



# Netherlands Biodiversity Data Services

**Naturalis**  
Biodiversity  
Center

# Naturalis Specimen Collection



# Digitisation on a Large Scale

- 
- 1 PISCES
  - 2 REPTIELEN & AMFIBIEEN
  - 3 CRUSTACEA
  - 4 PETROLOGIE
  - 5 FOSSIELE MACROVERTEBRATA
  - 6 FOSSIELE MACROVERTEBRATA
  - 7 CEPHALOPODA
  - 8 PALAEOZOISCHE MOLLUSCA
  - 9 CRUSTACEA & MOLLUSCA
  - 10 COLEOPTERA LEIDEN
  - 11 HYMENOPTERA LEIDEN
  - 12 LEPIDOPTERA LEIDEN
  - 13 ENTOMOLOGIE ZMA
  - 14 AVES: EENDACHTIGEN EN SKELETMATERIAAL
  - 15 AVES: EENDEN, STELTLOPERS EN HALLEN
  - 16 AVES: ZANOVOGELS
  - 17 SCHATKAMER & ARCHIEF
  - 18 MAMMALIA
  - 19 MAMMALIA
  - 20 DROGE TYPEN



**~ 8 Million**  
specimen records  
(41 Million specimens  
including specimens in  
storage units)

# Publishing Natural History Collection Data

## Objectives

- Preserve scientific heritage
- Add value to existing data by making them **searchable** and **citable**
- Make data accessible for researchers anywhere
- Enable meta-analyses on large amounts of data
- Facilitate collaboration between institutes

## Key points for developing data services

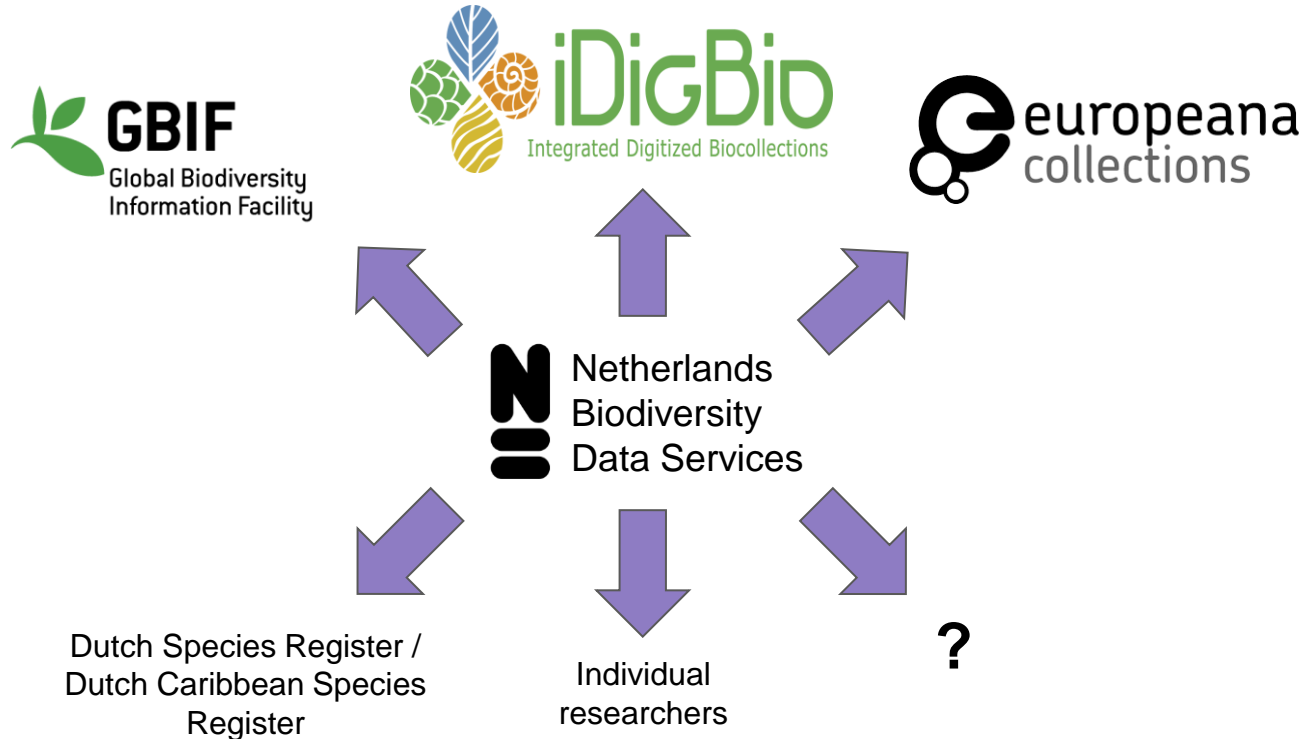
- Make data freely accessible (Open Content Policy)
- Apply non-restrictive license to digital content
- Standardised, machine-readable data formats
- Provide APIs for building custom software that uses the museum's collection data

# Netherlands Biodiversity Data Services

Two main types of data services:

- **Netherlands Biodiversity API (NBA)**
  - gives *automated* access to Naturalis collection data, i.e specimen and associated multimedia data, and also access to species data from Species 2000 and Naturalis
- **Persistent Uniform Resource Locator (PURL) services**
  - persistent URLs used as *permanent* and *citable* digital identifiers for specimen only

# Integration with International Initiatives



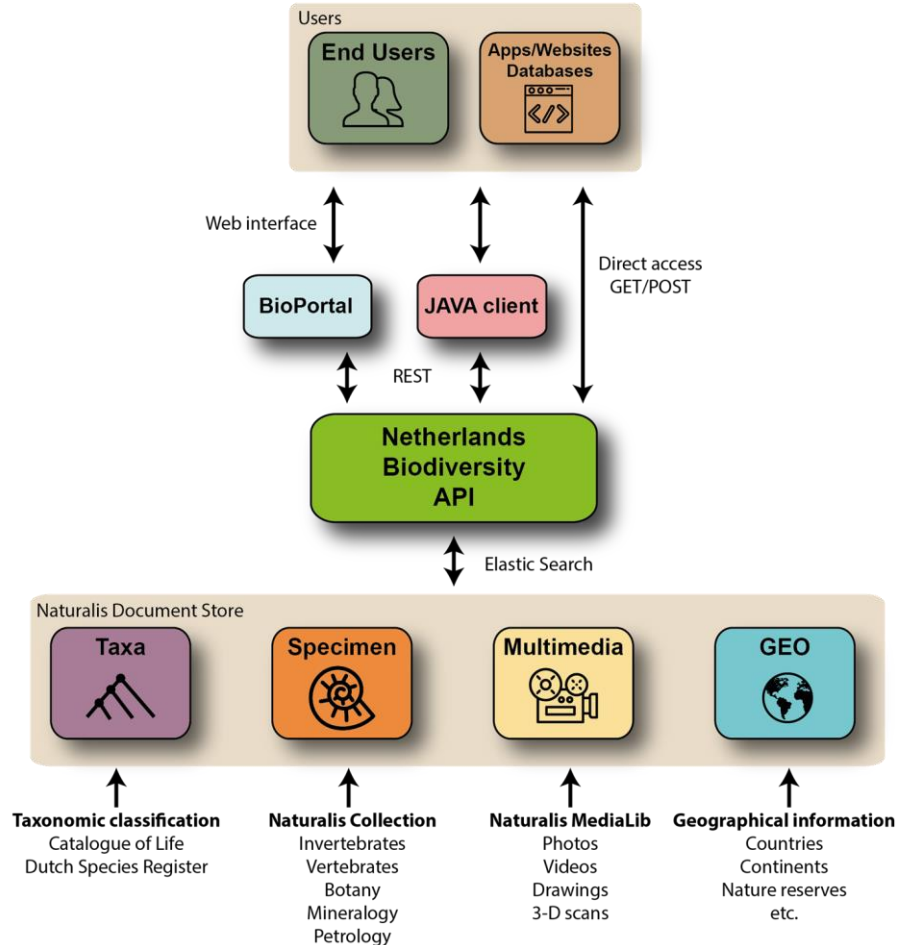
# NBA Overview

## NDS

- NoSQL database
- Four types of data in NDS (Naturalis Document Store)

## NBA

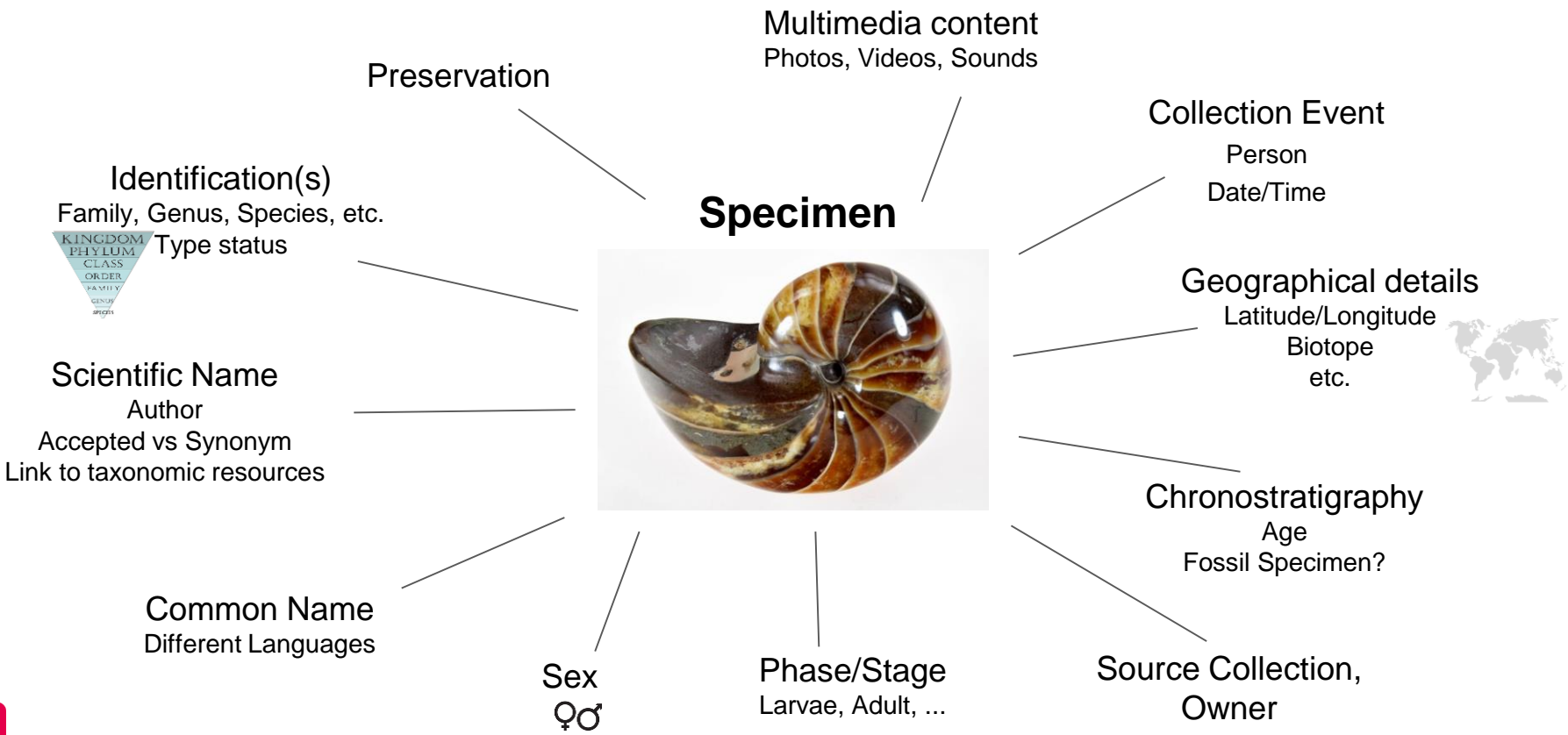
- near real-time search on NDS with Elasticsearch engine
- total of 98 endpoints
- includes java client (direct HTTP access) and a very helpful scratchpad
- underlying engine for Bioportal (Web Entrance to Naturalis Collection)



# Specimen Data Type

- Plant and fungi specimen  
*National Herbarium of the Netherlands (Leiden, Utrecht Wageningen)*
- Animals, arts, rocks, meteorites, ...  
*Rijksmuseum van Natuurlijke Historie (RMNH)*  
*Rijksmuseum van Geologie en Mineralogie (RGM)*  
*Zoologisch Museum Amsterdam (ZMA)*



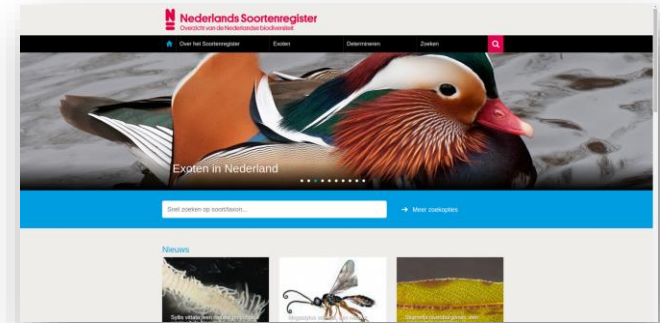


# Taxon Data Type

- Species & their taxonomic classification
- Scientific & common names
- Accepted names, synonyms, etc.

**Species checklists**, hosted at Naturalis:

- *Dutch Species Register*:  
Species occurring in the Netherlands
- *Catalogue of Life (CoL)*:  
Global species checklist published by the Species 2000 organisation



## Catalogue of Life



# Multimedia Data Type

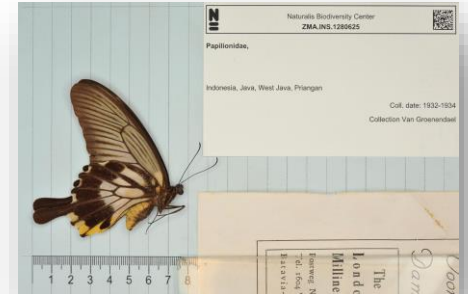
- Data types: Images / Videos
- Linked to
  - specimen object
  - taxon in Netherlands Species Registry
- Photos of species from Netherlands Species Register
- Mostly Creative Commons zero license (CC0)



*Acer palmatum*



*Hydrochoerus hydrochaeris*



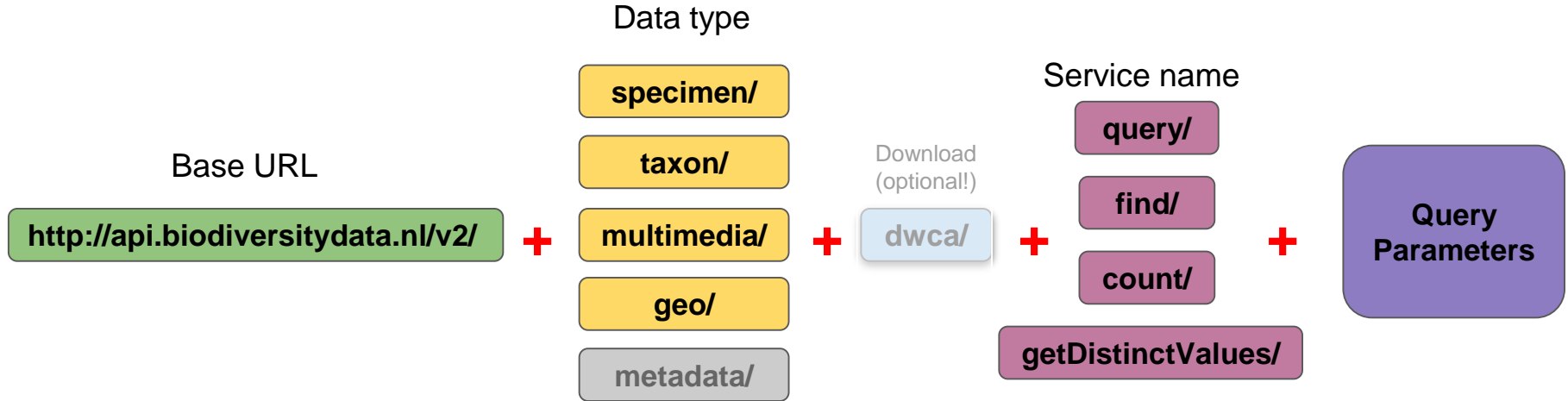
Unclassified *Papilionidae*

# GeoArea Data Type

- Not primary but *support* data type
- data imported from various sources
- Facilitating geographic search
- Polygons on different scales:
  - Country
  - Province (NL)
  - Municipality (NL)
  - Nature Reservations (NL)
- Polygons encoded in GeoJSON format



# HTTP Services Summary



**Example:** <http://api.biodiversitydata.nl/v2/specimen/query/?sex=female>

## NOTE:

- Not all above combinations possible
- Not all services take query arguments
- Download services: Darwin Core Archive (DwC-A) format
- Many more (data type-specific) services available

# Query Results

- JSON format
- Can be read by various tools and programming/scripting languages
- Output controllable by query parameters (e.g. `_fields`, `_size`, ...)

```
{
  "score": 0.7218451,
  "item": {
    "id": "ZMA.INS.747900@CRS",
    "fromCaptivity": false,
    "objectPublic": false,
    "multiMediaPublic": false,
    "gatheringEvent": {
      "country": "Switzerland",
      "provinceState": "Uri, Obwalden",
      "locality": "Engelberg, Herrenrüti",
      "localityText": "Switzerland, , Engelberg, Herrenrüti",
      "dateTimeBegin": "1880-08-24T00:00:00+0000",
      "gatheringPersons": [
        {
          "fullName": "Frey-Gessner, E."
        }
      ]
    },
    "identifications": [
      {
        "taxonRank": "subspecies",
        "scientificName": {
          "fullScientificName": "Bombus (Alpigenobombus) wurflenii ssp. mastrucatus Gerstaecker, 1869",
          "genusOrMonomial": "Bombus",
          "subgenus": "Alpigenobombus",
          "specificEpithet": "wurflenii",
          "infraspecificEpithet": "mastrucatus",
          "authorshipVerbatim": "Gerstaecker, 1869",
          "scientificNameGroup": "bombus wurflenii mastrucatus"
        },
        "defaultClassification": {
          "order": "Hymenoptera",
          "family": "Apidae",
          "genus": "Bombus",
          "subgenus": "Alpigenobombus",
          "specificEpithet": "wurflenii",
          "infraspecificEpithet": "mastrucatus"
        },
        "preferred": true
      }
    ]
  }
},
```

# Query Parameters (human readable)

Field	Operator	Value
sex	=	female
identifications.taxonRank	=	species
identifications.defaultClassification.family	=	Equidae

[http://api.biodiversitydata.nl/v2/specimen/query/  
?sex=female  
&identifications.taxonRank=species  
&identifications.defaultClassification.family=Equidae](http://api.biodiversitydata.nl/v2/specimen/query/?sex=female&identifications.taxonRank=species&identifications.defaultClassification.family=Equidae)

## NOTE:

- Query above: AND conjunction
- 'Human readable' queries have limited power
- E.g. partial matching, nested and/or constructs not possible
- Need complex query mechanism

# Complex Queries

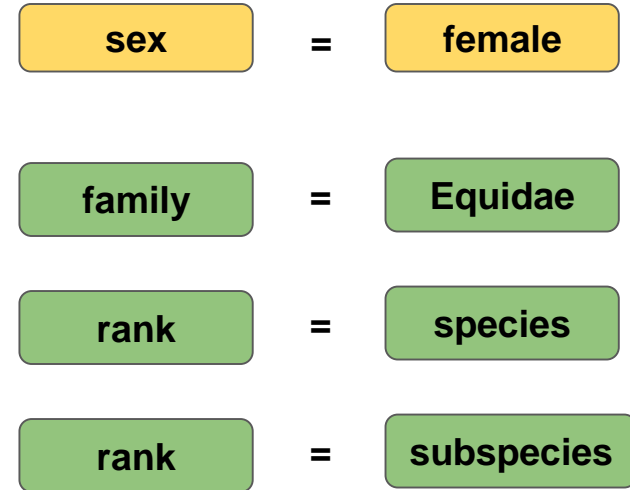
Field	Operator	Value
sex	=	female
identifications.taxonRank	=	species OR subspecies
identifications.defaultClassification.family	=	Equidae

- Model complex query using **\_querySpec** parameter
- Query term given in JSON format
- [http://api.biodiversitydata.nl/v2/specimen/query/?  
\\_querySpec= { <JSON query term> }](http://api.biodiversitydata.nl/v2/specimen/query/?_querySpec={<JSON query term>})



# Complex Queries example

```
{
  "conditions": [
    {
      "field": "sex",
      "operator": "EQUALS",
      "value": "female"
    },
    {
      "field": "identifications.defaultClassification.family",
      "operator": "EQUALS",
      "value": "Equidae"
    },
    {
      "field": "identifications.taxonRank",
      "operator": "EQUALS",
      "value": "species",
      "or": [
        {
          "field": "identifications.taxonRank",
          "operator": "EQUALS",
          "value": "subspecies"
        }
      ]
    }
  ]
},
"logicalOperator": "AND"
}
```



# Low-level access

- Copy encoded URL into web browser
- cURL (command-line http client):

```
curl -X POST http://api.biodiversitydata.nl/v2/specimen/query -d '{
  "conditions": [
    {
      "field": "identifications.defaultClassification.genus",
      "operator": "EQUALS_IC",
      "value": "Passiflora"
    }
  ]
}'
```



# NBA Client(s)

Programming languages:

- Java client ([http://naturalis.github.io/naturalis\\_data\\_api/javadoc/v2/all/](http://naturalis.github.io/naturalis_data_api/javadoc/v2/all/))
- R client (2018)

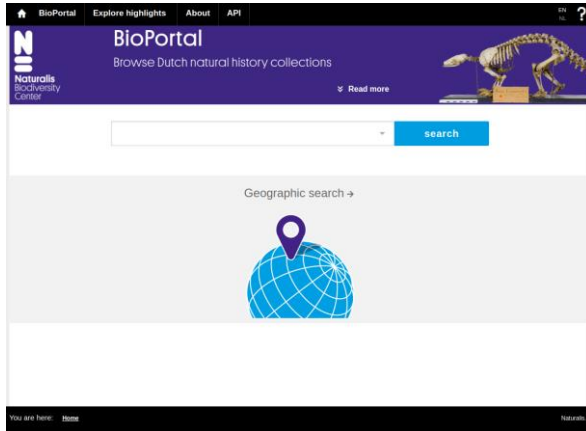
Applications:

- BioPortal
- Dutch Species Register (retrieves photos from NBA)
- Dutch Caribbean Species Register

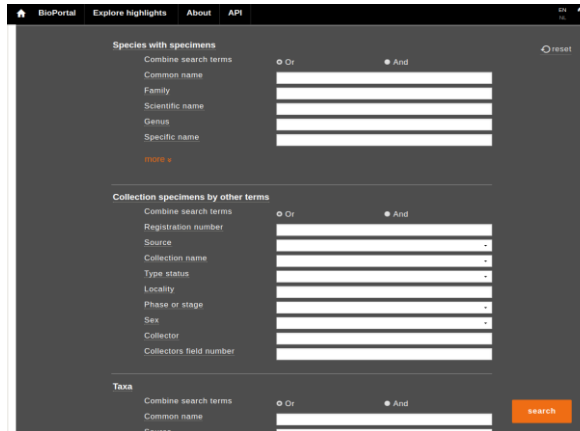


# BioPortal

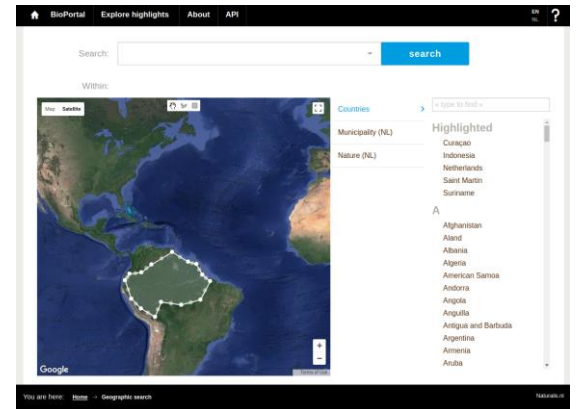
<http://bioportal.naturalis.nl/>



Simple search



Advanced search



Geo search



# NBA documentation

## <http://docs.biodiversitydata.nl>

### Netherlands Biodiversity Api <sup>v2</sup>

[ Base URL: /v2 ]

<http://api.biodiversitydata.nl/v2/reference-doc>

Access to the digitised Natural History collection at the Naturalis Biodiversity Center

[Naturalis Biodiversity Center - Website](#)

[Send email to Naturalis Biodiversity Center](#)

[GPL](#)

[Documentation and examples](#)

**specimen** Access specimens present in the collection >

**taxon** Access taxon information from taxonomic catalogues >

**geo** Access polygons of geographical areas >

**multimedia** Access multimedia objects associated with specimens ∨

**GET** `/multimedia/metadata/getPaths` Returns the full path of all fields within a document

**GET** `/multimedia/metadata/getSettings` List all publicly available configuration settings for the NBA

**GET** `/multimedia/metadata/getSetting/{name}` Get the value of an NBA setting

**GET** `/multimedia/metadata/getFieldInfo` Returns extended information for each field of a multimedia document



# Future directives (2018)

## More types of data

- Host occurrence data from [waarneming.nl](http://waarneming.nl) and offer access through NBA
- Host bird sounds from [xeno-canto.org](http://xeno-canto.org) and offer access through NBA
- Make Storage units available
- Molecular sequence data

## Better data

- Enrich data with geographical coordinates (to date ~5% referenced)
- Linked open data

## More clients

- Implement and publish R client





**nba@naturalis.nl**

**Naturalis**  
Biodiversity  
Center