



# The Netherlands Biodiversity Data Services and the R package nbaR: Automated workflows for biodiversity data analysis

Hannes Hettling, Maarten Schermer, Rutger Vos & Daphne Duin

**Naturalis**  
Biodiversity  
Center

# Naturalis Biodiversity Center

Leiden, the Netherlands

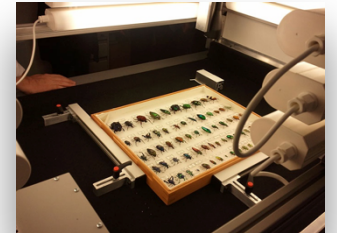


# Naturalis Biodiversity Center

Leiden, the Netherlands



# Naturalis Specimen Collection



# Digitisation on a Large Scale



- 00 PISCES
- 01 REPTIELEN & AMFIBIEËN
- 02 CRUSTACEA
- 03 PETROLOGIE
- 04 FOSSIELE MACROVERTEBRATA
- 05 FOSSIELE MACROVERTEBRATA
- 06 CEPHALOPODA
- 07 PALAEOZOÏSCHE MOLLUSCA
- 08 CRUSTACEA & MOLLUSCA
- 09 COLEOPTERA LEIDEN
- 10 HYMENOPTERA LEIDEN
- 11 LEPIDOPTERA LEIDEN
- 12 ENTOMOLOGIE ZMA
- 13 AVES: EENDACHTIGEN EN SKELETMATERIAAL
- 14 AVES: EENDEN, STELTLOPERS EN HALLEN
- 15 AVES: ZANGVOGELS
- 16 SCHATKAMER & ARCHIEF
- 17 MAMMALIA
- 18 MAMMALIA
- 19 DROGE TYPEN

**~ 9 Million  
specimen records  
(41 Million specimens  
including specimens in  
storage units)**

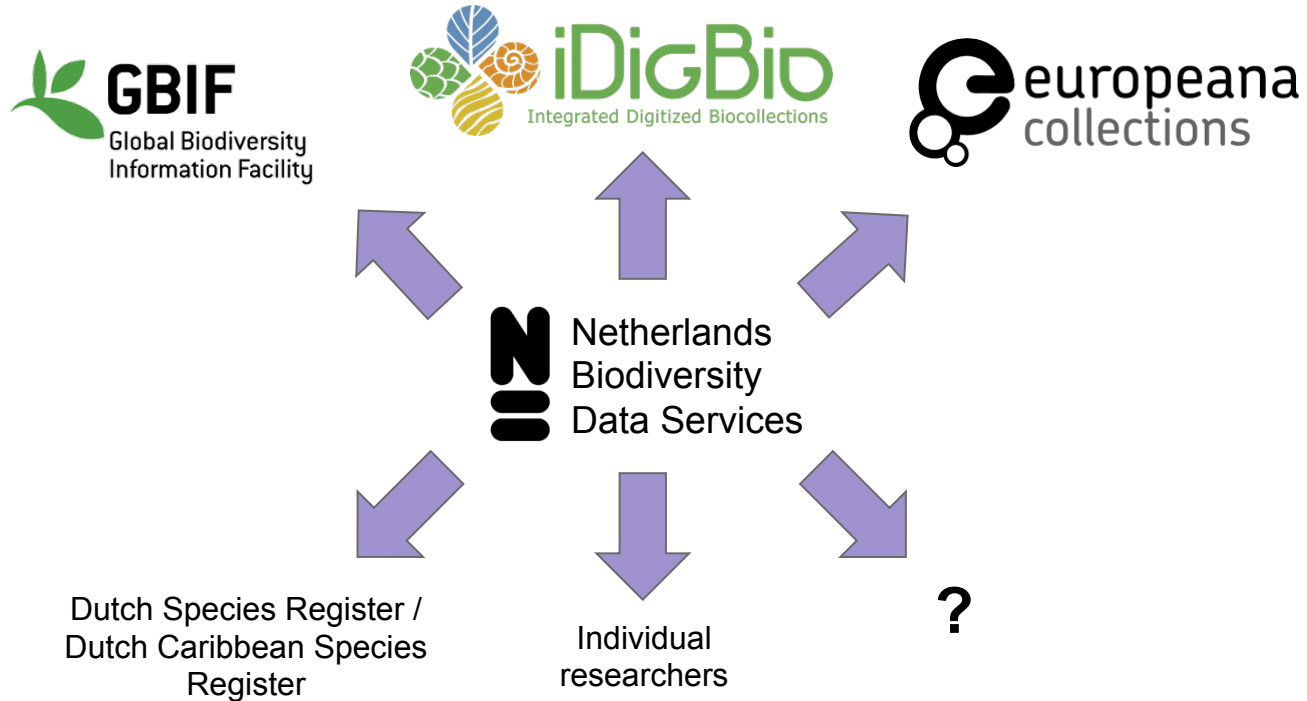
# Naturalis Open Content Policy

**“A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.”**

*(first sentence of Naturalis Open Content Policy)*



# Integration with International Initiatives



# 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



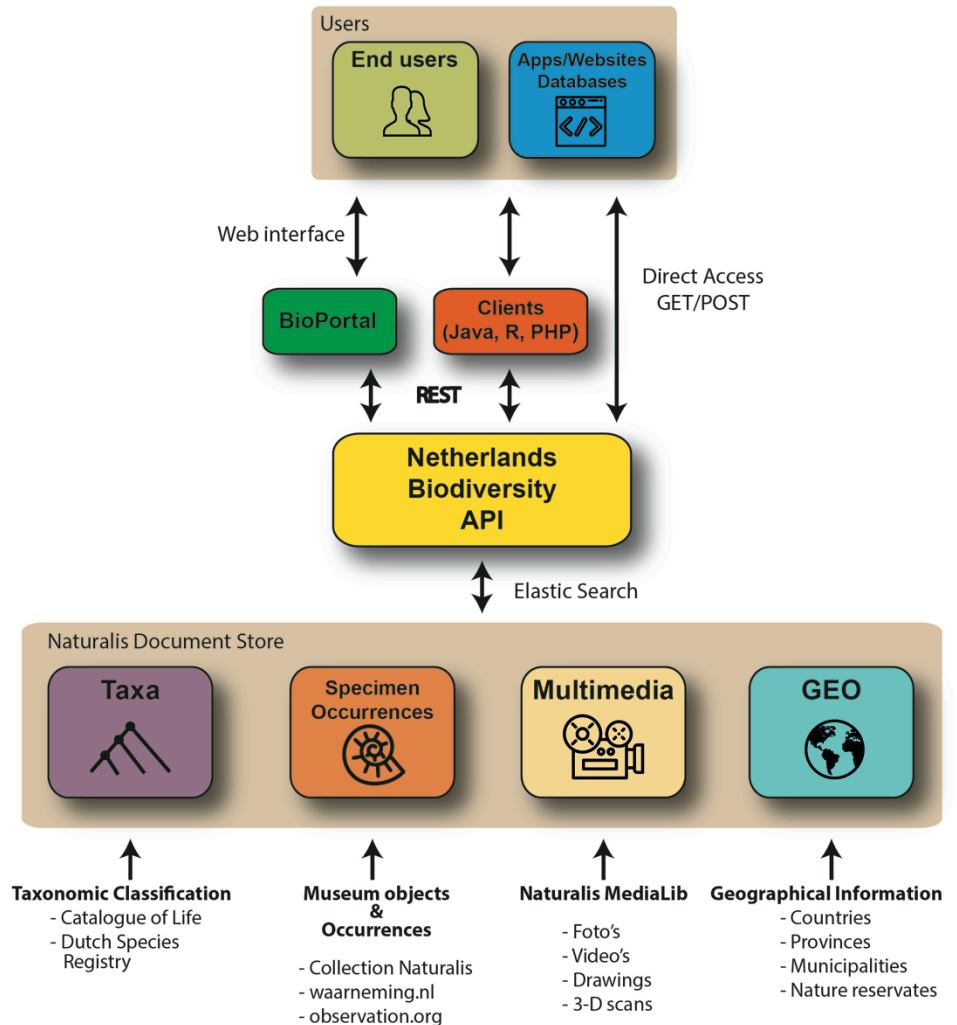
# NBA Overview

## NDS

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

## NBA

- near real-time search on NDS with Elasticsearch engine
- ~100 endpoints
- underlying engine for Bioportal (Web Entrance to Naturalis Collection)



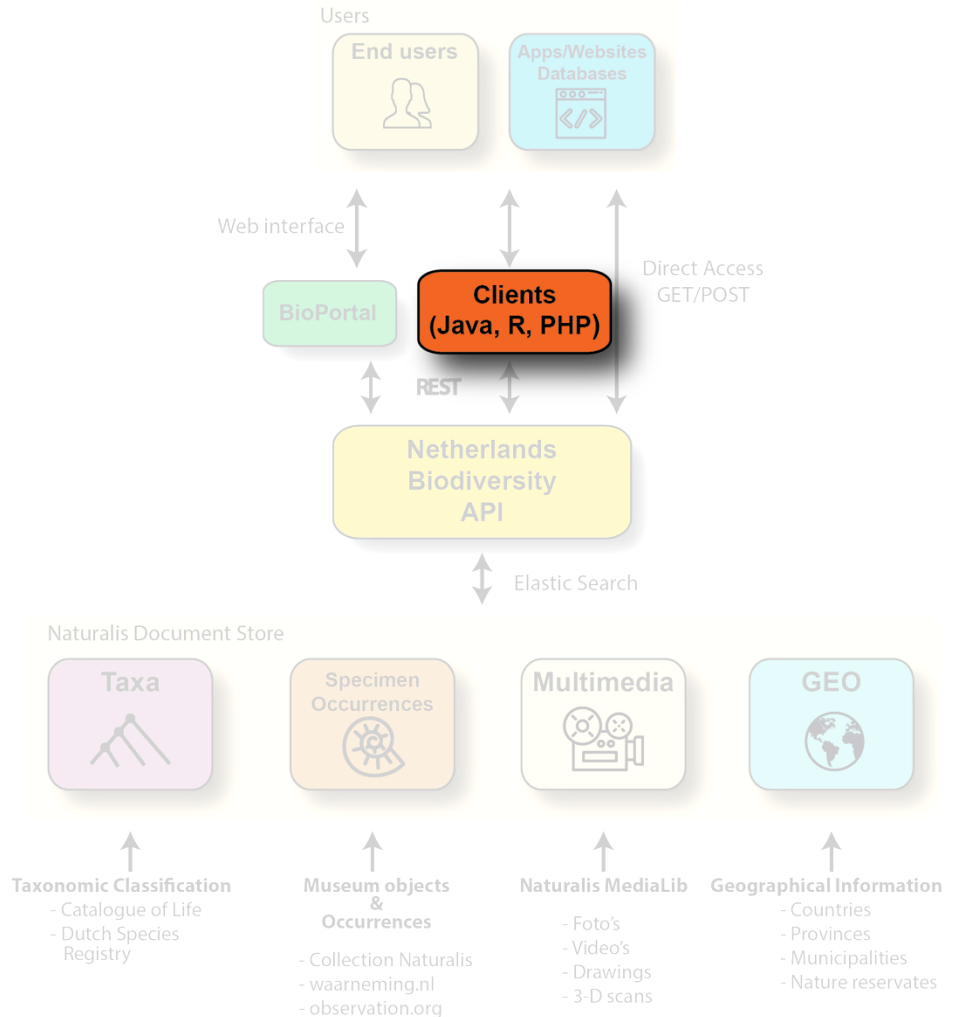
# NBA Overview

## NDS

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

## NBA

- near real-time search on NDS with Elasticsearch engine
- ~100 endpoints
- underlying engine for Bioportal (Web Entrance to Naturalis Collection)



# NBA R client: nbaR



Why R?

- Wide acceptance in the research community
- Easy to use, free of charge, open-source, platform independent
- Possible integration with other packages: Reproducible workflows for science

# nbaR

Aim:

- **Full client:** Implement
  - entire object model of NBA (e.g. Specimen, Taxon, SpecimenIdentification, GatheringEvent ...)
  - all end points
- Convenience functions: connect our data to popular packages
- Easily incorporate updates in NBA
- Submit to popular package repository: *R Open Science*

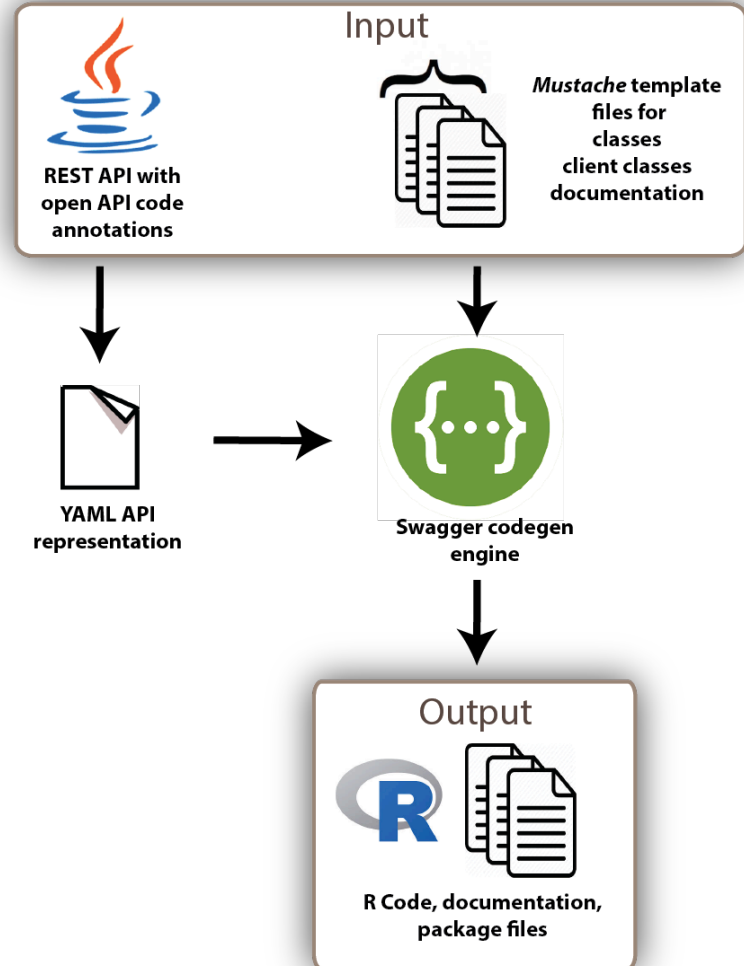
# Implementation

## Swagger open API specification

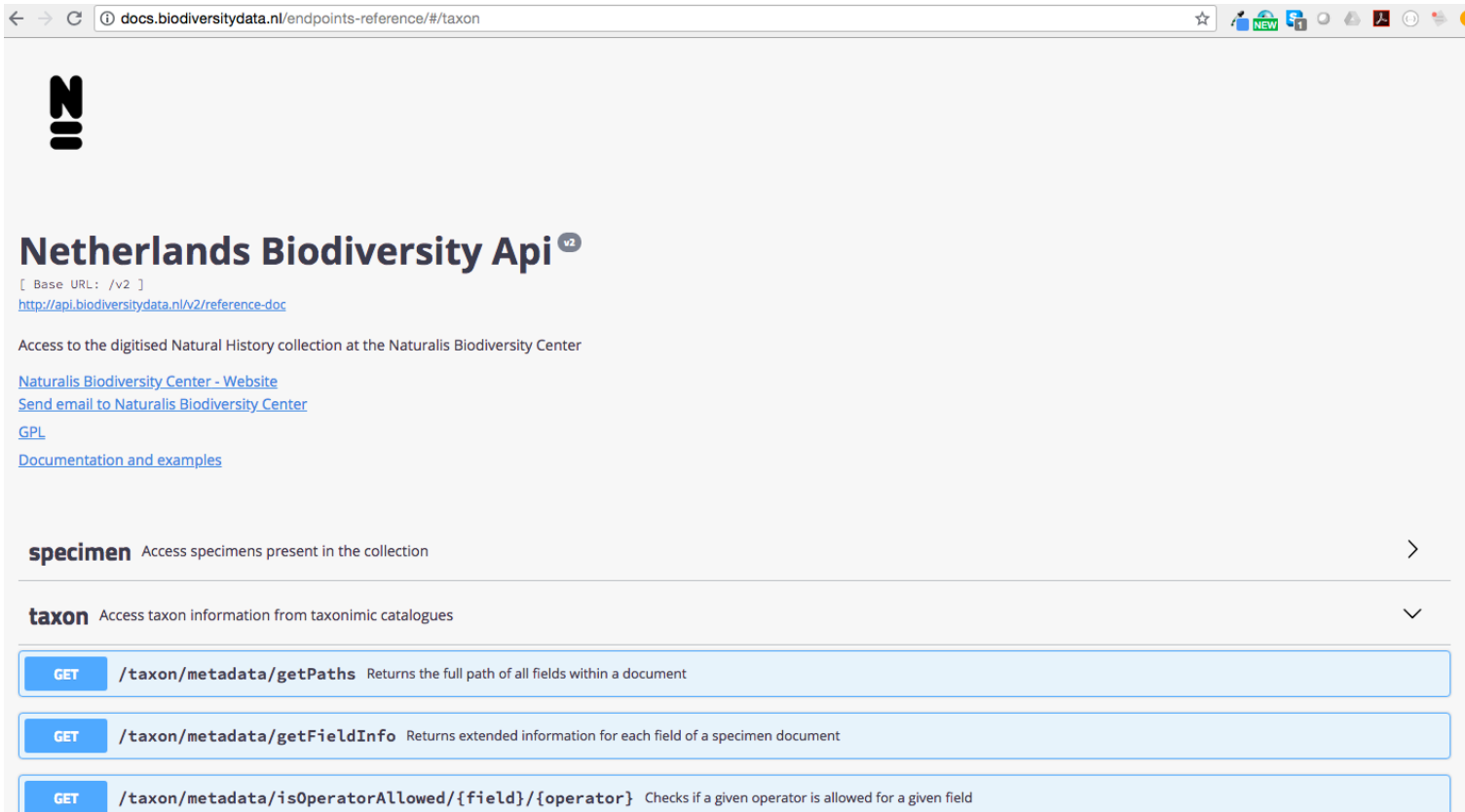
- <http://api.biodiversitydata.nl/v2/reference-doc>
- Parsed from Java annotations
- Complete API description in YAML format
- Models, endpoints, query parameters, return types ...

## Swagger codegen

- Input: Swagger YAML
- Code generation via *mustache* template files
- Configurable codegen engine



# Automatic generation of documentation



The screenshot shows a web browser window with the URL `docs.biodiversitydata.nl/endpoints-reference/#/taxon`. The page features the NIO logo in the top left corner. The main heading is "Netherlands Biodiversity Api v2", with a sub-note "[ Base URL: /v2 ]" and a link to `http://api.biodiversitydata.nl/v2/reference-doc`. Below this, there is a paragraph about access to the digitised Natural History collection at the Naturalis Biodiversity Center, followed by links for the website, email, GPL, and documentation. The page is organized into sections: "specimen" (Access specimens present in the collection) and "taxon" (Access taxon information from taxonomic catalogues). Under the "taxon" section, three API endpoints are listed in light blue boxes, each with a "GET" method label and a brief description of the endpoint's function.

**NIO**

## 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 ✓

- GET** `/taxon/metadata/getPaths` Returns the full path of all fields within a document
- GET** `/taxon/metadata/getFieldInfo` Returns extended information for each field of a specimen document
- GET** `/taxon/metadata/isOperatorAllowed/{field}/{operator}` Checks if a given operator is allowed for a given field



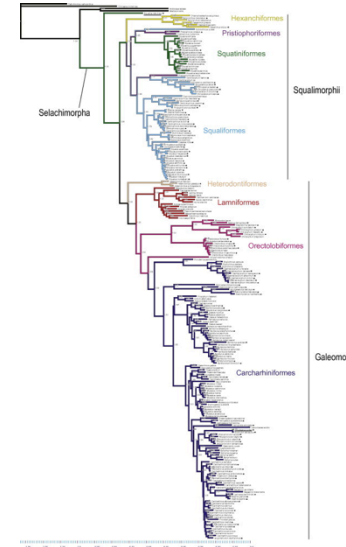
# Use cases in research (Ecology & Evolution)

- Stratigraphic level and certainty of fossil specimens:  
**time calibration** of molecular phylogenies
- Species Distribution modelling
- Past species distributions
- Abundances of collection of specimens through time
- Taxon name resolution
- Taxonomic research
- Image recognition for automatic classification of specimens

# Software demo session: Using nbaR to ...



... find the  
record for the  
oldest tomato  
specimen in  
the world



... calibrate a  
molecular  
phylogeny of  
the sharks





**Package:**

<https://github.com/naturalis/nbaR>

**Docs:**

<https://naturalis.github.io/nbaR/>

**Support:**

[nba@naturalis.nl](mailto:nba@naturalis.nl)

**Demo:**

**Tuesday, 17:00, Foyer**

**Naturalis**  
Biodiversity  
Center