



BIODIVERSITY BUILDING BLOCKS FOR POLICY

POLICY BRIEF

🏠 b-cubed.eu

✂️ [@BCubedProject](https://twitter.com/BCubedProject)

🌐 [B-Cubed Project](https://www.linkedin.com/company/b-cubed-project/)

📺 [@B-CubedProject](https://www.youtube.com/channel/UCB-CubedProject)

🌱 b-cubed.eu



Data pipelines for policy-relevant indicators

Quentin Groom¹, Nikol Stoykova², Andrew Rodrigues³, Jasmijn Hillaert⁴, Lina M. Estupinan Suarez⁵,
Sandra MacFadyen⁶ Ward Langeraeert⁷

¹ Meise Botanic Garden, Meise, Belgium

² Pensoft Publishers, Sofia, Bulgaria

³ Global Biodiversity Information Facility, Copenhagen, Denmark

⁴ Research Institute for Nature and Forest, Brussels, Belgium

⁵ German Centre for Integrative Biodiversity Research, Leipzig, Germany,

⁶ Mathematical Sciences, Stellenbosch University, South Africa

⁷ Research Institute for Nature and Forest, Brussels, Belgium

Key messages



Automated data pipelines make it possible to turn scattered biodiversity records into up-to-date indicators that support evidence-based policymaking.



Tools such as GBIF's data cube service, combined with a collection of open-source R packages provide standardised, interoperable, and FAIR-compliant workflows for analysing and sharing biodiversity data.



While these tools are mainly used by scientific and technical teams, the outputs they support, such as dashboards, reports, and indicators, deliver clear, accessible evidence to policymakers, improving the timeliness of biodiversity information.



Funded by
the European Union

B3 (Biodiversity Building Blocks for policy) receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101059592). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

Introduction

Moving from biodiversity records to reliable indicators that can inform policy remains a major challenge for data scientists and biodiversity informaticians. Although large volumes of biodiversity data are available, they are often fragmented, inconsistent, and difficult to integrate efficiently. As a result, producing indicators that are both scientifically robust and suitable for policy reporting requires considerable time and technical effort.

This brief outlines how automated, interoperable data pipelines can streamline this process.

We are developing clear, scalable pipelines that translate raw biodiversity data into decision-relevant indicators. By linking these to established infrastructures such as the [EBV Data Portal](#) and [GBIF](#), we ensure global visibility, interoperability, and suitability for a wide range of stakeholders.




These pipelines support the production of a broad suite of indicators, including species occupancy trends, invasive-species impact assessments, phylogenetic diversity metrics, trait-based invasibility estimates, and compositional dissimilarity/turnover measures across space and time. Together, they strengthen the evidence base for policy frameworks such as the EU Green Deal, the Habitats and Birds Directives, and Kunming-Montreal Global Biodiversity Targets under the Sustainable Development Goals.

To demonstrate how this approach can be applied in practice, four case studies in the B-Cubed project explore different ways these data pipelines can be used to generate policy-relevant insights. The case studies include real-time identification of invasive species risks, automated reporting under the Habitats Directive, and assessments of biodiversity trends in protected areas. While these applications will be under development until August 2026, they illustrate a future in which biodiversity data are not only richer but far more actionable for those making decisions on the ground.

Standardising biodiversity data

With the new [GBIF service](#) biodiversity occurrence data can be downloaded as “data cubes”. Occurrence data cubes are standardised structures, often stored as flat tables, that encode biodiversity records as multidimensional data (typically species × location × time) with occurrence measures (e.g., counts). This supports consistent indicator reporting and efficient assessment of biodiversity change for decision-making. For example, a cube can include occurrences of *Bombus humilis* in Belgium between 2012–2022, enabling rapid checks of trends and data gaps.

Cubes harmonise observations collected at different scales and resolutions, account for uncertainty, and retain much more information than traditional coarse-grid methods. This makes it easy to:




-  identify patterns across species, locations, and time;
-  convert highly diverse, uneven, and spatially uncertain species occurrence records into a coherent and lighter format suitable for analysis and modelling;
-  integrate occurrence data with environmental and landscape data with the same scale and resolutions.



GBIF service for data cubes

The new “cube” download format allows users to aggregate occurrences customised by the users themselves. For example, the total number of plant occurrences for all countries in Europe from preserved specimens can now be calculated with one query. This query will produce a small downloadable CSV file, which, previously, would have involved a much larger download with additional processing steps to derive those values.

Users can define:

-  the taxonomic level at which they want to aggregate data from kingdom down to species;
-  the temporal level at which they want to aggregate data, either by year, year and month or year, month and day;
-  a spatial reference grid to which they want to aggregate their data. Currently the European Economic Area reference grid, extended quarter degree grid, ISEA3H hexagonal grid, military grid reference system or by country/area.

By using pre-defined and widely-adopted spatial reference grids, download formats are then interoperable with spatial data from other sources used for monitoring changes in biodiversity. Users can also automate the process of dealing with common data quality issues when creating their cubes, including automated coordinate uncertainty management and exclusion of occurrences with common geospatial data quality issues, e.g. occurrences at country centroids.

Learn how to generate species occurrence cubes by following the [tutorials](#).

Turning data into actionable insights

Using biodiversity data cubes, indicators can be derived that distil complex multidimensional data into meaningful measures of biodiversity status and trends. The generated indicators range from species occupancy trends and biological invasion assessments to advanced metrics like phylogenetic diversity loss, which quantifies potential for loss of evolutionary history if endangered species are lost from an area. Automated indicators calculation enables repeatable, up-to-date monitoring, which is critical for dynamic threats like invasive species or dramatic climatic events. It also means that indicators can be recalculated immediately as new data are ingested.

Within B-Cubed, modelled cubes expand indicator workflows from description to inference. Rather than only summarising where and when biodiversity records occur, they help explain ecological patterns, compare communities, and anticipate emerging risks. Together, these cubes support a more policy-relevant evidence base by linking observed biodiversity change to environmental context, community reorganisation, and invasion processes. This makes it possible to move from retrospective reporting towards indicators that also support scenario analysis, early warning, and management planning.

While automated data pipelines and standardised biodiversity indicators greatly enhance our capacity to monitor and report on biodiversity, we recognise that these outputs are only as reliable as the underlying data. Biodiversity datasets often contain biases, gaps, and uncertainties arising from uneven sampling efforts, methodological differences, and taxonomic or geographic imbalances. Interpreting these indicators therefore requires close collaboration between data scientists, experts in the biology and ecology of the organisms concerned, and specialists familiar with data collection methods. Such interdisciplinary engagement is essential to ensure that indicators are interpreted correctly and translated into effective, evidence-based actions for policy and management.

B3verse: A collection of R packages

The **b3verse** enables the streamlined production of biodiversity indicators based on occurrence cubes. It is an open-source software ecosystem containing a suite of interoperable R packages. They provide a modular workflow – from data access and exploration to indicator calculation and uncertainty estimation.

Although these workflows are mainly operated by scientific teams and data specialists, their results can be delivered to policymakers through reports, dashboards, and policy-relevant indicators. The workflows are modular and well-documented to facilitate technical setup. However, advanced analyses may require access to higher computing resources or data management expertise.

Users can derive cubes using the online web GBIF interface or the **rgbif** package. They can also simulate¹ cubes using the **gcube** package.

Cubes can then be processed with the **b3gbi** package. This ensures the data for the indicator packages is consistent and properly formatted for all analyses. The **dubicube** package helps explore biodiversity data by checking its quality and reliability. It also provides

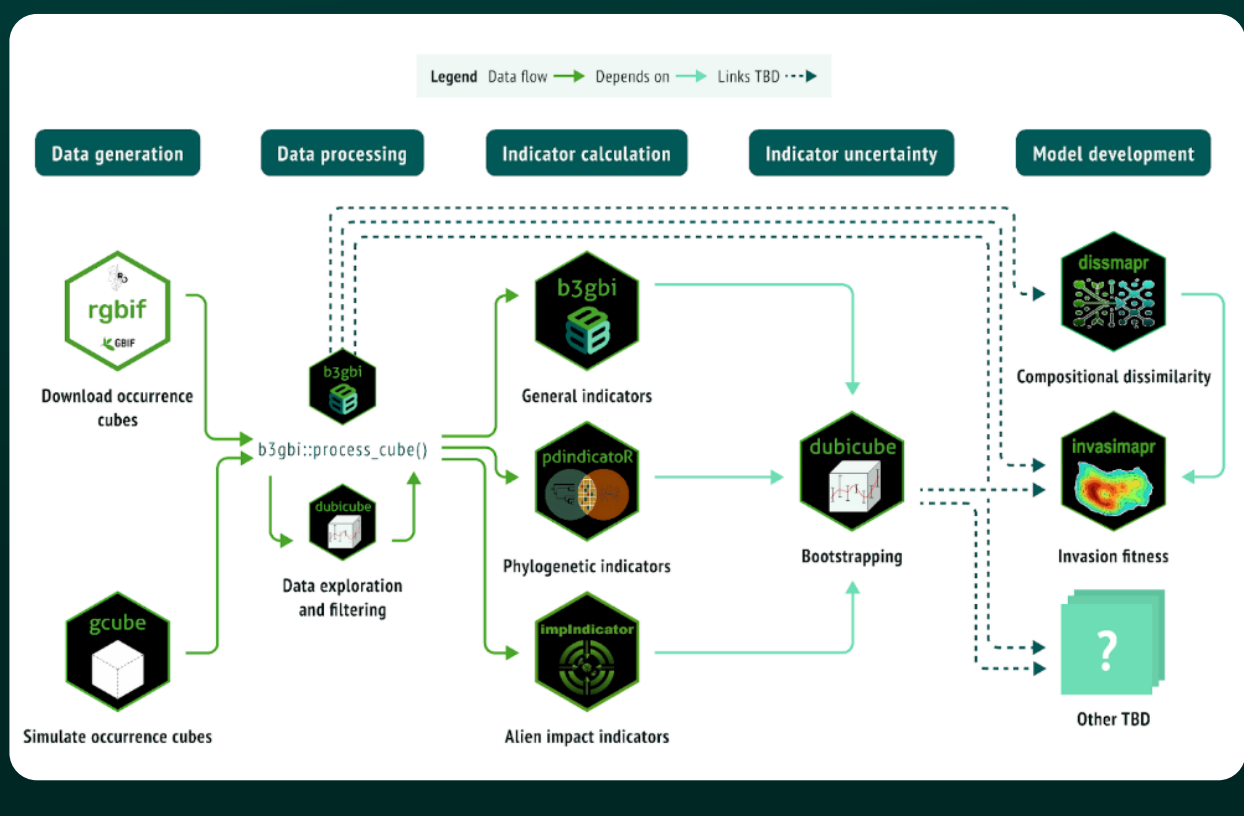
tools to calculate and visualise uncertainty in biodiversity indicators, making it easier to interpret the results.

Once the data cubes are processed, a wide range of indicators can be calculated, including those related to general biodiversity trends (**b3gbi**), phylogenetic diversity (**pdindicatorR**), and the impact of invasive species (**impIndicator**). The **b3verse** also supports modelled cube workflows that extend analysis beyond core occurrence cubes, including applications for environmental suitability, community turnover, and invasion risk. Uncertainty calculation and indicator interpretation are facilitated by the **dubicube** package.

Additional packages such as **ebvcube** and **trias** extend functionality toward integration with global data infrastructures and specific policy-relevant domains.

This is a practical user guide that explains the packages, shows how to install them, provides instructions for contributing to their development, and includes an example workflow for calculating indicators using several packages from the b3verse.

¹ Simulation studies help understand complex ecosystems by mimicking real-world scenarios and testing how different factors—like species distribution, clustering, and prevalence—affect biodiversity outcomes.



Modelled data cubes

Beyond standardised occurrence cubes, B-Cubed defines **modelled data cubes** as harmonised, multidimensional products that integrate species occurrences with environmental and ecological context to support modelling, indicator development, and decision use. The **Suitability Cube** links occurrences to environmental space to generate habitat-suitability diagnostics (including measures of environmental distance and applicability), improving interpretation of distribution modelling under current and future conditions. The **Dissimilarity Cube** operates at the community level, quantifying compositional turnover and reorganisation across space (and scenarios) to provide consistent measures of biodiversity contrast.

The **Invasibility Cube** focuses on biological invasions by combining occurrences with invasive-alien-species information and traits to estimate invasion fitness and derive indicators of species invasiveness and site/community invasibility. Built with open workflows, these cubes are modular and transferable across regions and taxa, supporting scalable, reproducible analyses aligned with FAIR principles.

These modelled cubes are implemented through dedicated R packages with guided workflows available online. **dissmapr** provides the end-to-end workflow for generating and analysing **Dissimilarity Cubes**, including compositional turnover modelling outputs intended for mapping and comparison across space and time. **invasimapr** implements the **Invasibility Cube** workflow, providing a reproducible pipeline for trait-informed invasion fitness estimation and the resulting invasiveness/invasibility indicators, supported by worked, step-by-step documentation.

Integrating occurrence data with geospatial data

Standardised biodiversity data cubes and the indicators derived from them provide critical insights, but they remain limited without a spatial context. Remote sensing can show us where and when landscapes and ecosystems are changing, while biodiversity records tell us which species are affected and how. By integrating these sources, we can move beyond fragmented and outdated datasets and instead generate timely, spatially explicit evidence that supports biodiversity targets under the EU Green Deal, SDGs, and global frameworks.

This integration is possible through modelled data cubes and tools like B3verse, deployed by B-Cubed, and is complemented with the Essential Biodiversity Variables (EBVs) Data Portal¹ to ensure that biodiversity indicators are not only scientifically robust but also accessible and interoperable across disciplines. The portal is grounded in the EBVs Framework proposed by GEO BON² and serves as a platform to share harmonised biodiversity datasets using a hierarchical system that organises geospatial data on species and ecosystems into the so-called EBVCube format³. Each published dataset is assigned a unique DOI and corresponding version number. Transparency and reproducibility are further strengthened through rich, embedded metadata compliant with metadata standards⁴. Together, these features support the harmonisation of biodiversity data and ensure accessibility, interoperability, and alignment with FAIR principles (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable).

The Portal provides a user-friendly interface to search, explore, visualise, and download biodiversity datasets, while authentication is only required for publishing data. Available resources span a wide range of spatial and temporal scales, from local to global, and include aggregated species occurrence data, single snap shots in time, and model outputs under future climate or socio-economic scenarios. The `ebvcube` R package complements the portal by enabling the full creation of EBVCube datasets – including metadata embedding and data structuring – while the portal itself supports metadata review and publication.

To enhance findability, accessibility, and interoperability, the EBV Data Portal integrates features such as DOI assignment, dataset versioning, Creative Commons licensing, and multiple download formats (netCDF, JSON-ACDD, XML-EML). It also provides programmatic access through a documented JSON-based API with integrated Swagger UI, allowing users to retrieve datasets and associated taxonomic or habitat information directly. Together, the portal and the R package streamline FAIR-compliant workflows and strengthen the open dissemination of biodiversity information.

1 Hosted at the German Centre for Integrative Biodiversity Research (iDiv) and a pivotal part of the GEO BON development tools.

2 <https://geobon.org/ebvs/what-are-ebvs/>

3 NetCDF-based format

4 Climate and Forecast Metadata Convention, Attribute Convention for Data Discovery, and Essential Biodiversity Variables framework

Calculating regional indicators in Europe using the B-Cubed workflows

B-Cubed is applying its workflows across four case studies that span diverse environments, regions, and challenges to test the limits of its tools. One of these studies focuses on Flanders, a European region with exceptionally rich species occurrence data. The work assesses a suite of biodiversity indicators, including general metrics such as cumulative and relative species richness, as well as phylogenetic and invasive species indicators. The objective is to produce an interoperable, publicly accessible dashboard.

Although the case study is ongoing, the example below illustrates the type of insights being generated. An indicator of species richness was created using the `b3gbi` with an angiosperm data cube for Belgium. This variable measures the total number of species recorded within a spatial unit in Flanders. The time series illustrates trends in the rate of invasive species establishment in Flanders since 1970. This is the headline indicator for Target 6 of the Kunming–Montreal Global Biodiversity Framework, which aims to reduce the introduction of invasive alien species by 50% by 2030.

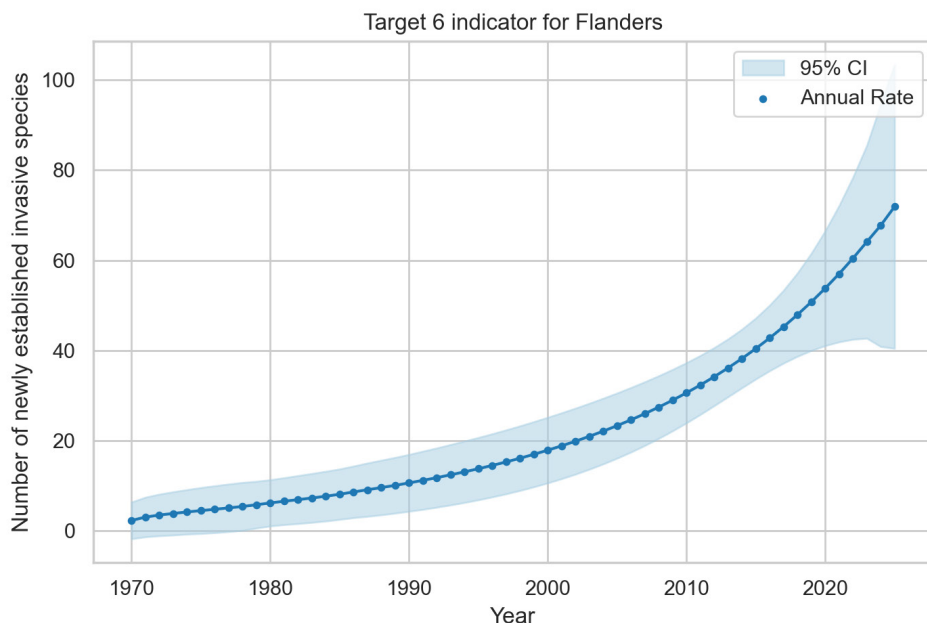


Figure 1: Target 6 CBD indicator for Flanders based on the package `b3alien` and the GRIIS checklist for Belgium

Doi: <https://doi.org/10.15468/xoidmd>, also uploaded on GBIF as a checklist: <https://www.gbif.org/dataset/6d9e952f-948c-4483-9807-575348147c7e>

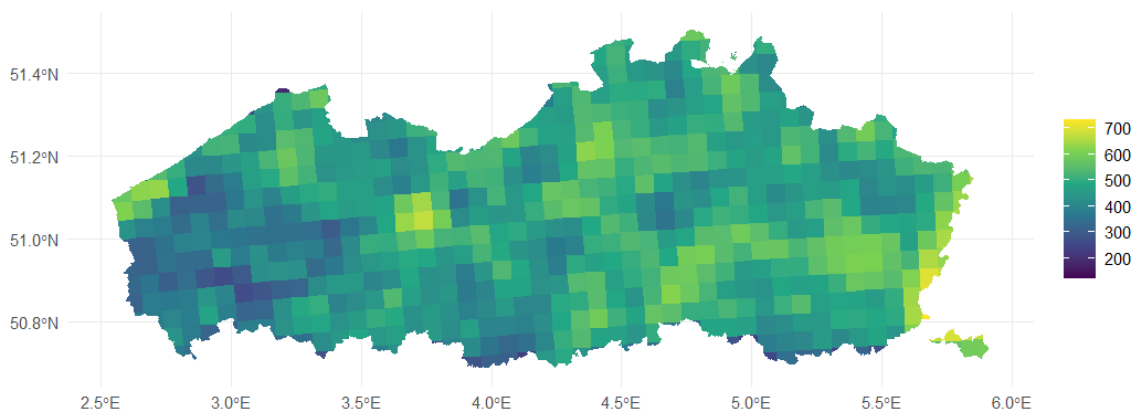


Figure 2: Map with total number of observed angiosperm species in Flanders (Belgium) at 5-km grid level since 1975. Reference of data: GBIF Occurrence Download <https://doi.org/10.15468/dLy9hpfa> Accessed from R via `rgbif` (<https://github.com/ropensci/rgbif>) on 2026-03-17

Conclusion

Automated, interoperable data pipelines provide a practical route from fragmented biodiversity observations to timely, policy-relevant indicators. By combining GBIF's occurrence data cube service with open, FAIR-aligned tooling in the **b3verse**, B-Cubed demonstrates how standardised workflows can deliver repeatable metrics on biodiversity status and trends, including general and phylogenetic indicators, invasive-species impacts, and uncertainty estimates. The addition of modelled cubes further extends these capabilities from documenting change to interpreting drivers and anticipating risks, supporting both monitoring and forward-looking assessment.

The value of these pipelines is not primarily technical; it lies in shortening the time between data collection and decision-ready evidence. When indicators can be recalculated as new records arrive, reporting can better reflect current conditions, early-warning signals can be detected sooner, and interventions can be evaluated more consistently over time. At the same time, robust use requires transparency about data limitations and close collaboration between data scientists, domain experts, and practitioners to ensure indicators are interpreted appropriately.

Together, the infrastructures and workflows described here establish a scalable foundation for operational biodiversity intelligence: standardised data products, reproducible indicator calculations, and mechanisms for publication and reuse through platforms such as the EBV Data Portal. Continued investment in interoperability, documentation, and community uptake will be essential to translate this foundation into routine, trusted decision support across regions, taxa, and policy contexts.





BIODIVERSITY
BUILDING
BLOCKS FOR
POLICY

PROJECT COORDINATOR

Dr Quentin Groom
Meise Botanic Garden
quentin.groom@plantentuinmeise.be

FUNDING SCHEME

HORIZON_HORIZON-CL6-2021-
GOVERNANCE-01-17

DURATION

1 March 2023 – 31 August 2026

BUDGET














€ 4 853 333,75

WEBSITE

b-cubed.eu



Consortium

-  Meise Botanic Garden
-  Global Biodiversity Information Facility
-  Research Institute for Nature and Forest
-  University of Bologna
-  Justus-Liebig-University Giessen
-  Ovidius University of Constanța
-  South African National Biodiversity Institute
-  Stellenbosch University
-  Pensoft Publishers
-  German Center for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
-  French Institute for Research in Computer Science and Automation
-  University of Aveiro
-  University of L'Aquila



**Meise
Botanic Garden**



GBIF



**RESEARCH INSTITUTE
NATURE AND FOREST**



SANBI
South African National Biodiversity Institute
Biodiversity for Life



**Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT**



PENSOFT



iDiv

German Centre for Integrative
Biodiversity Research (iDiv)
Halle-Jena-Leipzig



Inria



**universidade
de aveiro**



**UNIVERSITÀ
DEGLI STUDI
DELL'AQUILA**