



DATABEHEER

Invoeren van data

Hoe zal data ingevoerd worden:

- Locatie
- Folder structuur (Folders, subfolders etc.)
- File naam structuur (namen van files, namen van foto's en andere data)
- Controle van ingevoerde data
- Importeren van Excel, CSV of andere file format.

Data Cleaning

Data cleaning zal gedaan worden met Microsoft Excel en Openrefine. De verschillende datasets zullen volgens de GBIF standaarden gecorrigeerd worden op onder andere:

- Juistheid.
- Volledigheid.
- Consistentie.
- Gebruik van whitespaces.

Het controleren van data zal plaatsvinden met behulp van:

- GBIF Data Validator (<https://www.gbif.org/tools/data-validator>)
Hier kan een dataset worden geupload en gecontroleerd, alvorens deze wordt gepubliceerd op de GBIF site (<https://gbif.org>). De validator geeft dan aan waar er correcties aangebracht moeten worden.
- GBIF Test IPT. (<https://ipt.gbif.org/>)
Dit is een (identieke) testversie van de GBIF website. Hier kunnen datasets gepubliceerd worden en kan de gebruiker zien hoe de data eruit zal zien op de GBIF website. De Test IPT kan dus gebruikt worden om de laatste aanpassingen te doen alvorens de dataset gepubliceerd wordt op de GBIF site.

Opslag van data

Momenteel worden alle data van BBS opgeslagen op twee externe harde schijven. (3TB main en 4TB backup)

Het is de bedoeling dat er een eigen server komt en dat de data daarop opgeslagen gaat worden. De huidige harde schijven zullen dan als backup gebruikt worden. (Fase 2)

De foto's die gepubliceerd worden op de site van GBIF worden ook geupload op Internet Archive (<https://archive.org/>). De foto's op Internet Archive zijn een (in size) verkleinde versie van de originele foto's. De verschillende occurrences in de datasets krijgen dan een link naar hun desbetreffende foto(s) en zijn op die manier zichtbaar op de website van GBIF.

Backup

Tot 12 November werden er geen backups gemaakt van de data op de externe harde schijven. Voorstel is dus gedaan om direct een one-to-one backup te maken op twee nieuwe harde schijven, waarbij de twee nieuwe harde schijven elders dan het complex worden opgeslagen. (Fase 1) Van belang is dus ook om een encryptie te hebben op de backups voor het geval van verlies of diefstal. Dit is nog niet het geval in fase 1.

Het beste is om een 3-2-1 backup te implementeren. (3 backups, 2 verschillende media, 1 offsite backup)

Het beste voor Fase 2:

3 backups: 1x door datamanager, 1 x op 1set externe harde schijven , 1 x op tweede set harde schijven.

2 verschillende media: externe harde schijven en lokaal op de server of eventueel NAS.

1 offsite: 1 set externe harde schijven worden meegenomen.

Een cloud backup is ook een optie. Maar gezien de grootte van de data zou dat misschien niet haalbaar zijn. Evt. cloud optie bekijken.

Access management

Opstellen matrix voor toegang tot data.

Output van data

Databeheer kan worden onderverdeeld in een aantal processen, t.w.:

1. Voorbereiding specimens
2. Kalibreren fotocamera
3. Fotograferen specimen
4. Dataverwerking

Titel: Procesbeschrijving voorbereiding specimens

Beginpunt: De specimens worden gehaald vanuit hun opslagplaats.

Eindpunt: De specimens worden geplaatst op het rek om gefotografeerd te worden.

Doel: Dit proces beschrijft op welke manier specimens worden geprepareerd voor fotografie.

Procesverloop:

1. Het specimen wordt uit de kast of andere opslagplaats gehaald.
2. De specimens worden gesorteerd op alfabetische volgorde.
3.
 - a. De specimens worden gecontroleerd op juistheid van de naam en overige gegevens.
 - b. Indien de naam en/of overige gegevens niet juist zijn, worden deze verbeterd.
 - c. Voor het geval de herbarium-assistent twijfelt of problemen ondervindt, escaleert hij/zij deze naar een curator.
4.
 - a. De herbarium-assistent gaat na of de specimens gerestaureerd dienen te worden.
 - b. De herbarium-assistent pleegt restauraties aan de nodige specimens.
5.
 - a. Er wordt nagegaan of het specimen ook losse delen heeft.
 - b. De losse delen worden in een enveloppe geplaatst.
6.
 - a. Er wordt nagegaan of er meerdere sheets aanwezig zijn per specifiek specimen.
 - b. Indien er meerdere sheets aanwezig zijn worden zij genummerd.
7. Het specimen wordt voorzien van een stempel en barcode.
8. De barcode wordt ter registratie gescand met een barcode reader dat is verbonden met een laptop.
9. Specimens die voorbereid zijn worden op de juiste rek geplaatst om gefotografeerd te worden.

Uitvoerders en verantwoordelijkheden:

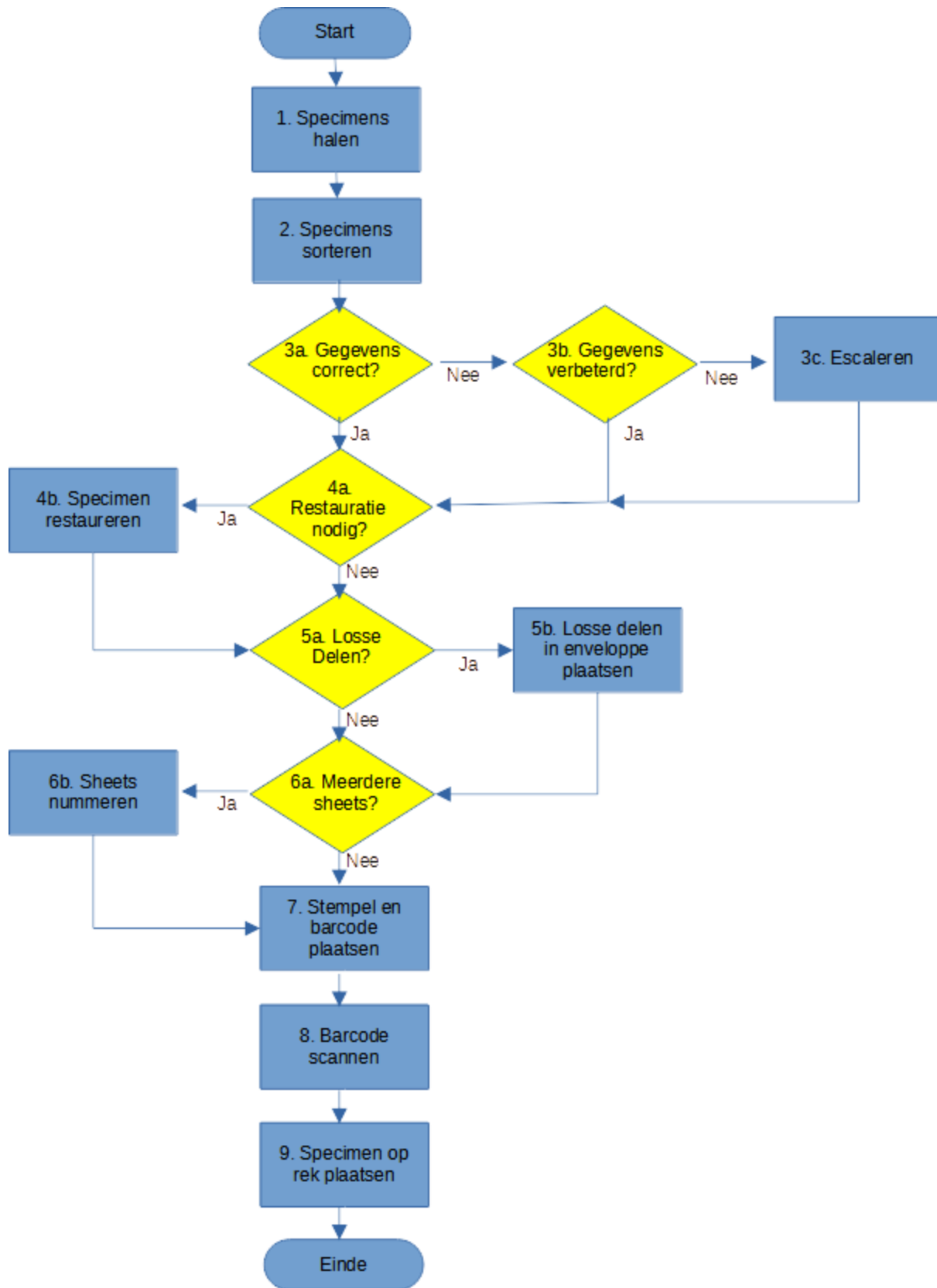
Herbarium-assistent.

Controle en meetmomenten:

- Controleren of naam en/of andere gegevens juist zijn.
- Controleren of restauratie nodig is.
- Controle op losse delen.
- Controle op aantal sheets.

Referenties naar andere processen:

- Dit proces sluit aan op proces fotograferen specimens.



Titel: Procesbeschrijving kalibreren fotocamera.

Beginpunt: Nagaan of de camera loodrecht is geplaatst.

Eindpunt: Het aanmaken van de camera.

Doel: Dit proces beschrijft op welke manier de camera wordt gekalibreerd.

Procesverloop:

1. A. Medewerker gaat na of de camera loodrecht is geplaatst boven het platform van de cameraopstelling.
B. Indien nodig verandert de medewerker de stand van de camera, zodat deze loodrecht staat.
2. A. De uitvoerder gaat na of de camera op de juiste hoogte is afgesteld.
B. Indien nodig wordt de hoogte bijgesteld met de knoppen die daarvoor dienen.
3. Hierna worden de lichten aangemaakt met behulp van het voetpedaal.
4. Vervolgens wordt de camera aangedaan.

Uitvoerders en verantwoordelijkheden:

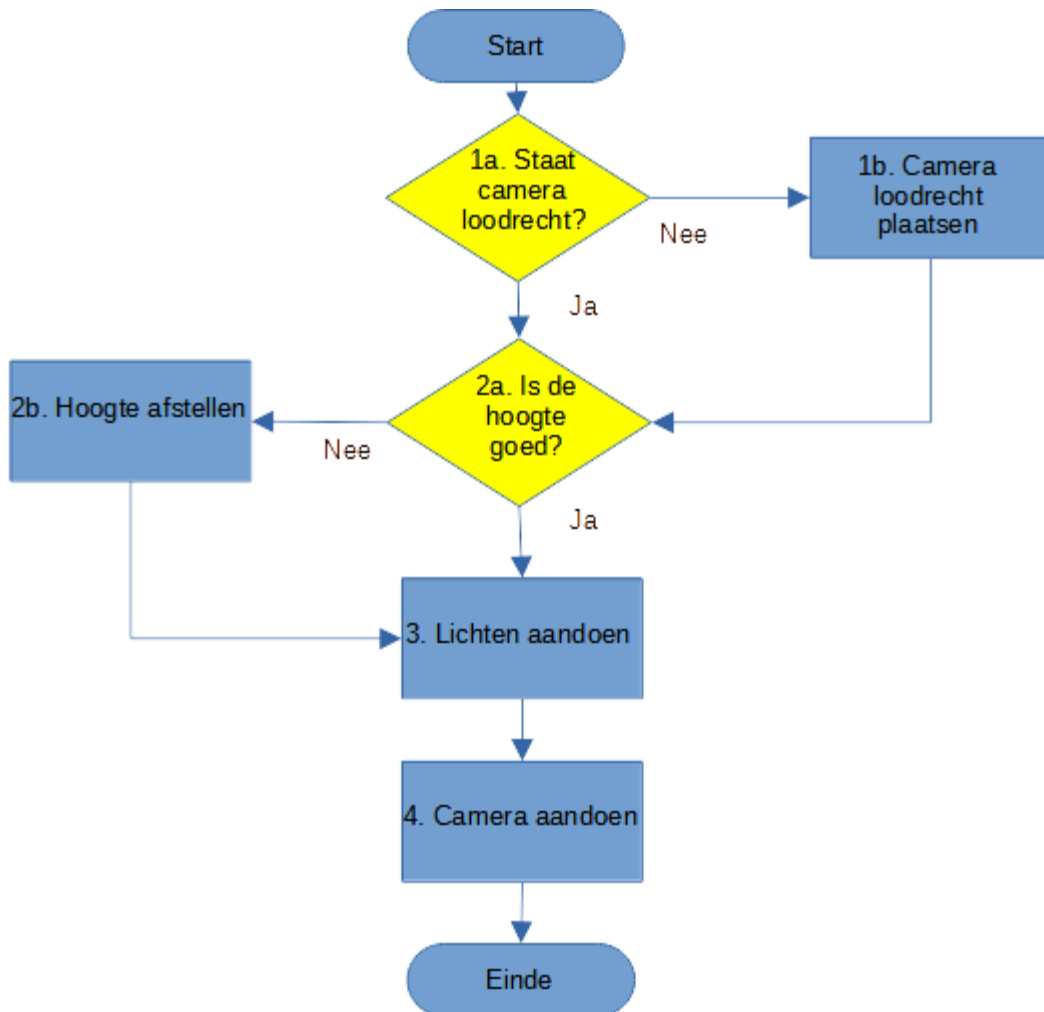
- Herbarium-assistent. (Aansprakelijke)
- Technische medewerker. (Verantwoordelijke handlanger)

Controle en meetmomenten:

- Controleren of de camera loodrecht staat.
- Controleren of de camera op de juiste hoogte staat.

Referenties naar andere processen:

- Dit proces sluit aan op proces voorbereiding specimen.
- Dit proces wordt opgevolgd door proces fotograferen specimen.



Titel: Procesbeschrijving fotograferen specimen.

Beginpunt: Nagaan of de computer aan is.

Eindpunt: Het opslaan van de gemaakte foto.

Doel: Dit proces beschrijft op welke manier specimen worden gefotografeerd.

Procesverloop:

1. A. De medewerker gaat na of de computer aan is.
B. De medewerker maakt de computer aan.
2. De externe harde schijf wordt gekoppeld op de computer.
3. De software 'camera', 'Capture NXD' en 'View' worden opgestart.
4. A. De medewerker gaat na of er een folder aanwezig is op de externe harde schijf waarin de foto's worden opgeslagen. Folders krijgen de naam van plantenfamilies en daarbinnen komen er folders met de plantengeslachten voor.
B. Voor het geval de folder er niet is, wordt deze aangemaakt.
5. De medewerker opent de folder waarin de foto's opgeslagen dienen te worden.
6. A. De medewerker gaat na of de camera aan is.
B. De medewerker maakt de camera aan.
7. Het specimen dat gefotografeerd moet worden wordt op tafel gelegd.
8. A. De medewerker gaat na of de hoogte van de camera goed is afgesteld.
B. De hoogte van de camera wordt juist afgesteld.
9. Er wordt een foto gemaakt van het specimen.
10. De foto wordt renamed naar dezelfde naam als de barcode.
11. De foto wordt op de externe harde schijf opgeslagen in de juiste folder.

Uitvoerders en verantwoordelijkheden:

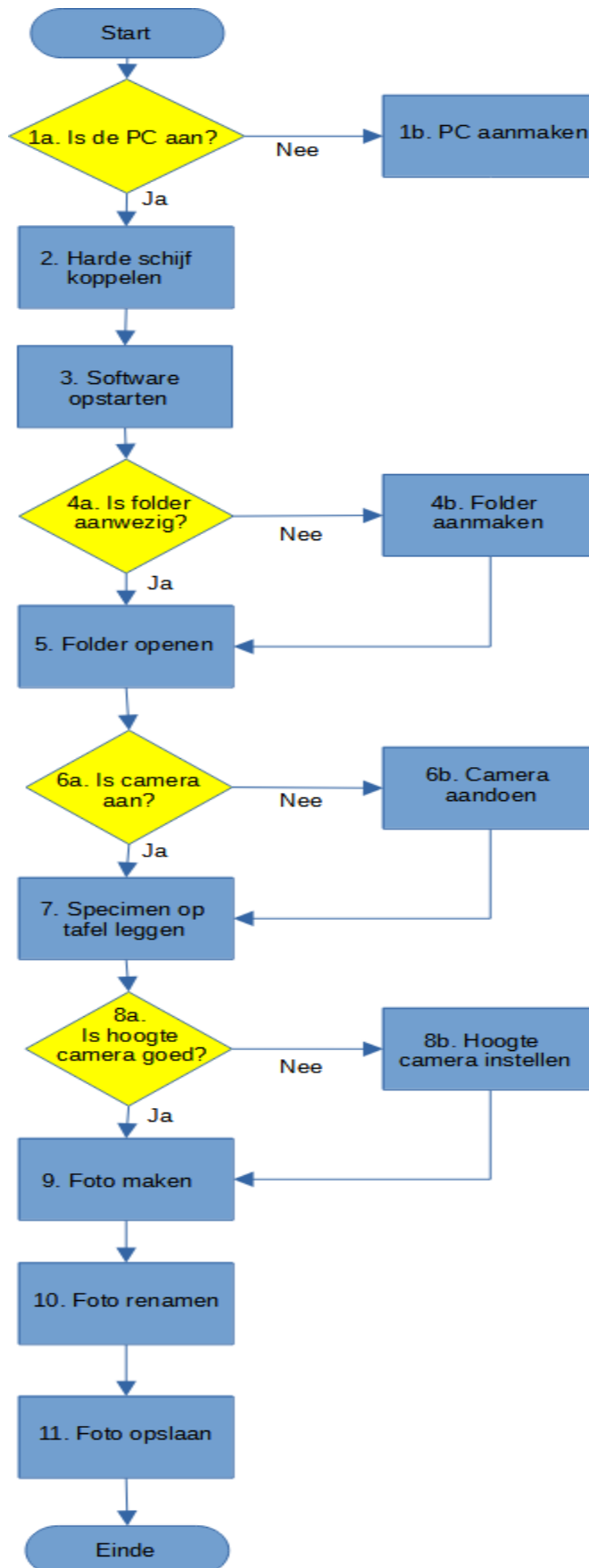
- Herbarium-assistant. (Aansprakelijke)
- Technische medewerker. (Verantwoordelijke handlanger)

Controle en meetmomenten:

- Nagaan of de computer aan is.
- Nagaan of de juiste folder aanwezig is.
- Nagaan of de camera aan is.
- Nagaan of de camera juist is ingesteld.

Referenties naar andere processen:

- Dit proces sluit aan op proces voorbereiding specimen en proces kalibreren fotocamera.



Titel: Procesbeschrijving Dataverwerking.

Beginpunt: Het verkleinen van foto's van specimen.

Eindpunt: Het publiceren van de dataset op de website van GBIF.

Doel: Dit proces beschrijft op welke manier data wordt verwerkt en gepubliceerd op de website van GBIF.

Procesverloop:

1. De databeheerder verkleint de foto's van de specimen naar JPEG formaat.
2. De verkleinde foto's worden opgestuurd naar Alembo. Deze kunnen online worden verstuurd of via een fysieke medium worden gebracht naar het kantoor van Alembo. Alembo plaatst op haar beurt de gegevens van de foto's in DETA.
3. De verkleinde foto's worden geupload op de website van Internet Archive. (archive.org)
4. Nadat Alembo de gegevens heeft verwerkt worden deze gecontroleerd in DETA.
5. Indien de data niet correct is, wordt deze gewijzigd.
Na de controle en eventuele wijzigingen hebben plaatsgevonden, wordt de data vanuit DETA geëxporteerd in CSV formaat. De data export wordt gedaan door de databeheerder.
6. De databeheerder formatteert de geëxporteerde data volgens Darwin Core standaarden en voegt eventuele tijdelijke kolommen toe. Vervolgens stelt de databeheerder de data beschikbaar aan de Herbarium-assistent of Assistent-Curator en onderzoeker(s) of curator in MS Excel formaat. De data wordt beschikbaar gesteld in datasets, waarbij de rijen de verschillende records (occurrences) aangeven en de kolommen de verschillende gegevens aangeven.
7. De databeheerder voert een data cleaning uit op de dataset. Hierbij worden onder andere dubbele records en reeds gepubliceerde records verwijderd.
8. De Herbarium-assistent en onderzoeker gaan na of de data correct is. De Herbarium-assistent en onderzoeker geven in de daarvoor bestemde kolom aan of de data juist en volledig is. Twijfelgevallen of gevallen waar er meer onderzoek voor nodig is kunnen ook in de kolom worden aangegeven.
Indien de data onjuist en/of onvolledig is dan passen de Herbarium-assistent en onderzoeker deze aan in hun daarvoor bestemde kolom.
9. Nadat de Herbarium-assistent en onderzoeker de data hebben gecontroleerd/gewijzigd (tier 1 en tier 2) doet de databeheerder nogmaals een data cleaning. Bij deze derde (tier 3) data cleaning gaat de databeheerder na of de records voldoen aan de vereisten van GBIF. Tijdens deze stap worden ook de links naar de foto's op de website van Internet Archive geplaatst in de dataset.
Ook worden zaken zoals blanco spaties, verschillende schrijfwijzen van namen, juiste datum en locatie aanduidingen etc. in deze fase conform gemaakt.
10. De dataset wordt door de databeheerder gepubliceerd in de test omgeving van GBIF.
11. Er wordt nagegaan of de dataset op de juiste manier is opgemaakt en geformatteerd en of de data en foto's op de juiste manier worden weergegeven.

Indien de data niet juist wordt weergegeven in de test omgeving, past de databeheerder deze aan, eventueel in overleg met de onderzoekers.

12. Nadat de data juist en correct te zien is in de test omgeving van GBIF, gaat de databeheerder over tot het publiceren van de data op de live website van GBIF.

Uitvoerders en verantwoordelijkheden:

- Databeheerder.
- Herbarium-assistent en Onderzoeker.

Controle en meetmomenten:

- Controleren van data in DETA.
- Nagaan of de data in de datasets correct is.
- Nagaan of de dataset voldoet aan de GBIF voorwaarden.
- Nagaan of de publicatie correct is op in de GBIF test omgeving.

Referenties naar andere processen:

- Dit proces sluit aan op proces fotograferen specimen.

Volgende proces:

Specify

1. Importeren van data van Excel/Access naar specify
2. Data editen (bijv. naamsveranderingen* en foute identificatie)
3. Opnieuw foto's maken na de correcte naamsverandering bij foute identificatie bijv. bij verkeerde determinatie van familie, genus, soort
4. Bij foute identificatie publiceren via GBIF en IBOSS en oude foto(s) verwijderen
- 5.

*Naamsveranderingen

1. de naamsverandering wordt door de onderzoeker gevalideerd (Aan GBIF vragen hoe de verandering op de website wordt aangegeven)
2. de correcte naam op de specimen (sheet) wordt door de HA opgeschreven
3. HA voert de gecorrigeerde namen in een aparte Excel file met de naam <HA_ddmmyyyy_naamsverandering.xlsx> (Barcode, oude naam, nieuwe naam, datum van naamsverandering en identifier, datum record gewijzigd in BBS database)
4. de aparte Excel sheet wordt aan de databasebeheerder gegeven
3. De naam wordt in Specify gewijzigd

Allereerste stap:

1. Verzamelen van data (via smart phone* of verzamelboek)
2. Data via smart phone wordt overgezet in Excel
3. Dat via verzamelboek wordt ingevoerd in Excel
- 4.

*D.m.v. Open Data Kit software:

- dropdown lijsten voor verzamelen van planten
- dropdownlijsten voor opnames

Tentatief 2^{de} week nov. training ODK : GM, WK, AG, TS, IM, SD, ND

Fotograferen: Houtcollectie, mossen, fungi, alcoholcollectie

Alcoholcollectie wordt ingevoerd door AG en TS. Daarna specerijen, fungi en mossen.

