

\*\*\*\*\*

# Archief van het aardse leven gaat digitaal

Het Nederlands Centrum voor Biodiversiteit Naturalis werkt aan een veelomvattend digitaliseringsproject. Maarten Heerlien en Jon Peter van den Oever over 7 miljoen collectieobjecten, 80 nieuwe arbeidsplaatsen en 9 werkstraten.

Maarten Heerlien en Jon Peter van den Oever

.....

‘Ga maar eens 800.000 mollusken sorteren’ is bij binnenkomst in de zaal Live Science van museum Naturalis in Leiden te lezen. Eronder staat geprojecteerd: ‘Nog 534.861 te gaan’. Live Science, geopend eind april 2011, is geen doorsnee museumzaal. Het is een experiment waarmee Naturalis de grenzen van het open museum wil opzoeken: een museaal concept waarbij de bezoeker niet langer alleen de ontvanger is van de kennis die het museum uitzendt, maar waarin het bouwen van bruggen tussen museumbezoekers, amateurwetenschappers en vakspecialisten centraal staat. Dat wordt in Live Science bewerkstelligd met bijvoorbeeld bij te wonen prepareersessies, contactdagen voor amateurverenigingen en de gelegenheid tot participatie in collectie- en onderzoekserelateerde activiteiten.

## Labelgegevens overtypen

Bijzonder onderdeel van Live Science vormt de Werkstraat, ingericht met faciliteiten waarmee onderzoekers en collectiebeheerders werk aan de natuurhistorische collectie kunnen verrichten. Alles onder toezicht van het museumpubliek en, tamelijk uniek in de museumwereld, slechts van dat publiek gescheiden door een hekje op kniehooft, om de afstand tussen bezoekers en medewerkers zo klein mogelijk te maken. Op deze manier kan de bezoeker zelf een oordeel vormen over de aard en relevantie van de werkzaamheden van het instituut. Tevens wordt zo de drempel om met vragen op medewerkers af te stappen, zo laag mogelijk gehouden. Naast museaal experiment is de Werkstraat vooral een goed geoliede productiemachine. Hier komen de 800.000 mol-

lusken (weekdieren) om de hoek kijken. In de Werkstraat worden de collecties recente mollusken van museum Naturalis en het Zoölogisch Museum Amsterdam, beide zo’n 400.000 exemplaren, gedigitaliseerd en geïntegreerd. Zeven dagen per week zijn vaste en tijdelijke medewerkers bezig om gegevens van de labels die aan elk object uit de twee collecties zijn bevestigd, over te nemen in een centraal registratiesysteem. Een deel van de objecten en hun labels wordt bovendien gefotografeerd, waarna het materiaal voorzien van barcodes in nieuwe bewaarkasten wordt geplaatst, ingedeeld op taxonomische groep. In lijn met het open museum kan het publiek bij deze werkzaamheden aanhaken door in een daarvoor ontwikkelde crowdsourcingapplicatie gegevens van gefotografeerde labels over te typen, in

\*\*\*\*\*  
**Wie zijn...**  
\*\*\*\*\*



**Maarten Heerlien** is werkzaam als projectleider op de afdeling Informatiediensten van NCB Naturalis en coördineert collectiegerelateerde webprojecten.



**Jon Peter van den Oever** is adviseur digitalisering en als projectleider betrokken bij de FES digitalisering van NCB Naturalis.



Live Science op een van de zes beschikbare iPads en thuis via de website van Naturalis.<sup>1</sup> En hoewel deze applicatie vooral bedoeld is om bezoekers zelf de digitaliseringswerkzaamheden te laten ervaren en niet zozeer om via de crowd grote hoeveelheden data te digitaliseren, zijn het afgelopen jaar toch zo'n 20.000 labels op deze manier overgetypt.

## Biodiversiteit in 37 miljoen objecten

De Werkstraat is onderdeel van het programma FES Collectie Digitalisering (FCD). Dit programma ging in de tweede helft van 2010 van start nadat eind 2009 een bedrag van 30 miljoen euro werd toegekend uit het Fonds Economische Structuurversterking aan het nieuw op te richten instituut Nederlands Centrum voor Biodiversiteit Naturalis. NCB Naturalis, gelanceerd op 28 januari 2010, is een intensieve samenwerking tussen het Zoologisch Museum Amsterdam, Nationaal Herbarium Nederland, de universiteiten van Leiden en Wageningen en Nationaal Natuurhistorisch Museum Naturalis.

De missie: een open archief van de diversiteit van het leven op aarde, waarmee de stamboom van het leven kan worden gereconstrueerd en begrepen. Deze kennis

## 'Slechts een hekjescheidt onderzoekers en collectiebeheerders van het publiek'

van de biodiversiteit draagt het instituut op allerlei manieren en niveaus uit om mensen bewuster te maken van het belang van de natuur en van een duurzaam gebruik van grondstoffen die zij levert. De collecties van de verschillende partners zijn samengevoegd, waarmee een totaalcollectie is ontstaan van 37 miljoen objecten. Qua omvang behoort NCB Naturalis hiermee mondiaal tot de top vijf in het natuurhistorische domein.

Van de toegekende FES-gelden is onder andere een laboratorium voor DNA-barcoding opgezet, waar de komende jaren van minstens 27.000 soorten het DNA in kaart wordt gebracht.<sup>2</sup> 13 miljoen euro is bestemd voor collectiedigitalisering. Met dit budget moeten medio 2015 twee doelstellingen zijn behaald: de digitalisering van minimaal 7 miljoen relevante collectieobjecten, waarvan 2 miljoen binnenshuis en de rest via outsourcing, en de

inrichting van een permanente infrastructuur binnen NCB Naturalis die de voortzetting van de collectiedigitalisering na afloop van het FCD-programma faciliteert.

## Onderzoeksprioriteiten leidend bij digitalisering

Het succesvol op dergelijke schaal en volgens een haast industrieel stramen digitaliseren van collecties – FCD levert tijdelijk tachtig extra arbeidsplaatsen op – is binnen de natuurhistorische sector een primeur die internationaal veel aandacht trekt. Wat het programma bovendien bijzonder maakt, is de pluriformiteit van de te digitaliseren collecties. Digitaliseringsprojecten binnen de natuurhistorische sector richten zich doorgaans op één specifieke soortgroep of collectietype. In het FCD-programma wordt de keuze voor te digitaliseren collecties vooral bepaald door de mate waarin zij aansluiten bij door NCB Naturalis geprioriteerde onderzoeksprogramma's. Daarnaast wordt gekeken naar de economische of maatschappelijke relevantie van collecties en welke collecties vanuit het oogpunt van collectiebeheer het beste kunnen worden gedigitaliseerd.

De collecties die voor digitalisering in aanmerking komen kunnen de meest uiteenlo-

## ‘Met Live Science wil Naturalis de grenzen van het open museum opzoeken’



pende objecten bevatten, en die objecten kunnen op verschillende wijzen geconserveerd worden. Naast de molluskenstraat zijn daarom nog acht digitaliseringstraten ingericht, elk met een focus op een specifieke objectsoort of bewaarmethode. Deze straten registreren gegevens van glaspreparaten, droge vertebraten, exemplaren op alcohol, herbariumvellen, entomologische collecties, houtsamples en geologische en paleontologische objecten. In foto- of scanopstellingen worden van veel van deze objecten op hoge resolutie digitale images geproduceerd.

Ook delen van de bibliotheek, de mediatheek en het archief van NCB Naturalis worden met een speciale boekscanner gedigitaliseerd en gemetadateerd. Dit is minstens even belangrijk als de driedimensionale objecten, want in deze papieren neerslag is de kennis over die objecten beschreven.

### Variabele inrichting werkproces

De inrichting en uitvoering van de digitaliseringprocessen, de outsourcing en de uitrol van de permanente infrastructuur vallen binnen het FCD-programma onder de procesprojecten, ofwel deelprojecten die gericht zijn op de uitvoering van digitalisering. In deze projecten wordt het werkproces voor elk specifiek digitaliseringstraject uitgewerkt en geïmplementeerd, maar worden ook de overkoepelende lijnen vastgesteld, zoals uniforme nummering van objecten. De mate van planvorming voor deze projecten varieert. Op basis van het type project wordt de planvorming gebaseerd op een businesscase, een projectplan of een ‘plan van aanpak’. Afhankelijk van het type en de omvang van elk project zijn voor de wijze van planvorming enige, mede door pragmatisme ingegeven vrijheidsgraden van toepassing.

Met het oog op tijd en kosten wordt per straat beoordeeld welke objectgegevens wel en niet worden geregistreerd en wat wel en niet wordt gefotografeerd of gescand. Uitgangspunt hierbij is telkens de minimale set aan gegevens die nodig zijn om op objectniveau te kunnen beheren en onderzoek te kunnen doen. Per object alle gegevens registreren is binnen de looptijd van het programma en het budget onmogelijk, maar kan zo nodig na 2015 worden gedaan in de permanente infrastructuur. Wel worden van speciale collecties, zoals de wereldberoemde Siebold-collectie en de collectie type-exemplaren, meer gegevens vastgelegd.<sup>3</sup>

### Registratiesystemen

Van elk te digitaliseren object wordt hoe dan ook de bijbehorende wetenschappelijke naam geregistreerd. Daarnaast zijn de verzameldatum en -locatie van belang en de namen van de verzamelaar en de determinant, ofwel degene die de bijbehorende wetenschappelijke naam aan de vondst heeft toegekend. Het gros van die gegevens wordt letterlijk overgenomen van de objectlabels, want voor interpretatie van bijvoorbeeld dubbelzinnige plaatsnamen en datums is weinig ruimte in het proces. Collectiebeheerders hebben een aparte set gegevens nodig, zoals de bewaareenheid en het soort object. Deze data wordt in de digistraten in eerste instantie ingevoerd in het daarvoor intern ontwikkelde Basis Registratiesysteem Digitalisering, in de wandelgangen BRD of Bardigi genoemd. Het BRD, geschreven in Delphi 7 en SQL Server 2005, wordt ingezet om nieuwe basisregistraties van objecten aan te maken, maar ook om bestaande sets met collectiegegevens in te lezen, op te schonen en te converteren. Bijvoorbeeld de data die via de crowdsourcingapplicatie worden verkregen, maar ook verschillende van de ruim 180 databases met collectiegegevens die Naturalis in de loop der jaren heeft opgebouwd en de Mac-databases van het Zoölogisch Museum Amsterdam. Het datamodel achter het BRD voldoet aan de voorwaarden van ABCD 2.06<sup>4</sup> en Darwin Core.<sup>5</sup> Deze standaarden, onderhouden door Biodiversity Information Standards (TDWG),<sup>6</sup> zijn ontwikkeld om natuurhistorische collectiegegevens eenduidig vast te leggen en uit te wisselen en worden breed gedragen door de internati-

onale natuurhistorische gemeenschap. Uiteindelijk worden de collectiedata opgenomen in het pas uitgerolde centrale Collectie Registratie Systeem van NCB Naturalis. De gedigitaliseerde herbariumcollecties worden geïntegreerd in Brahms, een systeem dat binnen de botanische sector is ontwikkeld ter ondersteuning van onderzoek aan en management van herbariumcollecties.<sup>7</sup>

### Duurzamer collectiebeheer

De ontwikkeling van het BRD wordt binnen het FCD-programma aangemerkt als systeemproject. Dit zijn deelprojecten gericht op de ontwikkeling van systemen ten behoeve van alle procesprojecten. Dat betreft naast de opzet en ontwikkeling van het BRD bijvoorbeeld ook de definiëring en uitrol van een Digitale Beeldbank en ondersteunende systemen voor DNA-barcoding. Deze projecten staan als het ware dwars op de andere projecten, zijn ook cruciaal voor de digitalisering en hebben daarmee een sterke afhankelijkheid.

Ten slotte is er in het programma een aantal randvoorwaardelijke projecten benoemd, projecten die ook van belang zijn voor het succes van de procesprojecten maar waarvoor de verantwoordelijkheid elders in de organisatie is belegd. Wel worden door het FCD-programma de financiële middelen beschikbaar gesteld voor de uitvoer van deze projecten. Het gaat bijvoorbeeld om de ontwikkeling van een systeemarchitectuur binnen NCB Naturalis die de verschillende componenten moet laten communiceren met zowel elkaar als met de buitenwereld en om het inkopen van storage- en reken capaciteit. Dat laatste is niet onbelangrijk, want op basis van intern onderzoek wordt geschat dat NCB Naturalis over vijf jaar ruim één petabyte aan collectie- en onderzoeksdata zal hebben geproduceerd.

Met al deze data en systemen kan NCB Naturalis op termijn een duurzamer beheer voeren op onze nationale natuurhistorische collectie. Dit beheersaspect is niet zozeer het einddoel van het FCD-programma, maar vormt eerder een middel om doelen te realiseren die breder zijn dan NCB Naturalis. Een van die te bereiken doelen is waardevermeerdering van collecties. Niet alleen de objecten, maar ook alle informatie rondom die objecten vertegenwoordigen een waarde. Elke wij-

ziging in de collectie voegt waarde toe. Een collectiebeheerder bijvoorbeeld die een deelcollectie toegankelijk heeft gemaakt volgens een systeem, heeft waarde toegevoegd. Zolang dit systeem alleen aan hem bekend is, is de toegevoegde waarde niet geborgd. Hetzelfde geldt voor een onderzoeker die metingen aan objecten heeft verricht.

Het vastleggen van deze informatie in een centraal beheerde databank borgt de waardevolle informatie voor toekomstige gebruikers. De waarde wordt het grootst als de gegevens worden gedigitaliseerd van de objecten die ook relevant zijn voor de (beleids)prioriteiten van NCB Naturalis. Op het laagste niveau wordt dit doel bereikt door informatie over delen van de collectie te uniformeren en digitaliseren in centrale systemen.

## Hommels en waterkevers

Met deze verrijkte collecties zijn vervolgens tal van toepassingen mogelijk die economisch of maatschappelijk interessant of zelfs van groot belang zijn. Zo worden in de Entomologiestraat de data van de deelcollectie Hommels gedigitaliseerd en worden de geconserveerde hommels zelf per lade op hoge resolutie gefotografeerd. De digitale hommels vormen vervolgens een referentiecollectie op basis waarvan onderzoekers trends in voorkomen en verspreiding van hommelsoorten in kaart brengen. Aangezien hommels voorname bestuivers van voedselgewassen zijn, raakt digitalisering van deze collectie aan de staat van de Nederlandse landbouwsector. Hetzelfde geldt voor de digitalisering van de collectie waterkevers, indicatief voor onze waterkwaliteit.

De 2D-straat richt zich onder andere op de digitalisering van boeken en tijdschriften over de Nederlandse flora en fauna. Dit gebeurt in het kader van het organisatiebrede programma Natuur van Nederland, in 2011 in NCB Naturalis gestart. Het programma heeft het vergroten van kennis over de Nederlandse natuur ten behoeve van wetenschappelijke en maatschappelijke vraagstukken en ten behoeve van voorlichting van het publiek als doelstellingen. De gedigitaliseerde werken worden online toegankelijk gemaakt. Uiteraard maakt digitalisering van de collecties het voor NCB Naturalis gemakkelijker om in te haken op grotere

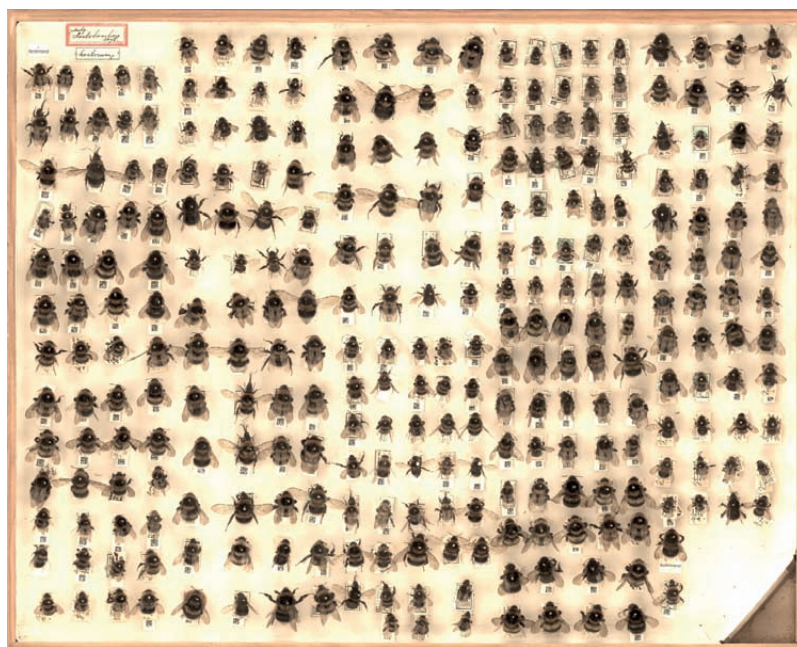
internationale data-aggregatoren, zoals de digitale bibliotheek voor natuurhistorische literatuur BHL-Europe, Europeana en GBIF, het wereldwijde netwerk voor biodiversiteitsdata.

## Draagvlak

Om de doelstellingen van alle 27 proces-, systeem- en randvoorwaardelijke deelprojecten te behalen vóór juli 2015, heeft NCB Naturalis nog een lange weg te gaan. De vaart zit er evenwel goed in. De teller voor de intern te digitaliseren collectieobjecten staat momenteel op ruim 800.000 en wekelijks komen daar gemiddeld zo'n 15.000 stuks bij. Ook de meeste andere projecten bevinden zich momenteel in de uitvoerende fase.

De grootste uitdaging die NCB Naturalis nog rest is: alle medewerkers digitaal laten werken met het Collectie Registratie Systeem en andere centrale systemen. Alleen zo kan in de toekomst, als het FCD-programma overgegaan is in een standaardproces, de digitalisering van steeds weer andere groepen uitgevoerd en ingepast blijven worden. En kan reeds gedigitaliseerd materiaal verder verrijkt worden door er steeds weer nieuwe informatie aan toe te voegen.

Deze manier van werken wijkt af van de huidige. Daarom wordt het enthousiasme ervoor zorgvuldig gekweekt door het FCD-team en het management van



Een van de gedigitaliseerde laden met hommels

Foto: NCB Naturalis

NCB Naturalis. Dat doen ze vooral door wetenschappers, collectiebeheerders en medewerkers van de sector Publiek te helpen steeds weer nieuwe toepassingen van de gedigitaliseerde content te ontdekken. Uiteindelijk is het succes van FCD, zoals dat geldt voor elk groot digitaliseringproject, maar van één factor afhankelijk: het draagvlak dat het geniet onder de medewerkers van NCB Naturalis en haar belanghebbenden.<sup>8</sup>

## Noten

- 1] Zie: [www.naturalis.nl/nl/het-museum/livescience/crowdsourcing/](http://www.naturalis.nl/nl/het-museum/livescience/crowdsourcing/).
- 2] Zie voor meer informatie hierover [science.naturalis.nl/research/dna-barcoding](http://science.naturalis.nl/research/dna-barcoding).
- 3] Over de collectie van Philipp Franz von Siebold (1796-1866) zie [www.sieboldhuis.org/hetsieboldhuis/siebold](http://www.sieboldhuis.org/hetsieboldhuis/siebold). Type-exemplaren zijn exemplaren die zijn aangewezen als voorbeeld en naamdrager van de soort en zijn als zodanig van cruciaal belang voor wetenschappelijk onderzoek. Ze hebben daarom een aparte status binnen natuurhistorische collecties. Voor meer hierover, zie [science.naturalis.nl/collections/schatten-van-ncb-naturalis](http://science.naturalis.nl/collections/schatten-van-ncb-naturalis).
- 4] ABCD: Access to Biological Collections Data.
- 5] Deze standaard wordt onder andere gebruikt voor uitwisseling van informatie over geografisch voorkomen van soorten: [www.tdwg.org/activities/darwincore/](http://www.tdwg.org/activities/darwincore/).
- 6] TDWG: Taxonomic Databases Working Group.
- 7] Brahm's: Botanical Research and Herbarium Management System.
- 8] Bekijk voor een impressie van het project de promotiefilm op [www.naturalis.nl/nl/over-ons/wat-doen-wij/collecties/](http://www.naturalis.nl/nl/over-ons/wat-doen-wij/collecties/).