

(D)A(r)tenvielfalt: Datenstandardisierung am Beispiel floristischer Funddaten

Jonathan Ruhm, Fachgebiet II 1.2 Botanischer Artenschutz, Bundesamt für Naturschutz

Pollichia-Tagung 28.02.2026, Von der Natur in die Datenbank



www.bfn.de

Uneinheitliche Datenerfassung verhindert effektive Nutzung floristische Daten

- Unterschiedliche **Erhebungsmethoden**
- Unterschiedliche **Datenfelder** und **Inhalte**
- Unterschiedliche **Bedeutung** gleicher Felder
- Unterschiedliche **Erfassungsintensität**
- Fehlende **gemeinsame Struktur**



KI-generiertes Bild (ChatGPT / OpenAI), 2026

Meyer et al. 2016 ; Kissling et al. 2018; Isaac & Pocock 2015

Heterogene Daten erschweren Austausch, Integration und Auswertung



- Datenaustausch zwischen Institutionen ist **aufwendig**
- Automatisierte **Verarbeitung** ist kaum möglich
- Datenharmonisierung erfordert **manuelle Nacharbeit**
- **Fehleranfälligkeit** steigt deutlich
- Analysen werden **erschwert** oder **unsicher**

Artname	Ort	Datum	Floristischer Status	Quelle
Plantago major	MTB 3545/2	12.08.2021	spontan	Floristische Kartierung
Plantago major L.	Nähe Jena	August 2021	Normalstatus	Gutachten
P. major	50.927, 11.589	2021	etabliert	Monitoringprogramm

Meyer et al. 2016; Isaac & Pocock 2015



Artname	Ort	Datum	Floristischer Status	Quelle
---------	-----	-------	----------------------	--------

Taxon	Raster	Funddatum	Status	Projekt
-------	--------	-----------	--------	---------

Plantago major	MTB 35			
----------------	--------	--	--	--

Name	Latitude	Longitude	Jahr	Status	Bemerkung
------	----------	-----------	------	--------	-----------

MTB 3545/2	2021-08-12	wild	FlorKart
------------	------------	------	----------

Plantago major L.	Nähe			
-------------------	------	--	--	--

Plantago major	50.927	11.589	2021	spontaneous	spontan, FlorKart
----------------	--------	--------	------	-------------	-------------------

3545/2	2021	etabliert	Monitoring
--------	------	-----------	------------

P. major	50.9 11.5			
----------	--------------	--	--	--

Plantago major	50.928	11.590	2021	established	etabliert
----------------	--------	--------	------	-------------	-----------

—	12.08.21	spontan	Ehrenamts erfassung
---	----------	---------	---------------------

P. major	50.9271	11.5889	2020	spontaneous	Monitoring
----------	---------	---------	------	-------------	------------

(D)A(r)tenvielfalt: Datenstandardisierung am Beispiel floristischer Funddaten, 28. Februar 2020

Ein Datenstandard definiert Struktur, Inhalte und Bedeutung der Daten

- Ein Standard ist wie eine **gemeinsame Sprache ... oder Legosteine**:
 - **Einheitliche** Definition von **Datenfeldern**
 - Klare **Bedeutungen** und erlaubte **Werte**
 - Automatisierte **Verarbeitung** wird möglich
 - **Austausch** und **Integration** werden erleichtert



KI-generiertes Bild (ChatGPT / OpenAI), 2026



Standards schaffen Klarheit, Vergleichbarkeit und Nutzbarkeit



Artname	Ort	Datum	Floristischer Status	Quelle
Plantago major	MTB 3545/2	12.08.2021	spontan	Floristische Kartierung
Plantago major L.	Nähe Jena	August 2021	Normalstatus	Gutachten
P. major	50.927, 11.589	2021	etabliert	Monitoringprogramm

Wissen. Name	MTB-Quadrant	Koordinate	Datum (ISO-Format)	Floristischer Status	Quelle
Plantago major L.	3545/2	50.90, 11.00	2021-08-12	etabliert	Floristische Kartierung
Plantago major L.	3545/2	50.300, 11.678	2021-08	etabliert	Gutachten
Plantago major L.	3545/2	50.246, 11.999	2021	etabliert	Monitoringprogramm

Die floristische Kartierung in Deutschland benötigt einen gemeinsamen Standard



- Bedarf nach einer **Standardisierung von Daten** in der floristischen Kartierung
(vgl. Abschlusserklärung der Konferenz zur Floristischen Kartierung in Deutschland 2023, Natur und Landschaft 99: 303–304, 2024; s.a. Metzging et al. 2021)
 - **Erleichterter Austausch und Datenharmonisierung**
 - **Datenqualität**
- Workshop zu den **Inhalten** eines solchen Standards im Jahr 2022
in Zusammenarbeit mit dem Netzwerk Phytodiversität Deutschland e. V. (NetPhyD)
- Wiederaufgreifen des Themas im **Jahr 2024**



KI-generiertes Bild (ChatGPT / OpenAI), 2026

DUFLOR – Datenstandard für Uniforme FLORistische Funddaten



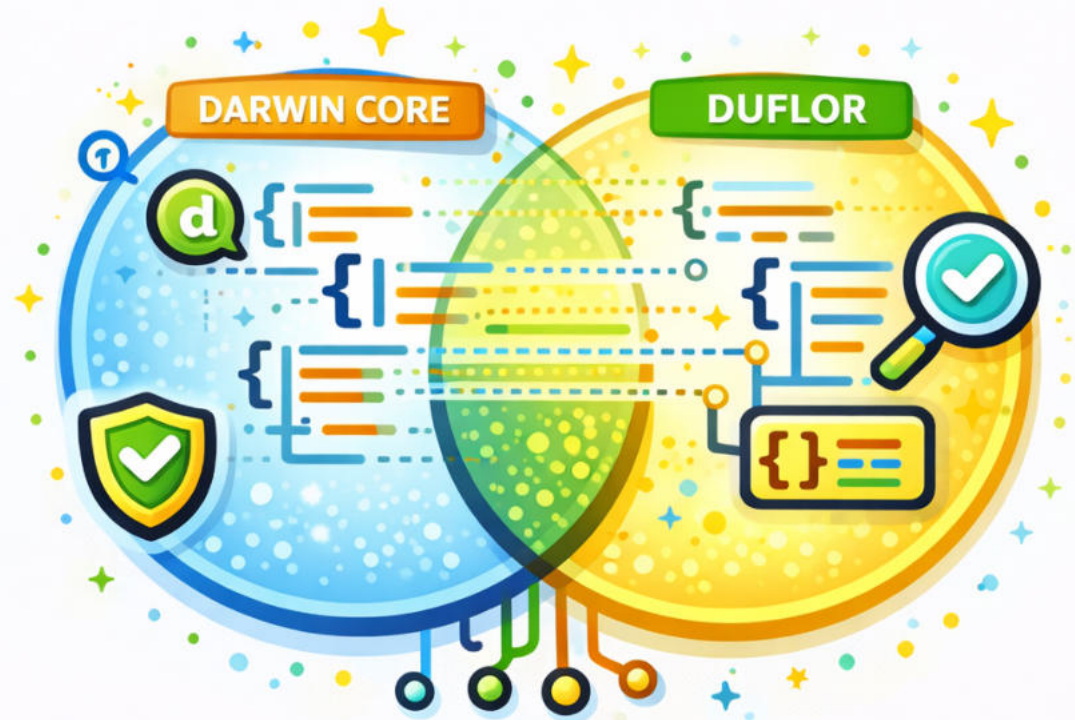
- **Dokumentationsstandard & Austauschformat**(kein Methodenstandard!)
- Wird über ein **Online-Repository** bereitgestellt
- **66 Felder / Attribute**, davon **18 obligatorisch** = Mindeststandard
- Technische Umsetzung im **JSON-Schema**
 - Detaillierte Spezifikation der Felder (z. B. zulässige Zeichen, Datumsformate, Typenzuweisung: Freitext, Zahl, Werteliste etc.)
 - **Abhängigkeiten zwischen Feldern** (z.B. WENN die Art des Nachweises ein Herbarbeleg ist DANN muss auch eine Beleg-ID angegeben werden.)



KI-generiertes Bild (ChatGPT / OpenAI), 2026

DUFLOR adressiert die spezifischen Anforderungen in Deutschland

- Basierend auf dem internationalen Standard **Darwin Core** (GBIF)
- **Spezialisierung** für (floristische) Funddaten in Deutschland
- **Erweiterung** um zusätzliche floristisch relevante Attribute
- Definition von **Beziehungen** und **Abhängigkeiten** zwischen Feldern
- **Vollständig in Darwin Core überführbar**



KI-generiertes Bild (ChatGPT / OpenAI), 2026

DUFLOR wurde breit abgestimmt und gemeinschaftlich entwickelt



- **1× Kommentierungsphase + 3x Online-Abstimmungsrunden**
- **Insgesamt 39 Beteiligte** aus Wissenschaft, Verbänden/Vereinen, Landesbehörden, Kartierung etc. (darunter **8 Bundesländer**)
- Austausch mit **Zentrum für Biodiversitätsinformatik und Sammlungsdatenintegration (ZSB)**



KI-generiertes Bild (ChatGPT / OpenAI), 2026

duflorR ermöglicht die automatische Validierung von Funddaten

- R-Paket zur **Validierung** gegen den Standard
- Automatische **Identifikation** von **Fehlern**
- Erstellung synthetischer **Testdaten**
- Bereitstellung von **Beispieldatensätzen**

Pre-release auf GitHub



(D)A(r)tenvielfalt:

rdaten, 28. Februar 2026

⚠ Status: EXPERIMENTELL / NICHT FÜR PRODUKTION

Dieses R-Package ist eine frühe Vorschau und noch nicht einsatzbereit. Funktionen, Namen und Ergebnisse können sich jederzeit ändern. Bitte aktuell nicht verwenden – Feedback über Issues ist willkommen.

README – duflor

✅ duflor – DUFLOR Data Validator

Beschreibung

duflor ist ein R-Paket zur Validierung floristischer Funddaten gemäß dem offiziellen DUFLOR-Datenstandard (Datenstandard für uniforme floristische Funddaten, Bundesamt für Naturschutz).

Das Paket bietet Funktionen zur:

- strukturierten Prüfung von Beobachtungs- und Nachweisdaten gegen ein JSON-Schema
- Erstellung synthetischer Testdatensätze
- Generierung typischer Validierungsfehler zur Testabdeckung

Autor: Jonathan Ruhm (Bundesamt für Naturschutz)
Lizenz: MIT

Installation

```
# Lokal installieren (nach Klonen oder Download)
devtools::install("pfad/zum/duflor")

# Oder direkt von GitHub
devtools::install_github("JonathanRu3000/duflor")
```

Inhalte

Hauptfunktionen

Funktion	Beschreibung
validate_floristic_occdata()	Validiert floristische Daten gegen das DUFLOR-Schema
generiere_floristic_occdata()	Erstellt vollständigen, verknüpften Beispieldatensatz
generate_test_cases()	Erstellt gezielt fehlerhafte Testfälle

DUFLOR wird implementiert, getestet und breit verfügbar gemacht



- **Technische Finalisierung**
(letzte Anpassungen, Validierung, Dokumentation)
- **Praxistest anhand realer Datensätzen**
- **Veröffentlichung des Standards**
→ über **Fachzeitschrift, GitLab (OpenCode)** und die **BfN-Homepage**
- **Ableitung von Standards für andere Gruppen / oder sogar gruppenübergreifender Standard ??**

```
"Lokalitaet": {
  "allof": [
    {
      "then": {
      }
    }
  ]
},
"Person": {
  "description": "Die kartierende Person als Quelle des Nachweises. Nicht obligat. Ueber personID koennen Zuordnungen von Nachweisen anonym vorgenommen w",
  "type": "object",
  "properties": {
    "personID": {
      "type": "string",
      "description": "Eine innerhalb eines Erfassungssystems eindeutige ID einer Person",
      "examples": [
        "TR-30",
        "47586",
        "HTZLO-PP",
        "Karl-8447"
      ]
    },
    "x-comment": [],
    "x-dwc-term": "recordedBy",
    "x-dwc-term-consistency": "Gute Uebereinstimmung",
    "x-dwc-identifier": "http://rs.tdug.org/dwc/terms/recordedBy",
    "x-abc-term": "",
    "x-abcd-consistency": ""
  },
  "personName": {
    "type": "string",
    "pattern": "(?:\\s*[A-Z]\\.|\\s*[A-Z][a-zA-Z]*\\s+[A-Z][a-zA-Z]*\\b(?:\\s*\\s*(?:\\s*[A-Z]\\.|\\s*[A-Z][a-zA-Z]*\\s+[A-Z][a-zA-Z]*\\b))*",
    "description": "Namen von Beobachtenden. Wenn mehrere dann durch | getrennt. Entweder voll ausgeschrieben oder abgekuerzt (siehe Beispiel)",
    "examples": [
      "A. Mustermensche | F. Falkenberg | Sabine Mueller|.."
    ]
  },
  "x-comment": [],
  "x-dwc-term": "recordedBy",
  "x-dwc-term-consistency": "Gute Uebereinstimmung",
  "x-dwc-identifier": "http://rs.tdug.org/dwc/terms/recordedBy",
  "x-abc-term": "PersonName / FullName",
  "x-abcd-consistency": "Gute Uebereinstimmung"
}
},
"Nachweis": {
  "type": "object",
  "description": "Ein Nachweis über ein Vorkommen einer Art",
  "properties": {
    "nachweisID": {
      "type": "string",
      "description": "Eine innerhalb eines Erfassungssystems eindeutige ID eines Nachweises",
      "examples": [

```

Bei ungeeigneten Bildhintergründen kann der Urheber auch an der Seite genannt werden.

Daten sollten FAIR sein

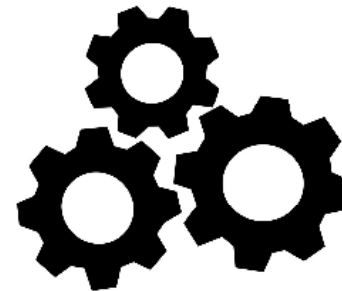


Findable

Accessible

Interoperable

Reusable



**Identifikationsnum
mer**

**Nicht auf Ihrer
Festplatte im Keller**

Standardisiert

Lizenziert

Nutzen Sie Datenstandards!



- **Biodiversitätsdaten sind wertvoll und oft aufwendig erhoben**
- **Ohne klare Struktur und Lizenz bleiben Daten ungenutzt**
- **Standardisierung macht Daten auffindbar und kombinierbar**
- **Lizenzangaben schaffen Rechtssicherheit für Nutzende und Datengebende**
- **Grundlage für langfristige Nachnutzung und wissenschaftlichen Mehrwert**

Literaturverzeichnis



- Isaac, Nick J. B.; Pocock, Michael J. O. (2015): Bias and information in biological records. In: *Biol J Linn Soc Lond* 115 (3), S. 522–531. DOI: 10.1111/bij.12532.
- Metzging, Detlev; May, Rudolf; Wolf, Daniel (2021): Die floristische Kartierung in Deutschland - Methoden, Ergebnisse, Herausforderungen und Chancen. In: *0028-0615* 96 (9+10), S. 426–433. DOI: 10.17433/9.2021.50153941.426-433.
- Meyer, Carsten; Weigelt, Patrick; Kreft, Holger (2016): Multidimensional biases, gaps and uncertainties in global plant occurrence information. In: *Ecology Letters* 19 (8), S. 992–1006. DOI: 10.1111/ele.12624.
- Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, I. Jsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific data* 3, S. 160018. DOI: 10.1038/sdata.2016.18.

Abbildungsverzeichnis



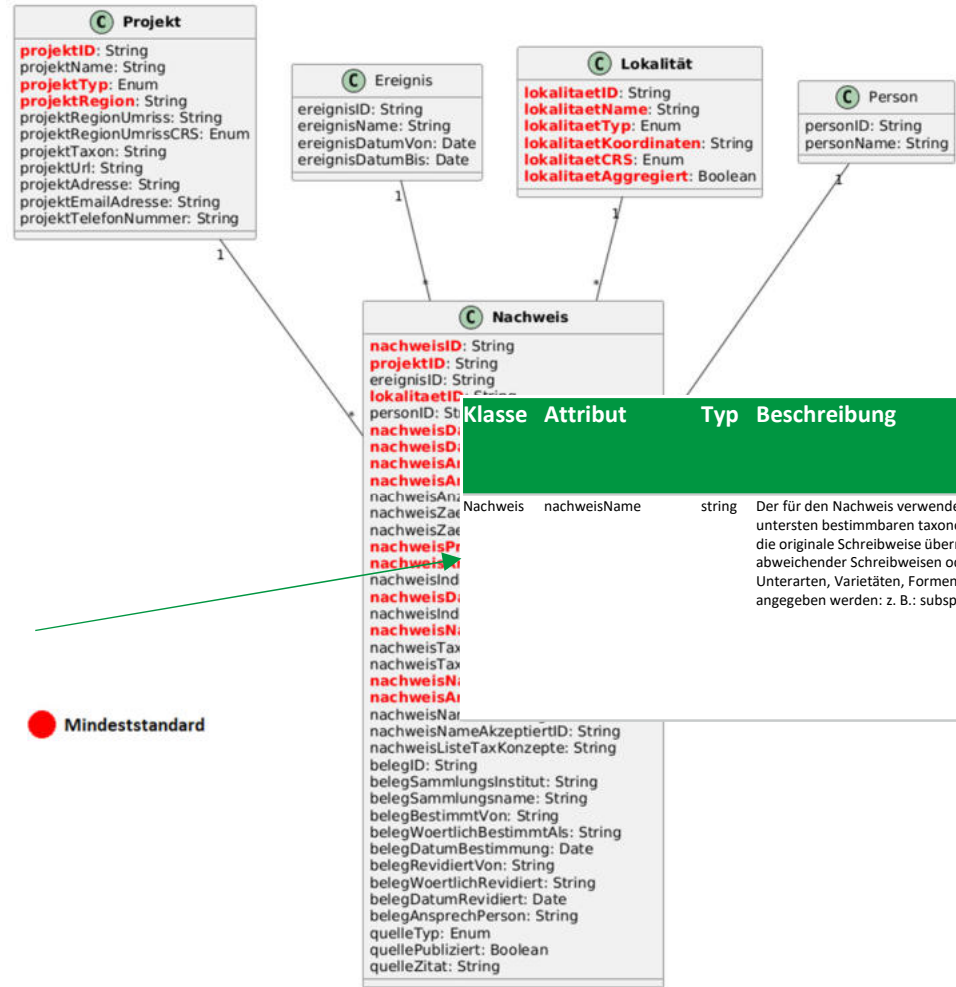
- Folie 4:
 - By SangyaPundir - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=53414062>
- Folie 13
 - Nick Youngson - link to - <http://www.nyphotographic.com/> - Do It Now, BY-SA 3.0, **Attribution**
Link: Pix4free.org - link to - <https://pix4free.org/>, **Original**
Image: <https://www.thebluediamondgallery.com/highway-signs/d/do-it-now.html>

Nutzen Sie Standards



Wissen. Name	MTB-Quadrat	Koordinaten	Datum (ISO-Format)	Floristische r Status	Bezugsraum	Quelle
Plantago major L.	3545/2	50.90, 11.00	2021-08-12	etabliert	Bayern	Floristische Kartierung
Plantago major L.	3545/2	50.300, 11.678	2021-08	Normalstatus	NRW	Gutachten
Plantago major L.	3545/2	50.246, 11.999	2021	spontan	Sachsen	Monitoringprogramm

Klassen und Attribute des Standards



● Mindeststandard

Klasse	Attribut	Typ	Beschreibung	Beispiel
Nachweis	nachweisName	string	Der für den Nachweis verwendete vollständige wissenschaftliche Name (inkl. Autor) des untersten bestimmaren taxonomischen Rangs. Es handelt sich um eine Freitext-Eingabe, in der die originale Schreibweise übernommen wird – inklusive eventueller Schreibfehler, abweichender Schreibweisen oder veralteter Nomenklatur. Unterarten, Varietäten, Formen und Hybride können durch die entsprechenden Kürzel angegeben werden: z. B.: subsp., var., f., × (bzw. x)	<ul style="list-style-type: none"> • Erica vulgaris L. • Aegilops speltoides subsp. ligustica (Savign.) Zhuk. • Triticum ovatum var. echinus (Godr.) Thell. • Agrostis vulgaris f. umbrosa (Schur) Junge • Asplenium septentrionale x trichomanes subsp. trichomanes • Atriplex longipes x prostrata