

PARELS VAN DATASETS

Project in het kader van het KNAW fonds voor een data-adviseur

Verslag van de archivering van de NIOO datasets

*‘Bird ringing data Netherlands 1960-1990 part 1’
&
‘Water Quality research Loosdrecht lakes (W.Q.L.)’*

*Onder redactie van:
Marianne van der Heijden en Saskia Woutersen-Windhouter*

NIOO-KNAW, 28 april 2016

Medewerkers:

1. Kiki Bink, Esther Brouwer, Maria Hundscheid en Joanne Malotau: verwerken ringgegevens
2. Henk van der Jeugd en Murad Maas: begeleiding, databewerking en validatie, Vogeltrekstation, NIOO-KNAW
3. Tim Walles MSc, projectmedewerker en Liesbeth Bakker, begeleiding, Aquatische Ecologie (AqE), NIOO-KNAW
4. Jorik Booij, junior informatiespecialist, Marianne van der Heijden, senior informatiespecialist NIOO-KNAW
5. Saskia Woutersen, stafmedewerker wetenschappelijke informatievoorzienig NIOO-KNAW en coördinator van het project
6. Marnix Berchum, projectmedewerker DANS en Valentijn Gilissen, data-adviseur DANS

NIOO-KNAW, 28 april 2016



Dit werk valt onder een [Creative Commons Naamsvermelding 4.0 Internationaal-licentie](https://creativecommons.org/licenses/by/4.0/).

Inhoudsopgave

Samenvatting.....	4
1. Achtergrond.....	5
1.1 Inleiding.....	5
1.2 Data	5
1.2.1 Ringgegevens.....	5
1.2.2 Waterkwaliteitsonderzoek Loosdrechtse Plassen (WOL)	7
1.3 Aanpak.....	7
2. Digitalisering.....	8
2.1 Ringgegevens.....	8
2.2 Waterkwaliteitsonderzoek Loosdrechtse Plassen (WOL)	8
3. Curatie	10
4. Zichtbaarheid/uitwisseling	12
5. Resultaat.....	15
Dankwoord	16
Referenties	17
Literatuur.....	17
Datasets.....	17
Websites.....	17

Samenvatting

In 2015 is in het kader van het project 'Parels van Data' gewerkt aan het gereed maken en publiceren van twee historische datasets. Het betreft de ringgegevens van het Vogeltrekstation uit de periode 1960 – 1990 en de gegevens van het Waterkwaliteitsonderzoek Loosdrechtse Plassen van de afdeling Aquatische Ecologie. De vogelringdata zijn gedigitaliseerd, gecontroleerd en van kenmerken voorzien (geografische locatie, taxonomische waarde) en daarna gepubliceerd bij DANS EASY. Om deze data ook een rol te geven in de Nederlandse Biodiversiteits Informatie Faciliteit (NLBIF) zijn de gegevens geconverteerd en aangepast aan de Global Biodiversity Information Facility (GBIF) gegevens en gepubliceerd in het NIOO-IPT-NLBIF-portaal. De waterkwaliteitsgegevens van de Loosdrechtse Plassen zijn geïnventariseerd, aangevuld en gedigitaliseerd en in het vereiste format gepubliceerd bij DANS EASY. Voor deze historische data geldt dat er soms een behoorlijke investering nodig was om de gegevens bruikbaar te maken en te laten aansluiten bij soortgelijke gegevens uit het veld. Met dit project 'Parels van data' is een start gemaakt met twee dataverzamelingen die zeker nog een vervolg gaan krijgen.

1. Achtergrond

1.1 Inleiding

Het Nederlands Instituut voor Ecologie (NIOO-KNAW) doet fundamenteel en strategisch ecologisch onderzoek van planten, dieren en microben op het land en in het water. Bijna alle onderzoekers produceren voor hun onderzoek digitale data. Het NIOO heeft daarom sinds 2006 actief geïnvesteerd in een data-infrastructuur, en sinds enkele jaren is daar beleid aan toegevoegd. Volgens dit beleid dienen de data die zijn verzameld en zijn gebruikt voor een publicatie, te worden opgeslagen in één van de NIOO-archieven, of bij een externe host zoals Dryad¹ of Pangea.² De NIOO-archieven bestaan uit (1) een data-portal³ opgezet in 2006 en gehost bij het VLIZ⁴ en (2) het DataverseNL⁵ netwerk dat sinds 2014 wordt beheerd door DANS.⁶ Hoewel het NIOO nu bijna 10 jaar actief is op dit gebied, heeft het NIOO ook geconstateerd dat nog weinig data worden hergebruikt, er onzekerheid is over de duurzaamheid van de opslag en de data voor potentiële gebruikers slecht vindbaarheid zijn op het internet. Met de KNAW subsidie 'Parels van data' is nader onderzocht hoe hergebruik, duurzaamheid en vindbaarheid kunnen worden verbeterd.

Op basis van twee unieke en historische 'Parels van datasets' is in nauwe samenwerking met de datamanagers van DANS, en onderzoekers en informatiespecialisten van het NIOO het volgende onderzocht:

1. Welke **datacuratie** noodzakelijk is om **hergebruik** beter mogelijk te maken. In het kader daarvan zijn de data op herbruikbaarheid getoetst en zijn indien nodig de data gecureerd zodat gegarandeerd is dat deze kan worden hergebruikt.
2. Welke (gedeelde) **standaarden en formaten** in het kader van **duurzame opslag** zijn aan te raden zodat de gegevens ook in de toekomst gegarandeerd toegankelijk zijn.
3. Hoe de gegevens eenvoudig kunnen worden uitgewisseld waardoor de data in meerdere systemen zichtbaar kunnen worden gemaakt, en welke internationale en disciplinaire **standaarden en formaten** daarvoor kunnen worden gebruikt zodat de metadata discipline-specifiek kunnen worden uitgewisseld (met het doel dat de data daardoor beter **zichtbaar en toegankelijk** worden).

1.2 Data

De twee datasets betreffen (i) ringgegevens en (ii) waterkwaliteit onderzoek Loosdrechtse plassen, respectievelijk uit de jaren 60 en 80 van de vorige eeuw.

1.2.1 Ringgegevens

Het Vogeltrekstation⁷ maakt onderdeel uit van het NIOO-KNAW en is hét expertisecentrum op het gebied van vogeltrek en vogeldemografie. Het Vogeltrekstation regelt de dagelijkse gang van zaken rond het vangen en ringen van vogels voor wetenschappelijk onderzoek. Door vogels te vangen en te voorzien van een metalen ring wordt informatie verkregen over de (veranderende) trek, de reproductie en de overleving van Nederlandse vogels. Het

¹ datadryad.org

² pangaea.de

³ data.nioo.knaw.nl

⁴ vliz.be

⁵ dataverse.nl

⁶ dans.knaw.nl

⁷ vogeltrekstation.nl

Vogeltrekstation verstrekt machtigingen tot het vangen en ringen van wilde vogels aan individuele ringers. Op dit moment zijn er ruim 530 ringers met zo'n machtiging actief in Nederland. Per jaar voorzien zij circa 300 duizend vogels van een metalen ring. Deze ringen worden uitgegeven door het Vogeltrekstation.

Het Vogeltrekstation beheert de ring- en terugmeldgegevens van vele miljoenen vogels, die al sinds 1911 in ons land van een ring zijn voorzien. Het Vogeltrekstation werkt daarbij als intermediair tussen verzamelaars en gebruikers van deze gegevens en kennis, ten behoeve van wetenschap, beleid en bescherming. De door het Vogeltrekstation beheerde databank bevat momenteel ruim 10 miljoen records en is beschikbaar voor analyse. Aanvragen voor het gebruik van de gegevens komen direct bij het Vogeltrekstation binnen of via de overkoepelende organisatie EURING.

Sinds 1991 zijn alle ring- en terugmeldgegevens digitaal beschikbaar. Uit de periode 1911-1991 zijn alle ringgegevens van vogels die later zijn teruggemeld eveneens digitaal beschikbaar, maar met name ringgegevens van algemene Nederlandse broedvogels uit deze periode die niet zijn teruggemeld, liggen opgeslagen in een papieren archief. Uit jaarlijkse ringverslagen is bekend dat circa 3 miljoen gegevens nog liggen te wachten op vertoetsing. De historische papieren gegevens zijn bijzonder kwetsbaar en er zijn geen kopieën.

Digitalisatie van **ringgegevens** van vogels die niet zijn teruggemeld is waardevol voor tenminste drie redenen:

1. De datum waarop jonge vogels in het nest werden geringd is een zeer bruikbare maat voor de *timing* van het broedseizoen. Wanneer alle ringgegevens zijn ingevoerd wordt het mogelijk om over een periode van ruim honderd jaar te analyseren in hoeverre verschillende soorten of ecologische groepen in staat zijn geweest schommelingen in temperatuur te volgen. Dergelijke unieke analyses stellen ons vervolgens in staat om te voorspellen in welke mate soorten of ecologische groepen in de toekomst in staat zullen blijken de snelle opwarming van het klimaat te volgen.
2. Ring- en terugmeldgegevens van algemene vogels zijn bij uitstek geschikt voor demografische analyses, met name van jaarlijkse overleving. Langjarige reeksen van de jaarlijkse overleving van vogels, gekoppeld aan populatietrends werpen licht op de sturende factoren achter aantalsveranderingen en kunnen helpen bij het formuleren van maatregelen die negatieve populatieontwikkelingen kunnen keren. Voor deze analyses is het nodig dat complete ringgegevens inclusief datum, locatie en leeftijd van de geringde vogels beschikbaar zijn.
3. Historische ringgegevens van Nederlandse broedvogels kunnen worden gebruikt om trends over zeer lange tijdsperiodes te construeren, aangezien de aantallen geringde vogels van verschillende leeftijdsklassen, gecorrigeerd voor vanginspanning, informatie geven over de aantallen die aanwezig waren.

Voor dit project zijn 69.000 **ringgegevens** uit de jaren 1960-1990 vertoetst en toegevoegd aan de database zodat deze gegevens kunnen worden gebruikt voor de nieuwe en meer volledige werkwijze. De data kunnen worden aangevraagd bij het Vogeltrekstation of via de overkoepelende organisatie EURING. Bij het beschikbaar stellen van de gegevens wordt gestreefd naar een samenwerking tussen aanvrager en het Vogeltrekstation en de gegevens worden om niet ter beschikking gesteld. De gegevens zijn benaderbaar via de

internetapplicatie griel.nl. Delen van de dataset zijn inmiddels open access beschikbaar gemaakt via DANS en GBIF.⁸

1.2.2 Waterkwaliteitsonderzoek Loosdrechtse Plassen (WOL)

Van 1982 tot en met 1990 is één van de eerste grootschalige onderzoeken uitgevoerd naar de effecten van herstelmaatregelen op de waterkwaliteit van de Loosdrechtse Plassen. Het onderzoek werd gecoördineerd vanuit het NIOO (destijds vanuit het Centrum voor Limnologie in Nieuwersluis). Het project was gefinancierd door de EU en een samenwerkingsverband tussen stakeholders uit het waterbeheer, overheden en wetenschappers. Doel was het effect van verminderde externe fosfaatbelasting op de waterkwaliteit en het aquatisch ecosysteem te monitoren. De gegevens zijn verspreid gerapporteerd in een veelheid van studentenverslagen, tussentijdse overzichten en jaarlijkse en samenvattende rapporten. Een coherent overzicht van de data ontbreekt, terwijl mede dankzij dit onderzoek het herstel van de Loosdrechtse Plassen als een van de best gedocumenteerde voorbeelden bekend is in de praktijkwereld. Echter de toegankelijkheid van de gegevens is slecht en de documentatie ad hoc. Doel van dit 'Parels van data' project is de rapporten, die in het bezit zijn van het NIOO, door te werken en de bruikbare gegevens te digitaliseren en op een toegankelijke manier te digitaliseren en openbaar beschikbaar te maken. Herstel van de waterkwaliteit en aquatische ecosystemen in ondiepe meren is wereldwijd een urgent thema, waarbij goede toegankelijkheid van basisdata van de kennis vergaard door het NIOO een groot verschil kan maken.

1.3 Aanpak

De twee datasets hierboven hebben de volgende werkwijze doorlopen:

1. digitaliseren en archiveren met behulp van gedeelde standaarden. In dit geval de bestandsbeschrijving en geprefereerde formaten van DANS.
2. datacuratie (zodat de data zijn geoptimaliseerd voor hergebruik).
3. uitwisseling van de (meta)data verbeteren het gebruik van internationale en discipline gerichte standaarden.

⁸ gbif.org

2. Digitalisering

2.1 Ringgegevens

Al bij een eerdere gelegenheid werd een dataset met historische ringgegevens (van 1910-1959) in DANS gepubliceerd.⁹ In het kader van het project 'Parels van datasets' zijn historische ringgegevens (van niet-teruggevonden vogels) uit de periode 1960-1990 ingevoerd. De invoer van deze historische gegevens is uitgevoerd door een groep ervaren invoerders bestaande uit ringers en studenten. Voor de invoer is een eenvoudig Excel-werkboek gebouwd met datavalidatie. Gegevens over ringnummer, vogelsoort, leeftijd, ringlocatie en ringdatum zijn ingevoerd, naast een wisselend aantal andere variabelen mits voorhanden in de originele data. De locatiegegevens zijn gematched met geografische coördinaten. Vervolgens zijn de gegevens gecontroleerd en toegevoegd aan de centrale database van het Vogeltrekstation in het Euring-format. De Euring-code is beschreven in het Euringcode 2000+ rapport.¹⁰ Vanuit deze database zijn de gegevens, voorzien van metadata, geëxporteerd naar DANS. De set van data is in december 2015 geüpload naar DANS EASY.¹¹ Tegelijkertijd zijn ook data ingevoerd uit de periode 1911-1959 om een eerder naar DANS geëxporteerde dataset verder te complementeren. Deze data zijn inmiddels ook gecontroleerd en worden nu toegevoegd aan de database.

Hoewel de invoer in Excel eenvoudig en snel is voor de invoerders zijn de validatiemogelijkheden beperkt, wat een uitgebreide foutencontrole en correctie nodig maakte. Bovendien was het nodig om geografische coördinaten bij de locatienamen te zoeken, hetgeen deels handmatig moest gebeuren. Daarom is bij de evaluatie van het eerdere project besloten een eenvoudig invoerscherm op de database te bouwen dat speciaal is toegesneden op de invoer van historische ringgegevens. Omdat de historische ringgegevens weinig gedetailleerde informatie bevatten is dit invoerscherm veel eenvoudiger en sneller in het gebruik dan de invoerschermen die ringers momenteel gebruiken voor de invoer van hun gegevens. Intussen is een eerste versie van het invoerscherm gereed en getest en wordt het binnenkort in gebruik genomen.

2.2 Waterkwaliteitsonderzoek Loosdrechtse Plassen (WOL)

Liesbeth Bakker was zelf in de jaren 80 niet betrokken bij de WOL-onderzoeken, maar doet wel al enige tijd onderzoek aan de Loosdrechtse Plassen, met name naar waterplanten. Voor het huidige onderzoek zijn de historische gegevens uit de WOL-onderzoeken een waardevolle aanvulling. De archieven van de WOL-data stonden op schijfjes (de digitale bestanden) en de papieren zaten ongeordend in een doos. Verder zijn er veel rapporten gepubliceerd en een eindverslag in boekvorm.¹²

⁹ van der Jeugd, H. P. (NIOO-KNAW) (2014). *Historical data on timing of ringing of nestling birds*. DANS. URL: dx.doi.org/10.17026/dans-zn5-733s.

¹⁰ Speek, G., Clark, J. A., Rohde, Z., Wassenaar, R. D., & Van Noordwijk, A. J. (2001). *The EURING exchange-code 2000*. Heteren. URL: euring.org/files/documents/data_and_codes/euring_exchange-code_2000.pdf.

¹¹ van der Jeugd, H. P. (NIOO-KNAW, Vogeltrekstation) (2015). *Bird ringing data Netherlands 1960-1990 part 1*. DANS. URL: dx.doi.org/10.17026/dans-2ch-6s6r.

¹² van Liere, L., & Gulati, R. D. (Eds.). (1992). *Restoration and recovery of shallow eutrophic lake ecosystems in The Netherlands : proceedings of a conference held in Amsterdam, The Netherlands, 18-19 April 1991*: Kluwer.

Deze WOL-data zijn door een junior medewerker, recent afgestudeerd in aquatische ecologie, gedigitaliseerd. Hij heeft in eerste instantie de materialen geïnventariseerd en de WOL-data gecontroleerd door alle bronnen naast elkaar te leggen en te checken op consistentie van de gerapporteerde parameters en de eenheden daarvan zorgvuldig terug te zoeken in de originele rapportages. Daarnaast heeft hij de monsterpunten op kaart gezet uit de verschillende bronnen, zodat duidelijk is welke van dezelfde monsterpunten kwamen. Na de inventarisatie van de meetgegevens heeft hij de geselecteerde gegevens omgezet in een database en aangevuld met eutrofieringgegevens¹³ en gegevens van Waternet. Vervolgens zijn de gegevens geconverteerd en gecontroleerd, in het bijzonder door een vergelijking te maken tussen de Waternet en WOL-data en de overlap daarin. Daardoor werd duidelijk wat voor parameters en eenheden de gemeten getallen precies weergaven. De database is als CSV-bestand geüpload naar DANS EASY. De dataset is aangevuld met digitale scans van (een deel van) de rapporten uit het WOL-project. De complete WOL-dataset is gedeponereerd en toegankelijk bij DANS.¹⁴

¹³ roelfpot.nl/projecten/eutrofieringsenquête.php

¹⁴ E.S. Bakker (NIOO KNAW) (2015). *Water Quality research Loosdrecht lakes (W.Q.L.)*. DANS. URL: dx.doi.org/10.17026/dans-xf2-y2vy

3. Curatie

DANS heeft in overleg met de dataproducenten de data gecureerd. Daardoor zijn de data nu beter geschikt voor hergebruik. Probleem bij het archiveren van data door de dataproducent (zonder tussenkomst van een data-archief) is dat de dataproducent zijn data doorgaans zo goed kent, dat er in zijn/haar ogen niets aan hoeft te gebeuren om de data voor hergebruik geschikt te maken.¹⁵ In het curatie-proces heeft DANS geadviseerd welke formaten het beste kunnen worden gebruikt voor duurzame archivering, welke metadata-standaarden kunnen worden gebruikt zodat de metadata volledig en begrijpelijk zijn en tot slot de metadata getest op volledigheid en begrijpelijkheid.

Veel dataproducenten slaan de onderzoeksdata alleen op in het formaat waarin de data zijn gecreëerd. Vaak zijn die formaten niet duurzaam. DANS adviseert het gebruik *preferred* formats.¹⁶ Dat heeft tot gevolg dat de WOL-publicaties in PDF en de Excelfile in CSV-formaat is gearchiveerd. Omdat het om scans van oude publicaties gaat, is PDF voldoende en heeft conversie naar het archiefformaat PDF/A nauwelijks tot geen meerwaarde. In het geval van het Vogeltrekstation is zowel het originele Excel-bestand als een versie in CSV gearchiveerd. De databases zijn begeleid door een codeboek, waarin de opbouw van de tabel en de gebruikte coderingen worden uitgelegd. Voor de Vogeltrekstation data voorziet de EURING code manual hierin; en voor de WOL-data zit deze informatie in de metadata of, voor wat betreft gebruikte methodologie van de monsternamen, in de originele rapporten.

Aangezien de meeste datasets niet automatisch van metadata worden voorzien moet de dataproducent zelf voor de metadatering zorgen. DANS gebruikt daarvoor de standaard 'Qualified Dublin Core'. Dit *Dublin Core* formaat is een gestandaardiseerd formaat en bestaat uit 15 veldbeschrijvingen. Het is tevens zeer geschikt voor het uitwisselen van metadata. De ringgegevens waren al gereed gemaakt voor uitwisseling in de Euring database, en daarmee was meteen aan een internationale archiveringsstandaard voldaan. Voor de WOL-data is voor de officiële aanbodfase overleg geweest met DANS over de juiste manier van aanleveren zodat alle benodigde gegevens bij het officieel uploaden op een juiste manier beschikbaar waren.

Tot slot test DANS de metadata op volledigheid en begrijpelijkheid voordat deze wordt gepubliceerd. Dat gebeurt volgens de principes van het *Open Archival Information System* (OAIS). Dat houdt in dat het door de dataproducent aangeboden informatiepakket, het zgn. *Submission Information Package* (SIP), handmatig gecontroleerd wordt op volledigheid en begrijpelijkheid. Deze *ingest*-functie zorgt ervoor dat de data klaar is voor archiveren en er een goede set beschrijvende metadata gereed is, het zgn. *Archival Information Package* (AIP). Deze set wordt in het archief gearchiveerd. De gebruiker kan vervolgens gebruik maken van een *Dissemination Information Package* (DIP). Het DIP is het pakket aan gegevens dat uiteindelijk naar de gebruiker wordt gestuurd. Dit pakket bestaat bij DANS-EASY uit de

¹⁵ Zie ook: Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLoS Biology*, 13(11). doi:10.1371/journal.pbio.1002295

¹⁶ dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreferredformatsUK.pdf

door de gebruiker gekozen dataset(s), de metadata in XML, *checksums* en de algemene bepalingen van DANS, de zgn. ‘*DANS General Conditions of Use*’.¹⁷

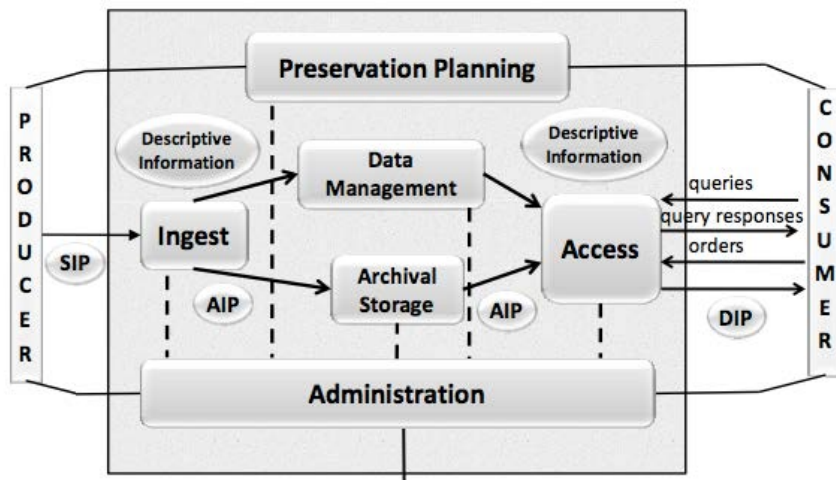


Fig. 1 *Open Archival Information System (OAIS)*¹⁸

Aan iedere dataset (de bestanden (*Data files*) inclusief de projectbeschrijving (*Description*) in *Dublin Core* metadata) is een duurzame identifieer ofwel een *Persistent Identifier (PID)* toegekend. De *PID* is de referentie en snelkoppeling naar de dataset. De gebruikte *PID* bij DANS is een *Digital Object Identifier (DOI)*. Deze identifieer staat bij DANS in de *Description* en leidt de gebruiker naar de voorpagina (*Overview*) van de dataset. Met een *PID* kunnen de data eenvoudig worden gevonden en geciteerd.

¹⁷ dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreservationpolicyUK.pdf

¹⁸ Consultative Committee for Space Data Systems (CCSDS). (June 2012). *Recommendation for Space Data System Practices: Reference Model for an Open Archival Information System (OAIS), Recommended Practice*, CCSDS 650.0-M-2, Magenta Book.

4. Zichtbaarheid/uitwisseling

Onderzoekdata is over het algemeen slecht te vinden. Uitgangspunt in dit project is dat de vindbaarheid van de onderzoeksdata kan worden verbeterd door de uitwisseling van de metadata te verbeteren door het gebruik van internationale en/of discipline gerichte standaarden zoals *Ecological Metadata Language (EML)*, *Biological Metadata Language (BML)* en de *EURING*-code.

DANS

De data is gearchiveerd bij DANS EASY. De dataset ofwel projectbeschrijvingen in DANS EASY worden opgeslagen conform het metadata formaat *Qualified Dublin Core*. DANS EASY is niet ingericht voor andere soorten metadata zoals *Ecological Metadata Language (EML)*, *Biological Metadata Language (BML)* en *EURING*-code. In de praktijk blijkt dat zelden een probleem aangezien voldoende informatie voor het vinden en begrijpen van de dataset in de *Qualified Dublin Core* velden geschreven kan worden. Andere soorten metadata laten zich over het algemeen goed naar *Dublin Core* 'mappen' en vice versa. Omdat voldoende projectgegevens direct bij het deponeren konden worden ingevoerd, bleek het niet noodzakelijk om EML en BML uitgebreid onder de loep te nemen. Specifieke coderingen of meta-documentatie voor databestanden binnen een dataset, konden bovendien worden opgevangen in de codeboeken. Deze codeboeken worden als aparte bestanden aan de dataset toegevoegd. Zo bevat de dataset van vogelringgegevens 'euring_exchange-code_2000.pdf'; en de WOL-dataset 'Codeboek_LdP-databae_WOL_GWA.pdf'. De EURING standaard (gebruikt voor de vogeltrekgegevens) maakt het daarbij mogelijk dat de data internationaal uitwisselbaar is door het gebruik van identieke velden en codes.

De uitwisselingsstandaard van DANS ('*Qualified Dublin Core*') zorgt er bovendien voor dat het Nederlandse Portaal voor Academische Informatie (NARCIS)¹⁹ de data kan *harvesten* volgens het OAI-PMH-protocol. De metadata van zowel DANS EASY als NARCIS zijn opengesteld zodat zoekmachines als Google deze kunnen indexeren. De WOL-data en de ringgegevens zijn dan ook eenvoudig via Google te vinden. Om de vindbaarheid van de data te vergroten zijn de gegevens in het Engels, maar daarnaast zijn ook trefwoorden in de Nederlandse taal toegevoegd.

¹⁹ narcis.nl

EASY offers sustainable archiving of research data and access to thousands of datasets.

nioo-knaw [Search help](#)

[Advanced search](#) [Browse](#)

3 RESULTS IN PUBLISHED DATASETS

List

Sort by:

REFINE 1

Search...

Advanced search

Audience

- Behavioural and educational sciences 1
- Interdisciplinary sciences 1
- Life sciences, medicine and health care 2

Access

- Open (registered users) 3

Water Quality research Loosdrecht lakes (W.Q.L.)

Date: 2015-04-17 Audience: Environmental studies
 Creators: Bakker, dr. E.S. (NIOO-KNAW) Access: Open (registered users)
 Relevance: 100% relevant Submitted: 2015-04-13
 Contributor: Walles, MSc T.J.W. (NIOO-KNAW)
 Source: NIOO-KNAW archive and old 'Waterkwaliteit onderzoek Loosdrechtse plassen' (WOU) - file

Historical data on timing of ringing of nestling birds

Date: 2014-02-17 Audience: Life sciences, medicine and health care
 Creators: Jeugd, dr H.P. van der (NIOO-KNAW) Access: Open (registered users)
 Relevance: 83% relevant Submitted: 2014-02-17
 Contributor: Jeugd, dr H.P. van der (NIOO-KNAW, Vogelrekstation)
 Relation: title=Jeugd, dr HP van der (NIOO-KNAW, Vogelrekstation) (2015): Bird ringing data Netherlands 1960
 Source: , Natural History Museum Leiden and Vogelrekstation (NIOO-KNAW)

Bird ringing data Netherlands 1960-1990 part 1

Date: 2015-12-16 Audience: Behavioural and educational sciences
 Creators: Jeugd, dr HP van der (NIOO-KNAW, Vogelrekstation) Access: Life sciences, medicine and health care
 Relevance: 76% relevant Submitted: 2015-12-16
 Relation: title=Jeugd, dr H.P. van der (NIOO-KNAW) (2014): Historical data on timing of ringing of nestling
 Access: Open (registered users)

Fig. 2 De datasets in Narcis.

NLBIF

In de zomer van 2015 is overleg gestart met NLBIF,²⁰ de Nederlandse tak van het internationale biodiversiteitsportaal Global Biodiversity Information Facility (GBIF). Er is afgesproken dat in eerste instantie de data van de ringgegevens zullen worden opgenomen in NLBIF. Daartoe is een speciale NIOO *Integrated Publishing Toolkit (NIOO-IPT)* ingericht²¹ waarmee de gegevens naar het NLBIF-portaal kunnen worden geconverteerd. NLBIF gebruikt als standaard het *Darwin Core Archive* formaat,²² wat is afgeleid van het *Dublin Core* formaat. De Darwin Core is ontwikkeld om taxonomische gegevens op een eenvoudige en eenduidige manier op te slaan en uit te wisselen. Ook kan in NLBIF voor de taxonomische namen de *Birds-of-the-World-standaard*²³ worden gebruikt, die nu ook voor de *Euring* als standaard geldt, en die gematcht kan worden met de GBIF-lijst. De data die in een eerder project zijn ingevoerd uit de periode 1911-1959 en waarvoor de invoer nog steeds doorgaat zijn al toegevoegd aan NLBIF. De gegevens die zijn ingevoerd in het kader van 'Parels van datasets' zullen aan NLBIF worden toegevoegd. Door het opnemen van de ringgegevens in NLBIF en GBIF zijn gegevens nog meer gestandaardiseerd en nog beter vindbaarder gemaakt voor het publiek dat op zoek is naar biodiversiteitsdata.

²⁰ nlbif.nl

²¹ ipt.nioo.knaw.nl/resource?r=vt-nioo-knaw-historic-ringdata

²² rs.tdwg.org/dwc

²³ ec.europa.eu/environment/nature/conservation/wildbirds/eu_species/index_en.htm

The screenshot shows the GBIF website interface. At the top, there is a navigation bar with 'Data', 'News', 'Community', and 'About' links, along with a search bar. Below this is a green banner for 'Vogeltrekstation - Dutch Centre for Avian Migration and Demography'. The main content area is titled 'Data publisher information' and contains the following details:

- FULL TITLE:** Vogeltrekstation - Dutch Centre for Avian Migration and Demography
- DESCRIPTION:** Vogeltrekstation - Dutch Centre for Avian Migration and Demography - coordinates all ringing activities of wild birds in the Netherlands. The extensive databank of the centre allows for detailed analyses of birds' survival rates, breeding success and movements. Vogeltrekstation stimulates the use of ringing data by initialising collaborations with research institutes and universities, as well as conducting its own independent research.
- ADDRESS:** Vogeltrekstation - Dutch Centre for Avian Migration and Demography, Droevendaalsesteeg 10, NL-6708 PB Wageningen, Netherlands
- WEBSITE:** <https://www.vogeltrekstation.nl/>
- TECHNICAL CONTACT:** [Henk P. van der Jeugd](#)
- ADMINISTRATIVE CONTACT:** [Henk P. van der Jeugd](#)
- POINT OF CONTACT:** [Henk P. van der Jeugd](#)
- ENDORSED BY:** Netherlands Biodiversity Information Facility

Fig. 3 De NLBIF-data wordt wereldwijd met GBIF en nationaal met NARCIS uitgewisseld.

Van de WOL-data zijn voornamelijk de fysisch-chemische gegevens opgeslagen. Er zijn destijds nauwelijks opnamen gemaakt van de macrofyten. Wel is het voornemen om als de WOL-data een goed overzicht van de ontwikkeling van de macrofyten kunnen geven, deze waterplantengegevens ook kunnen worden opgenomen in NLBIF. De metingen die tijdens het WOL-onderzoek zijn verricht, zijn gedaan op basis van de noodzaak van dat moment. Voor dit soort onderzoeken zijn geen vastgelegde standaarden, laat staan templates. Wel zijn de (in de aquatische wereld) bekende grootheden (bijv. totaal fosfaat) gemeten.

PR

Na publicatie van de data in DANS hebben zowel DANS als NIOO de toegankelijkheid van de datasets aangekondigd.²⁴ Op onderzoeksdata-pagina²⁵ van het NIOO staan links naar de datasets. Door opname in DANS, GBIF en Google zijn de data goed zichtbaar en verwachten we een toename in het gebruik ervan door derden en een toename van het aantal verzoeken om meer gedetailleerde gegevens, maatwerkproducten en samenwerkingsprojecten.

²⁴ dans.knaw.nl/nl/actueel/nieuws/parels-van-datasets-data-van-waterkwaliteitsonderzoek-loosdrechtse-plassen; dans.knaw.nl/nl/actueel/nieuws/parels-van-datasets-bird-ringing-data-netherlands-1960-1990-part-1; nioo.knaw.nl/en/news/pearls-data-making-water-quality-research-accessible; nioo.knaw.nl/en/news/historische-gegevens-vogeltrekstation-online-toegankelijk.

²⁵ nioo.knaw.nl/en/onderzoek/datasets.

5. Resultaat

De ringgegevens en de data van het WaterkwaliteitsOnderzoek Loosdrechtse plassen (de WOL-data) blijken enorm van elkaar te verschillen. De ringgegevens van zijn ordelijk en volgens de normen van de Europese gezamenlijke ringersverenigingen in gestandaardiseerde formaten bijgehouden; de hydrologen daarentegen bepalen zelf wat zij op dat moment van belang achten om te meten. Maar beide gegevens kunnen nu na digitalisering en curatie worden uitgewisseld en geïntegreerd met andere onderzoekdata. Voor de waterkwaliteitsgegevens betekende dat wel een duidelijke en meer intensieve inzet van de projectmedewerker, die alle materialen heeft geïnteriseerd en geordend. Voor een deel zijn oud-WOL-medewerkers benaderd met vragen naar ontbrekende gegevens/rapporten waarbij ook rapporten boven tafel zijn gekomen die nog niet bij het NIOO aanwezig waren. Daarnaast bleek er meer data aanwezig te zijn over de waterkwaliteit en waterplanten van de Loosdrechtse plassen, ook uit het WOL-tijdperk, bij zeer diverse instanties, met name de Universiteit Utrecht, provincies Noord-Holland en Utrecht, Waternet, Natuurmonumenten en Alterra (vegetatie database Synbosis en nog niet ingevoerde vegetatie opnamen in excursie boekjes). De gegevens uit het WOL-project zijn nu behoorlijk compleet gearchiveerd, maar een systematisch overzicht ontbreekt. Voor een compleet overzicht van de waterkwaliteit en waterplanten in de Loosdrechtse plassen over de laatste 30 jaar zijn er duidelijk meer diffuse databronnen voor handen.

Het door DANS aangeleverde *preferred* format (CSV en PDF) was met enige aanwijzingen goed te hanteren en door een codeboek erbij te leveren (bij ringgegevens de Euring-code) zijn de gegevens ook voor een niet-deskundige te begrijpen. De aangeleverde gegevens zijn door DANS gecontroleerd op correctheid en volledigheid. De bij DANS opgeslagen metadata is beschikbaar gesteld via *Qualified Dublin Core* en kan worden geharvest via het OAI-PMH protocol. De gegevens zijn nu ook via Google te vinden en aan opname van de gegevens in NLBIF en GBIF wordt nog gewerkt. De conclusie is dat er zoveel als mogelijk gewerkt moet worden met internationale en (indien beschikbaar) discipline specifieke standaarden (Euring, Darwin Core) en dat de herkomst van de gegevens duidelijk uitgelegd wordt in een meegeleverd document. Het voordeel van discipline specifieke standaarden is dat er meer gedetailleerde informatie kan worden meegegeven en data-integratie mogelijk wordt. Omdat de gegevens zijn voorzien van een persistent identifier zijn deze bovendien goed te citeren (en te vinden!). Omdat de data goed zichtbaar zijn, verwachten we een toename in het gebruik ervan door derden en een toename van het aantal samenwerkingsprojecten

Het zorgvuldig omgaan met data is ook lonend voor de mogelijkheden tot vervolgonderzoek. Beide datasets zijn onmiddellijk gebruikt: voor de ringdata betreft het de presentatie en de publicatie naar het NLBIF-NIOO-portaal. De WOL-data en de verdieping daarvan heeft ertoe geleid dat er, vanuit de meegeleverde bronneninventarisatie, verder gewerkt kan worden aan een tijdreeks van waterkwaliteitgegevens. De WOL-data bieden een goede basis om verder te kunnen uitbreiden voor specifieke onderzoeksdoelen. Voor beide datasets betekent de duurzame archivering een mooie stap vooruit in het onderzoek.

Dankwoord

De KNAW wordt bedankt voor het financieren van het project 'Parels van datasets' in het kader van het KNAW fonds voor een data-adviseur.

Het invoerwerk van de ringgegevens voor 'Parels van datasets' is gedaan door betaalde krachten Kiki Bink, Esther Brouwer, Maria Hundscheid en Joanne Malotaux. Daarnaast hebben Joop van Ardenne, Wouter Boere, Jan Klein, Aukje Kloppenburg, Tamar Reijnen en Esmee Schutgens op vrijwillige basis een belangrijke bijdragen geleverd aan het invoerwerk. Deze vrijwilligers gaan ook na afsluiting van het project verder met de invoer van historische gegevens.

Het inventariseren van de WOL-data voor 'Parels van datasets' is uitgevoerd door Tim Wallis. Lowie van Liere, Luc Mur en Herman ten Dam hebben rapporten en informatie gegeven met gegevens over de Loosdrechtse plassen en/of de beschikbaarheid daarvan. Jan Jansse heeft de ingevoerde gegevens van het WOL project voor de ontwikkeling van PCLake beschikbaar gesteld.

Xander van der Sar, Cees Hof en Valentijn Gilissen worden bijzonder hartelijk bedankt voor hun hulp bij het publiceren van de dataset in DANS en NLBIF.

Referenties

Literatuur

- Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLoS Biology*, 13(11). doi:10.1371/journal.pbio.1002295
- van Liere, L., & Gulati, R. D. (Eds.). (1992). *Restoration and recovery of shallow eutrophic lake ecosystems in The Netherlands : proceedings of a conference held in Amsterdam, The Netherlands, 18-19 April 1991*, Kluwer.
- Speek, G., Clark, J. A., Rohde, Z., Wassenaar, R. D., & Van Noordwijk, A. J. (2001). *The EURING exchange-code 2000*. Heteren.
- Consultative Committee for Space Data Systems (CCSDS). (June 2012). *Recommendation for Space Data System Practices: Reference Model for an Open Archival Information System (OAIS), Recommended Practice*, CCSDS 650.0-M-2, Magenta Book.

Datasets

- Bakker, E. S. (2015). *Water Quality research Loosdrecht lakes (W.Q.L.)*.
URL: <http://dx.doi.org/10.17026/dans-xf2-y2vy>
- van der Jeugd, H. P. (2014). *Historical data on timing of ringing of nestling birds*.
URL: <http://dx.doi.org/10.17026/dans-zn5-733s>
- van der Jeugd, H. P. (2015). *Bird ringing data Netherlands 1960-1990 part 1*.
URL: <http://dx.doi.org/10.17026/dans-2ch-6s6r>

Websites

DANS

- <http://dans.knaw.nl>
- <http://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreferredformatsUK.pdf>
- <http://dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreservationpolicyUK.pdf>
- <http://dans.knaw.nl/nl/actueel/nieuws/parels-van-datasets-data-van-waterkwaliteitsonderzoek-loosdrechtse-plassen>
- <http://dans.knaw.nl/nl/actueel/nieuws/parels-van-datasets-bird-ringing-data-netherlands-1960-1990-part-1>
- <http://narcis.nl>

NIOO

- <http://data.nioo.knaw.nl>
- <http://ipt.nioo.knaw.nl/resource?r=vt-nioo-knaw-historic-ringdata>
- <http://nioo.knaw.nl/en/news/pearls-data-making-water-quality-research-accessible>
- <http://nioo.knaw.nl/en/news/historische-gegevens-vogeltrekstation-online-toegankelijk>
- <http://nioo.knaw.nl/en/onderzoek/datasets>
- <http://vogeltrekstation.nl>

Overige dataportalen

- <http://datadryad.org>
- <http://dataverse.nl>
- http://ec.europa.eu/environment/nature/conservation/wildbirds/eu_species/index_en.htm
- <http://gbif.org>
- <http://nlbif.nl>
- <http://pangaea.de>
- <http://roelfpot.nl/projecten/eutrofieringsenquete.php>
- <http://rs.tdwg.org/dwc>
- <http://vliz.be>