BIODIVERSITY BUILDING BLOCKS FOR POLICY

M1 Delivery of a training strategy for B3

30/10/2024

Author(s): Sabrina Kumschick, Cristina Preda, Katelyn Faulkner, Sandra MacFadyen, Lina Estupinan-Suarez, Laura Abraham, Damiano Oldoni, Quentin Groom

Funded by the European Union

This project receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101059592). Views and opinions expressed are 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.

Table of contents

Summary

The main goal of this training strategy is to facilitate the uptake of the products developed through B3, to develop the skill set of the end-users and to increase knowledge exchange, ensuring the training resources outlast the lifetime of the project. To this end, we have developed a questionnaire to identify the training needs as seen by project partners, and a series of resources aimed at a broader audience.

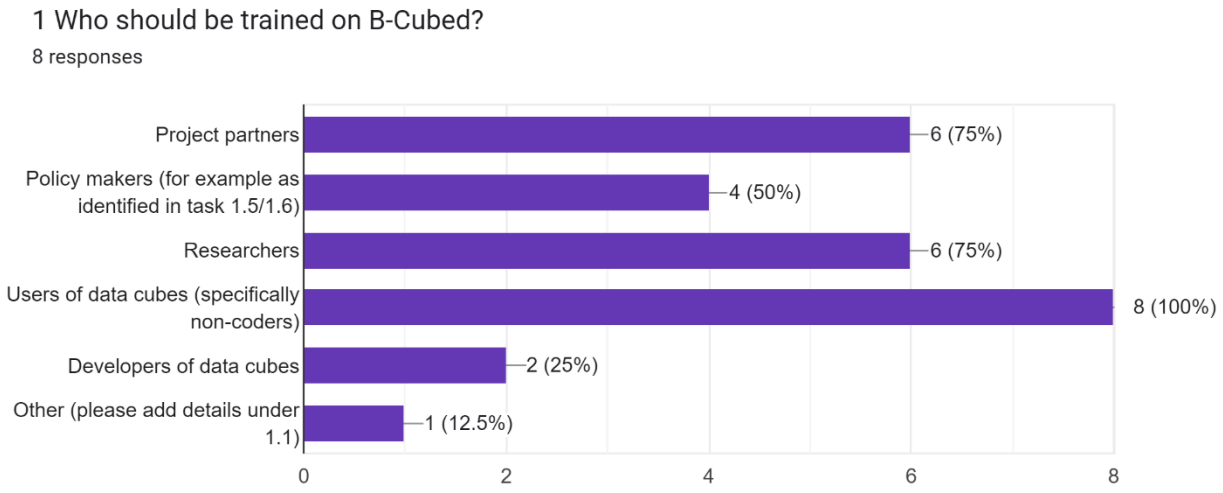
List of abbreviations

EU	European Union

1. Identifying training needs

A questionnaire was prepared to get input on which training the B3 partners deem necessary and useful, which stakeholders should be trained, and how the project partners are going to contribute to training activities (Supplementary Material, Appendix S1). It was sent out as a Google Form on 21 May 2024 to all project partners, who were asked to send replies by 28 May 2024 (the week of the General Assembly Meeting). We received eight responses on the Google forms, and two additional responses via email out of 56 recipients of the questionnaire. The results are presented below.

1.1. Who should be trained on B-Cubed?



All respondents agreed that users of data cubes should be trained, including people without coding experience. Internal training opportunities for project partners was also deemed relevant, as well as training of researchers (e.g., data scientists at national environment/biodiversity agencies; GBIF national nodes).

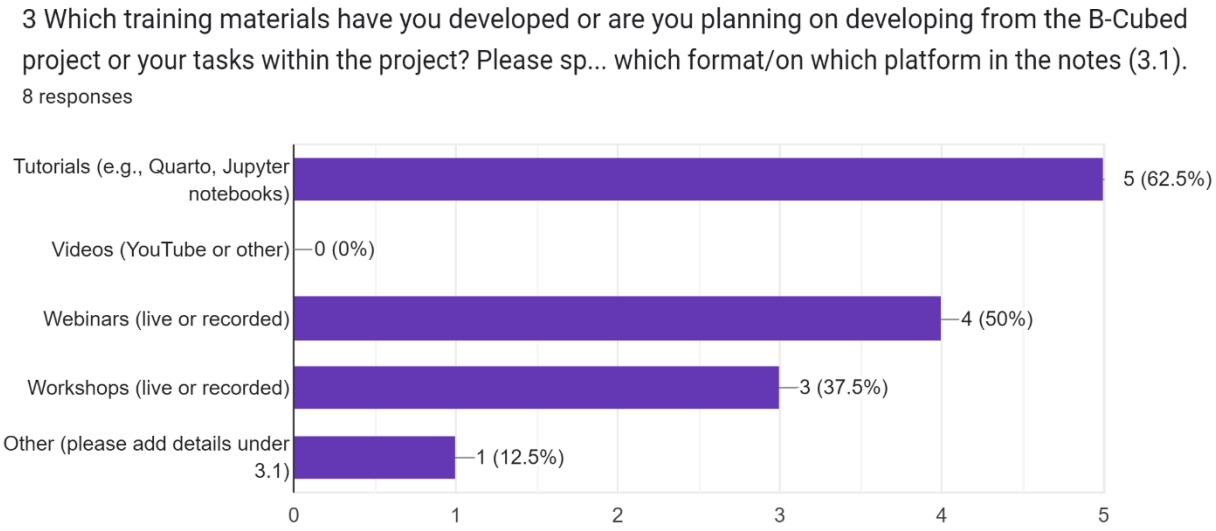
1.2.Which training would you like to receive from the B-Cubed project?

Respondents identified several topics that they would like to receive training in. Two that were mentioned most are the creation, analysis and use of data cubes, and how to set up workflows to create indicators, specifically using cubes.

- How to include grid systems?
- Understanding spatial, temporal and taxonomic inconsistencies of the data, and how to deal with them.
- Using GitHub and how to get/develop data cubes
- Design and development of R packages and online applications with AWS or ShinyR

- How can cubes be used to measure progress on Global Biodiversity Framework indicators?
- How can I use the GBIF cube/SQL API to answer questions (e.g. what are the top 10 observed species) per country that were previously hard to answer?
- To know all the possibilities (filtering data for example)
- A ‘how-to’ guide on cloud computing
- Tutorials on how to create occurrence cubes

1.3. Which training materials have you developed or are you planning on developing from the B-Cubed project or your tasks within the project?



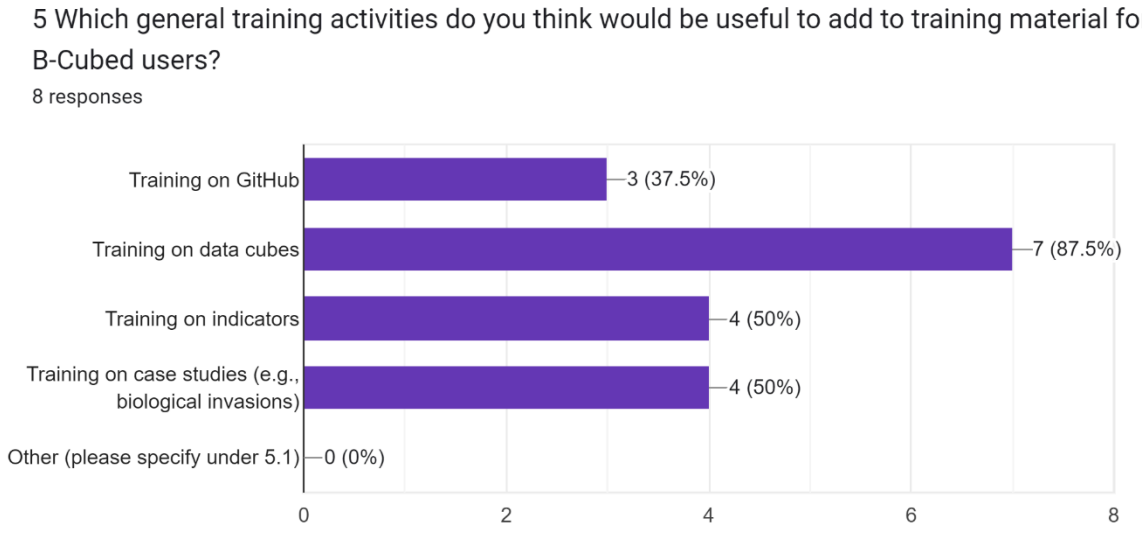
Many respondents are planning on developing training materials, and some are already available (e.g., https://biomath-lab.github.io/b-cubed/about/). Developers of R packages will create vignettes (similar to Jupyter notebooks) which explain the functioning of the R packages. Furthermore, tutorials and online workshops will be developed on data mobilisation from GBIF to the EBV data portal (under Task 3.3).

1.4. When would you be able to start working on these training materials, and by when will they become available?

Most respondents would be able to start working on the training materials in the second half of the project, in 2025. Some training material is already available, such as the software development guide (https://docs.b-cubed.eu/dev-guide/), and materials listed below. The Databricks webinar is not recorded, but material on how to use it is available at

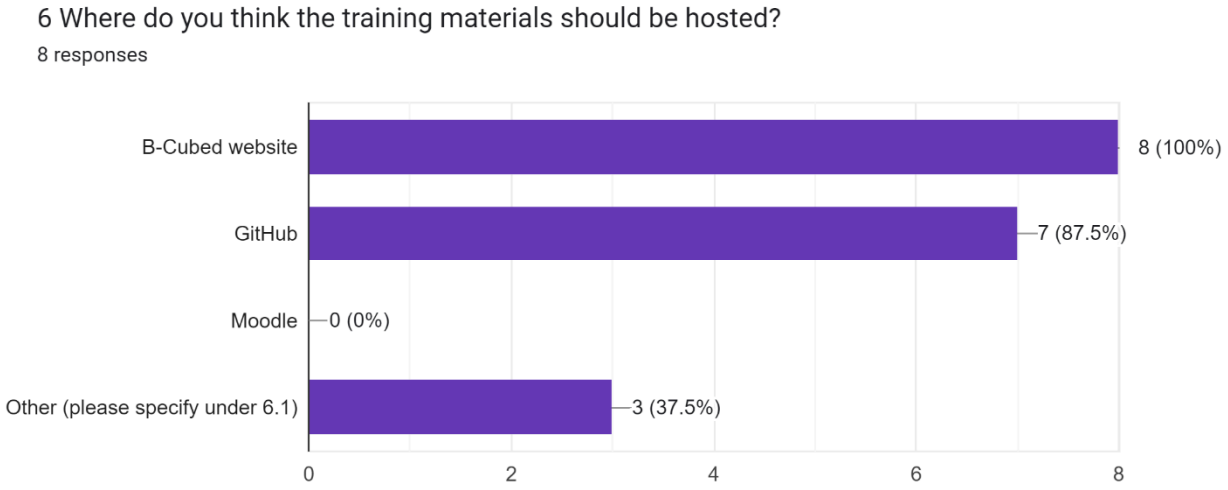
https://techdocs.gbif.org/en/data-use/b-cubed/generate-cube-databricks#generating-a-data-cube-using-microsoft-azure-databricks.

1.5. Which general training activities do you think would be useful to add to training material for B-Cubed users?



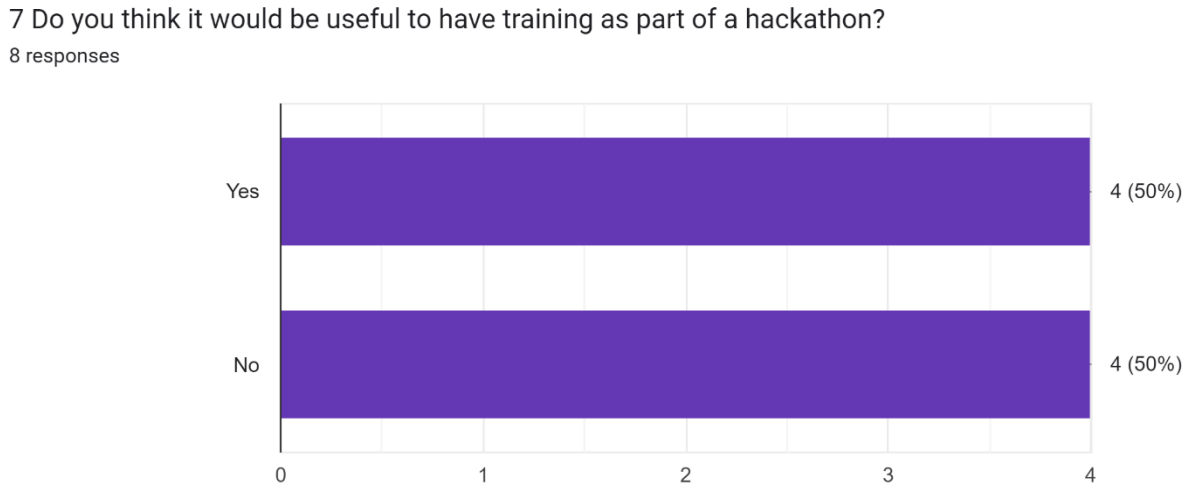
Training on data cubes was deemed most relevant, possibly training on indicators, case studies, and GitHub more generally (although some basics would be a prerequisite for much of the training envisaged). Other suggestions included training on developing tutorials for resulting B-Cubed workflows, and training on design and development of R packages and online applications with AWS or ShinyR.

1.6.Where do you think the training materials should be hosted?



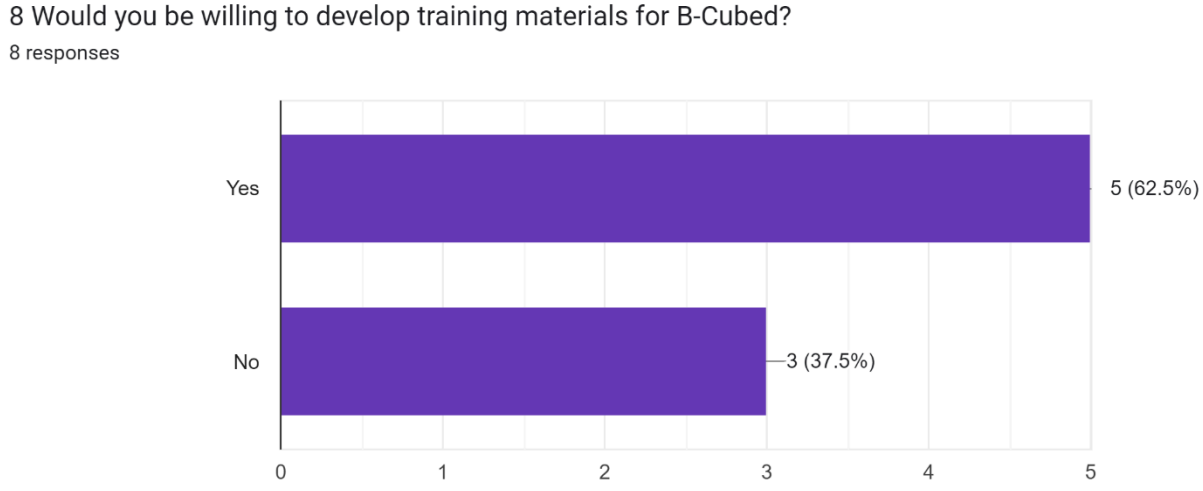
All respondents agreed that the B3 website should host the training materials (https://docs.b-cubed.eu/), and that Moodle is not a suitable platform for this project. GitHub was also deemed very relevant, and it was noted that scripts should be well documented on GitHub where they are maintained. There is a documentation website on GitHub which can be used to join everything together (https://github.com/b-cubed-eu/documentation). GBIF was also mentioned by several respondents as an important outlet, such as the GBIF website, https://techdocs.gbif.org/ (powered by GitHub), and the GBIF data blog (https://data-blog.gbif.org/), where training material can be reworked in a more narrative form and by which more users could be reached. The EBV data portal was also mentioned, and a YouTube channel recommended for videos.

1.7. Do you think it would be useful to have training as part of a hackathon?



Respondents were divided on whether training as part of a hackathon would be useful. It was mentioned that hands-on work can be useful for training. The format of a coding club or Q&A was recommended as an alternative format, as well as a workshop. The general assembly in 2025 was suggested as a suitable time for a possible training event. One respondent noted that in their opinion, training material is best consumed in text format at the participants own pace.

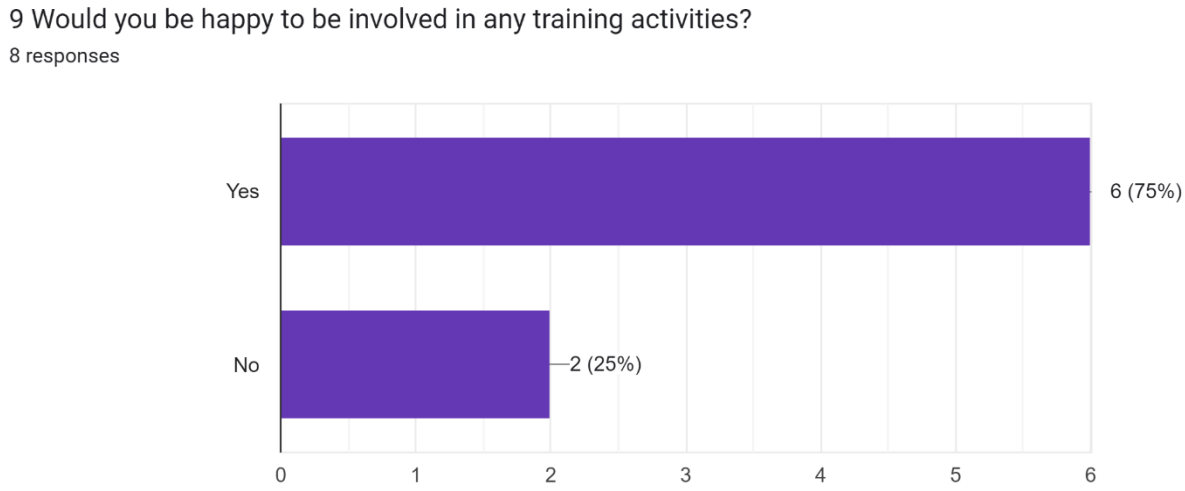
1.8. Would you be willing to develop training materials for B-Cubed?



Many respondents would be happy to develop training materials as needed. Some concrete materials offered are listed here:

- Case study (Task 6.2: biological invasions)
- The basics of spatial analysis with R
- The basics of Google Earth Engine
- Dissimilarity cube (Related to Task 4.2)
- Data mobilisation to the EBV Data Portal (Related to Task 3.3)

1.9.Would you be happy to be involved in any training activities?



Respondents would be happy to be involved in training activities in various capacities, such as presenter, trainer and developer of materials, reviewer, attendee, support and guidance.

2. Working groups at general assembly meeting

A session was held at the general assembly meeting in Montpellier on 28 May 2024 to discuss Task 1.4, specifically the training component thereof. Keeping in mind the main stakeholder groups that will benefit from B3 outputs, topics for potential training sessions were discussed with project partners, as well as potential trainers and the materials needed based on the agreed format of training sessions. Discussions were focused on the main stakeholder groups, to allow each partner to express their opinion. The results of the brainstorming session are summarised in Table 1.
The stakeholders for B3 are the following:

- Policy and governance
    - Parliaments and elected councils.
    - Executive government and administration
    - Other political actors
- Organisations
    - Organisations protecting biodiversity
    - Organisations working on public health surveillance and risk assessment and epidemiological intelligence
    - Organisations working on food security and sustainable agriculture

- Scientific community
  - Museums and herbaria
  - Informaticians and data scientist groups
  - Research bodies and actors
- Data collectors
  - Citizen scientists and civil society organisations
- Industry and practitioners
  - Land managers, farmers, and foresters.
  - Insurance companies, construction and urban planners.
  - Multilateral funding organisations
- General public
  - Citizens, indigenous peoples and local communities.

Table 1: Possible training activities of the B3 project

Target group	Theme	Format	Training materials	Trainers
Policy and governance	Not discussed	Not discussed	Not discussed	Not discussed
Organisations	learn how to get data appropriate to their needs using GBIF and B3 workflows/ tools	online	Webinar, self explanatory videos, forum for follow-up	GBIF experts & developers + internal users of B3 tools/ workflows
Scientific community	Open data & open science	short version (ecological/ conservation society conferences) – long version (online)	Mark down, GitHub	everyone within B3
Data collectors	cool stuff you can do in GBIF with your data	online	webinars, tutorials	someone from GBIF + organisation
Industry and practitioners	how to comply easily to biodiversity policies	very easy to use training tools, training sessions about policies	fully operational tools	label auditors, policy enforcers

Target group	Theme	Format	Training materials	Trainers
General public	Increase the awareness of importance of biodiversity for human life at local levels	design video for european citizens, possibility with subtitles in different languages, picture exhibition to highlight local biodiversity	national biodiversity portals that already exist → organise training on how to explore that portal	B3 partners engaging with local communities e.g. network of biology teachers

3. Training conducted and materials available

The B3 project has a documentation website on GitHub where information can be found on how to put together tutorials (https://github.com/b-cubed-eu/documentation). This should be followed by project partners when developing such materials.

3.1. E-learning and online resources

3.1.1.Webinars

A series of webinars are organised by and dedicated to B3 partners, aimed at conveying knowledge, exchanging ideas and gaining skills in specific areas related to the expertise of the consortium. The series is coordinated by Matilde Martini, University of Bologna, with a roughly monthly occurrence, throughout much of the duration of the project. Various subjects have been approached so far e.g. exploring Docker, simulation and visualisation of biodiversity data cubes, using Google Earth Engine (details below) with more sessions planned to take place by the end of the year (see 4.1 Planned webinars).

Date: 19th of March from 10:30am to 12:00pm
Title*:* Go see a Docker! First help with containerizing.
Presenter*:* Christophe Van Neste (Meise Botanic Garden)
Abstract*:* "Dependency hell" is a common challenge faced by developers and end-users alike. You find a promising library or program, only to discover that navigating its prerequisites and installation process leads to frustration and wasted time. This is where Docker comes to the rescue.
If you've never explored Docker before, or if you're struggling to deploy your tools across different systems, this workshop is designed for you. Discover how Docker containers can simplify software deployment and management, allowing you to focus on your work without getting bogged down in system configurations.
Whether you're a developer seeking seamless collaboration or a proponent of reproducible

science, this workshop offers practical insights and hands-on guidance. Join us on the Docker workshop and streamline your workflow.

Date: 15th of April from 10:30am to 12:00pm
Title*:* Unveiling Ecological Dynamics Through Simulation and Visualization of Biodiversity Data Cubes
Presenter*:* Ward Langeraert (The Research Institute for Nature and Forest (INBO))
Abstract*:* Simulation studies offer numerous benefits due to their ability to mimic real-world scenarios in controlled and customisable environments. Ecosystems and biodiversity data are very complex and involve a multitude of interacting factors. Simulations allow researchers to model and understand the complexity of ecological systems by varying parameters such as spatial and/or temporal clustering, species prevalence, etc.
During the B-Cubed hackathon, we have created a simulation framework for biodiversity data cubes using the R programming language. The framework starts from simulating a species distributed in a landscape over a temporal scope. In a second phase, the simulation of a variety of observation processes and effort generates the actual occurrence datasets. Based on their (simulated) spatial uncertainty, occurrences can then be designated to a grid to form a data cube.
The simulation framework can be used to assess multiple research questions under different parameter settings, such as the effect of clustering on occurrence-to-grid designation and the effect of different patterns of missingness on data quality and indicator robustness. Simulation studies can incorporate scenarios with missing data, allowing researchers to assess the impact of data gaps on analyses.
In this workshop, we will go over the results from the B-Cubed hackathon regarding this framework, show how users can utilise the R code in practise, and we can discuss how we can still improve the framework.

Date: 14th June 2024 from 10:30am to 12:00pm CEST
Title: Getting started with Google Earth Engine: Basic JavaScript rules, create, import, explore and visualise GEE datasets
Presenter: Sandra MacFadyen (Stellenbosch University)
Abstract: Sandra will present some Google Earth Engine basics with a practical course that should allow participants to:
1. Understand the basic layout of Google Earth Engine platform (incl. the Code Editor).
2. Understand basic JavaScript syntax rules.
3. Find and import datasets into the code editor.
4. Inspect a dataset in the console.
5. Visualise datasets in the interactive map explorer.
6. Use simple functions.
7. Know where to find help.

Date: 7th October 2024 from 11am. Practical session on 14th October 11am.
Title: EBVCube: Enhancing Biodiversity Data Sharing with Interoperable Geospatial Standards
Presenter: Lina Estupinan-Suarez, Henrique Pereira, Miguel Fernandez
Abstract: Gridded geospatial data on different dimensions of biodiversity, at different spatio-temporal scales are increasingly used to model and estimate biodiversity trends and their relationship with global change. There is an increasing need for an interoperable data standard

that covers the complexity and scope of biodiversity observations in a flexible and accessible way for the biodiversity researchers and practitioners. Building on efforts made by the climate community, we applied a hierarchical, multidimensional array structure based on the Network Common Data Form (netCDF) to the Essential Biodiversity Variables (EBV) framework promoted by GEO BON, which we call the EBVCube format. This data format allows for a more comprehensive data management and analysis and at the same time accommodates data diversity and scales efficiently. The EBVCube format is supported by the *ebvcube* R package which provides tools to create, access and visualise these datasets, streamlining the process and reducing complexity. Together with the EBV Data Portal, this approach offers a complete solution to manage biodiversity data efficiently and follow best practices for data sharing.

3.1.2. Tutorials

Online tutorials were provided on “Spatial data in R”, “Species distribution modelling in R”, and “Google Earth Engine for ecological modelling” in Javascript. These are available on the B3 website (https://docs.b-cubed.eu/tutorials/spatial-r/; https://docs.b-cubed.eu/tutorials/sdm-r/).

3.1.3.Coding club

The Research Institute for Nature and Forest (INBO) organises a coding club (https://inbo.github.io/coding-club/) on a monthly basis. It is an open science initiative with a clear mission: it unites all ecologists, researchers and coders who want to develop their (R) programming skills in a pleasant and supportive environment in order to replace the fear of coding with inspiration and motivation to learn. The INBO coding club is intended for INBO personnel, but external parties (including B3 partners) can ask for an invitation to join. This initiative was advertised during the B-Cubed General Assembly in May and other executive board meetings. People interested in participating can contact Damiano Oldoni by email or via Slack.

3.2. Training courses

3.2.1. Courses

A training course was held in South Africa on data cubes from 2nd to 23rd October 2024. The mini-school was aimed at academics, postgraduate students, managers of protected areas, environmental researchers, but is open to anyone interested in attending including to all partners of the B3 project.
Training materials are available through the NITheCS YouTube channel (https://www.youtube.com/@nithecs/videos).
Title: Empowering Biodiversity Monitoring through Data Cubes: Techniques & Applications for Open Science
Presenters: Sandra MacFadyen, Maarten Trekels
Abstract: Presented by the NITheCS Research Programme Advancing Biodiversity Informatics & Ecological Modelling Research (https://nithecs.ac.za/modelling-biodiversity), this 4-lecture Mini-school aims to empower researchers, students, and practitioners with cutting-edge skills in

biodiversity informatics and ecological modelling through the innovative application of data cubes. Participants will gain practical expertise in using data cubes for biodiversity monitoring and ecological analysis, fostering collaboration and openness within the scientific community. Each lecture is linked to one or more of the 17 Sustainable Development Goals (SDGs) (https://sdgs.un.org/goals).
LECTURE 1 (2 Oct 2024) Introduction to Data Cubes: Fundamentals and Applications
LECTURE 2 (9 Oct 2024) Building Data Cubes with GBIF: A Hands-On Guide for Biodiversity Monitoring
LECTURE 3 (16 Oct 2024) Analysing Data Cubes: Techniques and Tools for Ecological Modelling
LECTURE 4 (23 Oct 2024) Open Science: Sharing and Disseminating Results from Data Cubes

3.2.2.Hackathon

The hackathon held in Brussels in April 2024 explored the potential applications of the cubes. This four-day event, organised at the Herman Teirlinck building as a hybrid meeting, attracted 86 participants from diverse fields, including biology, ecology, and data science, as well as stakeholders from sister projects such as AD4GD and FAIRiCUBE. The event featured a combination of lectures and practical sessions, where participants had the opportunity to experiment with the cubes, identify opportunities, and address challenges with guidance from five B3 technical coaches. Five keynote speakers from related fields and sister projects presented informative and inspiring sessions, setting the stage for participants to explore cube applications and address real-world challenges in data science and environmental research.

Over the course of the event, participants worked in teams to develop and refine nine projects, all of which were shared on a dedicated GitHub repository titled “Hackathon-projects-2024.” Participants were encouraged to create separate repositories for their individual software solutions, fostering a collaborative and accessible environment.On the final day, each team presented its project to a jury panel of experts from universities and sister projects, allowing for constructive feedback and showcasing the diversity of innovative approaches explored. The PowerPoint presentations from the final session are available here.

The hackathon served as a training ground, promoting collaborative problem solving and knowledge sharing. This immersive learning experience helped to build capacity among participants while expanding the project’s knowledge base. For further details, checkout the hackathon webpage.

3.2.3. Live demonstration at COP16

During the 16th meeting of the Conference of the Parties (COP16) to the Convention on Biological Diversity (CBD), MLU lead the session titled “Advancing NBSAP Reporting Using GEO BON's EBV Data Ecosystem: Live Demonstration and Training” at the GEO-GEO BON Pavilion on October 21, 2024 (Blue Zone). In this session, we showcased the mobilisation of data from GBIF to the EBV Data Portal using the species occurrence software developed for B3. Additionally, the session highlighted the importance of FAIR indicators, a key focus of the B3 Policy Brief led by Senckenberg.

The session aimed to strengthen capacity building and facilitate technology transfer in line with Target 20 of the Global Biodiversity Framework. By breaking down barriers to data access, it aimed to improve the sharing and availability of high quality data and knowledge. In the process, we will support the broader objectives of Target 21 by ensuring better access to the best available information for biodiversity management and conservation.

4.Training plan

We aim to develop training built on the train-the-trainer principle to achieve a broader reach of the training and the products developed. Where possible, face-to-face training events will be tied to community events, workshops or conferences, and we will continue the development of training materials and the delivery of online training events. Specific efforts will be made to continue having a broad geographical coverage of the events, and for them to be gender inclusive.

4.1. Training materials

Based on the questionnaire, it was deemed important to train users of data cubes which are not specifically coders in how to apply the cubes. In the first half of the project, the focus of the B3 project was mostly on developing the B3 products and on training within the consortium. As the project matures, we will continue with the training of the B3 partners, using this opportunity to test the materials before opening training to individuals outside the consortium. On top of that, the partners involved in cube and indicator development are currently putting together tutorials and other training materials which can then be incorporated into training for users of cubes and indicators, and which will be made available on the B3 website or via GitHub.
When developing training courses, partners need to take into account the needs identified in the questionnaire, as well as the needs of stakeholders, for example the ones from the case studies studied in WP6 of B3, which will start in January 2025.

4.2. Possible training courses

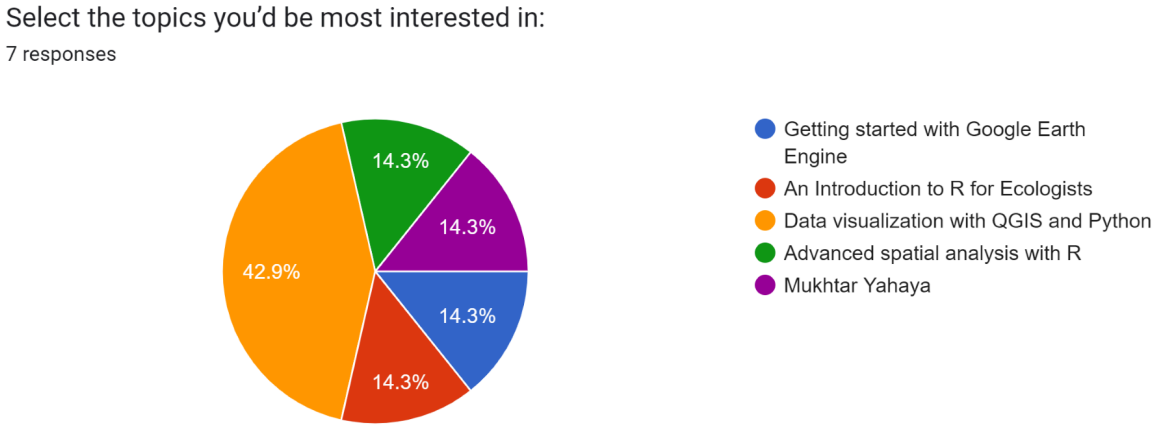
When discussing ideas for training with the project partners, a few suggestions were put forward for in person training. A suggestion is to widen the network of B3 and reach out to stakeholders in South America, potentially by organising a training session attached to a conference in South America in 2025 (e.g. https://www.livingdata2025.com/) or a webinar specifically for Latin America. Organising workshops and training events as satellites of conferences or other relevant meetings, such as the general assembly of B3 next year to be held in Leipzig, is considered an avenue worth exploring.

4.2.1. Use the cases studies as a training thematic focus

We will establish focus groups around themes emerging from the case studies, such as policy applications, conservation, agriculture, and public health. This can be a regular agenda item at individual case study meetings to share insights and develop domain-specific training materials.

4.2.2. Survey after training course on data cubes

A survey was conducted among the participants of the training course on Empowering Biodiversity Monitoring through Data Cubes: Techniques & Applications for Open Science to ask participants about further training needs. The results are shown in the figure below.



4.3. Planned webinars

The webinars have been well attended, and are planned to continue until the end of the project. Webinars are generally organised on a monthly basis. The title, abstract, presenter, along with the link, are provided to all B3 project partners through the Teamwork channel. Additionally, each webinar is recorded and made freely available for consultation.

4.4. Workshops

A B3 workshop about how to make R packages is planned for Early November 2024.The workshop will rely on the B3 software development guide. The workshop will be held online and will be hands-on. The workshop will be composed of two parts. First, a theoretical introduction will explain the basic concepts. A hands-on section with live coding will follow. The workshop will be recorded for increasing its outreach.

4.5. Evaluation and Feedback

We will include mechanisms to evaluate the effectiveness of training programs through feedback surveys and skill assessments. This feedback will be used to refine and improve future training sessions.
We are also planning on developing a report on the progress and impact of training initiatives, showcasing the development of skills across the B3 community.

5. Acknowledgements

All the respondents of the online survey are thanked for their inputs, as well as the participants of the general assembly meeting who took part in the working group discussions. We also thank the project partners for conducting and organising training events, webinars, and developing training materials.

6. Supplementary Material

Appendix S1: Questionnaire
B-Cubed, Task 1.4: Training activities
Your name and surname:_____
Your email address:_____

1 Who should be trained on B-Cubed?

- Policy makers (for example as identified in task 1.5/1.6)
- Project partners
- Researchers
- Users of data cubes (specifically non-coders)
- Developers of data cubes
- Other (please add) _____

Notes: _____

2. Which training would you like to receive from the B-Cubed project?
Please specify: _____

3. Which training materials have you developed or are you planning on developing from the B-Cubed project or your tasks within the project? Please specify for which task/workpackage, and in which format/on which platform in the notes.

- Tutorials (e.g., Quarto, Jupyter notebooks)
- Videos (YouTube or other)
- Webinars (live or recorded)
- Workshops (live or recorded)
- Other: _____

Notes: _____

4. When would you be able to start working on these training materials, and by when will they become available?
Start work on training materials:_____
Anticipated date of availability: _____

5. Which general training activities do you think would be useful to add to training material for B-Cubed users?

- Training on GitHub
- Training on data cubes
- Training on indicators
- Training on case studies (e.g., biological invasions)
- Other (please add) _____

Notes: _____

6. Where do you think the training materials should be hosted?

- B-Cubed website

- Github
- Moodle
- Other (please add) _____

Notes: _____

7. Do you think it would be useful to have training as part of a hackathon?

- Yes
- No

If yes, in which format?_____
Notes: _____

8. Would you be willing to develop training materials for B-Cubed?

- Yes
- No

If yes, in which format?_____
Please provide specifics here: _____

9. Would you be happy to be involved in any training activities?

- Yes
- No

If yes, in which format or capacity?_____
Please provide specifics here: _____