# BIODIVERSITY BUILDING BLOCKS FOR POLICY

# D1.3 Data Management Plan

30/08/2023

Authors: Nikol Yovcheva, Teodor Metodiev, Pavel Stoev, Flavia Roteda Ruffino, Francisco Javier Castro

**Prepared under contract from the European Commission**
Grant agreement No. 101059592
EU Horizon Europe Research and Innovation Action

| | |
|---|---|
| Project acronym: | **B3** |
| Project full title: | **Biodiversity Building Blocks for policy** |
| Project duration: | 01.03.2023 – 31.08.2026 (42 months) |
| Project coordinator: | Dr. Quentin Groom, Agentschap Plantentuin Meise (MeiseBG) |
| Call: | HORIZON-CL6-2021-GOVERNANCE-01 |
| Deliverable title: | Data Management Plan |
| Deliverable n°: | D1.3 |
| WP responsible: | WP1 |
| Nature of the deliverable: | DMP – Data Management Plan |
| Dissemination level: | Public |
| Lead partner: | PENSOFT |
| Recommended citation: | Yovcheva, N. et al. (2023). *Data Management Plan*. B3 project deliverable D1.3 |
| Due date of deliverable: | Month 6 |
| Actual submission date: | Month 6 |

Deliverable status:

| Version | Status | Date | Author(s) |
|---|---|---|---|
| 1.0 | Draft | 13 July 2023 | Nikol Yovcheva, Teodor Metodiev & Pavel Stoev, Pensoft Publishers |
| 1.1 | Ethics review | 21 July 2023 | Flavia Roteda Ruffino & Francisco Javier Castro, Plus Ethics |
| 1.2 | Internal review | 27 July 2023 | Quentin Groom, Agentschap Plantentuin Meise<br>Miguel Fernandez & Lina Estupinan Suarez, Martin-Luther-Universität Halle-Wittenberg |
| 1.3 | Partners review | 23 August 2023 | Representatives of all consortium organisations |
| 1.4 | Final v1 | 30 August 2023 | Nikol Yovcheva, Teodor Metodiev & Pavel Stoev, Pensoft Publishers |

# Table of contents

## Key takeaway messages

- Deliverable D1.3 Data Management Plan (DMP) provides an overview of the data and other research outputs B3 anticipates generating and reusing, as well as how these project resources will be managed in an open, effective and FAIR manner.
- The DMP elaborates on how data generation and reuse contribute to B3's objectives and describes the formats, licences, standards and data repositories to be utilised.
- D1.3 provides a set of data management recommendations for all B3 partners when collecting, reusing or generating research data or other research outputs in the project in order to uphold the FAIR data principles, while considering the ethical aspects of data management, such as the use of artificial intelligence.

## Executive summary

To ensure the findability, accessibility, interoperability and reusability of its research data and other research outputs, B3 developed a customised project Data Management Plan (DMP) in the form of D1.3. The DMP is based on the initial open science, open data and open source principles established in the project's description of action, as well as on insight collected from a consortium-wide consultation process. It starts with a concise overview (Chapter 2) of the data B3 expects to generate (Table 1) and reuse (Table 2) – its purposes, types, formats and sizes, as well as expected timeframes and potential users. Subsequently, the DMP details how B3 partners should implement the FAIR data principles to said data, outlining specific recommendations (Chapter 3).

The document also identifies the other research outputs that B3 will generate (Table 3) and elaborates on how the FAIR data recommendations can be applied to them (Chapter 4). Lastly, the plan goes over the administrative aspects of data management, including resource allocation (Chapter 5), data security (Chapter 6), intellectual property rights (Chapter 7), and ethics considerations (Chapter 8). It is important to note that the DMP is a living document that will be updated on demand to provide more detailed information as the project progresses and significant changes occur. At the very least, one DMP update is planned before B3's end.

## Non-technical summary

B3's Data Management Plan is a living document that outlines how data will be handled, stored, generated and shared throughout the project's lifecycle. It provides a roadmap for effectively managing the project's data to ensure its integrity and long-term reusability. The DMP begins by outlining why B3 will generate and reuse data, under what formats and licences, as well as where it will make it available for interested stakeholders. The deliverable then includes a set of recommendations on how B3 partners should handle the aforementioned data in order to ensure users can find it, access it, integrate it with their existing data and reuse it. Furthermore, the DMP illustrates how these data management recommendations can be applied to other research outputs, such as policy briefs and training materials. Lastly, the document goes over the administrative aspects of data management, such as resource allocation, data security, intellectual property rights and ethics. As a living document, the DMP will be updated on-demand whenever significant changes occur to B3's data, with at least one mandatory update before the end of the project.

## List of abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| APC | Article Processing Charges |
| API | Application Programming Interfaces |
| DMP | Data Management Plan |
| DOI | Digital Object Identifier |
| DPO | Data Protection Officer |
| EBV | Essential Biodiversity Variable |
| EML | Ecological Metadata Language |
| EU | European Union |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| GBIF | Global Biodiversity Information Facility |
| GDPR | General Data Protection Regulation |
| ORE | Open Research Europe |
| WP | Work Package |

# 1. Introduction

As a project focusing on providing standardised and fast access to biodiversity data, B3 embodies the principles of open science and findable, accessible, interoperable and reusable (FAIR) data. These pillars encourage knowledge sharing, reproducibility, and the potential for new discoveries through data-driven collaborations. In light of that, the Horizon Europe project B3 promotes the transparent and collaborative nature of scientific research by making research data openly available to the scientific community and society at large.

To put these principles into tangible action points, B3's work package (WP) 1 is responsible for developing D1.3 Data Management Plan (DMP) in month 6 of the project. The DMP covers B3's data collection, processing, handling during the project and preservation afterwards to ensure that research data can easily be discovered, accessed and reused by both humans and machines. It details the data types within B3 and elaborates on their preferred licences, methodologies, standards and metadata specifications. In addition, the DMP contains a set of recommendations on how to apply the FAIR data and open science principles to B3's research data and outputs. To provide additional practical support to project members, the DMP will be accompanied by an additional concise document with Data Management Guidelines containing to-the-point steps on how to manage B3's data, scientific publications and other research outputs in a FAIR and efficient manner.

The DMP was developed by building on the initial open science, open data and open source principles established in the project's description of action, as well as using insights collected from a consortium-wide consultation process. A detailed data management survey containing 15 questions (available in Annex 1) was distributed to B3's partners. It gathered information on the findability, accessibility, interoperability and reusability of the project's research data and other research outputs. Project members contributed to the development of the DMP by submitting a comprehensive answer to the questionnaire, identifying the datasets and other research outputs that they will generate and/or reuse and elaborating on their institutional data management practices. The survey was filled in by the 11 project organisations that will store, generate or reuse research data in relation to B3's work: MeiseBG, GBIF, EV INBO, UNIBO, JLU, SANBI, SUN, MLU, INRIA, UAVR and LTU. Moreover, the DMP was consulted by representatives from all project organisations prior to its submission.

# 2. Data summary

B3 will collect, generate and reuse data in support of the following project objectives:
- To enable a rapid, regular, sustainable product by simplifying, automating and accelerating the pipeline from data collection to biodiversity indicators.
- To improve the standardisation and comparability of biodiversity metrics globally by increasing the openness and FAIR provision of biodiversity data and its derivatives.
- To derive the biodiversity indicators that policy needs, making it easy for any stakeholder to use, access and understand, as well as enabling policymakers to commission new indicator workflows from 3rd parties to support their local needs.
- To improve data scientists' capacity to access model ready datasets for biodiversity and the environment that can be incorporated into manual, semi-automated and automated workflows for modelling of past, present and future biodiversity.

- To provide free and open source algorithms in a cloud computing environment for data exploitation and analysis, including machine learning, guaranteeing high robustness and reproducibility, while lowering the technical thresholds, efforts and costs.
- To provide functioning examples of workflows that can act as templates and building blocks for others to use and adapt to their own needs.

B3's work will rely on a combination of reused existing data and self-generated data, such as aggregated biodiversity observation data and models of biodiversity occurrence data. Serving as a preliminary overview, Table 1 presents a summary of the datasets that the project expects to generate, whereas Table 2 outlines the specific information about existing data intended for reuse in the project. Overall, new datasets will be generated through data analysis, aggregation, transformation and modelling, as well as by utilising B3's workflows. For its data production, the project will mainly reuse data mediated through GBIF.

Besides the data produced directly from the B3 project itself, a number of automated workflows and software packages will also be produced as part of the B3 products. These will enable users outside the B3 consortium to generate datasets tailored to their specific needs, such as different temporal coverage, resolution, target taxa, and indicators. These tailored datasets are also intended to be automatically minted with DOIs and deposited at selected online repositories with standard format and metadata. More information on the external dataset production pipeline using B3 workflows and software packages will be made available in the planned DMP update before the end of the project.

The vast majority of B3's generated datasets will contain quantitative data, with most of the newly-generated datasets being under 1 GB, except for the Species occurrence cubes in WP2, which are expected to be a couple of MBs each, amounting to a total of 10s-100 GBs when accumulated. More information on B3's Species occurrence cubes' format, parameterisation, hosting, underlying algorithms, as well as the metadata necessary to ensure FAIR principles can be found in *D2.1 Software specifications* (Desmet, Oldoni, Blissett, Robertson, 2023).

As of month 6, all of the datasets included in Table 1 are expected to be open access, with one dataset shared in an anonymised form in order to protect personal data. Most commonly, datasets will be made available under the CC BY 4.0 and the CC0 1.0 licence. Where relevant, documentation will be provided to facilitate data reuse.

B3's data will contribute to the development of the project, the specific needs of its work packages and the following stakeholder groups (information on the groups is available in *D1.2 Plan for Exploitation, Dissemination and Communication [Yovcheva, Metodiev, 2023]*):
- Scientific community
- International organisations
- Policymakers
- Relevant authorities
- General public

As B3 is still in its early stages, not all the information is currently available. Therefore, the DMP is considered a living document that will be updated on demand to accurately represent the development of B3's research data.

**Table 1: Summary of the data B3 anticipates generating\***

| No | Name of the dataset | Contact | Task/ WP | Generated via | Size | Format | Type of data | Sensitive/ Personal data | Delivery date | Type of access | Licence | Users | Point of access | Reuse support |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Species occurrence cubes | GBIF | T2.2, T2.3 | GBIF | 10s-100 GBs | CSV, EBV, NetCDF, Apache Parquet | Quantitative | No | Starting from M12 | Open | CC BY | Scientific community, International organisations | GBIF, EBV Data Portal, Zenodo | Technical document, Training material |
| 2 | Table of trends in the status of biodiversity | JLU | T5.1 | Data transformation | 10 MB | CSV | Quantitative | No | M34 | Open | CC0 | Scientific community, Policymakers | Zenodo | Not applicable |
| 3 | Species occurrence cube derived from structured monitoring schemes | INBO | T4.5 | Data aggregation, B3 workflow | <1 Gb | CSV, EBV NetCDF | Quantitative | No | M34 | Open | CC0 | Scientific community | GBIF, EBV Data Portal | Report from T4.5 |
| 4 | Species occurrence cube derived from opportunistic monitoring | INBO | T4.5 | Data aggregation, B3 workflow | <1 Gb | CSV, EBV NetCDF | Quantitative | No | M34 | Open | CC0 | Scientific community | GBIF, EBV Data Portal | Report from T4.5 |
| 5 | WP6 dataset (name TBD) | UAVR | T6.4 | Data aggregation, Modelling | TBD | TBD | Quantitative | No | M41 | Open | TBD | Scientific community, Policymakers, National authorities | TBD | TBD |
| 6 | Help desk reporting | MeiseBG | T1.4 | GitHub, Questionnaires | 1 MB | CSV | Quantitative | Name, User name, Email, Country | M41 | Open (Anonymised) | CC BY | B3 partners, Reporting | B3 website, Zenodo | Not applicable |

| 7 | Phylogenetic indicator | MeiseBG | T5.2 | Data aggregation and analysis | 100 MB | CSV | Quantitative | No | M22 | Open | CC0 | Policymakers | GBIF | Not applicable |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Dissimilarity Cube | SUN | WP4 | Data aggregation and analysis, Coding | 500 MB | GeoTIFF, R script | Quantitative | No | M21 | Open | CC BY, MIT License | Scientific community, Policymakers, General public | EBV Data Portal, GitHub, Zenodo, CRAN | R Markdown, Quarto Document |
| 9 | Invasibility Cube | SUN | WP4 | Data aggregation and analysis, Coding | 500 MB | GeoTIFF, R script | Quantitative | No | M22 | Open | CC BY, MIT License | Scientific community, Policymakers, General public, Relevant authorities | EBV Data Portal, GitHub, Zenodo, CRAN | R Markdown, Quarto Document |
| 10 | EU-stakeholders needs assessment | MLU | T1.5 | Exploratory analysis | <1MB | CSV | Qualitative/Quantitative | No | M12 | Open | CC BY | B3 partners | RIO, Zenodo | Not applicable |
| 11 | Suitability Cube | UNIBO | WP4 | Data aggregation and analysis, Coding | 500 MB | GeoTIFF, R script | Quantitative | No | M22 | Open | CC BY, MIT License | Scientific community, Policymakers, Relevant authorities | EBV Data Portal, GitHub, Zenodo | R Markdown |
| 12 | B3 Virtual Species | UNIBO | WP4 | Data transformation and analysis, Coding | 500 MB | GeoTIFF, NetCDF, Shapefile, R script | Quantitative | No | M22 | Open | CC BY, MIT License | Scientific community, Policymakers, Relevant authorities | GitHub, Zenodo | R GRASS GIS QGIS |
| 13 | Checklist of Alien Species in South Africa | SANBI | T6.2 | Data aggregation | 5 MB | CSV | Qualitative/Quantitative | No | M41 | Open | CC BY | Policymakers, General public | Zenodo | Not applicable |

*\* The information provided in Table 1 and Table 2 is based on an initial data mapping and is subject to change in order to reflect the project's development and arising data needs. Updated information, as well as adjustments to the current information, will be provided on an as-needed basis and at the latest during the planned DMP update before the end of the project.*

**Table 2: Summary of the data B3 expects to reuse\***

| No | Name of the dataset | Contact | Task/WP | Origin | Size | Format | Type of access | Licence |
|----|---------------------|---------|---------|--------|------|--------|----------------|---------|
| 1 | GBIF species occurrence data | GBIF | Several B3 tasks | GBIF | 0.1 - 100 GB | CSV, Darwin Core Archive | Open | CC0, CC BY, CC BY-NC |
| 2 | Alien species regional occurrences | JLU | T5.1 | Scientific articles, Reports, Unpublished material | 0.1 - 100GB | CSV | Open | TBD |
| 3 | ABV - Common breeding birds in Flanders, Belgium | INBO, UNIBO | T4.5, T4.1 | ABV breeding birds dataset (INBO) | 3 MB | Darwin Core Archive | Open (Aggregated on 1 km) | CC0 |
| 4 | Global Register of Introduced and Invasive Species (GRIIS) - Belgium | INBO | T6.3 | GBIF | 1 MB | Darwin Core Archive | Open | CC BY |
| 5 | GBIF species occurrence data for select invasive species from GRIIS checklist | INBO | T6.3 | GBIF | < 1Gb | CSV | Open | CC0, CC BY, CC BY-NC |
| 6 | Sentinel-2 (Copernicus) and Landsat | INRIA | T4.4 | Sentinel-2 (Copernicus) and Landsat | TBD | TBD | Open | TBD |

| 7 | Chelsa climate | INRIA | T4.4 | Chelsa | TBD | TBD | Open | CC BY |
|---|---|---|---|---|---|---|---|---|
| 8 | Phylogenetic trees | MeiseBG | T5.2 | Research publications | 20 MB | Newick | Open | CC0 |

*\* The information provided in Table 1 and Table 2 is based on an initial data mapping and is subject to change in order to reflect the project's development and arising data needs. Updated information, as well as adjustments to the current information, will be provided on an as-needed basis and at the latest during the planned DMP update before the end of the project.*

# 3. FAIR data

The ever-increasing volume and complexity of data generated by research projects create a need for a robust and standardised approach to data management. Recognising this important need, B3's DMP is designed to maximise the value of research data by making it findable, easily accessible, interoperable across different systems, and reusable by both internal and external stakeholders, as well as by humans and machines.

This chapter elaborates on how B3 will implement the four FAIR data pillars which are separable yet function synergistically to empower stakeholders to reuse research data. Within B3's DMP, research data encompasses all the data utilised throughout the project, incorporating primary (raw/input data) and processed data (output data) underlying project outcomes.

## 3.1. Making data findable

The first step in enabling stakeholders to reuse B3's data is making sure it is easily findable. To that end, the project will assign its finalised datasets with globally unique and persistent identifiers (e.g., Digital Object Identifiers). These play a pivotal role in achieving FAIR data by eliminating ambiguity, enabling meaningful interpretation by both humans and machines, and facilitating proper citation and reuse of data.

Furthermore, B3 will ensure its datasets are described with rich machine-readable metadata enabling automated systems to handle laborious and repetitive sorting and prioritising tasks that otherwise require significant involvement from researchers. The project will publish its datasets with metadata that clearly and explicitly includes the identifier of the data, offers descriptive and structural information and provides a rich representation of the dataset. This approach would enable stakeholders to locate data based on the information contained in the metadata itself, eliminating the requirement to have prior knowledge of the data's specific identifier.

When structuring its metadata, B3 will adhere to standardised formats that promote the discoverability, accessibility, interoperability, and reusability of the data. It is important to choose a standard that best aligns with the specific type of data to ensure compatibility with other datasets in the same field and facilitate its discovery. B3 has identified a number of suitable standards that can be used based on each partner's preference and specific data needs.

---

### Recommendation #1

B3 members are encouraged to employ one of the following metadata standards:
- Ecological Metadata Language (EML): includes modules for the spatial, temporal, taxonomic, and thematic extent of data, as well as research methods and protocols;
- DataCite Metadata: provides a standardised framework for data across disciplines;
- Darwin Core: a body of standards facilitating the sharing of biodiversity information;
- EBVs NetCDF format: following Quoss et al. (2023 *in prep.*), guidelines and standards for uploading data to the EBV Data portal building on the NetCDF Climate and Forecast (CF) Metadata Conventions in order to expand the climate standard to allow defining biodiversity related aspects.

---

**Recommendation #2**

If no specific metadata standard is applied, B3's datasets should be described by:
**Necessary**:
- Creator(s)
- Creation date
- Title
- Global persistent identifier
- Licence of use
- Embargo period
- Publisher
- Publication date
- Dataset description
- Input data/parameters used to generate the dataset
- Ontology-aligned keywords and meta-tags
- Horizon Europe funding: grant project name (Biodiversity Building Blocks for policy), acronym (B3) and number (101059592)

**Highly recommended**:
- Version or subset
- Dataset language
- Metadata language
- Date of metadata creation
- Character encoding
- Format version

Since identifiers and rich metadata alone do not guarantee the findability of digital research data on the internet, B3 will ensure its (meta)data is registered, indexed and can be harvested in a searchable resource. It will do so by uploading its finalised datasets to trusted repositories that offer search engines and indexing, such as Zenodo, as well as by making its data accessible through automated workflows on the GBIF Data Portal and EBV Data Portal. More information will be available in *D3.3 FAIR guidelines* which will outline how to ensure the FAIRness and interoperability of B3's results with the GEO BON Portal on Essential Biodiversity Variables.

**Additional information #1**

The GBIF Data Portal acts as a single point of entry to millions of data records. It offers browsing capabilities, enabling users to search for data by different parameters, such as occurrences, species, datasets and publishers. After performing their search, users have the option to download the data.

B3 members can contact GBIF partners for further questions on the portal's use.

## Additional information #2

The EBV Data Portal is an online platform that focuses on providing standardised access to essential biodiversity data. It offers: (i) search capabilities via a catalogue layout as the entrance point to find EBV datasets based on EBV metadata specifications; (ii) analytic capabilities to facilitate the reporting on EBVs and indicators by countries and regions; (iii) open machine-readable access via a REST JSON based API developed for the integration, sharing and use of the EBV datasets.

B3 members can contact MLU partners for further questions on the portal's use.

Lastly, B3 aims to facilitate the exchange of information among consortium members, which is why it will strive to adhere to a particular naming system for its files. Utilising a standardised and user-friendly approach to naming documents enhances effective collaboration among partners and simplifies the process of locating project files when needed.

## Recommendation #3

B3's files should adhere to the following unified naming convention: [B3_file-name_version_creation-date], whereby data format should be YYYYMMDD, number style version should be 01, 02, 03. Example: B3_ phylogenetic indicator_v01_20250206.csv

## Recommendation #4

When choosing a title for their EBV datasets, B3 members should consider the following:
- Be **clear** and **specific** in conveying the focus and content of your dataset.
- Strive for **brevity** while maintaining informative and meaningful titles.
- Use **descriptive keywords** highlighting important elements/variables in the dataset.
- Ensure **abbreviations are consistent** throughout the title and in the extended EBV dataset description.
- Tailor your title for the **target audience**, consider the intended users and adapt the title accordingly.
- **Avoid ambiguity**:
  - Only include specific methods, models, metrics, or indicators if it helps disambiguation.
  - Only include spatial and temporal information for context if it helps disambiguation.

Note that the title should serve as a teaser for your dataset and not replace the full metadata. A well-crafted EBV dataset title should provide enough information to captivate the user's interest and encourage further exploration of the details!

To ensure users have a centralised overview of B3's data, towards its end, the project will carefully check that all its datasets are linked in its OpenAire Explore page. If needed, additional measures will be considered, such as creating a simple CSV-file project catalogue, similar to Table 1, including the name, type, DOI and authors of the dataset.

## 3.2. Making data accessible

Once stakeholders are enabled to easily find data, they must also be informed about the regulations and processes related to its access. Accessibility in data management refers to ensuring that data and associated resources are readily available and easily obtainable to both humans and machines. This involves reducing barriers and restrictions that may impede access by depositing data in trusted repositories, ensuring that (meta)data is retrievable by its identifier using a standardised communications protocol and guaranteeing that metadata is accessible, even when the data itself is no longer available.

As a Horizon Europe project committed to promoting open science principles, B3 will ensure that its research results are openly accessible to the public. This will be accomplished by depositing the project's results in trusted repositories, as well as through the GBIF Data Portal and the EBV Data Portal, making them freely available for anyone to access and utilise.

---

### Additional information #3

Trusted repositories are either certified repositories (e.g., CoreTrustSeal) and/or disciplinary/domain repositories that are commonly used/endorsed by research communities (e.g., ELIXIR deposition databases). General-purpose (e.g., Zenodo) and institutional repositories could also be acceptable if they share these properties:

> Mechanisms to ensure integrity and authenticity of contents;
> Clear information about their policies/services;
> Broad, ideally open access to content (consistent with legal and ethical constraints);
> Assigned PIDs;
> Detailed, machine-readable and standardised metadata;
> Mid- and long-term preservation of contents, expert curation, quality assurance;
> National and/or international security criteria.

To select the most suitable repository, B3 partners can consult this Generalist Repository Comparison Chart.

---

All scientific publications by B3 will be published in gold open access journals** and immediately self-archived in trusted open access repositories, thus ensuring that they are freely available to the public. Partners will also consider using newly emerged publication venues such as Open Research Europe (ORE), where self-archiving is not needed, since manuscripts are automatically archived on a repository (Zenodo) once they successfully pass peer-review.

***It is important to keep in mind that only publication fees in full open access venues for peer-reviewed scientific publications are eligible for reimbursement. You can check the eligibility of the journal via this website.*

## Recommendation #5

In order to guarantee accessibility to scientific publications, B3 partners should:
- Deposit in a trusted repository, at the latest at the time of publication, a machine-readable electronic version of the published version or the final peer-reviewed manuscript accepted for publication;
- Provide **immediate open access via the repository**, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights;
  - for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g., CC BY-NC, CC BY-ND);
- Give information via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication;
- Include a Data Availability Statement (even when there is no associated data) with information on where data can be accessed, preferably not referring readers to contact authors in order to access the data.

B3 will also make its digital research data accessible through trusted open access repositories, as well as through the GBIF Data Portal and EBV Data Portal. Data underlying scientific publications will be made available in open access as soon as possible following the paper's publication and no later than the end of the B3 reporting period during which the paper was published, respecting any relevant embargo periods. Similarly, digital research data not directly linked to a publication will be shared as soon as possible through a trusted open access repository. In addition, partners will provide information via the repository regarding any potential research outputs or tools required for data reuse or validation.

## Recommendation #6

B3 partners should make their data open access by:
1. Uploading the data to an open access research data repository under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC0) or a licence with equivalent rights. They could use a generic repository such as Zenodo, a thematic one like GBIF or a trusted institutional repository.
   - Data underlying scientific publications has to be deposited as soon as possible after the paper has been published and no later than the end of the B3 reporting period during which the paper was published (respecting relevant embargo periods).
   - Other research data has to be deposited as soon as possible.
2. Providing information via the repository about any research output or any other tools and instruments needed to re-use or validate the data.

Partners should consider the additional step of describing datasets as open access data papers.

In terms of a general-purpose trusted repository, the majority of B3 partners expressed a preference for Zenodo as a suitable option, which they are familiar with and have previously explored appropriate arrangements regarding deposition.

---

**Recommendation #7**

When not using domain-specific repositories, B3 partners are encouraged to deposit their datasets to Zenodo as the project can verify that it covers all the requirements of FAIR data, most importantly:

- A DOI can be issued to every published record on Zenodo;
- Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing;
- Metadata of each record is sent to DataCite servers during DOI registration and indexed there;
- Metadata for individual records as well as record collections is harvestable using the standard, open, free and universal OAI-PMH protocol by the record identifier and the collection name;
- Metadata is publicly accessible and licenced under public domain. No authorisation is ever necessary to retrieve it;
- Data and metadata will be retained for the lifetime of the repository, ensuring that the metadata will be accessible, even when the data is no longer available;
- Metadata is stored in high-availability database servers at CERN, which are separate from the data itself.

---

B3 will follow the principle of data being "as open as possible as closed as necessary" and restricted access may be implemented for specific datasets, if deemed necessary. It is important to note that many datasets mediated through GBIF are done so under a "non-commercial" licence stipulation which is intent-based. Organisations sharing biodiversity data openly on GBIF comply with open dissemination of aggregated products and indicators, otherwise, their contribution to GBIF would be largely pointless, hence, as data cubes are not "...primarily intended for or directed towards commercial advantage or monetary compensation" licences can be assigned on a case-by-case basis.

Table 1 in Chapter 2 provides an overview of the datasets designated as restricted or closed (if any), along with the corresponding reasons. Partners will review the finalised datasets to assess the possibility of sharing them in aggregated and/or anonymised form to safeguard sensitive and/or personal information. In case datasets are anonymised, B3 will consider using the tool Amnesia which helps researchers anonymise their datasets in order to comply with the general data protection regulation (GDPR) requirements. Table 1 also includes the proposed embargo periods (if any) for certain datasets.

To the extent legitimate interests or constraints are safeguarded, and in line with the FAIR principles, machine-readable metadata for B3's data (including closed access datasets) will be available under a Creative Common Public Domain Dedication (CC0) or equivalent.

## 3.3. Making data interoperable

Once stakeholders have found and accessed B3's data, it is important that datasets are formatted and presented in a way which enables their interoperability with applications or workflows for analysis, storage, and processing. To that end, B3 will adhere to a set of endorsed best practices for interoperability within the community, such as adopting widely-used metadata vocabularies, standards, formats, and ontologies. In situations where less common ontologies and vocabularies have to be used, mappings will be provided to facilitate their conversion into more widely accepted ones. Additionally, if needed, partners will individually evaluate whether any newly developed ontologies or vocabularies should be openly published, enabling others to reuse, enhance, or expand upon them. This assessment will be conducted on a case-by-case basis, ensuring careful consideration of the potential benefits and implications.

### Recommendation #8

B3 partners are encouraged to follow these principles to ensure data interoperability:
- Use a formal, accessible, shared, and broadly applicable language for knowledge representation, such as the RDF extensible knowledge representation model, JSON LD, OWL, etc.;
- Use vocabularies such as Darwin Core that are documented and accessible to anyone using the dataset;
- Include qualified references to and properly cite other (meta)data by specifying if one dataset builds on another, if additional datasets are needed to complete the data, or if complementary information is stored in a different dataset.

To guarantee interoperability, B3 will utilise widely recognised file formats listed in Table 1.

### Recommendation #9

B3 members should use the following data formats:
- CSV for tabular data;
- netCDF for geospatial data and data cubes;
- GeoPackage or GeoJSON for vector data.

By employing such file formats, linking them to detailed, standardised, and machine-readable metadata (such as EML), and depositing them in trusted open-access repositories that offer programmatic access (like Zenodo), B3 ensures that its data can be accessed and retrieved by any programming language, eliminating the need for proprietary software.

Furthermore, B3 will establish mechanisms to ensure semantic, organisational and technical interoperability between its biodiversity data cubes and the EBV Data Portal. More information will be available in *D3.3 FAIR guidelines* which will contain guidelines on the FAIR and open depositing of data products to ensure that B3 data cubes are compatible with the EBV Data Portal and other outlets for data cube dissemination.

## 3.4. Increase data reuse

As previously established in Chapter 3.2, B3's data will be freely available in the public domain using standard reuse licences in order to permit the widest use by third parties possible, even after the end of the project. To enable users to easily determine if B3's data is useful for them in a particular context, project data will be published with standardised rich metadata (described in Chapter 3.1) that not only facilitates discovery but also describes the data's provenance, scope, limitations, generation process, technical details, related species, and more. Furthermore, to avoid ambiguity that could hinder data reuse due to challenges in complying with licence restrictions, B3's metadata will include licence conditions that are clear to both machines and humans. Where relevant, documentation needed to analyse, validate and reuse the data will be provided, such as project reports, information on methodology, data cleaning, analyses, variable definitions, units of measurement, etc.

**Recommendation #10**

To enhance the reusability of B3 data, project partners are encouraged to:
- Release data with a clear and accessible data usage licence;
- Describe their data not only with standardised metadata supporting discovery but also provide readme files or documentation needed for data analysis, validation and reuse with information on data provenance, scope, limitations, generation process, etc., such as the following:
  - Data origin;
  - Data provider;
  - Date of data retrieval;
  - Licence of use;
  - A URL or DOI for the original data;
  - Description of the input data, including its format, structure, and content;
  - Steps taken to clean and preprocess the data to ensure data quality and integrity.

# 4. Other research outputs

Besides ensuring FAIR data management, research projects should also consider and plan for the management of any other research outputs that they may generate or reuse. B3 will produce other outputs in the form of digital entities such as software and models, as well as documentation like policy briefs, reports, maps and training guidelines. Table 3 contains a summary of the other research outputs that the project expects to generate, as well as how they will be managed.

**Table 3: Summary of the other research outputs B3 anticipates generating\*\*\***

| No | Name of the output | Contact | Task /WP | Type of output | Delivery date | Type of access | Licence | Users | Point of access | Reuse support |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Phylogentic summaries for indicators | GBIF | T5.2 | Geographic surfaces and summaries | M26 | Open | CC BY, CC BY-NC | Scientific community, International organisations | GBIF | Technical documentation |
| 2 | Landscape analysis of stakeholder demands | JLU | T1.6 | Report | After M22 | Open | CC BY | Scientific community, Policymakers | Zenodo | No special tools required |
| 3 | Data quality assessment | INBO | T4.5 | Report | M34 | Open | CC BY | Scientific community | INBO's repository | TBD |
| 4 | Robustness assessment | INBO | T5.4 | Report | M26 | Open | CC BY | Scientific community | INBO's repository | TBD |
| 5 | Biodiversity indicator package B3 | INBO | T5.5 | R package | M36 | Open | MIT License | Scientific community | GitHub, CRAN, Zenodo | TBD |
| 6 | Case study 3 | INBO | T6.3 | Report | TBD | Open | CC BY | Scientific community | INBO's repository | TBD |

| 7 | Biodiversity indicators (TBC) | UAVR | T6.4 | TBD | M41 | Open | TBD | Scientific community, Policymakers, National authorities | TBD | TBD |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Deep Species Distribution Models | INRIA | T4.4 | Trained Deep Learning Models | TBD | Open | MIT License | Scientific community | GitHub, Zenodo | Not applicable |
| 9 | Policy Brief | MeiseBG | T1.9 | Policy Brief | M37 | Open | CC BY | Policymakers | Zenodo | Not applicable |
| 10 | SOC documentation | MeiseBG | T3.2 | SOC documentation | M37 | Open | CC BY | Scientific community | Zenodo | Not applicable |
| 11 | Report | MeiseBG | T5.2 | Report | M22 | Open | CC BY | Policymakers, Scientific community, National authorities | Zenodo and other similar open access publishing platforms | Not applicable |
| 12 | Regional Compositional Turnover Maps | SUN | WP4 | Publications | TBD | Open | CC BY | Scientific community, Policymakers, Practitioners | Zenodo, GitHub | Publications with supporting material |
| 13 | Regional Invasibility Maps | SUN | WP4 | Publications | TBD | Open | CC BY | Scientific community, Policymakers, Practitioners | Zenodo, GitHub | Publications with supporting material |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 14 | A report on the needs for European biodiversity policy related to biodiversity data cubes and monitoring, seeking synergies with EuropaBON | MLU | T1.5 | Report | M12 | Open | CC BY | B3 partners | Zotero | Will be provided if relevant |
| 15 | R scripts | UNIBO | WP4 | R Code to reproduce analyses | TBD | Open | MIT License | Scientific community | GitHub, Zenodo | R |
| 16 | The Status of Biological Invasions and their Management in South Africa in 2022 | SANBI | T6.2 | Summary report | After M42 | Open | CC BY | Policymakers, General public | Zenodo | Not applicable |

*** *The information provided in Table 3 is based on an initial data mapping and is subject to change in order to reflect the project's development and arising data needs. Updated information, as well as adjustments to the current information, will be provided on an as-needed basis and at the latest during the planned DMP update before the end of the project*

As detailed in Table 3, all of B3's other research outputs will be open access, with the majority of them being reusable under the CC BY and the MIT licences. They will be deposited in open access repositories such as Zenodo and GBIF. The outputs will be assigned globally unique and persistent identifiers (e.g., DOIs) and will be described with rich machine-readable metadata following community-endorsed standards such as Dublin Core and DataCite. If relevant, documents or tools needed to reuse the outputs will be included. Thus, where possible, B3 will try to align the management of its other research outputs with the FAIR principles detailed in Chapter 3, demonstrating a commitment to promoting the openness, accessibility, and reusability of these outputs.

---

### Recommendation #11

B3's members are encouraged to manage their other research outputs in adherence to the FAIR principles outlined in Chapter 3 of the DMP which can be condensed as:

- **Findable**: strive to deposit your research outputs described with rich metadata in trusted open access repositories which assign them globally unique and persistent identifiers and offer search engines and indexing;
- **Accessible**: aim to publish outputs in open access venues which accept the publication of less traditional research outputs or deposit them in trusted open access repositories;
- **Interoperable**: use standard formats, vocabularies and ontologies and accompany your outputs with a separate human-readable description of the output where needed;
- **Reusable**: strive to make your outputs accessible under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights.

---

The software produced within B3 will be published as open-source code under an MIT License (or equivalent), when appropriate as "software description papers" in relevant journals to ensure findability and reusability of all open-source resources. As part of T3.1, quality requirements will be specified and documented for all software developed in B3. More information on software management within the project will be available in *D3.3 FAIR guidelines*.

> **Recommendation #12**
>
> Creators of new software developed within B3 are strongly encouraged to make it available under the MIT License using the **full text and disclaimer of the licence**:
>
> MIT License
>
> Copyright (c) [year] [fullname]
>
> Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:
>
> The above copyright notice and this permission notice (including the next paragraph) shall be included in all copies or substantial portions of the Software.
>
> THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

As part of T3.2, the project will create a public, searchable and peer-reviewed documentation website for its toolbox, so that it can be used by the wide community. It will include descriptions of the accessible APIs, facilitating integration with external systems and services. B3 will also facilitate the use of its tools by creating exemplar workflows and specific tutorials, written in English as literate programming documents (e.g., Jupyter notebooks), but also included as source code snippets. Such an approach allows readers to easily test and adapt the described workflow to their needs.

All of B3's publications, public deliverables, policy briefs and training materials will be uploaded on the project's website which will be maintained for at least five years after the project's end. B3 will also make use of innovative open science practices recommended in the Horizon Europe Programme Guide (V3.0), such as participation in open peer-review and early sharing of research where journals allow it. The project will furthermore explore the possibility of publishing less traditional research outputs in suitable journals and platforms that support their dissemination such as Research Ideas and Outcomes (RIO) Journal, Open Research Europe, One Ecosystem, Knowledge4Policy and the EOSC training marketplace.

B3 will also establish an open access project collection in the Research Ideas and Outcomes (RIO) journal which will offer the possibility of hosting data papers, factsheets, policy briefs, project deliverables and infographics published with permanent DOIs to ensure B3's collective knowledge is available, citable and reusable beyond the project's lifetime. Furthermore, a

catalogue of B3's data cubes, workflows and services will be maintained after the project's end by GBIF and the GEO BON EBV Data Portal.

---

**Recommendation #13**

To increase the availability and transparency of B3's research outputs, project members should explore the possibilities offered by innovative practices in open science, including:
- Early and open sharing of research findings through methods like preregistration, preprint publication, etc.;
- Participation in open peer review;
- Leveraging platforms such as the Research Ideas and Outcomes (RIO) Journal and Open Research Europe to publish less traditional research outputs, such as project deliverables, software descriptions, methodologies, policy briefs and data papers.

---

Lastly, B3 will produce a total of 35 deliverables, 27 of which will take the form of a report which will be disseminated at a public level and available on the project's website. The project will also produce 38 milestones, some of which will be verified through a report. Public dissemination level deliverable and milestone reports containing insights of relevance to stakeholders will be considered for further publication options, which allow them to obtain a DOI, facilitating their reuse.

---

**Recommendation #14**

If relevant and appropriate, authors of B3 public deliverable and milestone reports should consider the option of publishing them in open access venues which accept less traditional research outputs (such as the project's RIO collection) or depositing them in trusted open access repositories (such as Zenodo), so that they can be assigned with globally unique and persistent identifiers, unless doing so would:
- be against the authors' legitimate interests;
- be contrary to any other constraints and obligations.

---

# 5. Allocation of resources

FAIR data management is associated with various costs, which can be classified into two main categories. Firstly, there are article processing charges (APC) associated with publishing in open access journals. Secondly, there could be fees related to deposition in trusted open access repositories. While some repositories, like Zenodo and GBIF, offer free deposition of data, others, such as Dryad, may charge users for publishing. Taking these costs into consideration, sufficient resources have been allocated to B3's members to ensure the FAIR and effective management of B3's data and other research outputs. It is expected that each project partner utilises its funding responsibly, prioritising open access practices. In addition, sufficient resources have been assigned to WP1 leaders, who are responsible for maintaining the DMP up to date, reflecting the current data needs and practices of the project and ensuring the data's long-term preservation.

# 6. Data security

As detailed in Chapter 3, once B3's open access datasets are fully compiled, they will be stored in trusted repositories, such as Zenodo, GBIF and the EBV Data Portal, that guarantee their preservation and curation over the long term. In the meantime, project partners will safeguard incomplete datasets on their respective servers, which are equipped with regular and/or automated backup protocols. Access to these servers is closely monitored and secured through processes such as two-factor authentication, VPN connection and/or firewall protection.

---

**Recommendation #15**

To ensure the security of B3's data, project members are encouraged to:
- perform regular backup procedures;
- enquire with full accuracy who the official Data Protection Officer (DPO) of their institution is and maintain contact with them.

---

# 7. Intellectual property management

Intellectual Property Rights management within B3 will be conducted in accordance with the Consortium Agreement and the Grant Agreement. Results are owned by the Party that generates them. In cases of joint ownership, the shares of ownership will be proportional to the co-owners' interests, work packages and contributions. The terms of the joint ownership, transfer and dissemination of results are outlined in the Consortium Agreement, Section 8 and will be conducted accordingly to ensure optimal cooperation among the B3 partners.

# 8. Ethics

For all activities funded by the European Union, ethics is an integral part of research from beginning to end, and ethical compliance is seen as essential to achieving real research excellence. Ethical research conduct implies the application of fundamental ethical principles and legislation to scientific research in all possible domains of research. All activities developed in B3 must comply with ethical principles and relevant national, EU and international legislation, the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights.

In the current issue related to the development and implementation of Artificial Intelligence in the context of projects financed with European funds, it is necessary to refer to the document launched by the European Commission on *how to complete your ethics self-assessment* (EC, 2021a). This document addresses, among other ethical issues, the ethical reference framework to be followed by projects that include artificial intelligence (AI) in their development. In this sense, the principles that should guide research, development and implementation of AI are (EC, 2021a: 39-41):
1. Human Agency and Oversight;
2. Technical Robustness and Safety;
3. Privacy and Data Governance;

4. Transparency;
5. Diversity, Non-discrimination and Fairness;
6. Environmental and Societal well-being;
7. Accountability.

On the other hand, in the same document, the European Commission recommends the use of the document *Ethics Guidelines for Trustworthy Artificial Intelligence* (2019) as a means of detecting and advising on possible risks related to the use and implementation of AI. In addition, another document to monitor and systematise the previous one has been launched by the European Commission – European Commission. (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment.

The principles set out in the *Ethics Guidelines for Trustworthy Artificial Intelligence* and underpinning the concept of trustworthy AI are based on the same principles set out above. Lastly, and getting to a more specific level regarding the uses of AI that may pose a greater risk in the context of the development of EU-funded projects, the European Commission launched the following document *Identifying serious and complex ethics issues in EU-funded research.* As the Commission points out (2021b: 9):

The particular ethical issues raised by developing or deploying new technologies will depend on the context, use case or nature of the application. Specific caution is warranted for technologies that may pose a high risk of harm to people's health and safety or fundamental rights and freedom.

In this sense, different fields of application are also identified in which a high risk could exist, these are (EC, 2021b):
1. Biometric identification systems;
2. The administration of justice;
3. Law enforcement;
4. Migration, asylum and border control;
5. Democratic processes and institutions;
6. The operation of essential critical infrastructure and services;
7. Access to education and vocational training;
8. Employment and worker management;
9. Access to public or private services and social welfare programmes.

For B3, particular emphasis will be placed on the following potential impacts:
● Systems/techniques that have the potential to lead to significant negative social impacts (e.g., on democracy, media, labour market, freedoms, educational choices, mass surveillance).
● Significant negative environmental impacts – either through intended applications or plausible alternative uses (EC, 2021b).

# 9. References

Desmet, P., Oldoni, D., Blissett, M. & Robertson, T. (2023). *Specification for species occurrence cubes and their production*. B3 project Deliverable D2.1.

European Commission. (2019). Ethics Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

European Commission. (2021a). How to complete your ethics self-assessment: V2.0. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/how-to-complete-your-ethics-self-assessment_en.pdf

European Commission. (2021b). Identifying serious and complex ethics issues in EU-funded research. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/guidelines-on-serious-and-complex-cases_he_en.pdf

European Commission. (2023). Horizon Europe (HORIZON) Programme Guide: V3.0. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

Yovcheva, N. & Metodiev, T. (2023). *Plan for Exploitation, Dissemination and Communication*. B3 project deliverable D1.2.

# 10. Annex: Data Management Survey

# B3 Data Management Questionnaire
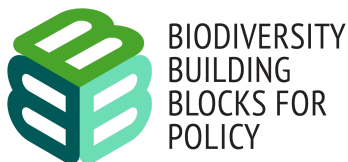


This questionnaire aims to collect information about the various types of data and other research outputs that will be collected, generated or stored by B3 members. Based on your feedback, we will develop the project's Data Management Plan (DMP). To obtain a better understanding of the DMP's content and structure, you can consult **this template** provided by the European Commission.

**We realise you probably don't have the exact answers to some of these questions yet, so feel free to give us your preliminary estimate, which you can modify/update either on your demand when needed or during the mandatory DMP update before the end of the project.**

There are 15 questions in this questionnaire and based on your answers, the DMP will:

1) **provide a data summary** of which project partners will generate and reuse data, for what purposes, with that type of access, and under what types, formats and sizes;

2) **list the other research outputs** which will be produced by the project, for example, software, workflows, models, policy briefs and training materials;

3) **provide recommendations** on applying the FAIR principles to B3 research data and to its other research outputs.

Please keep in mind filling out the questionnaire might take you longer than standard questionnaires do since it requires you to collect input from your colleagues on a number of different questions.

# B3 Data Management Questionnaire

To obtain a better understanding of why we collect this information, you can consult **this template** provided by the European Commission.

**We realise you probably don't have the exact answers to some of these questions yet. Feel free to give us your preliminary estimate, which you can modify/update either on your demand when needed or during the mandatory DMP update before the end of the project.**

**In the DMP, we will clarify that the information it contains is based on an initial data mapping and is subject to change in order to reflect the project's development and arising data needs.**

**\***1. First and last name

**\***2. Organisation/institution

Questions 3 and 4 concern the **data you will generate**.

3. Please provide the following *modifiable* information for the data you plan to **generate**:

1. **Name of the data output**: this refers to the finalised data product, for example, a single dataset.

2. **Contact person**: the person who can be approached with questions regarding how the data is stored, for example.

3. **Relevant task**

4. **Generated via**: for example, data aggregation, modelling, remote sensing, literature review, policy review, interview, surveys.

5. **Size**: a rough estimate which can be revised/updated when significant changes occur.

6. **Format**: for example, .csv; .netcdf .docx; .xlsx; .pdf; .mp4; .xml.

7. **Type of data**: qualitative or quantitative data.

8. **Sensitive data**: Yes/no. If yes, please specify, for example, racial, political, ethical, health, and more here.

9. **Personal data**: Yes/no. If yes, please specify, for example, name, surname, address, email, IP address, location data.

10. **Delivery**: a rough estimate of when the data can be publically shared. If there is an embargo period, please specify why and how long it will apply.

11. **Metadata**: will you use a common metadata format and/or what fields will you include in your metadata? For example, DOI, creator(s), title(s), date(s), size(s), format(s), rights, description(s), included datasets, applied filters.

12. **Users**: to whom might this data be useful?

13. **Type of access**: will it be open access? If not, please indicate the reasons, for example, ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related, contract.

14. **Licence**: what type of licence do you plan to use for your data?

15. **Point of access**: do you have a preferred route for making this data open access? For example, Zenodo, GBIF storage, GEO BON EBV Data Portal.

16. **Re-use**: potential documents or tools needed to re-use or validate the data.

❓ This question concerns the data you will **generate**.
❗ **Please fill in at least one answer**

| | Data output 1 | Data output 2 | Data output 3 | Data output 4 | Data output 5 |
|---|---|---|---|---|---|
| Name of the data output | | | | | |
| Contact person | | | | | |
| Relevant task | | | | | |
| Generated via | | | | | |
| Size | | | | | |
| Format | | | | | |
| Type of data | | | | | |
| Sensitive data | | | | | |
| Personal data | | | | | |
| Delivery date | | | | | |
| Type of meta-data | | | | | |
| Potential users | | | | | |
| Type of access | | | | | |
| Licence | | | | | |
| Point of access | | | | | |

| | Data output 1 | Data output 2 | Data output 3 | Data output 4 | Data output 5 |
|---|---|---|---|---|---|
| Re-use | | | | | |

**\***4. If your **generated data** contains personal or sensitive information, would you consider anonymising it and publishing it in an aggregated form so as to not disclose said information?

*If yes, please specify to which dataset(s) that would apply. If not, please use the box to explain why.*

❷ This question concerns the data you will **generate**.

❶ Choose one of the following answers

○ Yes

○ No

○ Not relevant

Please enter your comment here:

Question 5 concerns the **data you will obtain from elsewhere and reuse**.

5. Please provide the following *modifiable* information for the **existing data you will reuse**:

1. **Data name**

2. **Relevant task**

3. **Size**

4. **Format**: for example, .docx; .xlsx; .pdf; .mp4; .xml; .csv.

5. **Sensitive data**: Yes/no. If yes, please specify, for example, racial, political, ethical, health, and more here.

6. **Personal data**: Yes/no. If yes, please specify, for example, name, surname, address, email, IP address, location data.

7. **Metadata**: describe the metadata accompanying the data.

8. **Access**: open/restricted/closed access.

9. **Origin**: what is the origin of the data?

10. **Ownership**: who owns the data you will reuse?

11. **Licence**: what type of licence allows you to use the data?

12. **Re-use**: potential documents or tools needed to re-use or validate the data.

❷ This question concerns the data you will obtain from elsewhere and **reuse**.

| | Data 1 | Data 2 | Data 3 | Data 4 | Data 5 |
|---|---|---|---|---|---|
| Data name | | | | | |
| Relevant task | | | | | |
| Size | | | | | |
| Format | | | | | |
| Sensitive data | | | | | |
| Personal data | | | | | |
| Metadata | | | | | |
| Access | | | | | |
| Origin | | | | | |
| Ownership | | | | | |
| Licence | | | | | |
| Re-use | | | | | |

**According to the European Commission, in addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, training materials, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.). Beneficiaries should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.**

Question 6 concerns the **other research outputs you will generate**.

6. Please provide the following *modifiable* information for the **other research outputs you plan to generate**:

1. **Name of the output**

2. **Contact person**: the person who can be approached with questions regarding the development and publication of the output, for example.

3. **Relevant task**

4. **Type of output**: briefly describe the result you will produce.

5. **Delivery**: a rough estimate of when the output could be publically shared.

6. **Users**: to whom might this output be useful?

7. **Metadata**: will you accompany the publication of your result with metadata? If yes, will you use a common metadata format and/or what fields will you include in your metadata?

8. **Type of access**: will it be open access? If not, please indicate the reasons, for example, intellectual property, commercial, privacy-related, security-related.

9. **Persistent identifiers**: do you plan on sharing your output in a way that assigns it globally unique and persistent identifiers?

10. **Point of access**: do you have a preferred route for making your output accessible to end users?

11. **Interoperability**: how do you plan to make your output interoperable? For example, by using standard ontologies and accompanying it with a separate human-readable description of the output.

12. **Licence**: what type of licence do you plan to use for your output

13. **Re-use**: potential documents or tools needed to re-use the output.

❓ This question concerns the other research outputs you will **generate**.
❗ **Please fill in at least one answer**

| | Output 1 | Output 2 | Output 3 | Output 4 | Output 5 |
|---|---|---|---|---|---|
| Name of the output | | | | | |
| Contact person | | | | | |
| Relevant task | | | | | |
| Type of output | | | | | |
| Delivery date | | | | | |
| Users | | | | | |
| Metadata | | | | | |
| Type of access | | | | | |
| Identifier | | | | | |
| Point of access | | | | | |
| Interoperability | | | | | |
| Licence | | | | | |
| Re-use | | | | | |

Question 7 concerns the **other research outputs you will reuse**.

7. Please provide the following *modifiable* information for the **other research outputs you plan to reuse**:

1. **Name of the output**

2. **Relevant task**

3. **Access**: open/restricted/closed access.

4. **Origin**: what is the origin of the output?

5. **Ownership**: who owns the output you will reuse?

6. **Licence**: what type of licence allows you to use the output?

7. **Re-use**: potential documents or tools needed to re-use the output.

❷ This question concerns the other research outputs you will obtain from elsewhere and **reuse**.

| | Output 1 | Output 2 | Output 3 | Output 4 | Output 5 |
|---|---|---|---|---|---|
| Name of the output | | | | | |
| Relevant task | | | | | |
| Access | | | | | |
| Origin | | | | | |
| Ownership | | | | | |
| Licence | | | | | |
| Re-use | | | | | |

Questions 8-15 concern your **data management practices**.

**✱**8. Please provide a brief summary of your **institutional data management practices**, specifying:

1. **Data location**: where and how data will be stored before it is fully compiled, for example, institutional server or web hosting.

2. **Server location**: EU or non-EU; compliant or not-compliant with applicable data protection rules (for example, GDPR).

3. **Backup procedures**: type of backup procedures and their frequency.

4. **Protection**: how data security is ensured, for example, password or two-factor authentication.

5. **Data Protection Officer (DPO)**: name the person from your organisation who is responsible for data protection and can serve as a contact person if questions arise.

❷ This question concerns your **institutional data management**.

| | |
|---|---|
| Data location | |
| Server location | |
| Backup procedures | |
| Protection | |
| Data Protection Officer (DPO) | |

---

**\***9. Do you follow a specific naming convention for your files?

*If yes, please use the box to specify.*

❷ This question concerns your **data management**.
❶ Choose one of the following answers

◯ Yes

◯ No

Please enter your comment here:

**\***10. Do you use any standardised protocols to organise your database?
*If yes, please use the box to specify.*

❷ This question concerns your **data management**.

❶ Choose one of the following answers

○ Yes

○ No

Please enter your comment here:



**\***11. Will you be using standard vocabularies for all data types present in your data to allow inter-disciplinary interoperability?

*If yes, please use the box to specify.*

❷ This question concerns your **data management**.

❶ Choose one of the following answers

○ Yes

○ No

Please enter your comment here:



**\***12. Do you perform data cleaning to remove incorrect and unreliable information or uncomplete records?

*If not, please use the box to explain why.*

❷ This question concerns your **data management**.

❶ Choose one of the following answers

○ Yes

○ No

Please enter your comment here:

---

*13. Would you be interested in publishing your data in the form of data papers and/or software description papers?

*If yes, please give an example of a suitable dataset. If not, please use the box to explain why.*

❷ This question concerns your **data management**.
❶ Choose one of the following answers

○ Yes

○ No

Please enter your comment here:

---

*14. Do you have a preference for a trusted repository where to store your research data?

*If yes, please use the box to specify.*

❷ This question concerns your **data management**.
❶ Choose one of the following answers

○ Yes

○ No

Please enter your comment here:

---

**\*** 15. Can you identify potential obstacles (e.g., technical, social, policies) that would prevent delivering FAIR data during B3's lifetime and beyond? Information on FAIR data here.

*Iffiyes,fipleasefiusefithefiboxfitofispecify.*

❓This question concerns your **data management**.

❗ Choose one of the following answers

Yes
○

No
○

Please enter your comment here:

---

**Please do not hesitate to fill in the comments fields with any information that would help to clarify your answer or is not addressed in the list of questions!**