# **Collector voor WikiData**

Collector is een conversieprogramma dat specifiek werd ontworpen om gegevens uit de Brocade-databank om te zetten naar een door de gebruiker bepaald template dat een snelle upload naar WikiData toelaat. Veel van de stappen die bij het opstellen van een dergelijke ‘Quickstatement’ uitgevoerd moeten worden zijn al geautomatiseerd door Collector. Toch is er nog enige kennis van het proces vereist om manuele aanpassingen te maken die telkens specifiek zijn voor de collectie die je wilt verwerken. Een snelle gids voor het programma is beschikbaar in het eerste datasheet gelabeld ‘Start’ van de Excel file waarin tevens alle functionaliteit zich bevindt. Deze handleiding zal het gehele conversieproces echter stap voor stap overlopen en de nodige manuele aanpassingen aan de hand van korte voorbeelden toelichten.

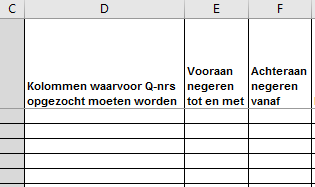
**Stap 1: selectie van de XML-file**

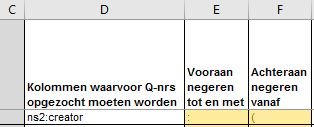
Door te klikken op de eerste grote grijze knop gelabeld “*Kies en verwerk een XML-bestand”* open je de file browser om een export-bestand van Brocade te openen. Dit wordt dan automatisch verwerkt en creëert binnenin het datasheet een overzichtelijkere weergave van alle informatie die beschikbaar is per record dan wanneer je dit handmatig zou doen met Excel of OpenRefine.

Wanneer je klaar bent met het verwerken van een collectie en je een nieuw bestand wilt verwerken dan kan je gewoonweg opnieuw op deze knop klikken. Het datasheet zal vervolgens leeggemaakt worden en opnieuw opgesteld worden voor de volgende collectie.

**Stap 2: de Q-nummers**

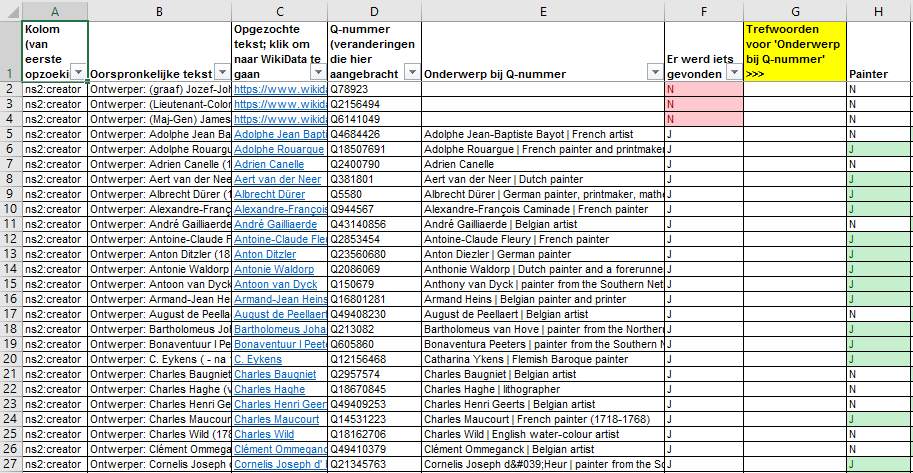
Q-nummers zijn van groot belang bij het opstellen van een template voor Quickstatements. Normaal gezien zou je deze manueel voor alle waarden moeten gaan opzoeken maar met Collector kan dit met een klik op de knop gebeuren. Een gelijkaardig programma dat deze functie uitoefent is OpenRefine, waarbij het ook mogelijk is om de Q-nummers van WikiData te exporteren. Het voordeel van Collector is echter dat na het ophalen van de Q-nummers deze ineens verwerkt worden in het datasheet, gelinkt worden aan de juiste waarden voor een controle en in een volgende stap automatisch verwerkt worden in het Quickstatements template.

Bij het opvragen van de Q-nummers is er een eerste manuele input vereist om te specifiëren welke waarden er juist opgevraagd moeten worden. Om deze aanpassing te maken moet je gebruik maken van het configsheet. Hier zal je alle nodige specificaties moeten invoeren die niet enkel belangrijk zijn voor het opvragen van de Q-nummers maar in verdere stappen voor het vormen van je template. Voor deze stap zijn we echter enkel genoodzaakt om aanpassingen te maken aan het eerste onderdeel van de configsheet. Dit ziet er normaalgezien zo uit:

In de eerste kolom zet je de titel van de kolom in je datasheet waarvan je de Q-nummers wilt opvragen. Er is hier tevens de mogelijkheid om te specifiëren over welk deel van de cel-waarden het hier precies gaat. Bij een XML-bestand afkomstig uit Brocade is het vaak zo dat de naam van de maker van een bepaald object die als ‘creator’ wordt vermeld, wordt voorafgegaan door zijn precieze rol en wordt opgevolgd door zijn levensjaren. Deze gegevens zouden de zoekopdracht die het programma in WikiData verricht echter helemaal verwarren. Daarom is er de mogelijkheid om Collector te laten weten dat deze informatie genegeerd mag worden. Bij een waarde zoals ‘Etser: Jan Theodoor Jozef Linnig (1815 - 1891)’ kan dus dit gedaan worden:

Waardoor enkel Jan Theodoor Jozef Linnig zal opgezocht worden op WikiData. Het verwerken van deze opdracht kan wel enkele minuten in beslag nemen, afhankelijk van de hoeveelheid records die verwerkt moeten worden en op hoeveel kolommen aan informatie het neerkomt. Wanneer Collector klaar is met het opzoeken worden er nieuwe kolommen aangemaakt achter elke waarde in het datasheet om ten eerste het gevonden Q-nummer weer te geven maar ook om de naam van het record dat werd gevonden weer te geven ter controle. Nadat de opzoeking voltooid wordt is het nodig om het Q-nummer sheet te overlopen om zeker te zijn dat er geen ontbrekende Q-nummers zijn of foute verwijzingen. Binnen deze tab worden al de waarden die opgezocht moesten worden, samen met hun opzoekingslink en het succes van de zoektocht weergegeven. Het is hier tevens mogelijk om een waarde ingegeven die moet vergeleken worden met het gevonden onderwerp om gemakkelijker fouten op te sporen. De kolom van het Q-nummer wordt automatisch gelinkt aan de kolommen die aangemaakt werden in het datasheet en kan dus verbeterd en aangevuld worden waar het nodig is. Het kan veel tijd en werk vergen om heel de lijst aan te vullen en eens je klaar bent maak je best een back-up in een notepad of een ander bestand. Wanneer deze stap is afgehandeld is het tijd om een passend template op te stellen.

Collector: het Q-nummers sheet



**Stap 3: het Quickstatements 2 Template**

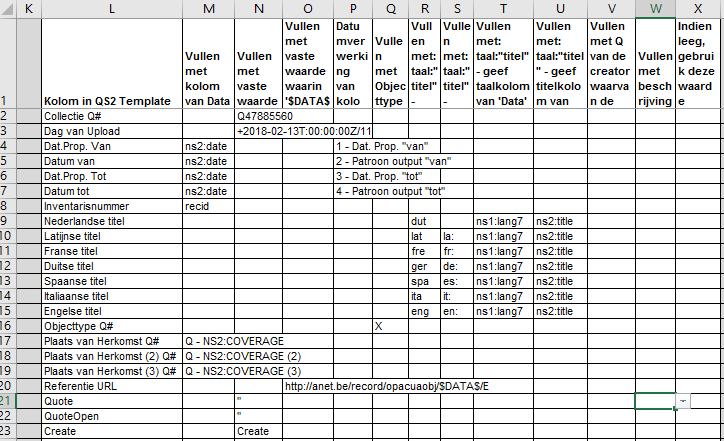
Zoals in de WikiData handleiding wordt uitgelegd maken we gebruik van Quickstatements 2 om alle informatie te uploaden. In deze stap ga je in het QS2Template-sheet een template aanmaken voor het uploaden van de data op dezelfde wijze als in de WikiData handleiding werd uitgestippeld. Het is hier dus de bedoeling dat je alle informatie, alle codes en waarden die je in het template wilt hebben op één lijn zet. Het is gemakkelijk om dit te doen maar het wordt aangeraden om op voorhand in een andere tekstverwerker zoals Notepad++ al eens uit te stippelen wat er allemaal in moet komen te staan. Door de werking van Collector is het hier wel niet nodig om aanhalingstekens rond de titels te zetten, wel rond andere ‘strings’ zoals bijvoorbeeld de beschrijving. Elk blok code dat een bepaalde waarde definieert moet tevens afgebakend worden door ‘blok begin’ en ‘blok einde’ om lege cellen automatisch te verwijderen. Het begin van de meeste templates zal er dus normaal zo uitzien:



**Stap 4: configsheet aanpassen aan de template**

Deze stap is nauw verbonden met de voorgaande. Nu je een template hebt opgesteld is het nodig om in de configsheet alle verschillende waarden die je in de template hebt gezet te definiëren. Dit gebeurt in de derde sectie van de sheet en is simpel om in te vullen mits je je grondig hebt voorbereid. Eens je een waarde hebt ingevuld kan je verduidelijken wat ermee moet gebeuren door een van de opvolgende kolommen in te vullen.

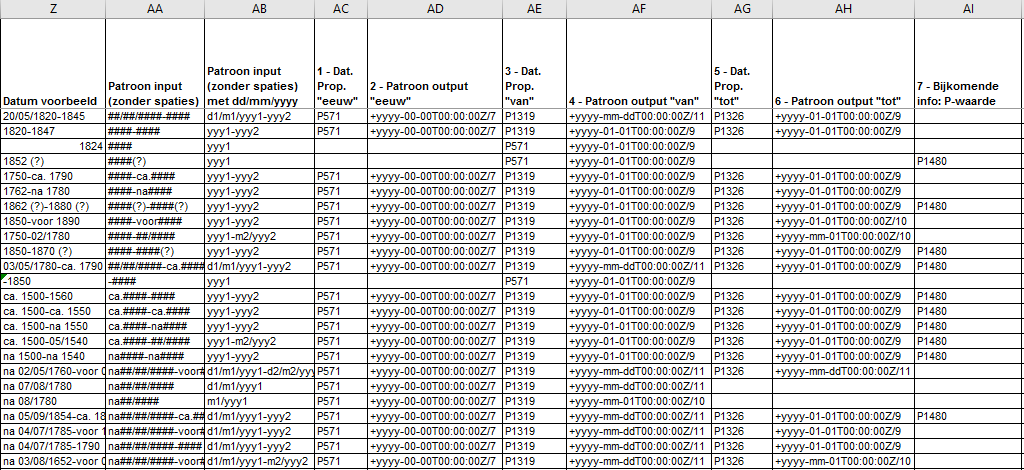
* Kolom **L**: zet hier de naam van **elke** soort kolom die je in QS2 Template hebt aangemaakt.
* Kolom **M**: zet hier de kolommen waaruit de informatie gehaald moet worden.
* Kolom **N**: zet hier een vaste waarde in.
* Kolom **O**: speciaal voor de URL’s naar de Engelse pagina’s. Maak de URL zoals in het onderstaande voorbeeld, en zet $DATA$ waar normaal het inventarisnummer komt.
* Kolom **P**: net zoals bij de Quickstatements is de datum een speciaal geval. Hier verwijs je naar kolommen die je moet aanmaken in de config voor data. (**Zie stap 5**)
* Kolom **Q**: het objecttype wordt tevens op een speciale manier gedefinieerd op basis van de creator-waarden. Zet hier gewoon een kruisje bij het objecttype en **zie naar stap 6**.
* Kolom **R**: zet hier de afkorting van de taal die in de data wordt gegeven.
* Kolom **S**: zet hier de taal unicode die je in de QS syntax moet gebruiken met dubbel punt.
* Kolom **T**: verwijs hier naar de kolom waarin de taal wordt verduidelijkt.
* Kolom **U**: verwijs naar het titel veld, hier moet slechts een veld vermeld worden, het programma zal de anderen zelf afzoeken.



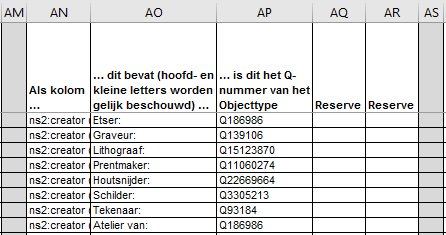
* Kolom **V**: geeft de mogelijkheid het Q-nummer van specifieke kunstenaars op te zoeken. Dit werd ontwikkeld om ontwerpers van andere kunstenaars te onderscheiden maar deze functionaliteit is al verwerkt in de dataverwerking. In plaats van ze hier te onderscheiden worden ontwerpers in een aparte kolom gezet bij de creatie van het datasheet.
* Kolom **W**: voor de Nederlandse en Engelse beschrijving zet je hier een verwijzing naar een laatste stap in de config, normaal EN en NL. (**Zie stap 7**)
* Kolom **X**: indien een kolom leeg is kan je hier een waarde ingeven die je er wilt laten invullen. Dit is handig indien je wilt aangeven dat het ontbreken van een auteur bijvoorbeeld moet leiden tot het vermelden van het Q-nummer voor anoniem.

**Stap 5: de datum**

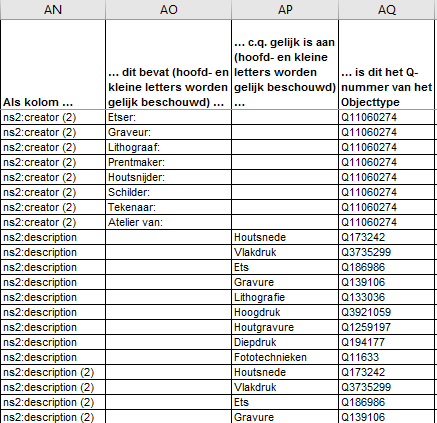
Alle datums moeten volgens de speciale syntax van QS2 opgesteld worden om correct geïnterpreteerd te worden door het programma. Het is hierbij belangrijk dat je een overzicht hebt van alle verschillende manieren waarop de datums in je databank worden weergegeven. Elk type moet immers zijn eigen speciale regel krijgen in Collector. Je neemt daarom telkens in de tweede kolom van dit deel een inputpatroon over dat overeenkomt met één soort weergave. Doe dit voor alle verschillende soorten en vul de opvolgende kolommen in op een gelijkaardige wijze zoals je hieronder ziet. In de genummerde cellen zet je de P-codes en datumpatronen die uit de input moeten voortkomen. De titels van deze kolommen moeten overeenkomen met de titels die je in de config hebt staan. In het onderstaande voorbeeld is er de mogelijkheid voorzien om jaarintervallen om te zetten in het correcte formaat en om eraan toe te voegen dat het gaat om een circa benadering.



**Stap 6: het opzoeken van de Q-nummers van objecttypes**

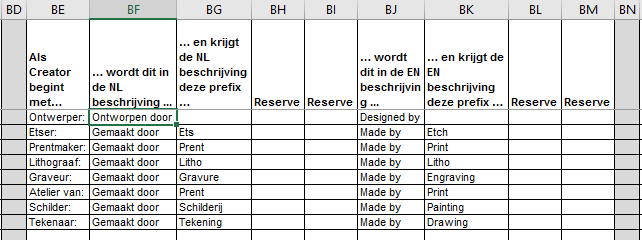
In de XML-bestanden werd er geen aparte vermelding gemaakt van de objecttypes. Daarom wordt er gebruik gemaakt van een omweg om aan de hand van de creators toch gemakkelijk een objecttype in te voegen. Hiervoor moet je kijken naar kolommen AN – AR waar je opnieuw zoals in de config je waarden moet definiëren. Verwijs telkens naar de tweede creator kolom waarin alle kunstenaars met hun werkomschrijving instaan en verduidelijk welke Q-nummers bij welke waarde moeten komen.

De vervaardigingswijze van een object is tevens nauw verbonden met het objecttype en wordt door Collector automatisch gefilterd uit de beschrijvende kolommen. Opdat het programma de correcte Q-nummers invult moet je hiervoor ook een verwijzing maken onder de config voor objecttypes. Je vermeld dan elke waarde die je mogelijk kan gebruiken voor de vervaardigingswijze en de verbonden Q-nummers.



**Stap 7: de beschrijvingen**

Je kan zelf handmatig de beschrijvingen toevoegen maar mits ze redelijk uniform opgesteld mogen worden kan dit ook grotendeels geautomatiseerd worden. In de laatste stap van de config kan je in kolommen BE-BM verduidelijken hoe je wilt dat je beschrijving eruitziet. De werking hiervan is redelijk vanzelfsprekend:

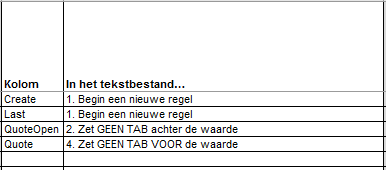


**Stap 8: voeg QS2 Template aan je datasheet toe**

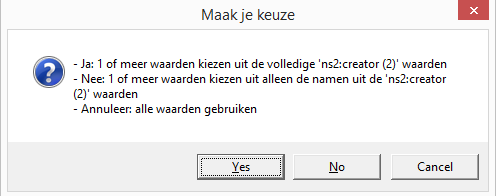
Na het afwerken van de template en het voorbereiden van de configsheet kan je via de derde grote grijze knop het template laten invullen. Dit wordt dan automatisch gegenereerd op het einde van je datasheet, ingevuld en bijna afgewerkt. Nu moet je enkel nog dit ingevulde template doorlopen om enige fouten recht te zetten of bepaalde waarden handmatig in te vullen.

Eens je ervan overtuigt bent dat alle data in je template compleet is kunnen we overgaan tot de finale stap.

**Stap 9: omzetten in tekstbestand**

In de configsheet is er nog een ding dat je kan doen om het jezelf gemakkelijker te maken voor de creatie van het tekstbestand en dat is het stellen van enkele regels in verband met de opmaak van je uiteindelijke Quickstatement. Normaal gezien zou je enkel dit nodig hebben:

Er zijn enkele mogelijkheden hier die je kan kiezen door op een cel van de tweede kolom te klikken. Probeer niet zelf nieuwe regels in te voeren zonder de nodige kennis van het programma omdat deze gedefinieerd moeten worden in de code. Eens dit gedaan is kan je op de laatste grijze knop klikken en het tekstbestand genereren. Je kan dan kiezen waar het opgeslagen moet worden, hoe het zal heten en welke informatie het bevat. Je kan namelijk kiezen uit een tekstbestand voor ALLE waarden, een tekstbestand betreffende EEN CREATOR of een tekstbestand voor een VERZAMELING CREATORS. Als alles goed verlopen is zal je een kant en klare Quickstatement batch hebben die geüpload kan worden naar WikiData.



# **Vragen in verband met Collector?**

**Q:** kunnen enkel Anet of Brocade XML-bestanden verwerkt worden?

**A:** nee, ook andere bestanden die op een gelijkaardige manier zijn opgesteld kunnen geformatteerd worden.

**Q:** kan ik zelf aanpassingen maken aan het programma?

**A:** indien je zelf ervaring hebt met Visual Basics for Applications kan dit, anders is het niet aangeraden. Maak altijd een back-up voor je begint te spelen met het programma!

Email:

[Arnaud.manise@student.uantwerpen.be](mailto:Arnaud.manise@student.uantwerpen.be)

[Arnaud.manise@gmail.com](mailto:Arnaud.manise@gmail.com)