



## COMMUNICATIES

### **Twee vensters op de zelfde waarheid: over een verantwoord gebruik van TWN.**

#### **Roelf Pot**

Al jaren is duidelijk dat er veel misgaat met het beheer, en vooral de uitwisseling, van data over biologische waarnemingen. De belangrijkste problemen zijn de toegankelijkheid van opgeslagen data, de vaak onduidelijke betekenis van het getal waarmee de hoeveelheid van de waargenomen soorten wordt aangeduid, en bovenal: de naamgeving van de soorten. TWN (Taxa Waterbeheer Nederland) is de recentste poging om orde te scheppen in de naamgeving in de hydrobiologie en iedereen wordt nu geacht alleen de namen te gebruiken van deze lijst. Op zichzelf een zeer waardevolle stap vooruit, maar de problemen worden er niet volledig mee opgelost en in combinatie met de huidige software soms zelfs verergerd. Ik bepleit in beperkte kring al jaren een goed werkende oplossing hiervoor, maar krijg de handen tot nu toe niet op elkaar. Ik heb daarom besloten zelf de oplossing uit te werken, maar vraag daarbij de hulp van de leden van de WEW. Op mijn website ga ik in op dit initiatief, met de voorlopige werknaam Thys<sup>3</sup> en de mogelijkheden tot deelname: [www.roelfpot.nl/thys3](http://www.roelfpot.nl/thys3). In dit stukje beschrijf ik de essentie van het probleem en de oplossing.

#### **Het venster van de taxonoom**

Een taxonoom refereert zijn taxonnamen aan de auteurs die de taxa hebben beschreven, door die auteursnamen achter de taxonnamen te zetten met het jaar van publicatie. De beschrijvingen staan in honderdduizenden verschillende publicaties en alleen specialisten binnen bepaalde groepen hebben het overzicht en een oordeel over welke namen en beschrijvingen correct zijn en welke niet.

Er is een wildgroei op internet van overzichtlijsten ontstaan waarin al deze namen staan opgesomd. Deze lijsten bevatten ook verouderde namen, maar die zijn dan verbonden aan een ander, geldig synoniem. De lijsten zijn voortdurend in beweging omdat er steeds meer namen bijkomen en de geldigheid van namen wordt herzien.

#### **Het venster van de waarnemer**

Een waarnemer refereert zijn taxonnamen aan een identificatiewerk. Dat kan een veldgids zijn, een sleutel voor een heel beperkte groep, of een encyclopedisch naslagwerk. Het zijn altijd sleutels tot het herkennen van individuen volgens de door de taxonomen gegeven beschrijvingen, maar gemodificeerd door de auteurs van deze identificatiewerken omdat ze de onderscheidende kenmerken van de soorten op de eerste plaats stellen.

Sommige individuen worden met de éne sleutel op een andere naam gebracht dan met een andere sleutel, niet alleen omdat er fouten of interpretatieverschillen in de sleutels kunnen zitten, maar vooral omdat de sleutels nooit compleet zijn. Ze bevatten altijd slechts soorten uit een bepaalde regio, zelfs als ze compleet voor die regio heten te zijn: er worden voortdurend nieuwe taxa ontdekt en beschreven. Alle namen zijn in combinatie met het identificatiewerk echter wel altijd herleidbaar.

#### **Dataverwerking**

Alle databases voor het opslaan van waarnemingen gaan uit van standaardlijsten van namen. Dat begon met lijsten van namen uit identificatiewerken en de registratie van waarnemingen was daarmee eenduidig en correct. De uitwisseling tussen verschillende systemen wordt echter een probleem als niet iedereen dezelfde identificatiewerken gebruikt. Hetzelfde probleem doet zich voor bij vergelijking van waarnemingsdata met indicatorlijsten zoals Ebeo-beoordeling en KRW-beoordeling die niet per sé de zelfde naamgeving volgen. Vergelijking met buitenlandse data is vrijwel altijd problematisch.

Er is een ontwikkeling geweest naar standaardlijsten die de namen van zo veel mogelijk identificatiewerken bevatten. Waarnemingen kunnen dan, ongeacht het gebruikte identificatiewerk worden opgeslagen onder de namen op zo'n standaardlijst. Maar dat gaat

ook fout, want daarbij wordt een stabiele en eenduidige relatie verondersteld tussen de identificatiewerken en de standaardlijst en die is er niet.

Met TWN is een poging gedaan de probleem die zich daarbij voordoen op te lossen door de meest gebruikte synoniemen en interpretaties ook op te nemen, maar ook dat blijkt toch op allerlei manieren fout te gaan, zeker na verloop van tijd. Als bijvoorbeeld een taxonnaam wordt opgesplitst, dan krijgt de oorspronkelijke naam twee betekenissen; de reeds ingevoerde data krijgen daarmee dus een dubieuze betekenis, tenzij ze meteen worden gecontroleerd en aangepast, het liefst door de oorspronkelijke waarnemers. Elke verandering geeft daarmee een risico op informatieverlies. Bij conversie van hele datasets blijkt dat helemaal vaak fout te gaan en vergissingen zijn in de praktijk zo goed als onherstelbaar.

### **Oplossing**

De oplossing is het loskoppelen van de taxalijsten voor waarnemingen van de standaardlijsten voor uitwisseling van gegevens. Waarnemingen moeten worden opgeslagen in relatie tot de gebruikte identificatiewerken en uitwisseling moet plaatsvinden op basis van standaardlijsten volgens de actuele (dynamische) taxonomische inzichten. De energie moet dan worden gestoken in het bijhouden van de koppeling tussen de identificatiewerken en de standaardlijsten omdat de eerste vast staan en de tweede veranderen. Deze oplossing maakt het mogelijk dat data altijd in de oorspronkelijke vorm worden ingevoerd en bewaard en er dus geen fouten meer insluipen. Bovendien is kwaliteitsborging dan mogelijk. Conversie vindt pas plaats vóór elk gebruik, altijd volgens de meest recente inzichten en naar een standaard naar keuze. In de hydrobiologie in Nederland is die standaard momenteel TWN, maar voor toepassing van Ebeo-beoordeling is dat de oudere TCN en voor intercalibratie van de KRW-maatlatten is dat het Europese Rebecca.

### **Software**

Om dit mogelijk te maken is een nieuwe generatie software nodig; software die in staat is dynamisch met meerdere soortenlijsten te werken. Het principe daarvan heb ik al in 1996 ontwikkeld (OpName2) en deze database vormt de basis van QBWat. QBWat koppelt de namen van verschillende taxalijsten aan de namen van de KRW-maatlatten op het moment dat dat gevraagd wordt. Als er dan een interpretatiewijzing plaatsvindt dan heeft dat geen gevolgen voor de oorspronkelijke data.

Omdat de databasesoftware nog onder DOS draait en de 64-bits versies van Windows Vista en Windows 7 geen DOS meer ondersteunen, ga ik dat databasesysteem updaten. In één moeite door kan ik er ook een zelfstandig werkend databasepakket van maken dat de oplossing biedt die ik hierboven beschrijf. De bedoeling is dat ik daarmee aantoon dat het kan, waarna hopelijk de echte softwarespecialisten het overnemen.

Het opbouwen van taxalijsten van alle gebruikte identificatiewerken en het leggen van de verbindingen tussen synoniemen en TWN is iets dat ik niet alleen kan. Veel specialisten hebben die lijsten echter al wel eens gemaakt en ik wil daar graag gebruik van maken. Bij de regiegroep van TWN is veel kennis verzameld over interpretaties en ook daar wil ik graag gebruik van maken.

Voor nadere informatie, ook over de onderbouwing van diverse beweringen in dit stuk, en vooral ruimte voor discussie en ontwikkeling: [www.roelfpot.nl/thys3](http://www.roelfpot.nl/thys3)

Met dank aan Herman van Dam en Ronald Bijkerk voor de vruchtbare discussies over dit onderwerp, een gezamenlijke publicatie is in de maak; verder met dank aan Frans Kouwets voor de discussies over TWN.