# Beleidsplan 2018-2020

#### Stichting Xeno-canto voor Natuurgeluiden

#### Versie

Definitief (april 2018)

#### **Auteurs**

Sander Pieterse (secretaris, Stichting Xeno-canto voor Natuurgeluiden) Willem-Pier Vellinga (voorzitter, Stichting Xeno-canto voor Natuurgeluiden) Bob Planqué (penningmeester, Stichting Xeno-canto voor Natuurgeluiden)

## Samenvatting

Sinds 15 januari 2018 wordt Xeno-canto gehost op de servers van Naturalis Biodiversity Center. De intensievere samenwerking met Naturalis en de verduurzaamde en schaalbare hosting, maken in het verleden geformuleerde ambities van Xeno-canto mogelijk. Dit beleidsplan 2018-2020 beoogt die ambities vorm te geven.

# 1. Xeno-canto in een notendop

De Stichting Xeno-canto voor Natuurgeluiden ("Xeno-canto", "XC") bouwt sinds 2005 met vrijwilligers aan een verzameling vogelgeluiden en aan <a href="www.xeno-canto.org">www.xeno-canto.org</a>, de website die toegang geeft tot de verzameling en bijbehorende database.

Ten tijde van de start van Xeno-canto waren wereldwijd de belangrijkste geluidsverzamelingen voornamelijk analoog. De mogelijkheden van digitaal opnemen en aanbieden van geluiden werden weliswaar onderkend, maar de invloed en mogelijkheden van het web werden niet op waarde geschat. Xeno-canto, aanvankelijk opgezet als een identificatiehulp voor vogelgeluiden uit Zuid-Amerika, werd juist opgezet om de zich steeds ontwikkelende mogelijkheden ten volle te gebruiken. Onder andere daardoor neemt Xeno-canto tussen de andere verzamelingen nog altijd een unieke positie in.

Xeno-canto hanteert een aantal leidende principes:

- iedereen met internettoegang kan geluiden toevoegen,
- correcte identificatie van de soort is de enige harde eis aan de geluiden,
- iedereen kan de identificatie van geluiden ter discussie stellen en becommentariëren,
- de geluidsopnamen worden gedeeld onder een Creative Commons licentie,
- de website maakt waar mogelijk gebruik van open source software.

Deze aanpak is succesvol gebleken: Xeno-canto beheert op dit moment een van de grootste openbaar toegankelijke digitale collecties van vogelgeluiden in de wereld: ~392.000 opnames van ~4250 opnemers. In maart 2018 waren geluiden van ongeveer 9.800 soorten

aanwezig, en waarschijnlijk is het in dat opzicht de grootste openbare verzameling ter wereld. De verzameling wordt veel gebruikt: in 2017 waren er naar schatting 2 miljoen unieke bezoekers en werden de geluiden gebruikt in minimaal 200 door Google Scholar geïndexeerde wetenschappelijke publicaties. De geluiden zijn tientallen miljoenen keren beluisterd.

De doelen van Xeno-canto zijn in de loop van de tijd geëvolueerd. Het doel is niet langer alleen om

- opnemen van vogelgeluiden populairder te maken,
- de toegankelijkheid van vogelgeluiden te verbeteren,
- en om de kennis van vogelgeluiden te vergroten,

maar ook expliciet om een verzameling aan te leggen van álle vogelgeluiden, dat wil zeggen

- van alle taxa,
- hun complete repertoire,
- de gehele geografische variatie,
- in alle ontwikkelingsstadia,
- in de loop van de tijd (temporele variatie; evolutie van zang).

# 2. Beleidsgrondslagen

Hoewel de collectie succesvol mag worden genoemd zijn er kanttekeningen te plaatsen. In 2004 werden door Richard Ranft van de British Library op basis van de praktijk een aantal eisen geformuleerd waaraan openbare geluidsverzamelingen idealiter moeten voldoen<sup>1</sup>. De belangrijkste daarvan luiden:

- 1. langetermijnvisie en planning
- 2. langetermijnondersteuning van een gevestigd instituut voor de organisatie, opslag van en toegang tot de collectie
- 3. de technische expertise en opslagfaciliteiten om punt 2 mogelijk te maken
- 4. een solide beleid om de collectie te behouden, inclusief een beleid om dataverlies door rampen te voorkomen
- 5. behoud van integriteit van de data voor de lange termijn

De hosting door Naturalis heeft ervoor gezorgd dat punten 2-4 naar tevredenheid zijn geregeld. De samenwerking met Naturalis heeft in eerste instantie een looptijd van 3 jaar, en wordt geëvalueerd in december 2020. Xeno-canto en Naturalis zijn overeengekomen te zullen zoeken naar projectmatige samenwerking in de toekomst betreffende doorontwikkeling van de site en het publiceren van de Xeno-canto data via de Naturalis infrastructuur. Een gedeelde ambitie is dat doorontwikkeling van de website opengesteld wordt voor bijdragen van derden en de site zo verder kan groeien met

<sup>&</sup>lt;sup>1</sup> Ranft. R. (2004). Natural sound archives: past, present and future. An Acad Bras Cienc. 76(2): 455-465. http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0001-37652004000200041

behulp van Xeno-canto, Naturalis en een internationale gebruikersgemeenschap. Dit beleidsplan 2018-2020 heeft voornamelijk betrekking op de mogelijkheden voor projectmatige samenwerking met Naturalis. Samenwerking of inititatie van samenwerking met derden wordt niet uitgesloten, maar Naturalis blijft ook dan een natuurlijke samenwerkingspartner.

# 3. Korte termijn, 2018

De volgende punten worden nog in 2018 gerealiseerd in het lopende project met Naturalis:

3.1 Xeno-canto data koppelen aan de Naturalis collectie.

Met behulp van de XC API brengen we XC metadata onder in de document store van de Netherlands Biodiversity API<sup>2</sup> ("NBA"). Ook configureren we een webservice-verzoek voor XC geluiden op het door Naturalis ontwikkelde en beheerde soortinformatiesysteem Linnaeus NG<sup>3</sup>.

Xeno-canto omvat een grote, waardevolle collectie die ingezet kan worden voor allerlei doeleinden. Het ligt voor de hand de collectie te koppelen aan de digitale collectie van Naturalis, waarbij Xeno-canto het *voorkeurs* geluidenplatform wordt. Wanneer werkpakket 1 is uitgevoerd en Xeno-canto op duurzame wijze draait binnen de cloud-infrastructuur van Naturalis, kunnen we de benodigde stappen maken op het gebied van data-uitwisseling.

Allereerst sluiten we Xeno-canto aan op de Netherlands Biodiversity API, waarmee de geluiden en metadata opvraagbaar worden via deze API en getoond worden naast ander collectiemateriaal in de Naturalis BioPortal<sup>4</sup>. We mappen het datamodel van Xeno-canto op die van de NBA en zorgen dat Xeno-canto data zo volledig mogelijk wordt opgenomen in de NBA. Mogelijk zijn daarvoor uitbreidingen nodig aan het datamodel.

In het soortinformatiesysteem Linnaeus NG configureren we de Netherlands Biodiversity API op zo'n manier dat soortinformatieprojecten op dynamische wijze verrijkt kunnen worden met geluiden uit Xeno-canto. Denk daarbij bijvoorbeeld aan mogelijkheid tot afspelen van vogelgeluiden op het Nederlands Soortenregister<sup>5</sup> en in de Dierenzoeker<sup>6</sup>. Deze webservice komt beschikbaar voor alle Linnaeus NG gebruikers.

### 3.2 Xeno-canto data publiceren op GBIF.

De basis voor het publicatiemechanisme naar GBIF wordt gevormd door de Darwin Core Archive publisher op de Netherlands Biodiversity API van Naturalis. Dit mechanisme levert nu op aanvraag DwCA zipfiles voor *occurrence* en *species checklist* data uit de NBA en zet

<sup>&</sup>lt;sup>2</sup> http://docs.biodiversitydata.nl/en/latest/

<sup>&</sup>lt;sup>3</sup> http://linnaeus.naturalis.nl

<sup>4</sup> http://bioportal.naturalis.nl

<sup>&</sup>lt;sup>5</sup> http://www.nederlandsesoorten.nl

<sup>6</sup> http://www.dierenzoeker.nl

deze klaar voor *harvest* door GBIF. Voor het correct naar GBIF aanleveren van de geluiden van Xeno-canto is een aantal acties en aanpassingen aan het publicatiemechanisme nodig:

- Aanmelden van Xeno-canto als organisatie (data publisher) bij GBIF
- Bepalen van de wijze van aanlevering: als occurrences, of als species checklist met multimedia extensie (simple multimedia<sup>7</sup> of uitgebreidere Audubon Media Description 8). Onderzocht wordt welke aanlevering het beste past bij de originele data.
  - o Format en velden van de DwCA file bepalen, EML file aanmaken.
  - Bestaande Darwin Core Archive publicatiemechanisme op Netherlands
     Biodiversity API uitbreiden met mediavelden, en configureren voor publicatie
     van Xeno-canto als dataset naar GBIF.
  - Testen/valideren van DwCA en EML files met behulp van de Darwin Core Archive Validator van GBIF<sup>9</sup>.
- Bepalen van frequentie van updates van Xeno-canto naar de NBA en van de NBA naar GBIF. Idealiter is de periode tussen beide updates zo kort mogelijk en zoveel mogelijk geautomatiseerd.

#### 3.3 Versteviging Stichting

Gezien de toenemende samenwerking met professionele partners is het noodzakelijk dat de Stichting Xeno-canto een steviger fundament krijgt. In 2018 neemt zal het bestuur daartoe stappen ondernemen. Fondsenwerving lijkt daar een noodzakelijke stap voor te zijn, en als de aanvraag van een ANBI status die het aantrekkelijker maakt om donaties aan de stichting te doen krijgt eerst prioriteit.

# 4. Lange termijn, 2019 en verder

Een aantal toekomstplannen kunnen in 2019-2020 worden uitgewerkt tot nieuwe projectvoorstellen en worden ingediend bij goed aansluitende subsidiegevers/-regelingen. Op dit moment is niet duidelijk wat de meest kansrijke plannen zijn.

## 4.1 Repertoires in kaart brengen

Het is al lang een ambitie van Xeno-canto om pagina's aan te bieden waarin het volledige repertoire aan typen zang en roepjes per vogelsoort wordt samengevat. Een voorbeeldontwerp daarvan is al eens uitgewerkt<sup>10</sup>.

Op dit moment is er een aantal categorieën waarin de geluiden door de gebruikers kunnen worden ingedeeld: zang, roep, vluchtroep, bedelroep, man, vrouw etc. Dit is een grove indeling die veel detail onzichtbaar laat. Vaak kent een soort veel meer verschillende roepjes, ieder met een bepaald doel, en in de zang kan een enorm repertoire verscholen

<sup>&</sup>lt;sup>7</sup> http://tools.gbif.org/dwca-validator/extension.do?id=gbif:Multimedia

<sup>&</sup>lt;sup>8</sup> http://tools.gbif.org/dwca-validator/extension.do?id=http://rs.tdwg.org/ac/terms/Multimedia

<sup>&</sup>lt;sup>9</sup> http://tools.gbif.org/dwca-validator/home.do

<sup>&</sup>lt;sup>10</sup> Mock-up ontwerp van repertoirepagina: <a href="http://www.xeno-canto.org/static/tmp/repertoire-wtsp.png">http://www.xeno-canto.org/static/tmp/repertoire-wtsp.png</a>

zijn. Voor soorten met veel opnames en een groot repertoire (denk bijvoorbeeld aan koolmees *Parus major*) is het op dit moment lastig om snel te zien wat voor verschillende vocalisaties er in de verzameling aanwezig zijn.

Om dit soort pagina's samen te stellen voor alle soorten is een ambitieus project. Het is nooit eerder geprobeerd, en zou een unieke toevoeging aan de verzameling, en aan de kennis van vogelgeluiden wereldwijd, zijn.

## 4.2 Open access repository voor ruw materiaal

In een van onze toekomstbeelden is Xeno-canto ook een *open access repository* voor natuurgeluiden, een soort GenBank voor natuurgeluiden ("SoundBank"). Er is ruimte voor opslag van lange onbewerkte opnamen, zoals veldopnamen die de bron vormen van gepubliceerde biologische en ecologische studies (naar voorbeeld van Tobias *et al.* 2013<sup>11</sup>). Die opnamen komen beschikbaar in de hoogst mogelijke kwaliteit en worden daarbij voorzien van een catalogusnummer en DOI. Op die manier leggen we de link tussen publicaties en bronmateriaal. Dit wordt overigens al op grote schaal gedaan voor opnames van individuele soorten, in talloze publicaties.

## 4.3 Van vogelgeluiden naar natuurgeluiden

In de hiervoor beschreven repository voor natuurgeluiden is plek is voor voor zowel opnamen van soundscapes (mits daar een biologische geluidenbron op staat) als opnamen van specifieke natuurgeluiden, onafhankelijk van soortgroep. Vogelgeluiden staan namelijk niet op zich; vogels maken onderdeel uit van een biotoop, waarin biotische en abiotische factoren bij elkaar een "soundscape" vormen.

Op de achtergrond van vogelopnamen op Xeno-canto hoor je regelmatig ook andere vocaliserende soortgroepen, zeker in de tropen. In uitzonderlijke gevallen blijken vogelopnamen soms - na discussie tussen gebruikers - eigenlijk afkomstig te zijn van een amfibie<sup>12</sup> of zoogdier<sup>13</sup>. Die dienen ook als zodanig geclassificeerd te worden in de collectie. Het ligt voor de hand klein te beginnen en de collectie uit te breiden naar andere soortgroepen waarin geluid een belangrijke rol speelt. Naast vogels zijn dat met name zoogdieren, kikkers en sprinkhanen.

## 4.4 Koppeling met andere citizen science projecten

De aansluiting bij Naturalis zal het naar verwachting ook gemakkelijker maken om deel te nemen in *citizen science* projecten waar vogelgeluiden centraal staan. Op dit moment fungeert Xeno-canto op bescheiden schaal als platform voor zulke projecten. Een voorbeeld daarvan is het Canadese project "White-throat song", waarin wordt gekeken naar de variatie in zang van *Zonotrichia albicollis*<sup>14</sup>. Ook is Xeno-canto in gesprek met de Universiteit van

<sup>&</sup>lt;sup>11</sup> Tobias J.A, R. Planqué, D.L. Cram & N. Seddon (2013). Species interactions and the structure of complex communication networks. PNAS 111(3): 1020–1025. <a href="http://www.pnas.org/content/111/3/1020.abstract">http://www.pnas.org/content/111/3/1020.abstract</a> (bijbehorende opnames op Xeno-canto: <a href="http://www.xeno-canto.org/set/3">http://www.xeno-canto.org/set/3</a>).

<sup>12</sup> http://www.xeno-canto.org/forum/topic/1502?lang=eng

http://www.xeno-canto.org/forum/topic/1500

<sup>&</sup>lt;sup>14</sup> Otter K., S. Ramsay & S. LaZerte (2016). White-throat Song. A citizen science project looking at song variation in white-throated sparrows. University of Northern British Columbia. <a href="https://www.whitethroatsong.ca/">https://www.whitethroatsong.ca/</a>

Leiden over ondersteuning van een grootschalig project voor het verzamelen en analyseren van opnamen van de zang van vrouwtjesvogels (vervolgstudie op o.a. Odom *et al.* 2014<sup>15</sup>).

## 4.5 Geautomatiseerde geluidsherkenning

Geautomatiseerde herkenning van vogelgeluiden m.b.v. algoritmen is om een aantal redenen een onderwerp dat in de belangstelling staat. Een van de drijfveren achter het onderzoek is dat het eenvoudig en goedkoop is om een biotoop met microfoons te bemonsteren. Op allerlei plekken in de wereld worden nu op grote schaal opnames van "soundscapes" gemaakt met als doel zoveel mogelijk te leren over de avifauna van bepaalde plekken. De analyse van dat soort opnames is een enorm karwei, waarbij algoritmes die gelijksoortige geluiden kunnen herkennen noodzakelijk zijn. Sinds 2014 neemt Xeno-canto voornamelijk als leverancier van data deel aan de jaarlijkse LifeClef "challenge" waarin onderzoeksgroepen worden uitgedaagd om vogelgeluiden te herkennen. De schaal van deze challenge is enorm: geluiden van 1.000 soorten moeten herkend worden, een aantal dat zonder het bestaan van Xeno-canto ondenkbaar zou zijn. Deze challenge heeft eerder al eens geleid tot een smartphone app voor het geautomatiseerd herkennen van Britse vogelgeluiden<sup>17</sup>.

Xeno-canto heeft zelf ook een belang bij deze algoritmes, omdat in feite maar slecht bekend is wat er eigenlijk aanwezig is in de collectie. Veel opnames bevatten niet geannoteerde achtergrondgeluiden en repertoire-elementen, waarvan naar verwachting vele "nieuw" zijn voor de verzameling wanneer ze worden geïdentificeerd. De algoritmes zouden dus geintegreerd kunnen worden om de geluiden in de collectie te classificeren. Daarnaast is het natuurlijk mogelijk om geluiden van vogelaars automatisch op naam te gaan brengen.

# 4.6 Uitbreiding mogelijkheden collectie-annotatie door gebruikers

Op dit moment kunnen gebruikers al op verschillende manieren een zinvolle bijdrage leveren aan de geluidsverzameling op Xeno-canto. Niet alleen door geluiden toe te voegen, maar bijvoorbeeld ook door een discussie over een geluidsopname te openen, door raadselgeluiden te determineren, de kwaliteit van een geluid aan te geven, door geluiden toe te voegen aan een openbare deelverzameling ("set") of door een artikel te schrijven/publiceren.

In het verleden is een aantal andere mogelijkheden om bijdragen te leveren de revue gepasseerd die, meestal vanwege benodigde ingrijpende wijzigingen aan de databasestructuur of vanwege prestatieproblemen (verlaagde sitesnelheid), niet zijn uitgevoerd of geïmplementeerd. Een paar ervan staan nog steeds hoog op de verlanglijst:

1. Toevoegen van achtergrondsoorten. Xeno-canto moedigt het toevoegen van opnames met daarop meer dan één soort geluid actief aan (zie ook eerdere punt

<sup>&</sup>lt;sup>15</sup> Odom K.J., M.L. Hall, K. Riebel, K.E. Omland & N.E. Langmore (2014). Female song is widespread and ancestral in songbirds. Nature Communications 5: 3379. <a href="http://www.nature.com/articles/ncomms4379">http://www.nature.com/articles/ncomms4379</a>
<a href="http://www.imageclef.org/lifeclef/2016">http://www.imageclef.org/lifeclef/2016</a>

<sup>&</sup>lt;sup>17</sup> Warblr: <a href="https://warblr.net/">https://warblr.net/</a>. Zie ook de publicatiepagina van onderzoeker Dan Stowell: <a href="http://www.mcld.co.uk/research/">http://www.mcld.co.uk/research/</a>

over "soundscapes"). De site geeft de mogelijkheid om de voorkomende achtergrondsoorten op te geven, maar op dit moment is het niet mogelijk om als willekeurige gebruiker direct een achtergrondsoort toe te voegen bij een opname waar de opnemer dat heeft nagelaten. Er zijn op dit moment per opname gemiddeld ongeveer twee soorten aangegeven, wat betekent dat er in feite 2 x 300.000 = 600.000 geluiden met identificatie in de verzameling staan. Het is niet duidelijk hoeveel achtergrondgeluiden er zijn, maar zeker een veelvoud van wat er nu is aangegeven.

2. Annoteren van specifieke geluiden in opnames. Het aangeven van achtergrondgeluiden is een goede zaak, maar het zou nog beter zijn als individuele geluiden op een opname aangegeven en geannoteerd zouden kunnen worden, met soort, type, etc., in feite alle velden die voor voorgrondgeluiden ook mogelijk zijn. Dit vereist een fundamentele verandering in de database, en de ontwikkeling van gereedschappen die het mogelijk maken om specifieke delen van geluiden te selecteren.

## 4.7 Verspreidingskaarten

Een van de aardigste neveneffecten van Xeno-canto is het feit dat de verspreidingskaartjes op de soortpagina's zeer vaak worden bekeken en dat fouten daarin veelvuldig worden gemeld. Een professionele kaartenmaker onder de gebruikers maakt van die toestand gebruik om de kaarten continu te verbeteren. Volgens velen vormt Xeno-canto de best bijgehouden verzameling vogelverspreidingskaarten op dit moment.

De verspreidingskaarten zouden veel beter gebruikt kunnen worden op de site bijvoorbeeld door tijdens het toevoegen van geluiden te controleren of een bepaalde waarneming binnen te verwachten geografische grenzen valt. Verder heeft het gezien de licentie de voorkeur om niet de .kmz files aan te bieden, maar een alternatieve techniek te gebruiken. Hier geldt ook weer dat een fundamentele aanpassing van de database nodig is. Tegelijkertijd zouden de locatie-invoer en de bron van de onderliggende kaarten (nu alleen Google Maps) herzien en verbeterd kunnen worden.

#### 4.8 Verbeteren van de API voor verbinden externe collecties

Behalve Xeno-canto is er een heel aantal andere sites die toevoegen van geluiden, of meer in het algemeen multimedia, aan waarnemingen van vogels mogelijk maken. In Nederland is Waarneming.nl/Observado.org het meest in het oog springende voorbeeld. Ondanks de mogelijkheden die worden geboden door het web (en door de licenties) wordt er maar in beperkte mate data uitgewisseld of hergebruikt tussen de verschillende sites. Verder zijn er bestaande archieven met een beperkte aanwezigheid op het web die manieren zoeken om hun verzameling onder de aandacht te brengen. In een enkel geval (eBird/Macaulay Library) is het tot een formele en innige samenwerking gekomen tussen een waarnemingensite (eBird) en een al bestaand multimedia-archief (Macaulay Library).

In het verleden is er contact geweest tussen een heel aantal partijen<sup>18</sup> om tot een betere uitwisseling van data te komen. De route via de Netherlands Biodiversity API naar GBIF ligt voor de hand, maar daarnaast kan een verbetering van de XC API hetzelfde doel dienen. XC heeft op dit moment een eenvoudige API die het mogelijk maakt om de verzameling door middel van een programma te benaderen, en te exporteren in JSON. Hiervan wordt bijvoorbeeld gebruik gemaakt in het R-pakket WarbleR<sup>19</sup> die kwantitatieve analyse van sets geluiden mogelijk maakt. Ook hier liggen kansen voor samenwerking met Naturalis.

## 4.9 Collectievorming en -digitalisering

Het gros van de collectie van Xeno-canto is *born digital* en zelf geüpload door gebruikers. Op verzoek biedt Xeno-canto de mogelijkheid voor bulk-imports aan aan gebruikers die in één klap een grote collectie willen delen. Bij hoge uitzondering helpt Xeno-canto gebruikers bij het digitaliseren van hun analoge verzamelingen. In het verleden is dat gedaan voor de unieke Aziatische geluidencollectie van ornitholoog Bas van Balen<sup>20</sup>. Mogelijk dienen zich in de toekomst nieuwe mogelijkheden aan.

<sup>&</sup>lt;sup>18</sup> Onder andere Karl-Heinz Frommolt (Museum für Natürkunde, Berlijn), Cheryl Tipp en Richard Ranft (British Library), Macaulay Library, Greg Budney, Felipe Toledo, Paulo Marques, Rafael Marquez, Borror Lab, Gianni Pavan en Klaus Riede.

<sup>&</sup>lt;sup>19</sup> warbleR: <a href="http://rpubs.com/marcelo-araya-salas/110155">http://rpubs.com/marcelo-araya-salas/110155</a>. Zie ook: Araya-Salas, M. & G. Smith-Vidaurre (2016). warbleR: an r package to streamline analysis of animal acoustic signals. Methods in Ecology & Evolution. <a href="http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12624/full">http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12624/full</a>

<sup>&</sup>lt;sup>20</sup> http://www.vantienhovenfoundation.com/P86-Digitisation-of-Indonesian-bird-song-recordings.html