

Opinie

Koppeling op wetenschappelijke naam of soortnummer?

Lodewijk van Duuren

1. Inleiding

Bij uitwisseling van bestanden met gegevens over planten en diersoorten is het noodzakelijk de soorten en andere taxonomische niveaus op de juiste wijze met elkaar te kunnen verbinden.

Om gegevens over soorten uit verschillende digitale bestanden aan elkaar te knopen is het noodzakelijk om een sleutelveld te hebben waarmee de soorten in beide bestanden op de juiste wijze gekoppeld kunnen worden. Het meest voor de hand liggend is een koppeling op wetenschappelijke naam. Een dergelijke koppeling kan echter problemen opleveren die voor een deel handmatig opgelost moeten worden. Een deel van de problemen kan opgelost worden met geavanceerde programmatuur. Toch ziet men op Internet op websites die namen uit meer bronnen samenvoegen, dat nog veel van de niet wezenlijke verschillen in de gepresenteerde gegevens blijven staan, zoals in de websites van ITIS en Ubio.

In Nederland hebben echter vrijwel alle soorten of andere taxa een uniek soortnummer waarmee ook een koppeling tussen verschillende bestanden op taxonniveau mogelijk is. In dit artikel worden de voor- en nadeel van beide typen koppelingen besproken. Twee voorbeelden worden in dit artikel gegeven: één tussen de Avifauna van Nederland (Van den Berg & Bosman 2001) en de vogellijst van de website Nederlands Soortenregister (www.nederlandsesoorten.nl), de tweede tussen de Flora van Nederland (Van der Meijden 1996) en de Flora Europaea (Tutin et al. 1964-1980). In de meeste gevallen zijn de problemen met koppelen voor zoölogische en botanische namenlijst vergelijkbaar.

2. Wetenschappelijke naam

De koppeling van bestanden op wetenschappelijke naam kan behoorlijk tijdrovend zijn als koppeling niet in alle gevallen mogelijk is en een aantal stappen doorlopen moeten worden om de probleemgevallen geautomatiseerd dan wel handmatig op te lossen.

De allereerste stap is het wegwerken van technische verschillen tussen bestanden.

2.1 Technische verschillen bestanden

De allereerste stap is het wegwerken van technische verschillen tussen bestanden.

Die kunnen bijvoorbeeld bestaan uit:

- Database-, spreadsheet- of tekstbestand naar hetzelfde databasesysteem overbrengen.
- Auteursnaam die in één veld met wetenschappelijke naam zijn opgenomen in aparte kolommen plaatsen.
- Genusnaam, soortnaam en ondersoortnaam die in verschillende kolommen zijn geplaatst in één kolom plaatsen.
- Namen die niet voluit geschreven zijn (bijvoorbeeld het niet herhalen van de genusnaam bij soorten) compleet maken.

- Code waarmee diakritische tekens zijn opgeslagen gelijkmaken.

2.2 Schrijfwijze wetenschappelijke namen

Ook in de schrijfwijze van wetenschappelijke namen komen ondanks de nomenclatuurregels van de ICZN en ICBN nog verschillen voor. Deze moeten eerst gelijk getrokken worden om een goede koppeling mogelijk te maken.

Mogelijke verschillen in bestanden zijn:

- Het aantal spaties tussen de onderdelen van de naam verschild.
- De ondersoorten zijn aangeduid met ssp. of met subsp.
- Aanduidingen van lagere taxonomische eenheden zoals subsp. of var. zijn geheel weggelaten.
- De cultivars zijn met of zonder aanhalingstekens en/of het voorvoegsel cv aangegeven.
- De vormen zijn aangeduid met forma of fo.
- De benaming van de bastaarden is niet gelijk (bijvoorbeeld met of zonder de namen van de 'oudersoorten').

De verschillen tussen de bestanden waren op dit punt niet groot.

Bij de vogelnamen speelden deze problemen in het geheel niet. De laatste drie van bovenstaande punten komen overigens vrijwel alleen in wetenschappelijke namen van planten voor. Bij de plantennamen bestond alleen het verschil in schrijfwijze van de cultivars: in de Flora van Nederland met aanhalingstekens geschreven en in de Flora Europaea zonder (*Populus nigra* cv. '*italica*' tegenover *Populus nigra* cv. *italica*).

2.3 Fouten in bestanden

Er moet rekening mee worden gehouden dat in de digitale bestanden echte fouten in de namen voorkomen, die bijvoorbeeld zijn ontstaan bij het invoeren van namen in bestanden. Bij de vergelijking van de Flora van Nederland en de Flora Europaea kwam in beide bestanden dergelijke fouten naar voren (9 wetenschappelijke namen met fouten en 4 auteursnamen met fouten). Dergelijke fouten in beide bestanden zijn steeds verbeterd en zijn verder niet in de kwantitatieve analyse (zie tabel 1 en 2) betrokken.

2.4 Synoniemen

De meeste verschillen tussen bestanden komen voort uit verschillende of voortschrijdende inzichten in de taxonomie van soorten, waardoor genera samengevoegd, dan wel gesplitst worden, niet geldige namen vervangen zijn door geldige namen etc. Voor een koppeling op wetenschappelijke naam is het noodzakelijk dat ook synoniemen in bestanden beschikbaar zijn en de relatie met de geldige namen gelegd kan worden.

2.5 Kleine verschillen in schrijfwijze

Sommige soorten in beide bestanden hebben kleine verschillen in de schrijfwijze die doorgaands niet als synoniemen worden beschouwd. Soms zijn het verschillen in de uitgang van de naam (*Melilotus albus* versus *Melilotus alba*), soms zijn het verschillen van één letter in spelling van de naam (*Asplenium foreziense* versus *Asplenium foresiense* of *Robinia pseudo-*

Tabel 1. Koppeling namen Flora van Nederland aan Flora Europaea

Typen koppeling	Aantal taxa	%
gevonden	2357	86.5
gevonden via synoniemen Flora Europaea	115	4.2
gevonden via synoniemen Flora van Nederland	71	2.6
gevonden via synoniemen FE en FN	9	0.3
gevonden via niet exacte koppeling en handmatige vergelijking	30	1.1
niet gevonden	144	5.3
Totaal	2726	100.0

acacia en *R. pseudacacia*). Dergelijke verschillen komen alleen aan het licht bij een handmatige vergelijking eventueel na een niet-exacte koppeling

Voorbeelden van kleine verschillen in schrijfwijzen die automatische koppeling onmogelijk maken zijn:

- ae in een naam i.p.v. e.
- ii op het eind van soorttoevoeging (eponiem) i.p.v. van i.
- verschillende uitgang -us, -um of -a afhankelijk van beoordeling van het geslacht van de wetenschappelijke genusnaam.
- gebruik van streepjes of niet.
- In oudere wetenschappelijke namen kunnen ook gebruik van diakritische tekens en hoofdletters verschillen opleveren.

2.6 Koppeling flora wetenschappelijke naam

Zie tabel 1 en 2. Bij de koppeling tussen Flora Europaea en de Flora van Nederland blijken de meeste wetenschappelijke direct, dan wel via de synoniemen gekoppeld te kunnen worden. Een zeer klein gedeelte (1.1%) bleek na handmatige controle en verbetering/verandering nog te koppelen te zijn. Ten slotte bleven nog 144 taxa over. Een groot deel daarvan bleek lager van rang dan soort of had betrekking op bastaarden. Tot slot moet op basis van deskundige expertise van deze overgebleven taxa vastgesteld worden of deze toch onder een andere naam in een van beide bestanden is opgenomen.

2.7 Koppeling fauna wetenschappelijke naam

Bij de koppeling van de vogellijst in Nederlandse soorten van Naturalis met de vogelnamenlijst in BioBase, die gebaseerd is op de Avifauna van Nederland, bleken twintig soorten niet op naam te koppelen te zijn

- Zes soorten die niet koppelen staan in de lijst van Naturalis onder Ongeldige naam of Andere combinatie, zodat ze wel indirect te koppelen zijn.
- Twee soorten van Naturalis staan niet in BioBase, het betreft echter incidentele import die niet in BioBase is opgenomen.
- Er blijven twaalf taxa over die niet op naam te koppelen zijn. Oplossing: deze als ongeldige naam of andere combi-

Tabel 2. Taxa uit de Flora van Nederland, die niet in Flora Europaea voorkomen

Niveau	Aantal
soort	1
ondersoort	22
variëteit	11
forma	1
microsoort	7
bastaard	50
uitheemse soort	39
cultivar	5
genus	8
Totaal	144

natie in de lijst van Naturalis opnemen.

- Overigens is er een groot aantal (onder)soorten die wel in de Avifauna van Nederland staan maar niet in Naturalis (hiervan heb ik geen overzicht gemaakt, want deze zijn voor de koppeling van geen belang).

2.8 Auteursnaam

In sommige gevallen betekent een gelijkkluidende wetenschappelijke naam toch een ander concept, afhankelijk van de auteur die het taxon heeft beschreven. Als de wetenschappelijke naam als code wordt gebruikt is het daarom noodzakelijk ook de auteursnaam op te nemen en ook daarop te koppelen.

Er is echter vrijwel geen standaardisatie in de gebruikte schrijfwijze van de auteursnamen. In praktijk betekent dat dezelfde auteursnamen op verschillende wijze zijn weergegeven: een voorbeeld is de auteursnaam van de libel *Ceriagrion tenellum* die op 16 verschillende wijzen geschreven is (tabel 3); en daarmee zijn niet eens alle mogelijkheden van spelling uitgeput. Overigens is het verschil tussen wel en niet gebruik van haakjes terug te voeren op slordigheden/fouten.

Voorbeelden van verschillen in auteursnamen zijn:

- naam voluit of afgekort: L. of Linnaeus.
- verschil in wijze van afkorten: Good. of Gooden..
- alle auteurs (drie of meer) of gebruik van et al of & al. (= et aliorum).
- bij twee of meer auteurs gebruik van &, et of and.
- naam met of zonder initialen: Gray of J.E. Gray, of Gray, J.E..
- naam met diakritische tekens of in plaats daarvan een extra e: Müller of Mueller.
- naam met of zonder diakritische tekens: Gùldenstàdt en Gùldenstàdt.
- verschil in schrijfwijze namen: Linnaeus, Linné; Medicus Medikus.
- verschil in schrijfwijze door verschil in omzetting van Russische, Chinese, Japanse en andere schriftsoorten naar het Latijnse alfabet.
- f. of fil. (= zoon van).
- namen los of aan elkaar: Du Roi of Duroi, Le Sueur of Lesueur.

Tabel 3. *Ceriagrion tenellum*, auteursnamen op Internet

(Villers)	(Villers, 1789)
(de Villers)	(Villers,1789)
de Villers	(VILLERS, 1789)
(De Villers)	De Villers, 1789
(DE VILLERS)	de Villers, 1789
(Villers 1789)	(de Villers, 1789)
de Villers 1789	(De Villers, 1789)
De Villers 1789	(DE VILLERS, 1789)

Tabel 4. Verschillen in auteursnamen planten.

Categorie	Aantal	Voorbeelden	
		Nederland	Europa
verschil in afkorten of schrijfwijze	369	Mill. Du Roi	Miller Duroi
verschil in gebruik 'ex'	35	Roth	Schrader ex Roth
echte verschillen	54	Alnus Gaertn.	Alnus Miller

Tabel 5. Verschillen in auteursnamen dieren.

Type verschil	Aantal	Voorbeelden	
		BioBase	Naturalis
wel/geen haakjes	11	(Latham, 1790)	Latham, 1790
jaartal verschillend	11	(Linnaeus, 1766)	(Linnaeus, 1758)
gebruik spaties	9	C.L.Brehm	C.L. Brehm
oe ipv ö	1	(Holböll, 1843)	(Holboell, 1843)
met/zonder initialen	4	(Gmelin, 1771)	(S.G Gmelin, 1771)
andere auteursnaam	2	Fischer, 1843	Nordmann, 1842
gebruik streepjes	1	(Guérin-Ménéville, 1843)	(Guérin Ménéville, 1843)
wel/niet diakritische tekens	2	(Güldenstadt, 1775)	(Güldenstädt, 1775)
wijze van afkorten	1	Conrad	Conrad von Balenstein

- hoofdletters of kleine letters: de Winston of De Winston; von Haast of Von Haast.
- met en zonder haakjes.
- met of zonder jaartal.
- met of zonder komma tussen naam en jaartal.
- met of zonder de nieuwe auteur (geplaatst achter de naam tussen haakjes).
- gebruik van spaties: J.E.Gray. J.E. Gray of J. E. Gray.
- gebruik van ex, non etc.: Roth of Schrader ex Roth.
- gebruik van in: Gronow in Gray, 1845 of Gronow, 1845.
- onbedoelde fouten in auteursnamen of geheel verkeerde auteursnamen.

Van de soorten die uiteindelijk gevonden zijn in de index van de Flora Europaea is vervolgens gekeken of ook de auteursnamen overeenstemmen. Dit is alleen gedaan voor namen die als geldige naam in de index van de Flora Europaea voorkomen (2357 taxa). Hier treden zoals verwacht gelijksoortige problemen op als bij de koppeling op wetenschappelijke naam (tabel 4). Ook in de auteursnamen van beide bestanden zaten fouten (4 taxa), deze zijn verbeterd.

Een goede vergelijking is pas mogelijk is, als de naamgeving van auteurs in gekoppelde bestanden aan dezelfde standaard voldoet. Voor de auteursnamen van planten bestaat er zo'n standaard, de index van auteursnamen van Kew (Brummit et al. 1992). De auteursnamen in het bestand van de Flora van Nederland zijn aan deze standaard aangepast. Echter niet voor de Flora Europaea. Dit zal dus eerst moeten geschieden om handmatige vergelijking zoveel mogelijk te voorkomen. Zoals uit tabel 2 blijkt is het grootste aantal soorten de naam gelijk of verschilt alleen de schrijfwijze van de auteursnaam. Toch blijven er nog 54 taxa (2.3%) over waarvan de auteursnaam wezenlijk verschilt. Een probleem bij de koppeling vormden ook de wetenschappelijke namen waar in de index van Flora Europaea 'group' staat vermeld. Aan de naam is namelijk niet direct af te lezen op welke groep van taxa het betreffende taxon betrekking heeft. Ook hier is een deskundig oordeel nodig om vast te stellen of de verschillen in auteursnamen op vergissingen berusten, of dat het echt om verschillende taxa gaat.

Hoewel de auteursnamen bij vogelsoorten zelden gebruikt worden, zijn de auteursnamen in de lijst van Naturalis toch vergeleken met die uit BioBase. De auteursnamen in BioBase zijn uit verschillende bronnen afkomstig. Op de 490 geldige namen zijn er 42 verschillen geconstateerd. De verdeling over de verschillende typen verschillen staat in tabel 5.

De verschillen zijn vergeleken met de naamgeving in Fauna Europaea en geconstateerd is dat de vogellijst van Naturalis daar geheel mee overeen komt. Daarom zijn vervolgens in BioBase de verbeteringen aangebracht.

2.9 Initialen auteursnamen

De namen John Edward Gray – George Robert Gray en Johann Friedrich Gmelin – Samuel Gottlieb Gmelin zijn voorbeelden van auteurs van wetenschappelijke namen met dezelfde achternaam. Bij vermelding van deze auteurs kan er verwarring ontstaan als geen initialen in de auteursnaam worden gebruikt. Toch ziet men heel vaak alleen Gray en Gmelin in plaats van J.E. Gray, G.R. Gray of J.F. Gmelin, S.G. Gmelin. Meestal laat men de initialen weg bij de meest gebruikte auteursnaam.

Gmelin staat zo meestal voor J.F. Gmelin, in de praktijk komt het toch regelmatig voor dat de minder bekend S.G. Gmelin ook tot Gmelin wordt afgekort. Eigenlijk verdient het aanbeveling in alle gevallen waarin een twijfel bestaat over de auteur, de initialen toe te voegen.

3. Soortnummer

In veel bestanden met gegevens van planten en dieren heeft iedere soort of taxon een uniek soortnummer, bijvoorbeeld in BioBase 2003. Dit soortnummer kan ook gebruikt worden om verschillende bestanden te koppelen.

Een dergelijk soortnummer moet gekoppeld zijn aan het concept van de soort of andere taxonomische eenheid. Een nummer niet aan een inhoud maar aan een naam koppelen heeft weinig zin, omdat dan net zo goed de naam als code gebruikt kan worden.

Het nummer is in principe een volgorde-nummer zonder betekenis, dat niet gewijzigd wordt. Uit het verleden zijn wel nummers overgenomen waarin geheel of gedeeltelijk een systematiek zat. Maar deze systematiek wordt al snel betekenisloos als het nummer ongewijzigd blijft. In principe worden nieuwe nummers steeds na het laatste hoogst uitgegeven nummer gekozen. Vervallen nummers worden alleen opnieuw toegekend als het vervallen concept weer wordt gebruikt.

Een dergelijk nummer kan alleen functioneren als er een autoriteit die garant staat voor de correcte wijze van nummering. Dit kan door afzonderlijke instanties per plant- of diergroep gebeuren, maar een organisatie die het beheer heeft over alle nummers is het meest geschikt.

4. Conclusies

Koppelen van bestanden op wetenschappelijke naam en auteur kan heel wat werk en problemen opleveren. Uit ervaring blijkt dat bij bestanden die op wetenschappelijke naam worden gekoppeld probleemgevallen overblijven die handmatig moeten worden opgelost. In sommige gevallen kan als hulpmiddel de koppeling op de Nederlandse naam helpen de juiste wetenschappelijke naam te vinden.

Bij meer geavanceerde manieren van koppelen kunnen bepaalde verschillen in schrijfwijze, zoals gebruik hoofdletters of kleine letters of gebruik van komma voor het jaartal nog wel overbruggen worden.

Als er echter een algemeen geaccepteerd nummer voor elk taxon aanwezig is kan die koppeling veel gemakkelijker tot stand gebracht worden. Aan beide mogelijkheden zitten echter voor- en nadelen (tabel 6).

Zolang er geen bindende afspraken zijn over de opbouw van de bestanden, de schrijfwijze van wetenschappelijke namen en auteursnamen en er geen bestanden zijn met de synoniemen van circa de afgelopen 50 jaar is een koppeling op soortnum-

Tabel 6. Voor- en nadelen gebruik nummers en wetenschappelijke namen bij koppeling bestanden.

Nummers	Wetenschappelijke namen
<i>Voordeel:</i> kort	<i>Nadeel:</i> lang
<i>Nadeel:</i> inhoud niet herkenbaar	<i>Voordeel:</i> inhoud goed herkenbaar
<i>Nadeel:</i> systematiek apart opslaan	<i>Voordeel:</i> deel systematiek in de naam
<i>Voordeel:</i> niet aan verandering onderhevig	<i>Nadeel:</i> aan verandering onderhevig
<i>Voordeel:</i> volledige koppeling met andere bestanden met nummers altijd mogelijk	<i>Nadeel:</i> volledige koppeling met andere bestanden alleen mogelijk als dezelfde schrijfwijze van namen wordt toegepast
<i>Nadeel:</i> fout in nummering leidt bijna altijd tot een verkeerd concept en daardoor slecht traceerbaar	<i>Voordeel:</i> fout in naam leidt meestal tot geen enkel concept en daardoor gemakkelijk traceerbaar
<i>Nadeel:</i> nummers alleen in Nederland in gebruik	<i>Voordeel:</i> wetenschappelijke namen overal in gebruik
<i>Nadeel:</i> lijst vervallen nummers bijhouden	<i>Nadeel:</i> lijst van alle synoniemen en ongeldige namen bijhouden

mer te verkiezen boven een koppeling op wetenschappelijke naam.

Het opstellen van een algemeen geaccepteerde nummering voor alle taxa op wereldschaal lijkt echter niet realistisch. Voor Nederland zijn echter voor veel dieren- en plantengroepen algemeen geaccepteerde nummeringen beschikbaar, zodat voor deze groepen een koppeling op wetenschappelijke naam achterwege kan blijven. Hoewel daar ook bezwaren aan kleven (zie tabel 6), is een toevoeging van deze taxonnummers aan soorten in de naamlijst op de website www.nederlandsesoorten.nl aan te bevelen. Tenslotte is op dit moment deze website de standaard voor Nederland als het gaat om de naamgeving van soorten.

5. Referenties

- Berg, A.B. van der & C.A.W. Bosman 2001. Zeldzame vogels van Nederland - met vermelding van alle soorten. Avifauna van Nederland 1. Tweede druk. – GMB Uitgeverij, Haarlem.
- Brummit, R.K. & C.E. Powell, 1992. Authors of plant names. A list of authors of scientific names of plants with recom-

- mended forms of their names, including abbreviations. – Royal Botanical Gardens, Kew.
- CBS & RIVM 2003. BioBase 2003. – In: Natuurcompendium 2003, Centraal Bureau voor de Statistiek en Natuur- en Milieuplanbureau.
- ICBN 2000. International Code of Botanical Nomenclature. – International Association for Plant Taxonomy, Vienna.
- ICZN 1999. International Code of Zoological Nomenclature. – International Commission on Zoological Nomenclature, London.
- Meijden, R. van der 1996. Heukels' Flora van Nederland. 22e druk. – Wolters-Noordhoff, Groningen.
- Tutin, T.G., V.H. Heywood, N.A. Burges et al. 1964, 1968, 1972, 1976, 1980. Flora Europaea Volume 1-5 and consolidated index Flora Europaea. – Cambridge.

Lodewijk van Duuren
21 november 2007

Publicaties

Door Roy Kleukers (RK) en Menno Reemer (MR), met medewerking van Berend Aukema (BA) en Bram Koese (BK).

Publicaties EIS-Nederland

Anonymus 2007. Waarnemingenverslag 2007, dagvlinders, libellen en sprinkhanen. – EIS-Nederland, De Vlinderstichting en de Nederlandse Vereniging voor Libellenstudie. Jaarlijks overzicht van de meest recente verspreidingskaarten van de drie insectengroepen. In het papieren rapport staan een paar voorbeeldsoorten, op de website van De Vlinderstichting en EIS-Nederland zijn pdf's van met kaarten met alle soorten te bekijken. [RK]

Boesveld, A. & V. Kalkman 2006. Inventarisatie van de Westenschouwse Inlaag op het voorkomen van de Nauwe

korfslak *Vertigo angustior*. – EIS-Nederland, Leiden. 8 p.

Boesveld, A. & V. Kalkman 2007. Verspreiding en habitat van de zeggekorfslak *Vertigo moulinsiana* in Zuid-Holland. – EIS-Nederland, Leiden. 28 p.

Verslag van een gedetailleerd onderzoek naar de verspreiding van deze Habitatrichtlijn-soort. De zeggekorfslak blijkt verrassend genoeg op behoorlijk veel plaatsen voor te komen in Zuid-Holland, zoals Kinderdijk, Kagerplassen en Nieuwkoopse Plassen. [RK]

Bouwman, J. & V. Kalkman 2007. Eindrapportage inhaalslag libellen 2006. – De Vlinderstichting / EIS-Nederland, Wageningen / Leiden. 22 p.