

# Geocoding van de Nederlandse botanische collectie van Naturalis met hulp van citizen scientists en crowdsourcing

## Resultaten 2019

Laurens Sparrus, Dion van der Hak (FLORON) & Jan Wieringa (Naturalis Biodiversity Center)



FLORON-rapport 2020.02

NLBIF project  
nlbif2018.2019.003

27-2-2020



## Colofon

Rapportnummer:	2020.02
Datum uitgave:	27-2-2020
Citeren als:	Sparrius, L.B., D.D. van der Hak & J. Wieringa, 2020. Geocoding van de Nederlandse botanische collectie van Naturalis met hulp van citizen scientists en crowdsourcing. Resultaten 2019. FLORON-rapport 2020.02
Projectnummer:	FL2019.053
Projectleider:	Laurens Sparrius
Opdrachtgever(s):	NLBIF

## Summary

The botanical collection of Naturalis Biodiversity Center contains about 550,000 specimens from The Netherlands. All specimens have been scanned and transcribed. However, reliable geographical coordinates were present in only 8,600 cases. Aim of this project is to enlarge this number by means of (1) extraction of grid square codes and coordinate pairs from text fields, (2) semi-automated and (3) manual geocoding based on location names using crowd sourcing by FLORON volunteers. We only took specimens into account with taxon names and collector names that could be mapped to taxon and people present in the National Database Flora and Fauna, about 230,000 specimens. By the end of the first year of the project, a precise location (less than 1000 m inaccuracy) had been assigned to 28,000 records. A tentative location at 2500 m inaccuracy has been assigned to the other 202,000 records.

## 1 Inleiding

In de zomer van 2019 startten FLORON en Naturalis de eerste fase van het project “Geocoding van de Nederlandse botanische collectie van Naturalis met hulp van citizen scientists en crowdsourcing”. Het belangrijkste doel van dit project is het toekennen van coördinaten aan de in eerdere jaren gedigitaliseerde botanische collectie van Naturalis. De collectie is ontsloten in het Bioportal van Naturalis en GBIF, maar door het ontbreken van coördinaten, kunnen de gegevens niet optimaal worden gebruikt. Het tweede doel was het toevoegen van de digitale herbariumcollectie aan de Nationale Databank Flora en Fauna waarmee de gegevens breed beschikbaar komen voor toepassingen in Nederland en aansluiten op de actuele in Nederland gehanteerde naamgeving van plantensoorten.

## 2 Aanpak

### Basisgegevens

De basis voor het project is een selectie van gegevens in de collectiedatabase BRAHMS. Hieruit is een export met 565.000 records gemaakt. Dit zijn alle records waar als land “Netherlands” is ingevuld.

### 2.1 Koppelen van taxa en personen

Koppelen personen en taxonomie: dit onderdeel leverde diverse problemen op. Omdat de tijd beperkt was, hebben we precies zoveel personen en taxa naar de huidige geaccepteerde namen omgezet, dat de helft van de half miljoen Nederlandse BRAHMS-records in de geocoding tool konden worden opgenomen. Voor de andere personen en taxa moet in volgende jaren een nieuwe slag worden gemaakt met hulp van enkele specialisten. In afwijking van het projectplan hebben we Gerard Dirkse niet ingeschakeld als specialist, omdat we in de vervolgproject meer tijd nodig hebben om de overige taxonamen in BRAHMS beter te kunnen begrijpen. Een groot deel van de namen betreft infraspecifieke taxa waarvan de namen die niet te vinden zijn in de lookup service TNRS. Bij personen komen veel afwijkende schrijfwijzen voor en namen van verzamelaars die soms niet meer leefden in het opgegeven jaar.

### 2.2 Coördinaten toekennen

In BRAHMS zijn locaties op verschillende manieren toegekend:

Als geografische coördinaat (een puntlocatie met een nauwkeurigheid), soms aan de hand van een plaatsnamenregister (gazetteer)

Een uurhok, kilometerhok of IVON-hok. Deze gegevens horen in speciale hiervoor bedoelde velden te staan in BRAHMS, maar zijn voor het gros van de records niet hierin ingevoerd maar in tekstvelden terecht gekomen, of helemaal niet overgenomen van het etiket. Dergelijke hok referenties staan in een bepaald format en konden met filters uit de tekst van een locatieveld of opmerkingenveld gehaald worden. Locaties met een uurhok worden in de komende jaren verder verbeterd via crowdsourcing door FLORON-vrijwilligers.

Een met GPS bepaalde locatie. In deze gevallen is de coördinaat al bekend en zijn de gegevens overgenomen. In BRAHMS waren 8.600 waarnemingen met nauwkeurige plaatsbepaling (LLRES 5 of hoger) beschikbaar.

Wanneer alleen een locatiebeschrijving (tekst) aanwezig was, dan is in de tekst gezocht naar bekende woonplaatsen, namen van polders en velden, parken, bossen, natuurgebieden en waterlichamen. Als bron voor de namen zijn referentiebestanden gebruikt uit open data bronnen, zoals de topografische kaart 1:10.000 (kadaster: TOP10NL), woonplaatsen (CBS) en de kaart Leven met Water. In alle gevallen is een voorlopige locatie toegekend met een onnauwkeurigheid van 2500 meter. In de NDFD is deze als een uurhok (5x5 km) opgeslagen. Deze locaties worden in de komende jaren verder verbeterd via crowdsourcing door FLORON-vrijwilligers.

## 2.3 Geocoding tool

De bestaande tool in de NDFF Verspreidingsatlas om waarnemingen van nauwkeurige coördinaten te voorzien, is voor dit project aangepast aan het werken met afbeeldingen/scans (Figuur 1) uit de Netherlands Biodiversity API (linked open data).



Figuur 1. Zoomfunctie van de geocoding tool.

**Atriplex prostrata - Spiesmelde**

informatie over deze soort

30 aug 1952

Dollart, achter Finsterwolde, op kwelder. 265-575 (atlasblok 08-52)

Aanwezig

leg. J.Mart. Duiven nr. 1161 in collectie van Naturalis

Bewijs in collectie

Bron-id L.3188962

Soortnaam (bron) Atriplex prostrata

Naam waarnemer (bron) Herb Duiven, J.M.

<http://floron.nl/observations/brahms/L.3188962>

bekijk waarnemingen uit dezelfde bron (opent in nieuw venster)

Alle eigenschappen van de waarneming aanpassen (opent in nieuw venster)

Toon waarnemingen uit hetzelfde jaar

Plaats opzoeken op de kaart

OSM

● punt ● 10 m ● 100 m ● 1 km ● 5 km

Coördinaten:   [Kaart bijwerken](#)

Naam gebied:

Verbeterde locatie opslaan

Figuur 2. Geocoding tool. De voorlopig toegekende locatie wordt getoond ("Finsterwolde"). De gebruiker bepaalt op basis van de locatiennaam (Dollart, achter Finsterwolde, op kwelder) wat de meest waarschijnlijke locatie op de kaart geweest moet zijn.

## 2.4 Rollen van de medewerkers

**Laurens Sparrius:** applicatie-ontwikkeling geocoding tool, kwaliteitscontrole, projectleider

**Dion van der Hak:** Dataverwerking NDFF, koppelen van BRAHMS-codes aan NDFF-taxa en -personen, handmatige geocoding

**Jan Wieringa:** Dataverwerking in BRAHMS, doorleveren locaties aan GBIF

## 3 Resultaten en aanbevelingen

### 3.1 Verwerkte gegevens

Aan 230.000 records uit BRAHMS is een coördinaat toegekend, waarvan in 28.000 gevallen een nauwkeurige coördinaat (tot 500 meter nauwkeurig). De gevonden hok-coördinaten zijn teruggeladen naar BRAHMS.

### 3.2 Gebruik van de geocoding tool

De 230.000 records uit BRAHMS en een koppeling naar de gescande herbariumvellen zijn geïmporteerd in de crowdsourcing tool van de NDFV Verspreidingsatlas.

Vanaf de tweede helft van december is de tool onder de aandacht gebracht van onze achterban. Na een testperiode zijn nog kleine aanpassingen in de gebruikersinterface doorgevoerd. Zo kan de gebruiker nu een overzicht bekijken van de laatste 25 beoordeelde herbariumvellen. Hiermee is het mogelijk om terug te gaan naar eerder beoordeeld vel. Ook kan worden ingezoomd op de herbariumvellen.

De voorbereiding van de dataset kostte veel meer tijd dan gepland. De crowdsourcing tool kwam pas laat beschikbaar en er zijn door vijf vrijwilligers in twee weken tijd 167 herbariumvellen beoordeeld.

### 3.3 Levering dataset BRAHMS en NDFV

De 230.000 records in de Geocoding tool komen na controle beschikbaar in de NDFV. Wanneer de locaties van de herbariumvellen gecontroleerd zijn, dan wordt de controle op de juistheid van de determinatie meegenomen in het validatieproces van de NDFV.

Medio februari 2020 heeft Naturalis de 202.000 voorlopige & 28.000 nauwkeurige (definitief toegekende) coördinaten ontvangen voor aanvulling van BRAHMS en ontsluiting in GBIF.

### 3.4 Aanbevelingen

In 2019 is een begin gemaakt met het toekennen van coördinaten aan de botanische collectie van Naturalis. In dit eerste jaar zijn vooral de eenvoudigste zaken uitgevoerd, die hebben geleid tot toewijzing van betrouwbare coördinaten aan 5% van de records in BRAHMS, en circa 50% voorlopig toegekende coördinaten. Aanbevolen wordt op in de komende jaren door te gaan op de ingeslagen weg, waarbij zowel professionele als vrijwilligersinzet nodig is om tot een kwalitatief hoog niveau te komen.

- Het maken van een koppeling tussen de taxonomie en personen in BRAHMS en personen en de moderne naamgeving van plantensoorten in de NDFV kostte veel meer tijd dan verwacht. Aanbevolen wordt om in de komende jaren jaarlijks in stapjes steeds enkele dagen



te besteden aan het koppelen van oude soortnamen en personen aan de juiste codering in de NDFD.

- Besteed meer tijd aan het werven van vrijwilligers, bijvoorbeeld met bijeenkomst om in groepsverband te werken. Vanuit de vrijwilligersachterban is enthousiast gereageerd op de ontsluiting van de herbariumcollectie in de NDFD maar dit heeft nog niet geleid tot een groot aantal gecontroleerde records.
- In komende jaren kan ook worden getracht op herbariummateriaal van (regionaal) zeldzame soorten te bepalen aan de hand van waarnemingen uit andere bronnen, zoals de historische FLORIVON-dataset. Door bestanden slim te koppelen kan mogelijk nog één keer een groot aantal records van betrouwbare coördinaten worden voorzien.
- In de geocoding-module kunnen succesfactoren uit vergelijkbare buitenlandse projecten worden overgenomen (zoals iDigBio, Herbaria@Home). Dat betreft vooral het in kleine pakketten aanbieden van herbariumvellen tot taken die een vrijwilliger één dagdeel kan voltooien. Dat kan nu al op basis van locatie, maar moet ook worden ontwikkeld voor één taxon, een verzamelaar of een bepaalde periode. Dat zijn relatief goedkope uitbreidingen van de bestaande software.

## **FLORON**

Natuurplaza  
Toernooiveld 1 - 6525 ED Nijmegen  
Postbus 9010 - 6500 GL Nijmegen

T: 024 - 7 410 660 (alg.)  
[www.floron.nl](http://www.floron.nl)

