

artikel

Inleiding

What's in a name? Persoonsnamen (en biografische basisgegevens) zijn vaak het startpunt van een zoektocht door digitale erfgoedcollecties. Afhankelijk van de beschikbare gegevens - ook metadata - zal de zoektocht meer of minder resultaten opleveren. **Dit artikel verkent aan de hand van het project Vrouwen van papier hoe beschrijvende metadata en biografische databanken (groepen van) personen in beeld brengen die anders onbenoemd en onvindbaar dreigen te blijven.**

Vrouwen van papier: Gezelles vrouwelijke correspondenten uit de schaduw van de geschiedenis

"They say Men have a cell more in their brains than Women but that we possess an additional fibre in our hearts." [\[Guido Gezellearchief, 7278\]](#) Dit opmerkelijke citaat komt uit een brief van Jemima Stroud, één van de 200 vrouwen waarmee Guido Gezelle correspondeerde. De bijna 700 brieven die vrouwen aan Gezelle schreven, staan vol opmerkelijke, ludieke of intieme passages. Twee jaar lang (2023-2024) zetten de medewerkers en vrijwilligers van het Guido Gezellearchief binnen het project Vrouwen van papier zich in om de brieven van of aan vrouwen te digitaliseren, transcriberen, annoteren en online beschikbaar te maken. Dat gebeurde via het editieplatform GezelleBrOn, maar ook via een magazine, een tentoonstelling en het online delen van de digitale persoonsdata die binnen het project verzameld werden.

Als onderdeel van de wetenschappelijke brieveneditie streeft het Gezellearchief ernaar een korte biografie te schrijven voor alle personen die vermeld staan in de brieven, dus niet alleen de 200 vrouwen waarmee Gezelle correspondeerde. In een personenregister registreren de medewerkers ook de naam en eventuele varianten, plaats en datum van geboorte en overlijden, beroep en de link met Gezelle. Gezelles correspondentes weerspiegelen een grote diversiteit aan nationaliteiten, talen, beroepen en maatschappelijke standen. Dat maakte het onderzoek niet vanzelfsprekend.

Er was binnen de Gezellestudie eerder al aandacht voor de vrouwen in de omgeving van de dichter, maar dat bleef beperkt tot familieleden, directe medewerksters en Britse migranten, en stond vaak volledig in functie van hun rol in het leven van Gezelle. Voor bijna iedere vrouw uit het personenregister was bijkomend onderzoek nodig, wat niet makkelijk was door de ondervertegenwoordiging van vrouwen in de studies en bronnen.

Er waren ook andere hindernissen. Vaak was het moeilijk om te weten over welke vrouw het precies ging: getrouwde vrouwen en weduwes hanteerden meestal de naam van hun man, kloosterzusters waren enkel bekend onder hun kloosternaam en sommige brieven waren ondertekend met enkel een familie- of voornaam. In de 19^{de} eeuw liep er natuurlijk meer dan één 'weduwe Janssens' of 'zuster Marie' rond. Gezelle schreef ook met groepen die zich in de marge van de maatschappij bevonden, zoals verpauperde migranten die in de Brugse Engelse Kolonie opgevangen werden. Sommige vrouwen uit deze gemeenschap bleven hun hele leven op de dool, waardoor het moeilijk was om biografische gegevens te verzamelen.

Toch konden de vrijwilligers en onderzoekers veel vrouwen succesvol identificeren en een (spreekwoordelijk) gezicht geven door in talloze archieven en gedigitaliseerde collecties te duiken. Dit project zou niet mogelijk geweest zijn zonder de vele digitale data die zij hebben kunnen uitpluizen. De ontbrekende puzzelstukken vonden ze vaak terug in digitale bronnen: van commerciële genealogische sites, over gedigitaliseerde registers van burgerlijke stand tot persoonsdatabanken.

Uit vorige deelprojecten waren ook al datasets opgebouwd over leraren van het Sint-Lodewijkscollege in Brugge, leden van de Koninklijke Academie Nederlandse Taal en Letteren en Brugse drukkers.

Van papier naar digitale databanken

De persoonsgegevens die binnen het project verzameld werden, kregen ook een plek in drie online databanken: ODIS, Wikidata, en binnenkort ook in de tool die binnen het project Visual Name Authority tot stand komt. Vrouwen zijn ondergerepresenteerd in databanken, dus door het delen van informatie over Gezelles vrouwelijke correspondenten wordt de genderkloof in databanken een beetje gedicht. Onderzoek naar vrouwen wordt zo toegankelijker. Hieronder komt

elk van de drie databanken aan bod. Wat onderscheidt het ene platform van het andere? Op welk publiek richten ze zich? En vooral, wat was de ervaring van Vrouwen van papier met de databank?

Wikidata: een vrije databank

Wikidata speelt een centrale rol als collaboratieve, meertalige en machineleesbare databank. Wikidata is een implementatie van Wikibase, een databank ontwikkeld door Wikimedia Duitsland in 2012. Een projectcoördinator of gemeenschap gebruikers kan impliciet een datamodel definiëren op basis van de hiërarchie van statements die de relatie tussen entiteiten bepaalt. Een statement is een predicaat, bewering, of verklaring waarmee de relatie, via een property, tussen een subject en een object of waarde wordt beschreven.

Iedereen kan items, zoals personen, toevoegen die vervolgens worden geïdentificeerd met een unieke en permanente Wikidata ID. Alle data worden gepubliceerd onder een CC0-licentie. Cultureelerfgoedorganisaties en -aggregatoren zoals ODIS kunnen authority data aan Wikidata toevoegen en verwijzen naar hun eigen permanente identifier, waardoor Wikidata een hub vormt die informatie verbindt en gebruikers toeleidt naar andere databronnen.

Aan een Wikidata-record kan je ook mediabestanden in Wikimedia Commons koppelen, bijvoorbeeld een portretfoto. Wikipedia-pagina's in verschillende talen over hetzelfde onderwerp worden als sitelinks gekoppeld via een uniek Q-nummer, waardoor een Wikipedia-lezer gemakkelijk kan doorklikken tussen de pagina's in de verschillende talen.

Het Gezellearchief bezorgde de data uit het personenregister aan meemoo, die instond voor de import in Wikidata. Meemoo maakte gebruik van OpenRefine bij de voorbereiding van de data, maar de import bleef deels ook manueel werk. Er werd beslist om alle personen uit de registers mee te nemen, dus niet alleen vrouwen, vanwege de onderlinge relaties in het netwerk van Gezelle. Zo werden meer dan 120.000 wijzigingen of aanvullingen in ruim 3000 Wikidata-lemma's uitgevoerd.

Voordelen van Wikidata :

- De collectie van een instelling wordt gemakkelijker vindbaar.
- De collectiebeheerder beslist welke data publiek worden vrijgegeven.
- De collectiebeheerder ontwerpt een passend datamodel.

- Wikidata Query beschikt over functionaliteiten om gegevens overzichtelijk te presenteren als een tabel, een graaf, een geografische kaart, of een lijst van afbeeldingen.
- Ontbrekende Wikipedia-pagina's schrijven of vertalen, wordt eenvoudiger. Wikipedia heeft een ingebouwde vertaaltool.
- Wikidata is eenvoudig in alle richtingen te koppelen met Wikipedia en Wikimedia Commons.
- Het nodigt instellingen uit om na te denken over een gepaste licentie om vrije kennis kenbaar te maken.

ODIS: een contextuele historische databank

De online databank ODIS werd in 2000-2003 opgericht door de culturele archieven ADVN, Amsab-ISG, KADOC-KU Leuven en Liberas. Sindsdien groeide ODIS uit tot een breed partnerschap van erfgoed- en onderzoeksorganisaties. Zij gebruiken de databank om gevalideerde kennis over de geschiedenis en het erfgoed van de samenleving in Vlaanderen, in zijn Belgische en internationale context, samen te brengen, op te bouwen en te delen. Begin 2025 bevatte ODIS ruim 318.000 records over organisaties, personen, families, gebeurtenissen, gebouwen en daarmee verbonden archieven en periodieken. Bijna de helft van de records is gepubliceerd in de ODIS-publiekscatalogus, die in 2024 108.173 keer bezocht werd. Partners en bezoekers gebruiken ODIS als encyclopedie, authority-databank, digital humanities tool en/of heuristisch instrument. Het datamodel is gebaseerd op internationale standaarden en bestaat uit een combinatie van tekstvelden en veldgroepen voor gestructureerde gegevensinvoer via keuzelijsten/thesauri.

ODIS nam in het Gezelle-editieproject een dubbele rol op. Enerzijds was de databank een belangrijke bron om informatie over personen - vooral vrouwelijke religieuzen - te verzamelen. Anderzijds stroomden de gegevens uit het Gezelle-persoonregister door naar ODIS. Er werd vertrokken vanuit een selectie records die voldeden aan minimale vereisten: een naam en een geboorte- en overlijdensplaats en -datum. Na 'datamassage' en de opmaak van een conversietabel werden twee batches met gegevens over 569 vrouwen ingeladen in ODIS, waar ze verknoopt werden met de bestaande datareeksen. De gegevensreeks werd via de ODIS-publiekscatalogus beschikbaar gesteld voor raadpleging, analyse en hergebruik. De overlap tussen de bestaande datareeksen

in ODIS en het personenregister was beperkt (5%). De conversie betekende dan ook een mooie netto-aangroei van het aantal ODIS-records over vrouwen.

Voordelen van ODIS:

- Het ODIS-datamodel biedt veel flexibiliteit, er zijn weinig verplichte onderdelen.
- ODIS biedt gevalideerde, in bronnen verankerde informatie aan.
- Door relaties te leggen tussen ODIS-records kunnen informatieclusters uitgebouwd worden.
- Datareeksen kunnen via geavanceerde zoekfuncties diepgaand geanalyseerd worden.
- De nieuwe databank en publiekswegsite die in 2022-2027 ontwikkeld worden (dankzij een investeringskrediet van FWO) zullen bijkomende analyse- en visualisatiemogelijkheden bieden. Ook wordt de interconnectiviteit van ODIS versterkt.

Visual Name Authority: een gedeeld beheer voor persoonsidentificatie

Visual Name Authority (VNA, 2023-2026) is een samenwerking van meemoo met 23 cultureelerfgoedpartners, waaronder ook ODIS en Wikidata. In VNA wordt samengewerkt aan een overkoepelende namenlijst met personen die relevant zijn voor het cultureel erfgoed in of uit Vlaanderen. Uniek is dat aan die persoonsnamenlijst ook een referentieset met gezichtsuitsneden wordt gekoppeld. Door het samenbrengen van biografische persoonsdata én gezichtsuitsneden in één tool wordt metadataverrijking gefaciliteerd op een manier die voordien niet mogelijk was, bijvoorbeeld met behulp van artificiële intelligentie. Bedoeling is dat de data in VNA gecombineerd doorzoekbaar zijn, hiaten in bestaande persoonsnamenlijsten gedetecteerd worden, data over en foto's van personen gestandaardiseerd uitwisselbaar zijn en ingezet kunnen worden in eigen initiatieven rond metadataverrijking. Dit resulteert in vlottere workflows bij diverse digitaal erfgoedprocessen.

Binnen het Gezellearchief is er een bescheiden fotoarchief van ongeveer 1000 foto's: van straatzichten van Gezelles geboortehuis tot familiefoto's van zijn zus. Tekenend voor de beperkte beschikbaarheid van foto's met vrouwen uit de 19de eeuw is dat er slechts 24 foto's zijn gevonden met duidelijk zichtbare en identificeerbare vrouwen. Die bruikbare portretfoto's zijn bezorgd bij de

voorbereiding van de VNA-databank. Het Gezellearchief heeft veel tijd geïnvesteerd in het plakken van gezichten op de vrouwen waarmee Gezelle correspondeerde. Dankzij de VNA blijven deze gezichten in de toekomst beschikbaar en kunnen ze opnieuw worden gebruikt, ook door andere organisaties.

Voordelen van VNA:

- VNA erkent het belang van veelgebruikte authorities zoals RKDartists&, VIAF en ISNI, maar die zijn vaak niet voldoende dekkend voor de diversiteit van het cultureelerfgoedveld in Vlaanderen. VNA wil die kloof dichten door een zo groot en divers mogelijke groep van personen op te nemen die relevant zijn voor het erfgoed in of uit Vlaanderen.
- De combinatie van geaggregeerde persoonsdata met gezichtsuitsneden van personen die relevant zijn, en zo een gezamenlijke referentieset opbouwen, bestaat nog niet!
- Een belangrijke doelstelling van VNA is de optimalisatie van beschikbare tijd en middelen van cultureelerfgoedorganisaties in Vlaanderen. Door data aan te leveren aan VNA, kan je hieraan als organisatie je steentje bijdragen.

Waarom data bijdragen aan overkoepelende databanken?

Bij gecontroleerde persoonsnamenlijsten denk je misschien aan internationale, veelgebruikte authorities zoals VIAF, ULAN of ISNI. Deze aggregatoren verzamelen authority records van verschillende organisaties op één centraal platform. Ze worden vaak gebruikt bij de registratie van Vlaamse cultureelerfgoedcollecties, maar lang niet alle personen die relevant zijn voor die collecties, zijn erin opgenomen of komen daarvoor zelfs niet in aanmerking.

Daarom gaan organisaties vaak hun eigen persoonsnamenlijst aanleggen, afgestemd op hun collectie. Die lijsten worden vaak enkel intern gebruikt, waardoor heel wat waardevolle informatie gefragmenteerd blijft. Meerdere organisaties kunnen ook informatie over dezelfde persoon beheren, maar met uiteenlopende gegevens. Dit belemmert kennisdeling en leidt vaak tot onnodig dubbel werk.

Samenwerken aan gecontroleerde termenlijsten, zoals person name authorities, kan de herbruikbaarheid van jouw data dus verhogen. Maar hoe ga je aan de slag? Een eerste stap is de creatie van een eigen persoonsnamenlijst. Daarbij ga je de

data standaardiseren, normaliseren en opschonen en idealiter een identifier toevoegen. Door ook externe identifiers op te nemen (bijvoorbeeld Wikidata, ODIS of VNA) voorzie je de data van meer context. Een handige tool hiervoor, die ook in Vrouwen van papier werd ingezet, is OpenRefine.

Je kan er bovendien voor kiezen om jouw data bij te dragen aan overkoepelende initiatieven, zoals in Vrouwen van papier deed aan ODIS, Wikidata en VNA. Zo wordt jouw persoonsnamenlijst, of jouw beeldmateriaal, onderdeel van een breder netwerk en wordt de informatie herbruikbaar door andere erfgoedprofessionals en onderzoekers.

De auteurs

Collyn Verlinde, functioneel beheerder digitaal erfgoed bij Openbare Bibliotheek Brugge

Joris Colla, afdelingshoofd Behoud en Beheer bij KADOC-KU Leuven en coördinator van ODIS

Geert Van Pamel, medeoprichter en bestuurslid van Wikimedia België

Axelle Vanmaele, medewerker expertise bij meemoo, Vlaams instituut voor het archief

overige tekst

Praktische tips

1. OpenRefine is een veelgebruikte tool om data op te schonen en te verrijken met OpenRefine, maar soms is manueel werk nodig. Evalueer voor je begint wat nodig is voor jouw dataset.
 - a. deelnemen aan een overkoepelend initiatief: stem af met de partners over o.m. de structuur van de data en de mate van details
 - i. Joris bekijkt documentatie uit Vrouwen van Papier
 - b. nakijken/beslissen/evalueren cost/benefit van automatisch/manueel opladen
 - c. reciproque of symmetrische statements, b.v. (huwelijks)partner, kinderen of hoeft slechts in 1 richting opgeladen te worden, en kan nadien via Pywikibot aangevuld worden.
2. Eens je dataset is opgeschoond ... verrijken
 - a. verrijken mbv reconciliation service van OpenRefine
 - b. maar je kan ook externe ids ophalen, bv. QID, ODIS, VIAF, ISNI ...
3. Data ontsluiten
 - a. eigen platform
 - b. of iets als Wikidata ??
4. Referenties:
 - a. identificatie/pid/URI/url creëren
 - b. via OpenRefine kan je deze data door laten stromen naar Wikidata
5. Er werd vertrokken van een spreadsheet met rijen en kolommen. Iedere rij bevat de beschrijving van een persoon; de kolommen bevatten de verschillende beschrijvende attributen. Deze spreadsheet wordt met OpenRefine opgeladen naar Wikidata op basis van properties en statements. Er zijn ook Pywikibot scripts ter beschikking om semi-automatisch aanpassingen te maken. Manuele correcties blijven altijd mogelijk.

Biografische databanken

Biografische databanken en person name authorities

Person name authorities zijn lijsten met persoonsnamen die gebaseerd zijn op bronbestanden. De lijsten zijn dus gecontroleerd, door de validatie op basis van bronbestanden kan een record dus de 'autoriteit' worden om een persoon te beschrijven. Authority records worden hierdoor beschouwd als kwalitatief en betrouwbaar.

Daarnaast dragen person name authorities ook bij aan efficiëntere processen bij digitale collectieregistratie, vormen ze doelgerichte zoekingen naar informatie over een persoon, waardoor gegevens over en materiaal overmet die persoon beter vind- en doorzoekbaar is, en kunnen ze gebruikt worden om links te leggen tussen de verschillende databanken met

informatie over die personen (of in relatie tot die persoon). Door autoriteiten in een gestandaardiseerd formaat te beheren, wordt de uitwisselbaarheid van data verhoogd.

uniek identificeren - eddy en eddy dedupliceren

Als we het hebben over gecontroleerde persoonsnamenlijsten denken we vaak aan internationale en veelgebruikte autoriteiten zoals VIAF, ULAN, ISNI ... Deze initiatieven zijn aggregatoren - authority records van verschillende organisaties worden op één centraal platform verzameld en toegankelijk gemaakt. De data op deze platformen worden vaak gebruikt bij de registratie van Vlaamse cultureelerfgoedcollecties, maar niet alle personen die voor die collecties relevant zijn, zijn erin opgenomen én komen vaak ook niet in aanmerking om erin opgenomen te worden. Organisaties gaan daarom vaak zelf een gecontroleerde persoonsnamenlijst aanleggen, die relevant is voor hun eigen collectie. Vaak worden die lijsten niet open ontsloten en enkel voor interne processen ingezet, in andere gevallen maken organisaties de lijsten publiek herbruikbaar op hun eigen platformen of via overkoepelende initiatieven, zoals Wikidata, ODIS en Visual Name Authority, waarover we het verder in dit artikel over hebben.

Biografische databanken

Biografische databanken en person name authorities

Person name authorities zijn lijsten met persoonsnamen die gebaseerd zijn op bronbestanden. De lijsten zijn dus gecontroleerd, door de validatie op basis van bronbestanden kan een record dus de 'autoriteit' worden om een persoon te beschrijven. Authority records worden hierdoor beschouwd als kwalitatief en betrouwbaar.

Daarnaast dragen person name authorities ook bij aan efficiëntere processen bij digitale collectieregistratie, vormen ze doelgerichte zoekingen naar informatie over een persoon, waardoor gegevens over en materiaal over met die persoon beter vind- en doorzoekbaar is, en kunnen ze gebruikt worden om links te leggen tussen de verschillende databanken met informatie over die personen (of in relatie tot die persoon). Door autoriteiten in een gestandaardiseerd formaat te beheren, wordt de uitwisselbaarheid van data verhoogd.

uniek identificeren - eddy en eddy dedupliceren

Als we het hebben over gecontroleerde persoonsnamenlijsten denken we vaak aan internationale en veelgebruikte autoriteiten zoals VIAF, ULAN, Wikidata ... Deze initiatieven zijn aggregatoren - authority records van verschillende organisaties worden op één centraal platform verzameld en toegankelijk gemaakt. De data op deze platformen worden vaak gebruikt bij de registratie van Vlaamse cultureelerfgoedcollecties, maar niet alle personen die voor die collecties relevant zijn, zijn erin opgenomen én komen vaak ook niet in aanmerking om erin opgenomen te worden. Organisaties gaan daarom vaak zelf een gecontroleerde persoonsnamenlijst aanleggen, die relevant is voor hun eigen collectie. Vaak worden die lijsten niet open ontsloten en enkel voor interne processen ingezet, in andere gevallen maken organisaties de lijsten publiek herbruikbaar op hun eigen platformen (bijvoorbeeld UGent Memorialis, een authority over Gentse professoren) of via

overkoepelende initiatieven, zoals Wikidata, ODIS en Visual Name Authority, waarover we het verder in dit artikel over hebben.

Titels

TITELS

1. Van Vrouwen van papier naar digitale databanken
2. Vrouwen van papier, digitale databanken in vizier
3. Vrouwen van papier gelinkt, gezocht en gevonden
4. Vrouwen van papier gedeeld, gelinkt en herkend
5. Van Vrouwen van papier naar dames in digitale databanken
6. Vrouwen van papier, een reis van archiefstukken tot digitale databanken
7. Gezelle en digitale databanken? De data uit 'Vrouwen van papier' is hier!
8. Vrouwen van papier, een gedeeld project
9. Vele handen maakten van Vrouwen van papier een plezier. Het coöperatief ontsluiten van persoonsgegevens in digitale databanken.