STAP 2: *Bewerk het bestand in Word*

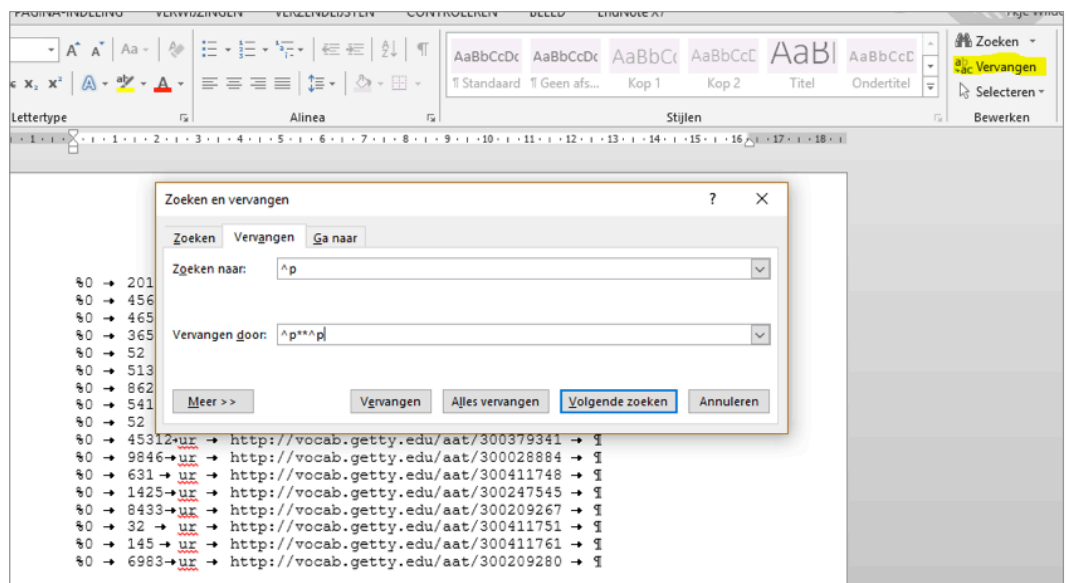
- Open dit bestand in **Word**. Verwijder de bovenste regel, de header.

Tag-Prifref	Prifref	Tag-URI	URI
%0	2015 ur	http://vocab.getty.edu/aat/300129420	
%0	4565 ur	http://vocab.getty.edu/aat/300204952	
%0	465 ur	http://vocab.getty.edu/aat/300264039	
%0	365 ur	http://vocab.getty.edu/aat/300209996	
%0	52 ur	http://vocab.getty.edu/aat/300210387	
%0	51331 ur	http://vocab.getty.edu/aat/300243776	
%0	8621 ur	http://vocab.getty.edu/aat/300379348	
%0	541 ur	http://vocab.getty.edu/aat/300266639	

- Klik op de knop 'alles weergeven', hieronder geel gemarkeerd (of **Ctrl-8**) om Alineamarkeringen weer te geven. Hierbij staat het icoontje → voor een tab en ¶ voor een harde return.



- Nu kun je via 'zoeken en vervangen' (Ctrl-H of de knop Vervangen rechtsboven) eenvoudig de structuur van dit document aanpassen zodat dit een door Adlib herkenbaar .dat-bestand wordt.
- Voer de volgende acties uit:
- Vervang **^p** door **^p\*\*^p**. Zo wordt de harde return vervangen door harde return\*\*harde return. De sterretjes komen op de volgende regel te staan, apart.

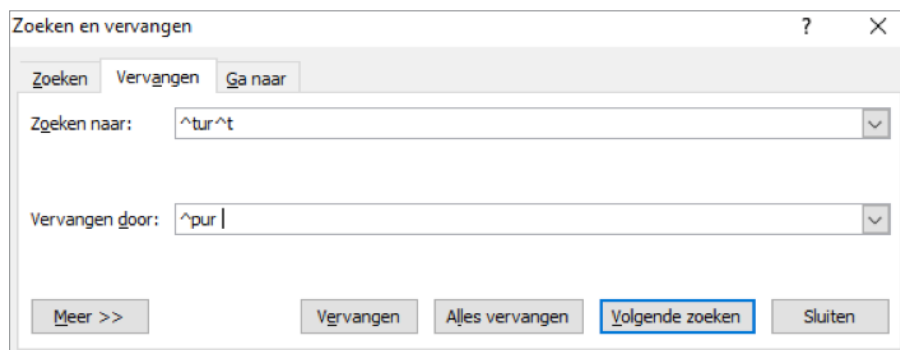




*Resultaat:*

```
%0 → 2015→ur → http://vocab.getty.edu/aat/300129420 → ¶
**¶
%0 → 4565→ur → http://vocab.getty.edu/aat/300204952 → ¶
**¶
%0 → 465 → ur → http://vocab.getty.edu/aat/300264039 → ¶
**¶
%0 → 365 → ur → http://vocab.getty.edu/aat/300209996 → ¶
**¶
%0 → 52 → ur → http://vocab.getty.edu/aat/300210387 → ¶
**¶
%0 → 51331ur → http://vocab.getty.edu/aat/300243776 → ¶
**¶
%0 → 8621→ur → http://vocab.getty.edu/aat/300379348 → ¶
**¶
%0 → 541 → ur → http://vocab.getty.edu/aat/300266639 → ¶
**¶
```

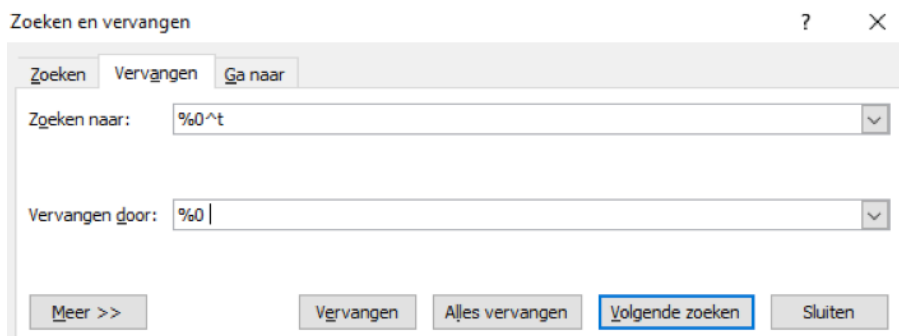
- Nu ga je een nieuwe regel maken met de URI-tag en de URI zelf.
- Vervang '^tur^t' door '^pur' (let op de spatie na 'pur'). Zo wordt ur tab naar de volgende regel verplaatst en vervangen door ur spatie.



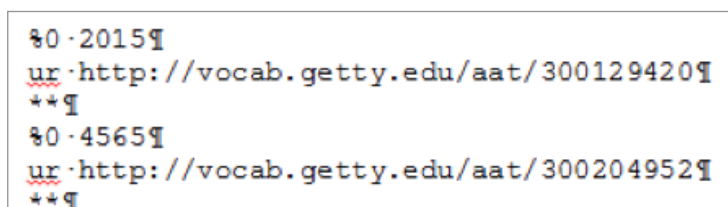
*Resultaat:*

```
%0 → 2015¶
ur http://vocab.getty.edu/aat/300129420¶
**¶
%0 → 4565¶
ur http://vocab.getty.edu/aat/300204952¶
**¶
%0 → 465¶
ur http://vocab.getty.edu/aat/300264039¶
**¶
%0 → 365¶
ur http://vocab.getty.edu/aat/300209996¶
**¶
%0 → 52¶
ur http://vocab.getty.edu/aat/300210387¶
**¶
%0 → 51331¶
ur http://vocab.getty.edu/aat/300243776¶
**¶
```

- Nu ga je de tab tussen de priref-tag en de priref door een spatie vervangen. **Vervang ‘%o^t’ door ‘%o ’ (let op de spatie achter %o).**

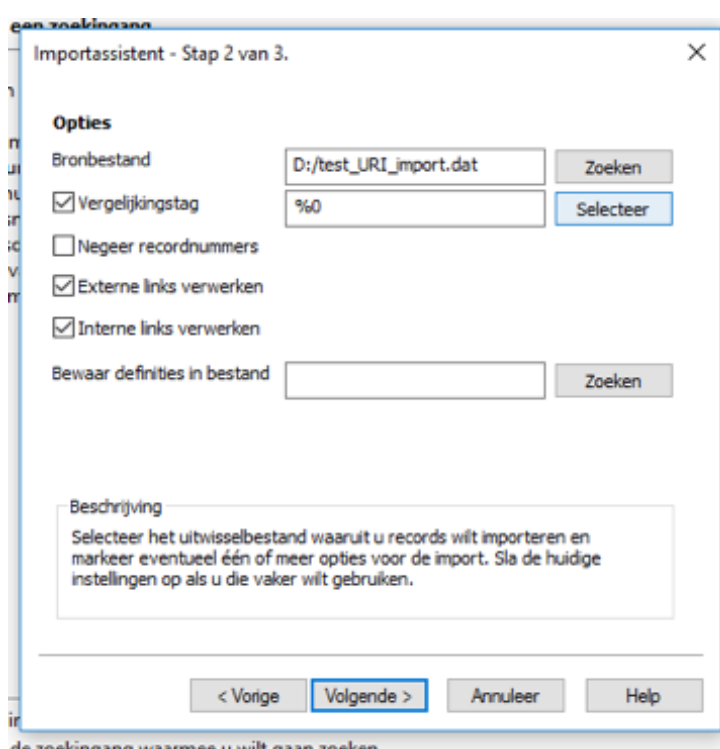
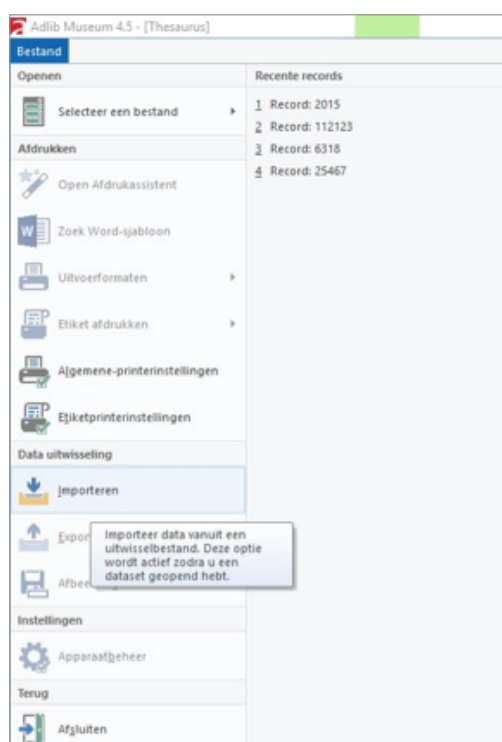


*Resultaat:*

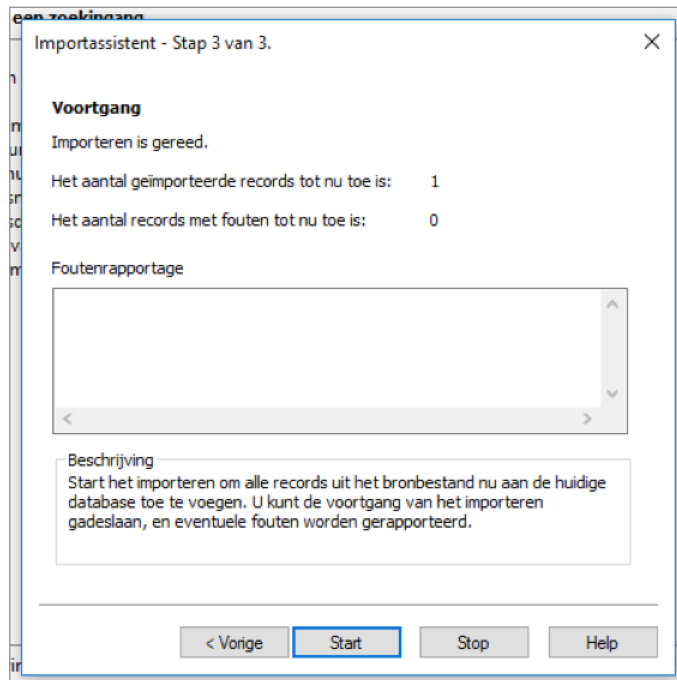


### STAP 3: *Importeer het bestand in Adlib*

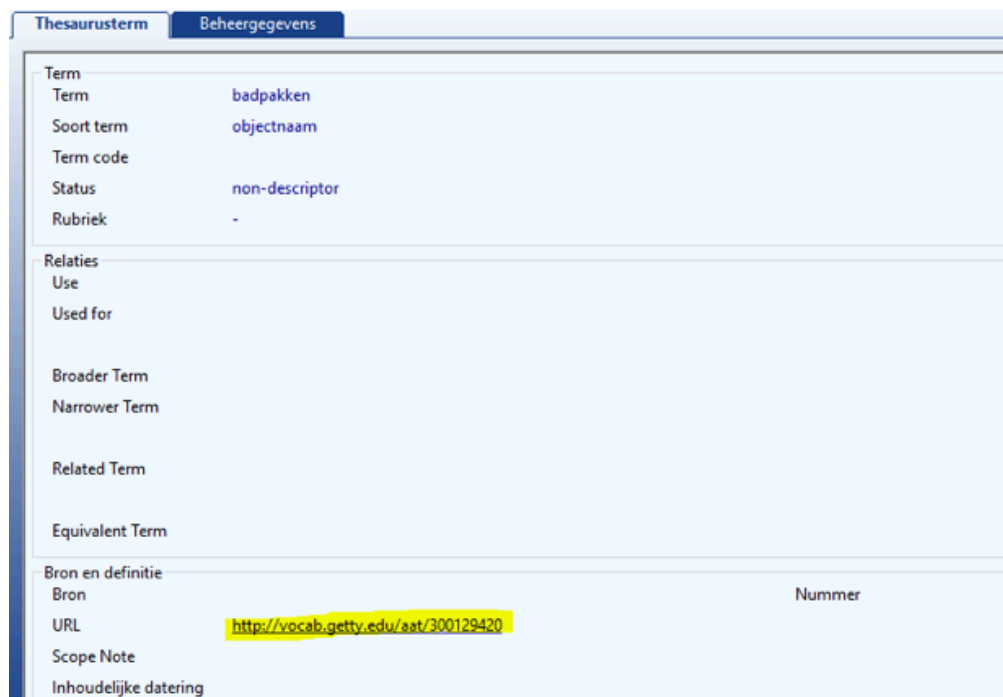
- Open **Adlib**.
- Ga naar de **Thesaurus**.
- Ga naar **Bestand, Importeren**.



- Kies bij **Methode** voor 'Adlib Tagged-bestand'.
- Kies bij **Bronbestand** voor het **dat.bestand** dat je hebt aangemaakt.
- Vink 'Vergelijkingstag' aan en vul daar in: '%o'.
- Zie de afbeelding hierboven; de andere vakjes kunnen blijven aan- of uitgevinkt zoals standaard.
- Klik op **Start**. In het voorbeeld is er geheel toevallig sprake van één enkel record.
- Klik **zodra de import is afgerond** op **Stop**.



- Controleer **steekproefsgewijs** de geïmporteerde informatie.
- Gefeliciteerd! Je bent de eigenaar van URI's!



*Figuur 10:*  
Het is aan te raden om de import van URI's in Adlib steekproefsgewijs te controleren.

## 5 LITERATUUR

### *Adlib: Importeren en exporteren, externe terminologiebronnen*

- <http://www.adlibug.nl/2011/05/importeren-en-exporteren-vanuit-excel-door-jeroen-de-meester/> (over importeren en exporteren)
- [http://documentation.axiell.com/alm/en/index.html?ds\\_generaltopicsadvancedextsources.html](http://documentation.axiell.com/alm/en/index.html?ds_generaltopicsadvancedextsources.html) (over externe terminologiebronnen in Adlib)

### *Linked Data, Opschoon en linkproces*

- [Freeyourmetadata.org](http://freeyourmetadata.org)
- Tutorial Linked Data Stap voor Stap: <http://www.den.nl/pagina/334/den-lab>
- Hooland, Seth van, en Ruben Verborgh, *Linked Data for Libraries, Archives and Museums. How to clean, link and publish your metadata*. Londen: Facet, 2014.

### *Thesauri/terminologiebronnen en (Getty) Vocabularies*

- <http://www.aat-ned.nl> en [aat@rkd.nl](mailto:aat@rkd.nl).
- Harpring, Patricia, *Introduction to Controlled Vocabularies. Terminology for Art, Architecture and Other Cultural Works*, Los Angeles: Getty, 2010.
- <http://www.getty.edu/research/tools/vocabularies/lod/>
- <https://groups.google.com/forum/#!forum/gettyvocablod> (hulpforum)
- Riesthuis, Gerrit en Gerard Wijnands (eds.), *Thesaurusbouw: handboek voor opleiding en praktijk*, Den Haag: Nederlands Bibliotheek en Lektuur Centrum, 1992.

### *CultuurLINK*

- <http://cultuurlink.beeldengeluid.nl/app/>

### *OpenRefine*

- [https://www.projectcest.be/wiki/Bestand:20150430\\_Handout\\_Open\\_Refine\\_workshop.pdf](https://www.projectcest.be/wiki/Bestand:20150430_Handout_Open_Refine_workshop.pdf) (handleiding voor het opschoonen van data door Inge van Stokkom en Alina Saenko)
- [openrefine.org](http://openrefine.org)
- <http://freeyourmetadata.org/cleanup/> (om te oefenen)
- <https://github.com/OpenRefine/OpenRefine/wiki> (algemene informatie)
- <https://github.com/OpenRefine/OpenRefine/wiki/Clustering> (over Cluster & Edit)
- <https://groups.google.com/forum/#!forum/openrefine> (hulpforum)

### *Relevante academische artikelen*

- Hildebrand, Michiel, et al., 'Supporting subject matter annotation using heterogeneous thesauri. A user study in Web data reuse.' *International Journal of Human Computer Studies* 67, 2009, pp. 888-903.
- Hooland, Seth van, et al., 'Evaluating the success of vocabulary reconciliation for cultural heritage collections.', *Journal of the American Society for Information Science and Technology* 64, 2013, pp. 464-479.

## 6 COLOFON

### *Woord vooraf:*

Kaat Debo, directeur, ModeMuseum Antwerpen (MoMu)  
Annette Gaalman, voorzitter Digitaal Erfgoed Bruikbaar, NDE en adviseur, Erfgoed Brabant  
Wilbert Helmus, Domeinmanager Digitaal Erfgoed Bruikbaar, NDE

### *Auteurs:*

Shannon van Muijden, Zuiderzeemuseum (§3.1, §3.2, §4.2, §4.3, Addendum 2 (CSV))  
Inge van Stokkom, Netwerk Oorlogsbronnen (Addendum 1, §3.2, Addendum 3)  
Ykje Wildenborg, ModeMuseum Antwerpen (Hfd 3, §4.1, §4.3, Addendum 3)

### *Met bijdragen van:*

Jeroen De Meester, Stad Antwerpen, Adlib Gebruikersgroep (§4.4, Addendum 2 (Tagged))  
Rolf Blijleven, freelance Adlib-ondersteuning, Adlib Gebruikersgroep (Addendum 4)

### *Redactie:*

Ykje Wildenborg, ModeMuseum Antwerpen  
Bert Lemmens, PACKED vzw

### *Vormgeving:*

Studio Marise Knegtmans

### *Met dank aan:*

Joop Vanderheiden, Rijksdienst voor het Cultureel Erfgoed  
Trineke Kamerling, Rijksmuseum Amsterdam  
Judith van Gent, Amsterdam Museum  
Eva Coudyzer, KMKG-MRAH  
Rony Vissers, PACKED vzw  
Michiel Hildebrandt, Spinque  
Henk Vanstappen, Datable  
Reem Weda, Nederlands Instituut voor Kunstgeschiedenis (RKD)  
Wilbert Helmus, Netwerk Digitaal Erfgoed  
Dieter Suls, ModeMuseum Antwerpen  
Enno Meijers, Koninklijke Bibliotheek

### *Mogelijk gemaakt door:*

Vlaamse Overheid  
Netwerk Digitaal Erfgoed

### *Afbeelding voorblad:*

Wildschieter, Collectie Zuiderzeemuseum, Enkhuzen



## Over de projecten

### *Netwerk Oorlogsbronnen (NOB)*

Het Netwerk Oorlogsbronnen (NOB) heeft tot doel de fysiek verspreide WO2-collecties digitaal beter vindbaar en bruikbaar te maken. Zo wordt het gebruik van de oorlogsbronnen gestimuleerd en de herinnering aan de Tweede Wereldoorlog levend gehouden. NOB wordt gefaciliteerd door het NIOD Instituut voor Oorlogs-, Holocaust- en Genocidestudies en gefinancierd door het Ministerie van VWS, Vfonds en VSBfonds.

### *Netwerk Zuiderzeecollectie*

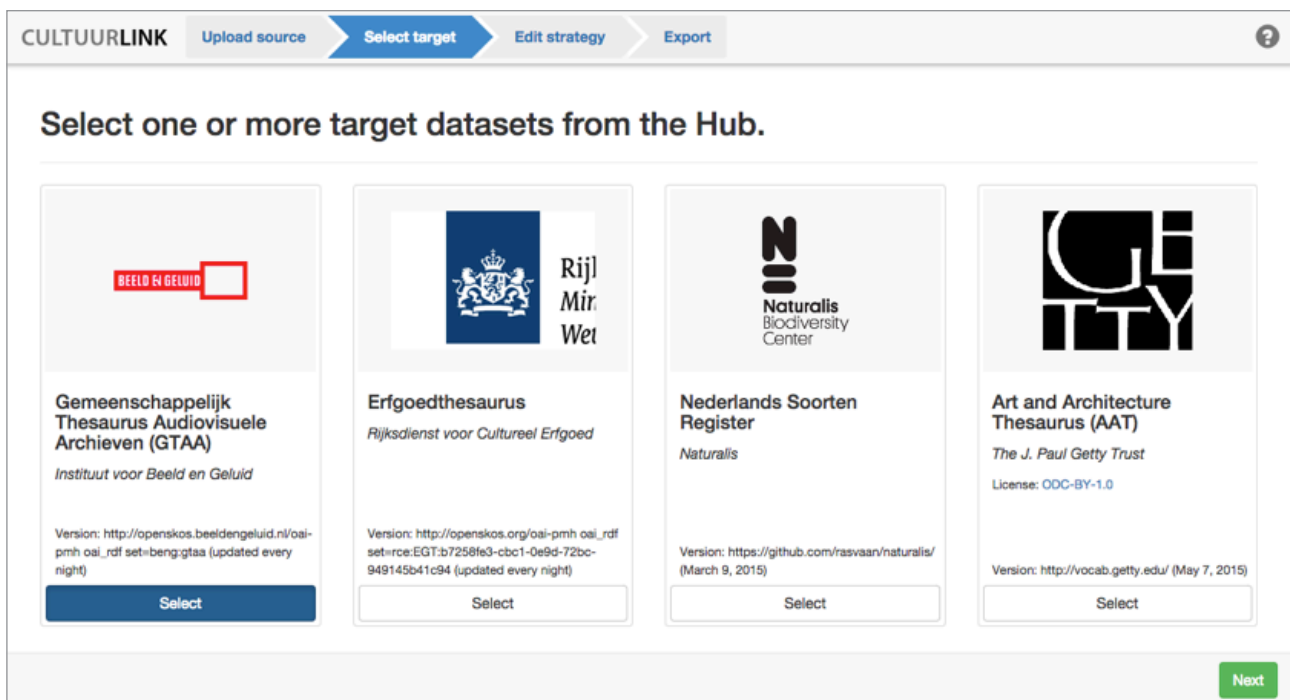
Netwerk Zuiderzeecollectie werd begin 2016 opgericht door 21 instellingen die een directe link hebben met het erfgoed van de Zuiderzee. De activiteiten van het netwerk richten zich op praktische projecten van beperkte omvang. Het doel van Netwerk Zuiderzeecollectie is het realiseren van een aantrekkelijke Zuiderzeecollectie met uitingen in het digitale domein, reizende tentoonstellingen, levendig bruikleen verkeer en een actieve uitwisseling van kennis en ervaring die in het netwerk voorhanden is. De projecten worden mogelijk gemaakt dankzij het Mondriaan Fonds, VSBfonds, Netwerk Digitaal Erfgoed (OCW), Gemeente Hoorn, Gemeente Edam-Volendam, Gemeente Bunschoten-Spakenburg en Gemeente Waterland.

### *Visuele Thesaurus voor Mode & Kostuums (VTMK)*

Het project Visuele Thesaurus voor Mode & Kostuums (november 2015 - september 2017), getrokken door het ModeMuseum in Antwerpen (MoMu), werd onder meer geïnspireerd door het portaal voor mode-erfgoed Modemuze. De VTMK heeft tot doel het gebruik van *controlled vocabularies* te motiveren binnen Vlaamse en Nederlandse musea met een collectie mode, en biedt daarvoor zowel een demonstrator, verbetering van de terminologie en een handleiding: 'Verbind je termen'. Het samenwerkingsverband werd mogelijk gemaakt door de Vlaamse Overheid en als pilotproject ondersteund door Netwerk Digitaal Erfgoed (NDE).

## ADDENDUM 1 • KOPPELEN MET CULTUURLINK

CultuurLINK is ontwikkeld door het bedrijf **Spinque**, in opdracht van het Nederlands Instituut voor Beeld & Geluid. Het is een programma, waarmee je een **lokale terminologiebron (je thesaurus) kunt verbinden aan een externe bron**. Het bij elkaar zoeken van termen gaat op woordbeeld. In CultuurLINK zijn de volgende externe terminologiebronnen opgenomen: GTAA, Erfgoedthesaurus, Soortenregister, AAT, MIMO, Iconclass, WO2-thesaurus. De eerste stap is het **omzetten van de lokale termenlijst naar SKOS**. Dit skos-bestand upload je in CultuurLINK, waarna er verschillende manieren zijn om de twee terminologieën aan elkaar te knopen.



Figuur 11: In CultuurLINK zijn de terminologiebronnen al aanwezig.

### Verskossen van je thesaurus

SKOS – of Simple Knowledge Organisation System – is een **datamodel** ontwikkeld door het World Wide Web Consortium, W3C, voor het online gebruiken van onder andere thesauri. **Een datamodel is een 'set velden' waaronder informatie kan worden geregistreerd**: in dit geval velden voor een terminologie of andere gecontroleerde overzichtslijst. De SKOS-velden zijn dan ook bestemd voor bijvoorbeeld voorkeurstermen (preferred terms), de hoger liggende termen in de thesaurus (broader terms) en de unieke termcode (identifiers) en de URI's.<sup>16</sup>

Door je eigen thesaurus te 'verskossen' geef je **skos-veldnamen** (ook wel: **attributes** in IT-taal) aan de velden van je eigen thesaurus. Alleen de voorkeursterm (woordt: skos:prefLabel) kan hiervoor al genoeg zijn. Maar je kunt het beste ook je term-ID of priref (skos:identifier) en niet-voorkeursterm (skos:altLabel) meenemen. In CultuurLINK zorgt dit ervoor dat je je lokale terminologie kunt linken met andere, gedeelde terminologiebronnen die beschikbaar zijn in SKOS. Termenlijsten (lokale terminologieën) kunnen in OpenRefine omgezet worden in SKOS door deze drie stappen te volgen:

## STAP 1: *Installatie*

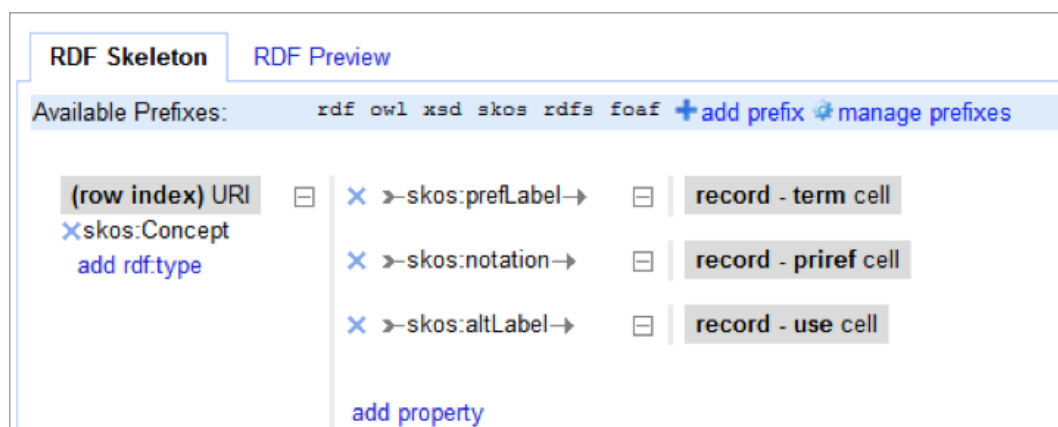
- Installeer OpenRefine; versie GoogleRefine 2.5 op <http://openrefine.org/download.html>. Installeer de extensie DERI versie 0.8 (scroll naar beneden voor de extensieslijst). Deze combinatie werkte voor ons het best.

★

Zie ook: <https://github.com/OpenRefine/OpenRefine/wiki/Installing-Extensions>

## STAP 2: *RDF/SKOS maken*

- Maak een nieuw project aan: **create project**.
- Upload je termenlijst (zie 3.2), met header (kolomnamen).
- Klik in het project rechtsboven op **RDF**, en **Edit RDF skeleton** (dit is een nieuw uitklaplijstje, geïnstalleerd door de DERI-extensie).
- **Add prefix**: tik in 'skos'; er verschijnt een bijbehorende URL waar je op klikt.
- Links klik: **add rdf:type**. Maak hiervan **skos:Concept**, dat verschijnt wanneer je het intikt.
- Midden klik: **add property**: maak hiervan **skos:prefLabel**.
- Rechts: kies voor de **naam van de kolom 'voorkeursterm' in jouw bestand**.
- Links klik: **add rdf:type**. Maak hiervan **skos:Concept**.
- Midden klik: **add property**: maak hiervan **skos:identifier**.
- Rechts: kies voor de **naam van de kolom 'identifier' in jouw bestand**.
- Doe hetzelfde voor de broader term als deze in je export aanwezig is (skos:broader).
- **Save**.



Figuur 12:  
De mapping van eigen thesaurusvelden naar SKOS-velden in OpenRefine

## STAP 3: *Export*

- Klik op Export en kies **RDF turtle**.
- Het programma opent je thesaurus in skos (controleer of je al je velden erin terugziet) in een browser.
- Klik rechtermuisknop: **Opslaan als**. Hij maakt er automatisch een Turtle-bestand van.

## Op weg met CultuurLINK

Er zijn grofweg **vijf vaste stappen** in het werkproces te onderscheiden: de externe thesaurus kiezen, een linkstrategie maken, resultaten bekijken, je linkstrategie uitbouwen en exporteren.

★

Zie ook de filmpjes op [cultuurlink.beeldengeluid.nl](http://cultuurlink.beeldengeluid.nl).

### STAP 1: De externe thesaurus kiezen

- Ga naar [cultuurlink.beeldengeluid.nl](http://cultuurlink.beeldengeluid.nl), waar je try-out sessies kunt doen. **Upload het Turtle-bestand** hier. Het kiezen van de **externe terminologiebron** kan nu, in de stap 'select target', maar ook later, in het werkblad.
- Klik **Volgende** tot je in het **werkblad** zit (zie Figuur 13).
- Om in het werkblad een externe terminologiebron toe te voegen, klik op **data source** en sleep het blokje van jouw keuze (AAT, GTAA, Iconclass...) naar het werkvlak.
- Voor de WO2-thesaurus sleep je het blokje custom naar het werkblad. Klik op het **radartje** (instellingen) van **custom** en vul bij **CHUNK** in: 'wo2'. Nu wordt de WO2-thesaurus opgehaald.

### STAP 2: Linken: string match

- De meest basale strategie is om, onder **create mapping**, het blokje '**string match**' links (hier staan de acties) naar het werkvlak te slepen (in het midden). Zie Figuur 13.
- Vervolgens sluit je de twee terminologieën aan middels het slepen van hun **outputvakje** ('**result**', onderaan de blokjes) naar de **inputvakjes** van het **string match-blokje** (A en B). Zie de Figuur: er komt een lijn te staan tussen de blokjes.

Hier kunnen verschillende acties gekozen worden

eigen thesaurus

CULTUURLINK Upload source Select target Edit strategy Export

1. Data Source >  
2. Filter source >  
3. Create mapping ▾  
String RegEx match +  
Word overlap +  
String match +  
String fuzzy match +  
String prefix match +  
4. Filter mapping >  
5. Combine >  
6. Select candidate >

Custom c1490696880710 RESULT  
Custom wo2 RESULT

String match  
skos:prefLabel  
skos:prefLabel  
RESULT

Koppelactieblokje 'String match' wordt uit de lijst geselecteerd en naar het werkvlak gesleept

inputvakje  
instellingen  
koppelactieblokje 'String match'

De inhoud van het geselecteerde blokje in het werkvlak

id	skos:prefLabel
1 http://data.niod.nl/WO2_Thesaurus_test/4893	Ohrbeck (nl)
2 https://data.niod.nl/WO2_Thesaurus/1561	Naoorlogse periode (nl)

Figuur 13:  
Interface voor het samenstellen van een matching-strategie in CultuurLINK.

- Klik op het **radartje** rechts boven op het **string match-blokje** om de instellingen van de koppelactie vorm te geven, zie Figuur 14. Deze instellingen bepalen welke koppelingen gelegd worden.
- In dit scherm kies je de velden die gekoppeld worden: bijvoorbeeld voor het matchen van de prefLabel-termen van de lokale terminologiebron aan de prefLabel-termen van de externe terminologie, en geen ‘stemming’ toepassen. Dit zijn de ‘strengste’ instellingen, waarmee je alleen de termen vindt die **exact** hetzelfde zijn.
- **OK**

Figuur 14:  
Instellingen string  
match

String match

ATTRIBUTE\_A skos.prefLabel < Het veld van thesaurus A dat gekoppeld moet worden

ATTRIBUTE\_B skos.prefLabel < Het veld van thesaurus B dat gekoppeld moet worden

USE\_SUBPROPERTIES

CASESENSITIVE

LANGUAGE\_A ANY

LANGUAGE\_B ANY

STEMMING snowball-dutch < De 'stemming' (stam) die toegepast kan worden

COMPARISON equals

OK Cancel

### STAP 3: *Eerste resultaten bekijken*

- Klik op het **result blokje van je string match-blokje**: onderaan in het beeld verschijnen dan de gemaakte links. De gekoppelde termen staan steeds onder elkaar in (meestal) twee rijen die dezelfde achtergrondkleur hebben. De bovenste rij is de term uit blokje A en de onderste rij is de hieraan gekoppelde term uit blokje B.
- Deze **koppelingen kun je beoordelen** door op het tekenje links van de rijen te klikken: je kunt kiezen voor exactMatch als het gaat om termen die exact hetzelfde betekenen.  
**Let op: je hebt aan deze controle alleen iets** als je met de RDF (Linked Data) export gaat werken. Als je voor een CSV-export kiest (wat wel het makkelijkste is) dan komen alle matches door, ook matches die niet goedgekeurd zijn. Je kunt dan dus ook pas ná de export de verkeerde matches verwijderen, in Excel bijvoorbeeld.

### STAP 4: *Je strategie verder uitbouwen*

- Meer matches kunnen gevonden worden door andere instellingen en andere ‘blokjes’ te gebruiken. **Sleep een nieuw string match-blokje** naar het werkvlak, onder het string match blokje dat al in gebruik is.
- **Verbind het nieuwe blokje met het oude** blokje door een verbindingslijntje uit de outputvakjes van het oude blokje **not A** en **not B** te slepen naar de inputvakjes van het nieuwe blokje.

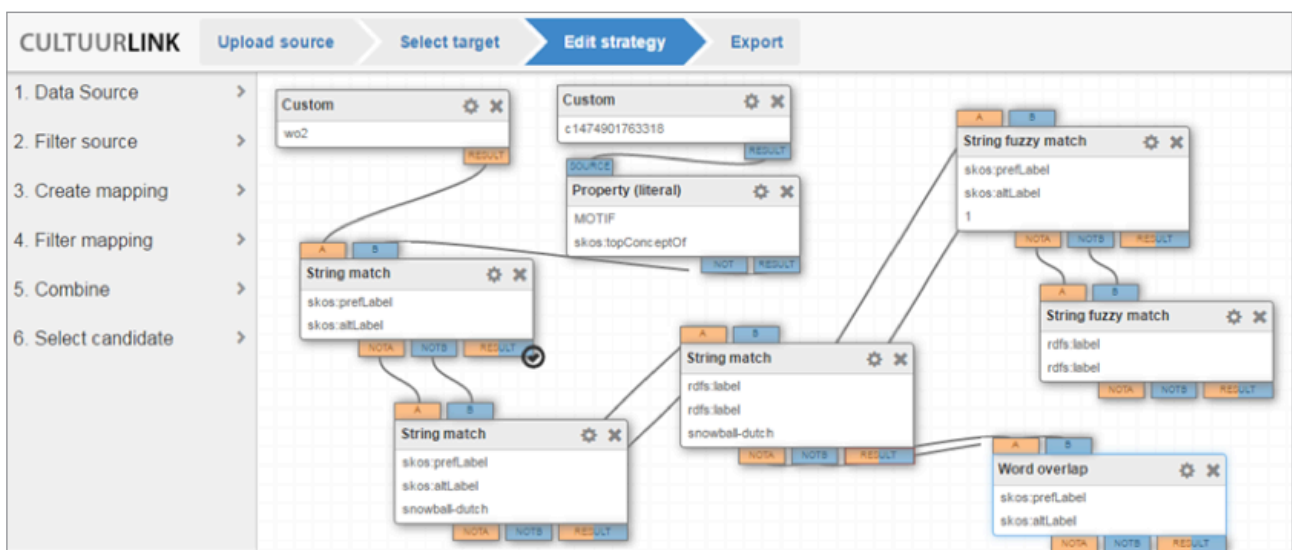


- Het nieuwe **string match-blokje** kan anders ingesteld worden, door bijvoorbeeld de niet-voorkeurstermen (altLabels) aan elkaar te matchen, of zowel de voorkeurstermen als de niet-voorkeurstermen (via rdfs:label, waarmee je de inhoud van beide velden meeneemt).

★

Als de ene thesaurus 'kazemat' gebruikt als voorkeursterm met 'bunker' als niet-voorkeursterm, en de andere thesaurus heeft 'bunker' als voorkeursterm, worden deze twee concepten niet aan elkaar gekoppeld als er alleen voorkeurstermen gematcht worden. Ze worden wel aan elkaar gematcht als de instelling 'rdfs:label', of 'altLabel aan prefLabel' gebruikt wordt.

- Een andere instellingsoptie is om **stemming** toe te passen, dat wil zeggen dat de stam van het woord gekoppeld wordt. In situaties waarin de ene term een meervoudsvorm is (bunkers) en de andere term enkelvoud (bunker), worden ze dan toch verbonden.
- Er zijn ook andere match-blokjes beschikbaar, zoals **string fuzzy match (levenshtein = 1)**. Daarbij worden ook vergelijkbare (niet exact dezelfde) termen aan elkaar verbonden.
- **Word overlap** geeft soms ook nog heel goede resultaten, maar gaat met afkortingen vaak de mist in: die worden allemaal aan elkaar gematcht. In een dergelijk geval dien je dus goed te controleren.



Figuur 15:  
Voor elke koppelmethode kan de gebruiker van CultuurLINK middels de instellingen (klik: radartje) precieze matchingscriteria bepalen.

★

Het resultaat in het werkvlak bestaat uit meerdere blokjes die aan elkaar verbonden zijn, en die elk hun eigen resultaten weergeven. De instellingen van het eerste blokje zijn het 'strengst' en worden idealiter steeds loss: het laatste blokje zal daardoor gewoonlijk de meeste foutieve matches bevatten.

★

Er zit geen overlap in de resultaten van de blokjes omdat elk volgende blok alleen de termen gebruikt die nog niet eerder gematcht zijn (dat is de 'not A' en 'not B').

Zie hier een uitgebreide matchingsstrategie van een bepaalde koppelactie:

STAP	TYPE MATCHING	MATCHINGVELDEN	INSTELLINGEN
1	String match	prefLabel: prefLabel	
2	String match	prefLabel: prefLabel	Stemming: Dutch
3	String match	prefLabel: prefLabel	Stemming: English
4	String match	rdfsLabel: rdfsLabel	
5	String match	rdfsLabel: rdfsLabel	Stemming: Dutch
6	String match	rdfsLabel: rdfsLabel	Stemming: English
7	String fuzzy match	prefLabel: prefLabel	Levenshtein = 1
8	String fuzzy match	rdfsLabel: rdfsLabel	Levenshtein = 1
9	Word overlap	prefLabel:prefLabel	
10	Word overlap	rdfsLabel:rdfsLabel	Geen stemming

Figuur 16: Schematische weergave matchingstrategie: van streng naar coulant

### *Handig: je externe doelterminologie filteren*

Zoals te zien is in het CultuurLINK-werkblad zijn er nog veel meer blokjes te selecteren. Wanneer er bijvoorbeeld aan de AAT gekoppeld wordt, is het handig om alleen het deelgebied van de AAT te selecteren waar jouw termen ook in vallen. Dit doe je via **filter mapping**. Ook de eigen, lokale terminologie kan gefilterd worden, bijvoorbeeld omdat je alleen een bepaald type termen wilt koppelen (inhoudstrefwoorden en geen geografische trefwoorden). In Figuur 17 is er gefilterd op 'topConceptOf' met de waarde 'MOTIF', omdat de inhoudstrefwoorden in die collectie het type 'MOTIF' hebben. In de vergelijking worden dan alleen die trefwoorden gebruikt.

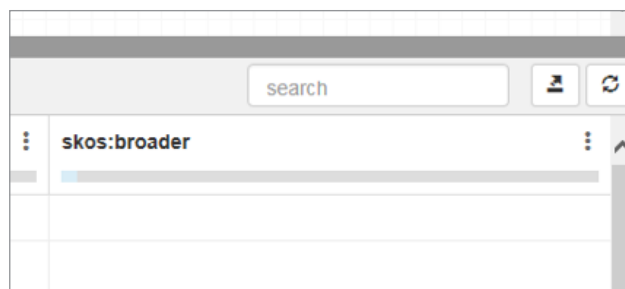
- Klik de lijst met termen (onder in het scherm) aan. Ga naar het **radartje** (linksboven de lijst).
- Selecteer in de lijst met *attributes* het veld met de term waarop je wilt filteren. Als dat veld al geselecteerd is, doe je niets.
- Klik **ok**. Ga in de termenlijst naar het **attribute** (bijvoorbeeld skos:prefLabel, of skos:broader) waarop je wilt selecteren. Klik op **de drie stipjes**, klik vervolgens **show distinct values**. Kies uit de lijst met waarden degene (het hiërarchische onderdeel van de terminologie) die je als filter wilt gebruiken en kies **add as filter**. Maak een connectie tussen het blokje van de termenlijst naar de filter en vervolgens van de filter naar de koppelacties.

id	skos:hasTopConcept	skos:prefLabel
1 <a href="http://localhost:3333/0">http://localhost:3333/0</a>	GEOKEYW	-
2 <a href="http://localhost:3333/1">http://localhost:3333/1</a>	MOTIF	
3 <a href="http://localhost:3333/10">http://localhost:3333/10</a>	PERIOD	
4 <a href="http://localhost:3333/11">http://localhost:3333/11</a>	MOTIF	1944
5 <a href="http://localhost:3333/12">http://localhost:3333/12</a>	MOTIF	1945

Figuur 17: Door een filter aan te brengen zorg je ervoor dat CultuurLINK met een beperkter deel van de terminologiebron vergelijkt.

## STAP 5: *Exporteren en gebruiken*

- Je exporteert het resultaat het best door middel van de **CSV-export**. Die functie staat rechts boven de resultaten, het pijltje met het streepje eronder.



- De **export doe je het best per blokje**: als er dus bijvoorbeeld vier verschillende blokjes gebruikt zijn, exporteer je vier keer. Je weet dan bij de evaluatie van je matches welke exact zijn en welke mogelijk controle behoeven. Als je toch in één keer wilt exporteren, kun je een blokje toevoegen met de functie 'mix mappings'.
- Open de CSV's één voor één in Excel. **Verwijder de kolommen die je niet nodig hebt**: je bewaart dus alleen je term-ID's, de URI's en eventuele andere informatie die je wilt importeren.

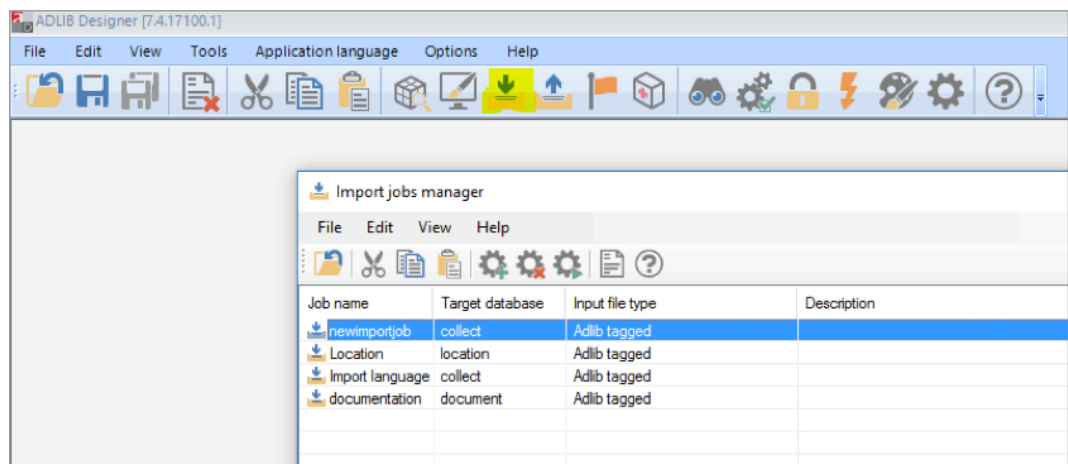
- Controleer de matches op correctheid, **verwijder de onjuiste matches**. Je kunt er dan voor kiezen om de resultaten samen te voegen in één bestand, dat is gemakkelijker importeren.
- Ga naar §4.4 of Addendum 2 voor het importeren van de links in je collectiebeheersysteem.
- Er is ook een **Linked Data export (RDF-triples)**. Dat is de beste optie als het resultaat na de export verwerkt wordt in een linked data-toepassing, maar minder handig als er een import voor een collectiebeheersysteem van gemaakt moet worden.

## ADDENDUM 2 • URI'S IMPORTEREN IN ADLIB DESIGNER

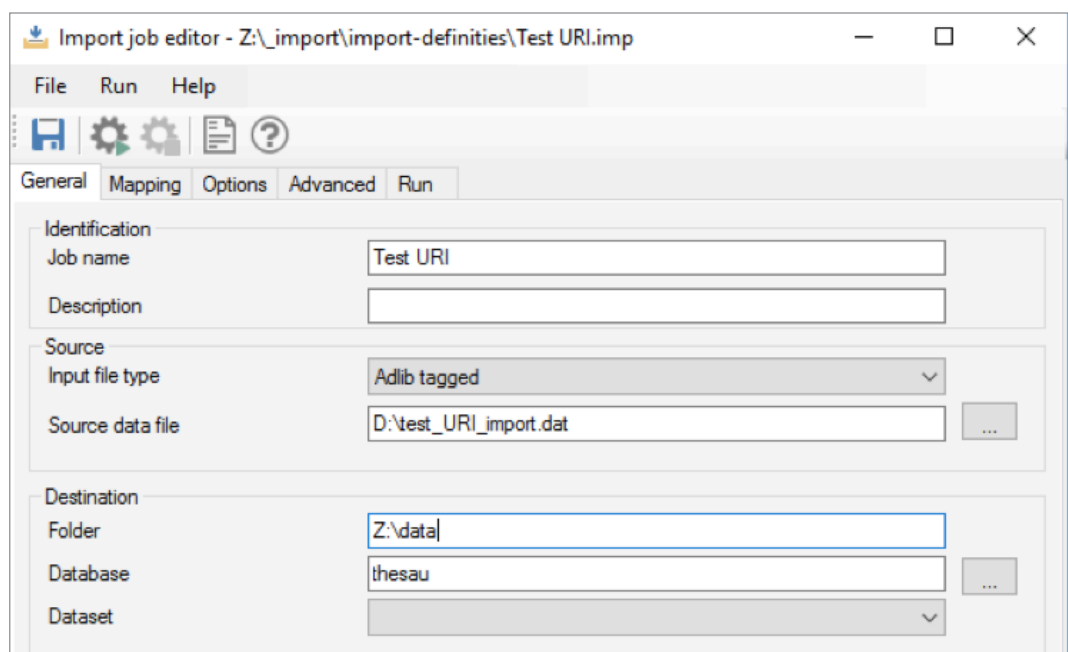
In Adlib Designer is het mogelijk om met een **CSV** ofwel met een **Adlib Tagged** imports te draaien. Dit is een meer geavanceerde werkwijze dan die uit de casus in hoofdstuk 4. Bijvoorbeeld kan zo een occurrence worden toegevoegd met behoud van de eventueel bestaande invulling van het veld. **Hieronder leggen we beide mogelijkheden stap-voor-stap uit: Adlib Tagged en CSV.**

### URI's importeren met een Adlib Tagged in Adlib Designer

- Maak een **Adlib Tagged** bestand zoals beschreven in § 4.4.
- Open **Adlib Designer**.
- Klik op **Manage import jobs**; de **Import jobs manager** opent. Als je hiermee eerder werkte, staan hier misschien al import-definities (opgeslagen eerder uitgevoerde opdrachten).

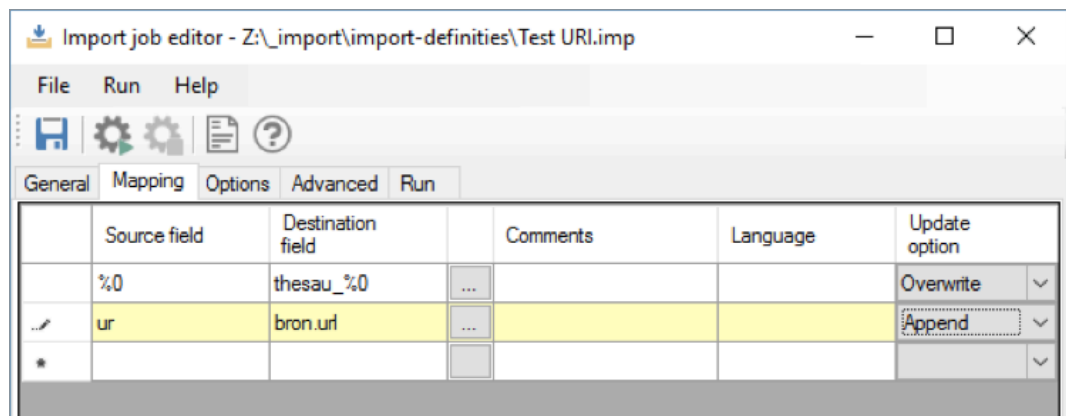


- Klik op **Create new import job**: radartje rechts boven. Dan verschijnt dit scherm:

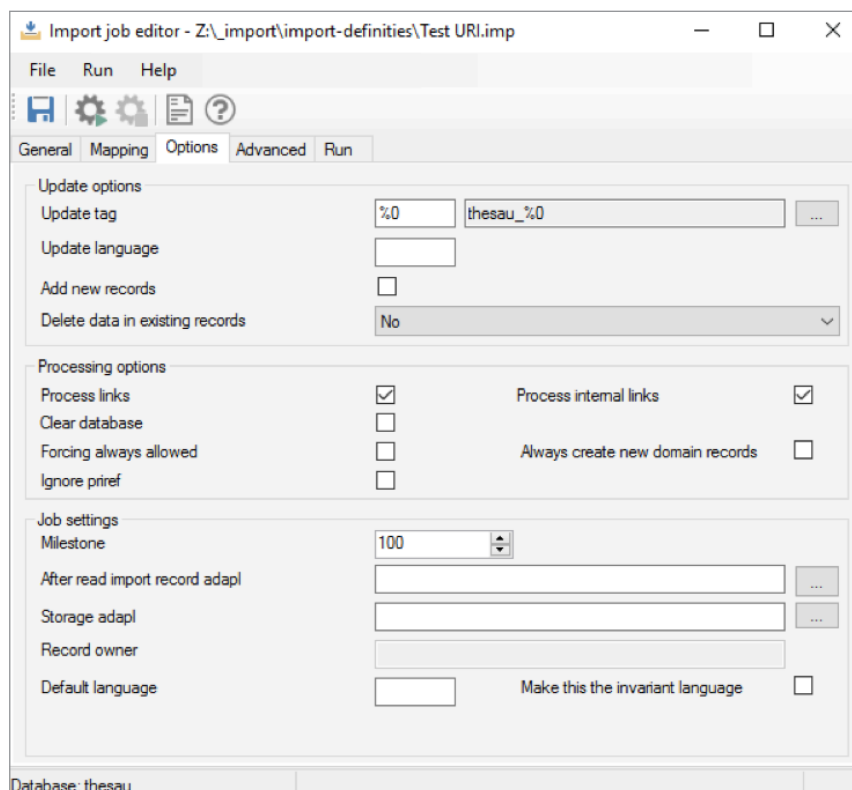




- Vul informatie over je import job in op het eerste tabblad van de Import job editor:  
**Name, Source:** klik aan **Adlib Tagged**, Kies bij **Source Data file** voor het bestand dat je zojuist maakte. **Destination:** ga naar datamap en selecteer **thesau.inf**.
- Als je al informatie in je doelveld hebt staan kies dan tabblad **Mapping**, vul dan de velden waarin je in het bestand verwijst in, en kies bij **Update option** bij het veld in kwestie voor **Append** (dan vult hij de bestaande metadata aan, met een extra *occurrence*, zonder die dus te overschrijven) of **Append if not present** (dan vult hij de metadata alleen aan als het veld nog niet is ingevuld).



- Vul op **Tabblad Options** in bij **Update tag: %0**, Vink **Add new records** uit. Laat verder de standaardinstellingen staan. **Blijf uit de buurt van het veld Clear Database, dit zou uw database geheel leegmaken!**



- Klik op Tabblad **Run**. Klik op het icoon voor **Run import job** (radar). Klik op **Ok** bij **Save current job**.
- Controleer de import **steekproefsgewijs**.

## URI's importeren met een CSV in Adlib Designer

Zie hieronder het Excel-document waarin werd gewerkt om de termen te matchen:

	A	B	C	D	E	F	G	H	I	J
1	preref	term	broader_t	preref	term	erfgoedth	erfgeo.uri	latitude	longitude	Similarity
2		9 aankoop		4185 aankoop		http://data.cultureelerfgoed.nl/semnet/				00.001
3		294 aardewerk		73 aardewerk		http://data.cultureelerfgoed.nl/semnet/				00.001
4		1110 affiche		96 affiche		http://data.cultureelerfgoed.nl/semnet/				00.001
5		2656 akte		115 akte		http://data.cultureelerfgoed.nl/semnet/				00.001
6		1413 aluminium		3752 aluminium		http://data.cultureelerfgoed.nl/semnet/				00.001
7		2462 Amersfoort		127 Amersfoort		http://sw:	52,155	5,3875		00.001
8		55 Amsterdam		134 Amsterdam		http://sw:	52,37403	4,88969		00.001
9		1013 Apeldoorn		4549 Apeldoorn		http://sw:	52,21	5,96944		00.001
10		100 aquarel		169 aquarel		http://data.cultureelerfgoed.nl/semnet/				00.001
11		2148 archeologie		175 archeologie		http://data.cultureelerfgoed.nl/semnet/				00.001
12		1065 armband		188 armband		http://data.cultureelerfgoed.nl/semnet/				00.001
13		1377 Arnhem		192 Arnhem		http://sw:	51,98	5,91111		00.001
14		265 asbak		195 asbak		http://data.cultureelerfgoed.nl/semnet/				00.001
15		1527 baadje		213 baadje		http://data.cultureelerfgoed.nl/semnet/				00.001
16		803 Baarn		220 Baarn		http://sw:	52,21167	5,2875		00.001
17		1796 bakeliet		3753 bakeliet		http://data.cultureelerfgoed.nl/semnet/				00.001

### STAP 1:

#### Voorbereiding

- **Open je document met gematchte termen** in Excel. Dit kan je export uit CultuurLINK zijn, of je export uit Fuzzy Lookup. Hier is gewerkt met een export uit Fuzzy Lookup.
- Verwijder de kolommen die je niet nodig hebt zodat je eindigt met de **prerefs van de termen uit jouw eigen Adlib-systeem** en de **URI's die je wilt gaan importeren**. Verwijder ook de bovenste rij. Het ziet er dan als volgt uit:

	A	B	C	D	E	F
1	9	http://data.cultureelerfgoed.nl/semnet/721003ba-91ff-4bbd-8641-ed2823c1e625				
2	294	http://data.cultureelerfgoed.nl/semnet/dfbf8107-09f9-49c9-8118-8e79b7613723				
3	1110	http://data.cultureelerfgoed.nl/semnet/6b41d93f-58ed-47ce-9aa4-0eb51a326d89				
4	2656	http://data.cultureelerfgoed.nl/semnet/e3cde916-fba3-4cd9-967a-fa975dd76aea				
5	1413	http://data.cultureelerfgoed.nl/semnet/da1106de-0aa0-4591-9cf7-11ac08b5bfc				
6	2462		http://sws.geonames.org/2759821/	52,155	5,3875	
7	55		http://sws.geonames.org/2759794/	52,37403	4,88969	
8	1013		http://sws.geonames.org/2759706/	52,21	5,96944	
9	100	http://data.cultureelerfgoed.nl/semnet/152a6b74-0549-4e53-aec0-f8209db88b86				
10	2148	http://data.cultureelerfgoed.nl/semnet/c32e9420-6176-42cc-a65c-25f671d1ec23				
11	1065	http://data.cultureelerfgoed.nl/semnet/fc8a2ab3-6ee5-4395-b3ec-e2e2fb68aa49				
12	1377		http://sws.geonames.org/2759661/	51,98	5,91111	
13	265	http://data.cultureelerfgoed.nl/semnet/d7db75b6-d16a-43e2-b377-91f5a2612505				
14	1527	http://data.cultureelerfgoed.nl/semnet/859cf884-35d9-407f-bd33-6fd14a9dd9f6				
15	803		http://sws.geonames.org/2759544/	52,21167	5,2875	
16	1796	http://data.cultureelerfgoed.nl/semnet/1769524f-8124-4d54-b6ad-845444089857				
17	668	http://data.cultureelerfgoed.nl/semnet/ff7f121a-5ee7-4f7f-96b7-946eb06aa502				
18	1504		http://sws.geonames.org/2759407/	52,14	5,58472	
19	707	http://data.cultureelerfgoed.nl/semnet/6f447a2e-0b20-4ea4-bb88-12ad98032fcd				
20	1018	http://data.cultureelerfgoed.nl/semnet/2035cba8-d10c-47a5-b10f-2a5295a396b2				
21	1554	http://data.cultureelerfgoed.nl/semnet/46197e42-144e-4d6f-8f16-a00dc48bb09f				
22	2662	http://data.cultureelerfgoed.nl/semnet/0741920b-94c2-4d93-9deb-71701a77f520				
23	1153	http://data.cultureelerfgoed.nl/semnet/9837eab4-12cf-4d4b-85e2-a4f2237e2a43				
24	1551	http://data.cultureelerfgoed.nl/semnet/58abf0fc-d62b-45b6-af92-77caa797b263				
25	211	http://data.cultureelerfgoed.nl/semnet/ceb91b0c-75e8-4063-a01d-eeb27e4e7941				
26	642					
27	746		http://sws.geonames.org/2759197/	51,99833	5,67639	
28	1708	http://data.cultureelerfgoed.nl/semnet/c2cb71db-70a0-4a85-b8b7-da6fc5bc48c5				
29	1454	http://data.cultureelerfgoed.nl/semnet/7ebaf5a0-7a95-4ae0-81ee-2db87df90041				
30	1978	http://data.cultureelerfgoed.nl/semnet/d6edff13-b994-48e8-aad9-31a7c5c42572				

- Voer altijd eerst een test uit van 3 of 4 regels! Mocht er iets aan de import niet kloppen dan zie je dit meteen en zit je niet gelijk vast aan honderden foute records. Ook is het handig om een kopie van je applicatie te hebben als 'test'-versie om dergelijke importjobs uit te testen. **Maak dus een kopie** van de eerste vier regels van je bestand.
- Sla dit bestand op als een **CSV (gescheiden door lijstscheidingstekens)** en sluit het bestand (importeren lukt niet als het bestand open staat).
- Open **Adlib Designer**

## STAP 2: *De import configureren*



- Ga naar het icoontje **Beheer import procedures**:
- Ga naar **Bestand** en dan **Nieuw**.
- Geef de importjob een **naam** en een **beschrijving**.
- Geef bij **Soort bronbestand** aan dat het een **ASCII gescheiden (\*.csv)** is.
- Kies bij **Bronbestand** de .csv die je zojuist hebt gemaakt.
- Bij **Bestemming** ga naar **Database** en klik op het **icoontje met drie puntjes**. Selecteer daar de map waar de Adlib applicatie staat opgeslagen. Ga naar de map **Data** en selecteer het bestand **thesau.inf**. We willen de URI's namelijk gaan importeren in de database Thesaurus.
- Ga naar het **tabblad Mapping**. Hier geef je aan welke gegevens je naar een bepaald veld in Adlib wilt gaan importeren. Met de cijfers o1, o2, o3, enzovoorts geef je de kolommen (A, B, C, enzovoorts) in je .csv aan.
- Door op het **icoontje met de drie puntjes** achter **Bestemmingsveld** te klikken geef je aan naar

welk veld je de informatie in die specifieke kolom wilt gaan importeren. Hierboven zien we dat de 1e kolom de priref bevat. Dan wordt dan dus **Bronveld 01** en het Bestemmingsveld wordt **%0**, want dat is in Adlib de tag voor de priref.

- **Bronveld 02**, met de URI, importeer je naar het veld waar je de URI wilt terugzien.
- Heb je dit alles gedaan dan sluit je af met een puntkomma als field separator <FS> en oxDA als rule separator <RS> en ziet het er als volgt uit:

Bronveld	Bestemmingsveld	Commentaar	Taal	Bijwerk optie
01	%0			Vervangen
02	erfgoedthesaurus...			Vervangen
03	erfgeo.uri			Vervangen
04	latitude			Vervangen
05	longitude			Vervangen
.	<FS>			Vervangen

- Ga naar het tabblad **Opties** en vul de velden als volgt in:

Opties voor bijwerken

Vergelijkingstag: %0

Update taal: [ ]

Nieuwe records toevoegen:

Wis alle gegevens in een record vóór bijwerken: Ja, alleen velden uit de veldenlijst

Opties tijdens de verwerking

Koppelingen verwerken:       Interne koppelingen verwerken:

Wis database vooraf:

Forceren altijd toestaan:       Maak altijd nieuwe domein records aan:

Priref negeren:

Instellingen van de taak

Voortgangsteller: 10

Adapl na lezen import record: [ ]

Adapl vóór opslag: [ ]

Record eigenaar: [ ]

Standaardtaal: [ ]      Stel in als invariante taal:

Database: thesau

### STAP 3:

### *Import & nazorg*

- Het valt aan te bevelen om een **adapl** voor na lezen import record toe te voegen waarin je vermeldt welke wijzigingen je hebt aangebracht (met naam gebruiker, datum, naam CSV-bestand).<sup>17</sup> Een andere optie is om in de mapping, en dus ook het .csv, nog een extra kolom toe te voegen waarin je melding maakt van de import en die te mappen naar het veld 'Opmerkingen' of 'Bijzonderheden'.
- Ga vervolgens naar **Uitvoeren** en klik op het **radar-icoontje** om de import te starten.



- **Controleer de import** steekproefsgewijs.

## ADDENDUM 3 • EENVOUDIG OPSCHONEN

Elke instelling heeft **eigen invoerregels** en elke persoon een eigen interpretatie. Zo kan dezelfde term in **vele varianten** in de objectrecords terechtkomen. In veel lokale terminologieën staan termen met spelfouten en extra spaties, verschillende spellingsvarianten van dezelfde term en termen die bestaan uit meerdere woorden, die als losse trefwoorden bedoeld zijn. In jargon spreekt men terecht van ‘messy data’.

**Metadata opschonen** kan voor sommigen van ons dan ook zomaar leiden tot een sprong terug in de tijd. Het is immers net als je kamer opruimen - het **kost tijd en het lijkt niet urgent**. Het is geen geheim dat je spullen sneller vindt wanneer je hebt opgeruimd. Hetzelfde geldt voor data. Er is **geen goed overzicht** terwijl de kans groter wordt dat er records ‘vergeten’ worden omdat ze niet naar boven komen in zoekacties. Bovendien: **links zijn gemakkelijker te leggen met consistente metadata**.

Bij het opschonen van de thesaurus worden spelvarianten van dezelfde term aan elkaar gekoppeld, extra spaties verwijderd, termen met spelfouten verbeterd of verwijderd en samengestelde termen gesplitst. Het kan ook een **goed moment zijn om inhoudelijk naar de toekenning van termen te kijken**: zijn de invoerregels consistent en zijn ze ook goed toegepast? Wat is bijvoorbeeld het inhoudelijk verschil tussen het veld ‘materiaal’ en het veld ‘techniek’? Worden de kleuren van een werk in een gecontroleerd veld beschreven, en is dat echt nodig? Hoe is de naamsopbouw bij personen: achternaam eerst, met voorletters of voornamen en waar komen titels? Zo kan men door het eenmalige opschonen van data **in de toekomst fouten voorkomen**.

### Opschonen voorbereiden

**Opschonen gaat het gemakkelijkst als je al een opgeschoonde lijst hebt**, waarmee je je eigen data kunt vergelijken (dit kan ook een terminologie zijn waarmee je later wilt linken). Een vergelijking, bijvoorbeeld met Fuzzy Lookup, leidt immers tot een overzicht van gebruikte termen, die niet voorkomen in de opgeschoonde lijst. Dat **‘residu’** geeft een duidelijk zicht op waar de ruis, **de mogelijke vervuiling**, zich bevindt in je dataset. Dit betekent dus, dat een deel van het linkproces kan dienst doen als opschoonproces. Maar **ook als je liever handmatig, of zonder vergelijking, de opschoning aangaat, heb je iets aan wat hieronder volgt**.



*Het kan helpen, om het werk in eerste instantie per veld én per deelcollectie aan te pakken.*



*Deze acties vergen soms samenwerking tussen databeheerder en domeinspecialist.*



*Hoe dit proces exact verloopt is afhankelijk van de collectie, het veld, de huidige inhoud van het veld, de nieuwe regels die je wilt toepassen, enzovoort. De mate waarin dit aangepakt wordt is ook afhankelijk van de staat van de digitale registratie en de wensen van de instelling.*

---

Heb je geen referentielijst, maar wil je er eentje maken? We adviseren om jouw lijst dan van tevoren te reduceren met OpenRefine’s Cluster & Edit-functie.<sup>18</sup> OpenRefine is een open source tool die op het eerste gezicht lijkt op Excel, maar meer kan met tekstuele data en zeer goed is gedocumenteerd.<sup>19</sup> De **‘Cluster & Edit’-functie** in dit programma is ideaal als je lijst te lang is om overzichtelijk te zijn, omdat hiermee op elkaar gelijkende woorden naar één van de spelvarianten



kunnen worden overgezet (zie Figuur 18). Dit programma is hierdoor relevant voor musea die geen opgeschoonde lijst ter beschikking hebben. Voordat we op de stappen ingaan geven we eerst enkele wenken die kunnen helpen bij het proces.

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
8	6290	<ul style="list-style-type: none"> <li>• lid verzet (5321 rows)</li> <li>• Lid verzet (957 rows)</li> <li>• Lid Verzet (5 rows)</li> <li>• lid Verzet (3 rows)</li> <li>• Lid verzet (1 rows)</li> <li>• Lid verzet ? (1 rows)</li> <li>• lid verzet (1 rows)</li> <li>• lid verzet. (1 rows)</li> </ul>	<input type="checkbox"/>	lid verzet
6	76	<ul style="list-style-type: none"> <li>• Hulpbesteller P.T.T. (63 rows)</li> <li>• Hulpbesteller PTT (9 rows)</li> <li>• Hulpbesteller P.T.T. (1 rows)</li> <li>• Hulpbesteller PTT. (1 rows)</li> <li>• P.T.T. Hulpbesteller (1 rows)</li> <li>• P.T.T. hulpbesteller (1 rows)</li> </ul>	<input type="checkbox"/>	Hulpbesteller P.T.T.

Figuur 18:  
De 'Cluster&edit'-functie van OpenRefine.

## Veel voorkomende 'vervuiling'

Musea bepalen zelf hoe zij hun database invullen. Er zijn echter enkele regels die voor alle instellingen van toepassing kunnen zijn :

- **Gebruik verplichte richtlijnen** voor de invulling van de velden.
- **Trailing whitespaces:** probeer geen spaties achter of voor een term in te voeren. Hoewel onzichtbaar voor het oog, ziet een computer de term als nieuw en anders.
- **Field overloading/multi-valued cells.** Gebruik niet meer dan één waarde binnen dezelfde term (*occurrence*). Meerdere waarden moeten apart in een nieuw veld komen.

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
3	3	<ul style="list-style-type: none"> <li>• DUIKEN-LO JODEN (1 rows)</li> <li>• DUIKEN/LO JODEN (1 rows)</li> <li>• JODEN DUIKEN/LO (1 rows)</li> </ul>	<input type="checkbox"/>	DUIKEN-LO JODEN
2	2	<ul style="list-style-type: none"> <li>• EDUCATIE MUSEA-VMF (1 rows)</li> <li>• EDUCATIE MUSEA/VMF (1 rows)</li> </ul>	<input type="checkbox"/>	EDUCATIE MUSEA-VMF
2	2	<ul style="list-style-type: none"> <li>• BESTUUR-MG ZUIVERING-POD (1 rows)</li> <li>• BESTUUR/MG ZUIVERING/POD (1 rows)</li> </ul>	<input type="checkbox"/>	BESTUUR-MG ZUIVERING-POD
2	2	<ul style="list-style-type: none"> <li>• OVW-BATALJON FRIESLAND (1 rows)</li> <li>• OVW/BATALJON FRIESLAND (1 rows)</li> </ul>	<input type="checkbox"/>	OVW-BATALJON FRIESLAND
2	2	<ul style="list-style-type: none"> <li>• NA 1945 VERZET (1 rows)</li> <li>• verzet na 1945 (1 rows)</li> </ul>	<input type="checkbox"/>	NA 1945 VERZET
2	2	<ul style="list-style-type: none"> <li>• BEVRIJDING LEGERS (1 rows)</li> <li>• LEGERS BEVRIJDING (1 rows)</li> </ul>	<input type="checkbox"/>	BEVRIJDING LEGERS
2	2	<ul style="list-style-type: none"> <li>• FOTOGRAFIE PERSONEN HAYE</li> </ul>	<input type="checkbox"/>	FOTOGRAFIE PERSONEN HAYE

Figuur 19:  
Samengestelde termen kunnen beter vooraf gesplitst worden. 'Legers' en 'bevrijding' zijn twee verschillende concepten.



Let betreft field overloading ook op samengestelde termen. Het lijkt één term, maar het zijn er eigenlijk twee. Hierbij is scheidslijn, van wat goed en verkeerd is, erg smal. Trouwjurk is een samenstelling die vaak voorkomt; het staat in woordenboeken en kan daarom worden gebruikt. Dit is een **begripseenheid**. Voor trouwpantalone, trouwoverhemden of trouwondergoed is dat niet het geval. Daarvoor gebruik je beter twee waarden: 'trouwkleding' en 'pantalon' enz.



Riesthuis en Wijnands zeggen hierover in Thesaurusbouw (1992): 'In het algemeen kan men stellen dat samenstellingen die in het mondelinge en/of schriftelijke verkeer steeds als eenheid worden gepresenteerd, aangemerkt worden als **begripseenheid**.'

- **Homoniemen:** Dit zijn woorden die precies hetzelfde geschreven worden, maar iets anders betekenen. In geval van meertaligheid is dit extra belangrijk. Gebruik in dat geval altijd een 'qualifier' tussen haakjes.

OBJECTNAAM	AANTAL OBJECTEN
Hemd (mouwloze onderkleding)	12
Hemd (bovenkleding)	210

- **Spelvarianten of -fouten:**

OBJECTNAAM	AANTAL OBJECTEN
T-shirt	53
T-shirts	65
tshirts	1

- **Waarde die in een ander veld thuishoort**, zoals hier respectievelijk 'beschrijving' en 'technieken':

OBJECTNAAM	AANTAL OBJECTEN
kraag (in uitvoering)	1
bedrukt textiel (ikat)	1

- **Spelvarianten** wanneer het om hetzelfde gaat:

OBJECTNAAM	AANTAL OBJECTEN
Hes	32
Hesje	1
hesjes	6
hessen	48

- **Let op! Soms lijken synoniemen geenszins op elkaar.** Voor deze gevallen is het handig om de hulp van een domeinspecialist aan te spreken. Hier wordt steeds hetzelfde type object bedoeld:

OBJECTNAAM	AANTAL OBJECTEN
salopet	1
salopette	3
salopettes	2
tuinbroek	10
tuinbroekje	2

## Stappenplan opschonen

### STAP 1: *Zorg voor een referentielijst*

★

Buiten je collectiebeheersysteem werken is handiger omdat je dan aantekeningen kunt maken en je voortgang kunt bijhouden. Dit doe je bijvoorbeeld in Excel.

★

Je kunt ook een Excel-lijst genereren met het aantal keer dat een term wordt gebruikt, en die alfabetisch of op aantal sorteren. Weinig gebruikte termen duiden immers vaak op de ongeldigheid ervan.

- Maak een **CSV-export** van **prirefs** en **thesaurustermen** uit je collectiebeheersysteem, die je wilt aanpakken (zie §4.2). Open deze in Excel.
- Zorg dat de lijst waarmee je wilt vergelijken ook in een Excel-bestand staat.
- Deze lijsten vergelijk je met elkaar (bijvoorbeeld met Fuzzy Lookup, zie §4.3).
- **De termen die niet voor 100% zijn gematcht**, gebruik je als basis voor **Stap 2**.
- **Wil je niet werken met een referentielijst?** Kopieer je kolom met termen dan, en ga in die tweede kolom de termen die je niet wilt houden handmatig aanpassen, naar de gewenste voorkeursterm. Ga daarvoor meteen door **naar punt 3 van Stap 2: Analyseer de vervuiling**.

### STAP 2: *Analyseer de vervuiling*

- Controleer de ‘bijna-matches’ (1-90%): dit is een deel van je opschoningswerk! **Verwijder** de term in de tweede kolom, wanneer dit niet de beoogde opgeschoonde term is. Zoals in het voorbeeld hieronder, waarin verschillende matchings onjuist zijn.

baby draagband	draagband	0,9167
smoking	smoking (suit)	0,9077
kuitlaarzen	kuitlaars	0,8753
zomerjurk	zonnejurk	0,8386
baboesjka	babushka	0,8386
schooldas	schooltas	0,9643
trainingsjack	trainingspak	0,9431
mouwloos	mouwloos jasje	0,9333
trouwkleding	rouwkleding	0,9327
multiple (kostuum)	kostuum	0,9231

- Zet alle termen waarbij dan nog geen voorkeursterm staat apart.
- **Neem een kijkje in de lijst overgebleven termen** om te zien welk type fout veel voorkomt (zie ook hierboven voor veel voorkomende onjuistheden). Plaats deze bijvoorbeeld op verschillende tabbladen.
- Maak een **overzicht van de algemene aanpassingen** en overleg dit met de collega's met kennis van het domein dat je aanpakt.

**STAP 3:** *Maak je eigen stappenplan en bedenk nieuwe regels*

- **Bepaal een eigen stappenplan:** de stappen waarmee je de gevonden fouten zal verbeteren in je collectiebeheersysteem.
- Bedenk op basis hiervan goede regels, **opdat je in de toekomst consistente invoer krijgt** (stem af met je collega's). Pas als nodig de **lokale registratieregels** aan.

**STAP 4:** *Vul de opgeschoonde termen in*

- **Plaats, in de kolom achter de vervuilde termen de juiste voorkeursterm**, geholpen door de **nieuwe registratieregels en de referentielijst** (of vraag hulp aan een domeinspecialist/conservator).



Een deel van het overzicht kan gaan over *meervoudige, te splitsen termen*.

---

**STAP 5:** *Voer de opschoning door in je collectiebeheersysteem*

**Je voert je stappenplan uit in je collectiebeheersysteem: pas de huidige objectenrecords/thesaurus aan** (zie Addendum 4 voor uitleg voor het collectiebeheersysteem Adlib). Zorg ervoor, dat je aanpassingen staan geregistreerd (bijvoorbeeld in je Excel-bestand) en houd je voortgang bij.

## ADDENDUM 4 • OPSCHONING VERWERKEN IN ADLIB

Als je een Excel-document hebt, kun je de opschoning **handmatig** of **automatisch** uitvoeren.

**Handmatig** opschonen is een veilige en overzichtelijke methode als je **gevalideerde termen** wilt opschonen; hieronder staat deze uitgelegd. In Adlib zijn **onderstreepte velden** gevalideerd, namelijk in de **Thesaurus** of in de **Personen en Instellingen**.



*Gecombineerde termen moeten worden opgelost door de term te zoeken en een tweede occurrence aan te maken voor het tweede deel van de gecombineerde term, zie Figuur 20.<sup>20</sup> Hierop gaan wij hier niet verder in.*

*Figuur 20:*

*Field overloading bij dit museum (kolom links) zorgt dat de termen niet aan een terminologie kunnen worden gelinkt (kolom rechts). Fluweel en plakkaatverf dienen bijvoorbeeld gesplitst te worden; fruit/compôtebord ook, in de aparte waarden fruitbord en compôtebord.*

611	1469	fluweel & plakkaatverf	3802	fluweel
612	1421	fluweel&plakkaatverf	3802	fluweel
613	1434	fluweel&plakkaatverf	3802	fluweel
614	2346	foto (kleurenfoto)	881	foto
615	2338	foto (kleurenfoto)	881	foto
616	2380	foto/ krantenartikel	881	foto
617	2298	fotopapier (prentbriefkaart)	2368	prentbriefkaart
618	2301	fotopapier; prentbriefkaart	2368	prentbriefkaart
619	1858	fruit/compôtebord	900	fruit
620	297	gebakken & geglazuurd	948	geglazuurd
621	2333	gebeurtenis & topografie	3043	topografie

**Automatisch opschonen met een import** leent zich voor niet-gevalideerde termen, zoals datumvelden of als je bijvoorbeeld voornaam, achternaam, en tussenvoegsels en e-mailadres in aparte velden in de Personen en Instellingen wilt zetten. Als je er al enige ervaring mee hebt, kun je het best een **CSV** of **Adlib Tagged-bestand** maken van je resultaat in Excel of OpenRefine. Met het recordnummer als vergelijkingstag (%o) kun je verbeteringen importeren. Dit gaat vergelijkbaar aan het importeren van URI's: zie §4.4.



*Kun je een Excel- of Adlib-tagged-bestand met geschoonde termen importeren in de Thesaurus of de Personen & Instellingen? Ja, maar dat is gespecialiseerd werk. Die import moet bij namelijk de juiste relatie leggen, maar bij nog niet bestaande termen ook nog een nieuw record aanmaken met het juiste domein. Je zult hiervoor een import-job in Adlib Designer moeten maken. Wil je meerdere alternatieve termen per voorkeursterm importeren, dan ontcom je niet aan een import-adapl bij die importjob. Bereid dit soort acties goed voor of roep de hulp in van mensen met ervaring.*

### Handmatig opschonen: Voorkeurstermen en Niet-voorkeurstermen

**Handmatig opschonen** houdt in dat je een relatie legt tussen één of meer vervuilde termen en de juiste term voor dat begrip. Het voordeel van deze aanpak is dat je gegevens beter vindbaar worden, én dat je niet alleen bestaande invoer opschooft, maar dat ook toekomstige invoer schoner wordt.

## Hoe het werkt achter de schermen van Adlib: een voorbeeld

Je hebt bijvoorbeeld een **schets van Vincent van Gogh** in je collectie, en je organisatie registreert vervaardigers altijd in de vorm 'Achternaam, voornaam tussenvoegsels'. Maar zowel publiek als registratoren weten dat niet altijd. Het publiek zoekt bijvoorbeeld op *Ven Go* of op *Vincent*. Een registrator kan *Van Gogh* invoeren.

Nu eerst wat terminologie. De **juiste term** heet **voorkeursterm** of **preferred term (PT)**. In het voorbeeld is dat:

- *Gogh, Vincent van*

Een **vervuilde term** heet **niet-voorkeursterm** of **non-preferred term (NPT)**. We hebben drie NPT's:

- Vincent
- Ven Go
- Van Gogh

Elke niet-voorkeursterm kan maar één voorkeursterm hebben, maar een voorkeursterm kan meerdere niet-voorkeustermen hebben.

## Use en Used for

In Adlib werkt het zo: zodra je in een Thesaurus- of een Personen en Instellingenrecord het veld *Used for* (gebruikt voor) invult, dan gebeurt er een aantal dingen:

- Dat record wordt een voorkeursterm.
- De term die je invult in *Used for* wordt een niet-voorkeursterm.
- Als die term nog niet bestaat in Adlib, wordt er automatisch een record voor aangemaakt. Dat heet een **gespiegeld record**.
- Als die term al wel bestaat, wordt in alle records de niet-voorkeursterm vervangen door de voorkeursterm (mits het **domein**, de **soort term**, overeenkomt, in dit geval **vervaardiger**).

Omgekeerd werkt precies zo. Vul je *Use* (gebruik) in, dan wordt de term die je daar invult, de voorkeursterm, en het record waarin je werkt, de niet-voorkeursterm. En die wijziging wordt ook Adlib-breed verwerkt.

## Het resultaat van deze handeling: Van Gogh

De snelste werkwijze is dat je in het record met de term *Gogh, Vincent van* drie occurrences van **Used for** aanmaakt, één met 'Vincent', één met 'Ven Go' en één met 'Van Gogh'.

- Alle records die vervaardiger **Ven Go** of **Vincent** of **Van Gogh** hadden, hebben nu **Gogh, Vincent van** als vervaardiger. Je database is dus schoner!
- Als iemand in de toekomst ooit 'Ven Go' invult bij vervaardiger, wordt die term automatisch vervangen door 'Gogh, Vincent van'.
- **Een registrator die twijfelt over Van Gog**, ziet een stoptekentje bij **Van Gogh**. Kiest hij die, dan



krijgt hij automatisch de juiste term.

- Het maakt niet meer uit **met welke term iemand zoekt op vervaardiger**. 'Ven Go' of 'Vincent' of 'Van Gogh', alle records met vervaardiger 'Gogh, Vincent van' worden gevonden.

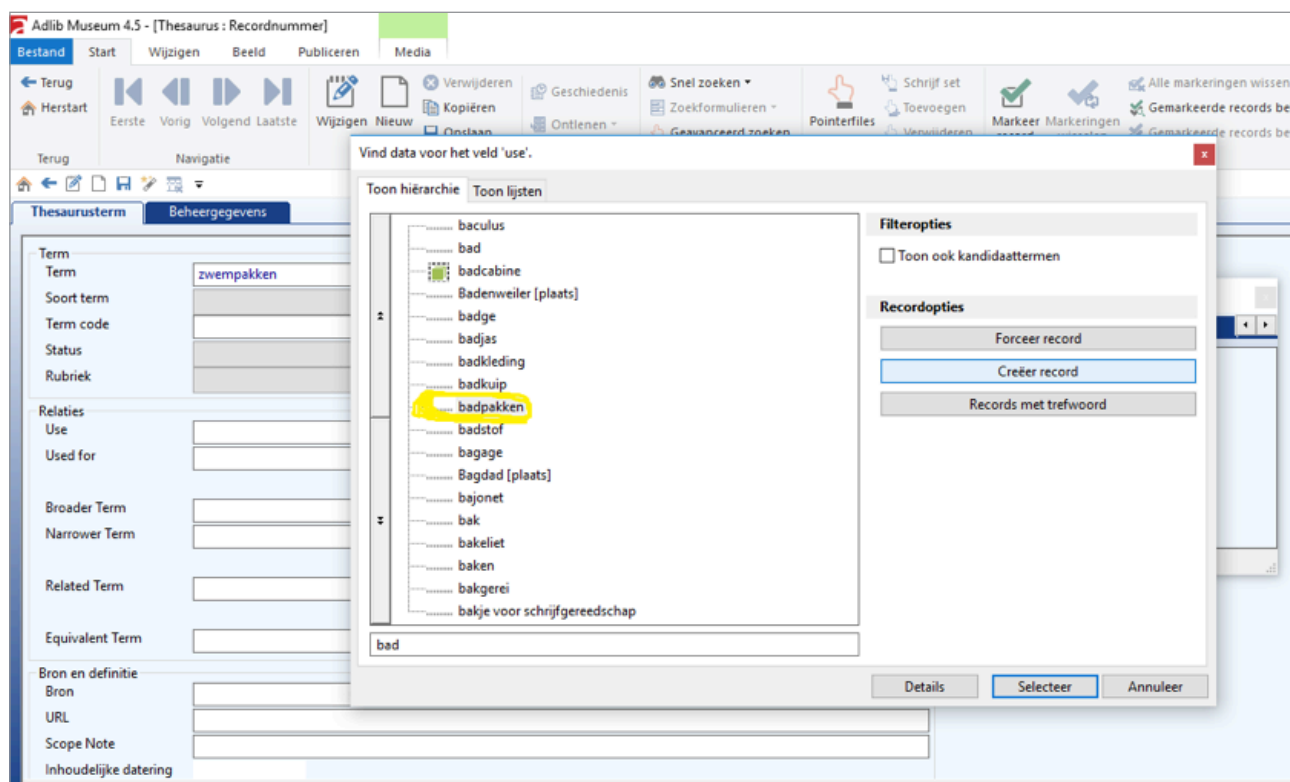
## Stappenplan handmatig opschonen

- Zorg, dat je de lijst met vervuilde termen en de bijbehorende geschoonde termen in handen hebt.
- Open de **Thesaurus**.
- **Ga naar** een vervuilde term, in ons voorbeeld 'zwempakken'.
- Klik bij **Use** in de balk.
- **Zoek in de lijst naar de juiste voorkeursterm en selecteer** deze. In ons voorbeeld 'badpakken'.
- **Sla op**.

★

*Nu gaat Adlib 'Feedbackitems berekenen'. Dit kan lang duren! Laat het systeem gewoon doorgaan, ook al kan het lijken of Adlib gecrasht is!*

- **Gefeliciteerd**, je hebt één thesaurusrecord geschoond.



## Noten

- 1  
'Non-techy people' is ontleend aan Marcia Lei Zeng's presentatie over microthesauri: [https://www.getty.edu/research/tools/vocabularies/microthesauri\\_zeng.pdf](https://www.getty.edu/research/tools/vocabularies/microthesauri_zeng.pdf). We verkiezen deze term boven het in gebruik zijnde 'non-experts'.
- 2  
Zie het colofon voor meer informatie over de projecten.
- 3  
Op de plaats van terminologiebronnen kun je ook lezen: lijsten van (kunstenaars)namen of plaatsen.
- 4  
Bijvoorbeeld: <https://rkd.nl/nl/explore/artists/66219>
- 5  
Het is al mogelijk om zelf een koppeling te implementeren in Adlib. Dit staat hier uitgelegd: [http://documentation.axiell.com/alm/en/index.html?ds\\_generaltopicsadvancedextsources.html](http://documentation.axiell.com/alm/en/index.html?ds_generaltopicsadvancedextsources.html). Rolf Blijleven heeft deze tekst gereviseerd ter facilitering van het gebruik.
- 6  
<http://projectcest.be>, <http://www.depotwijzer.be/thesauri>, <http://den.nl>.
- 7  
Het netwerk is in 2014 gestart op initiatief van het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW).
- 8  
Het technisch beheer ligt in Nederland in handen van vijf knooppunten: Koninklijke Bibliotheek, Nederlands Instituut voor Beeld en Geluid, Rijksdienst voor Cultureel Erfgoed, Koninklijke Nederlandse Akademie van Wetenschappen, Nationaal Archief.
- 9  
Het is ook mogelijk om andersoortige relaties op te nemen binnen je collectiebeheersysteem, zoals related match, narrow match en close match. Vanwege het karakter van deze introductiehandleiding kiezen we ervoor dit hier niet te behandelen.
- 10  
Zie [https://www.projectcest.be/wiki/Publicatie:Vocabularies\\_selecteren](https://www.projectcest.be/wiki/Publicatie:Vocabularies_selecteren)
- 11  
De AAT kan verrijkt worden door termen aan te bieden bij het Nederlands Instituut voor Kunstgeschiedenis (RKD). Dit beheert de Nederlandse versie en nieuwe termen met een redactieraad beoordeelt. Voor meer informatie zie <http://website.aat-ned.nl/toelichting-op-de-aat/richtlijnen>. In België kan contact worden opgenomen met het Koninklijk Instituut voor het Kunstpatrimonium (KIK-IRPA).
- 12  
Zie meer op: <https://datahub.io/dataset,w3.org/2001/sw/wiki/SKOS/Datasets> <https://datahub.io/dataset,w3.org/2001/sw/wiki/SKOS/Datasets> en [den.nl/terminologiebronnen](http://den.nl/terminologiebronnen)
- 13  
Daarbij is het ook in staat om automatisch te linken met terminologiebronnen op het web en heeft het diverse exportfuncties. Voor meer informatie zie Hoofdstuk 5, Literatuur.
- 14  
Op <http://www.getty.edu/research/tools/vocabularies/aat/> zoek je een term, klik op 'Semantic view'. In de zoekbalk en ook bovenin, direct onder de naam achter 'Source', vind je de URI.
- 15  
Prifref is de identifier (van de thesaurusterm) in Adlib.
- 16  
Zie <https://www.w3.org/2004/02/skos/>
- 17  
Zie voor meer informatie <http://www.adlibsoft.nl/producten/designer>.
- 18  
Zie <https://github.com/OpenRefine/OpenRefine/wiki/Clustering>.
- 19  
Zie o.a. de handleiding in de literatuurlijst.
- 20  
Lees meer over het gebruiken van *occurrence* in de Adlib gebruikersgids (p. 163): <http://www.adlibsoft.nl/Content/Documentation/Dutch/Adlib%20Gebruikersgids%20A5.pdf>