DELIVERABLE 1.7 - SOFTWARE BEST PRACTICES

FEBRUARY 26TH 2025

biont-training.eu

contact@biont-training.eu

Co-funded by
the European Union

About this report

Grant Agreement	101100604 - DIGITAL-2022-TRAINING-02
Action Acronym	BioNT
Action Title	Bio Network for Training
Deliverable	1.7 Software Best Practices
Work package	1 Training design and development
Dissemination level	Public
Authors	S. Di Giorgio, Alan O' Cais, L. Paladin, I. Paredes Cisneros, S. Razick, P. Samarakoon, approved by all BioNT partners
Delivery date	2025-02-26

Consortium members

Acronym	Partner
EMBL	EUROPEAN MOLECULAR BIOLOGY LABORATORY
BIOBYTE	BIOBYTE SOLUTIONS GMBH
HPCNOW	HPC NOW CONSULTING SL
UO	UNIVERSITETET I OSLO
UB	UNIVERSITAT DE BARCELONA
ZBMED	INFORMATION CENTRE FOR LIFE SCIENCE
RIcapacity	RICAPACITY GMBH
ALU-FR	ALBERT-LUDWIGS-UNIVERSITAET FREIBURG
EPFL	ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE

Table of Contents

Project Overview

The BioNT consortium is dedicated to providing a comprehensive training program and fostering a community for digital skills relevant to the biotechnology industry and biomedical sector. With a curriculum tailored for both beginners and advanced professionals, BioNT aims to equip individuals with the necessary expertise in handling, processing, and visualising biological data, as well as utilising computational biology tools. Leveraging the consortium's strong background in digital literacy training and extensive network of collaborations, BioNT is poised to professionalise life sciences data management, processing, and analysis skills.

This workshop at a glance

BioNT conducted its sixth training workshop titled "Code & Collaborate: The FAIRytale of Software Development" on February 4-6 2025. This intensive three-day workshop, part of BioNT's advanced curriculum, received 47 applications from participants across multiple countries (with 45 responses to the pre-workshop survey). The workshop successfully engaged both industry and academic participants, with representation from SMEs, research institutions and job seekers

Out of the 45 pre-workshop survey respondents, 89% of applicants reported using programming languages regularly, and 53% were already familiar with version control software, indicating a strong foundation for advanced training. Over 18 hours of interactive sessions, participants engaged in around 30 hands-on exercises, including collaborative code reviews, documentation practices, and the creation of individual repositories. With 22 participants completing the post-workshop survey, the impact was significant - 91% reported feeling confident to apply their new skills immediately, while 95% indicated they would recommend the workshop to colleagues. This high level of engagement and practical impact exemplifies BioNT's commitment to delivering hands-on, immediately applicable digital skills training for the biotechnology and biomedical sectors.

From individual coder to development team

The workshop guided participants through an intensive learning journey, transforming individual coding practices into collaborative development skills. Following The Carpentries and CodeRefinery hands-on models, each concept was immediately reinforced through practical exercises, allowing participants to learn by doing.

Day 1 focused on building foundations for collaborative work, introducing participants to FAIR software development principles and tools essential for team-based development. Through hands-on exercises with version control, participants practised collaborative workflows and repository management.

Day 2 deepened these collaborative skills by addressing code quality and reproducibility. Participants learned how to structure development projects effectively, create reproducible environments, and implement practices that make code more readable and maintainable. They applied these concepts to real-world scenarios, experiencing firsthand how good practices facilitate team collaboration.

Day 3 brought everything together with a focus on testing, documentation, and continuous integration. Participants learned how to implement automated testing strategies and create effective documentation - critical skills for maintaining sustainable software projects. The day concluded with practical sessions on open collaboration, preparing participants to contribute to and maintain shared codebases effectively.

Training Materials and Resources

The workshop was built upon established training materials from respected communities in scientific computing. The core content was drawn from The Carpentries ecosystem, with additional materials adapted from CodeRefinery's software development best practices curriculum. These materials were enhanced with industry-relevant examples and use cases to meet BioNT's mission of bridging academic and industrial practices. All materials are openly available and licensed under CC-BY 4.0, enabling future reuse and adaptation by other trainers and communities. The main lesson from the Carpentries Incubator was “Tools and practices for FAIR research software”. CodeRefinery lessons are available under CC-BY-4.0 license under https://github.com/coderefinery.

Day - Topic	Tutorial
Day 1 - Technology needed for Collaborative work	FAIR research software Tools and practices for FAIR research software development Version control Concepts around collaboration Collaborating within the same repository Practicing code review How to contribute changes to repositories that belong to others
Day 2 - Tools & practices for FAIR research software	Software licensing, Software citation, Sharing data Reproducible Research Organizing your projects Recording computational steps Recording dependencies Recording environments Reproducible development environment Code readability
Day 3 - Tools & practices for FAIR research software	Code structure Code correctness Continuous Integration for automated testing Code documentation Open code & collaboration

Technical infrastructure

Organisation of the workshop

This workshop ran for three days, from the 4th to the 6th of February 2025. The entire event was conducted virtually at no cost to participants. On all three days, the sessions were delivered from 09:00 to 16:00 CET.

Webpage and registrations

The CECAM event management platform, provided by the EPFL, was used to create a dedicated webpage for the workshop, which included the workshop description, learning objectives, requirements, program (Figure 1), and any further information relevant to potential participants. The webpage is accessible at https://www.cecam.org/workshop-details/-1447.

Figure 1 - Workshop program as displayed on the event page at the CECAM platform.

For registration, the CECAM platform was used to manage the applicant's information and communication. In parallel, the EMBL servers were used to collect pre- and post-workshop information through pseudo-anonymised surveys. The survey data was linked to the applicant's data only via a unique identifier, provided in the CECAM registration process, as well as in the EMBL-based survey. This ensured that only the workshop organisers accessed the applicants' personal data while still collecting information relevant to the workshop separately. To register, applicants had to: (i) register on the CECAM platform, (ii) complete and submit the pre-workshop survey, and finally (iii) complete the application on the CECAM platform using the unique identifier provided in the pre-workshop survey.

Applications were reviewed based on answers in the pre-workshop survey (containing no personal information). Applicants working in small and medium enterprises (SMEs) or who identified themselves as job seekers would have been prioritised if needed, but after a thorough assessment of technical and personnel capacity, all 47 applicants were accepted. The communication of the application outcome to all participants, as well as any additional communication, was performed via the CECAM platform.

Infrastructure for the workshop

Zoom

The workshop was delivered via Zoom, allowing participants to learn directly from the trainers with opportunities for real-time interaction through chat or voice. However, to facilitate video recording, participants were asked to keep their microphones muted and cameras off during presentations. All direct interactions were instead conducted in written form via a prestructured collaborative document, which allowed for anonymous participation. To serve the collaborative documents, a HedgeDoc collaborative space was set up by BIOBYTE, and was hosted on their server.

To further enhance the interactive experience, the instructors sometimes engaged in dialogue around the topics under discussion. The idea was to bring in some of the “real world” experience of the instructors when it came to the topics, and to show that the simplified examples of the hands-on can have practical impact in larger projects.

Shared documents

A **Main** collaborative document, set up by the instructors and organisers, was shared with the participants before the workshop. Each section of the workshop had dedicated *Hands-on* boxes to report on the task status, ask questions or raise issues. Helpers engaged and assisted participants by answering the questions and issues directly in this document. This Main document was updated live during the workshop. Separate boxes to answer questions were used to improve participant engagement and as an indirect learning assessment (Figure 2).

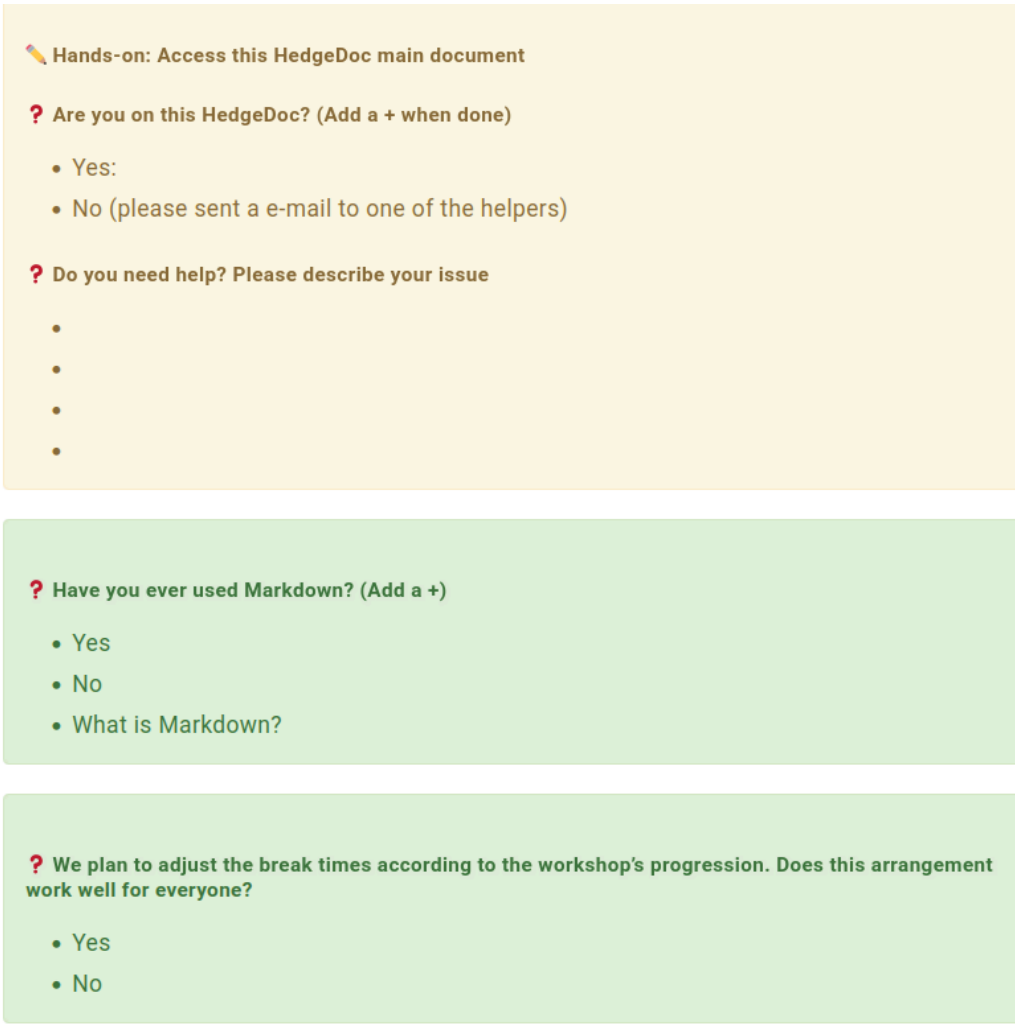

Figure 2 - Example of question and hands-on boxes in the collaborative document.

To help with the organisation, four HedgeDoc documents were used: (i) a **Template** with all instructions and boxes for hands-on, questions, etc; (ii) the **Main** document with the information for the participants during the workshop, filled with boxes related to the section covered by the instructor to help with the navigation and cleaned during each break to avoid an overcrowded document; and (iii) a **History** document collecting all the content from the Main document. This document was shared with participants during the workshop, to grant them access to all prior conversations. In addition, (iv) a document for Helpers & Instructors was created providing the workshop setup, interactions and explaining tasks of the helpers and instructors.

There was very lively interaction with the learners through the shared document. This was made possible by the extended set of helpers available to answer participant questions.

Instructors from time to time used the shared document to check the satisfaction with lesson delivery and made necessary adjustments. Type-along sessions where the learners tried to follow the instructor on their computers and demo sessions where the learners watched the instructor completing tasks were clearly communicated to the learners. End of the day

surveys conducted in the shared document revealed that many learners really enjoyed the type-along sessions.

Technical tools

The workshop required a significant set of development tools to be available to the users. These were defined as a set of prerequisite software which included Bash, Python and VS Code (with both Linux shell and Python experience being course prerequisites). It also required them to create and use ssh keys with an account on GitHub (for collaborative development).

During the workshop, the instructors used tools such as Shellshare to ensure learners could see/use the command history of the instructor in real time.

Trainees, advertising and engagement

Advertisement

The workshop was advertised via social media, several websites, mailing lists or Slack spaces of networks and communities (ELIXIR, Bioconductor, LifeSciTrainer, OLS, BioRN cluster, NFDI4Microbiota and de.NBI, among others). For the advertising of this workshop, a tailored image was generated, which included a QR code to facilitate access to the registration platform, as shown in Figure 3.

Figure 3 - Example of workshop advertisement through LinkedIn.

Trainees

Applications and pre-workshop survey

45 applicants completed the pre-workshop survey and 47 submitted their application forms via the CECAM platform. These 45 survey answers are therefore analysed and shown in the following sections. The pre-workshop survey comprised 21 questions covering skills, demographics, and miscellaneous topics. The majority of questions were optional for the successful completion and submission of the survey.

General information

Of the 45 applicants, 22 were male, 20 female and 2 preferred not to say. The country of origin and employment are summarised in Table 2.

Country	Nationality	Employment	EU / non-EU
Argentina	1		non-EU
Belgium	1		EU
Bulgaria	2		EU
Cameroon	1	1	non-EU
Cyprus	1	2	EU
Ecuador	1	1	non-EU
Egypt	1	1	non-EU
Germany	6	19	EU
Greece	3	2	EU
India	7	2	non-EU
Indonesia	1		non-EU
Italy	1		EU
Jordan	1		non-EU
Kenya	1	1	non-EU
Lithuania	1		
Luxembourg	1		EU
Myanmar	1		non-EU
Netherlands (the)	1	1	EU

Nigeria	2	1	non-EU
Norway			EU
Pakistan	1		non-EU
Portugal	1	1	EU
Qatar		1	non-EU
Serbia	1	1	non-EU
Singapore		1	non-EU
Somalia	1	1	non-EU
Spain	3	1	EU
Sweden		3	EU
Tunisia	1	2	non-EU
United Kingdom		1	non-EU
United States of America		1	non-EU
Uruguay	1	1	non-EU

Table 2 - Workshop applicants' nationality and country of employment from the pre-workshop survey.

Most applicants worked or studied in the fields of Genetics, Genomics and Bioinformatics followed by Biomedical or Health Sciences (Figure 4) and were mostly academic employees (Figure 5-A) in the category of Graduate students and research or support staff (Figure 5-B).

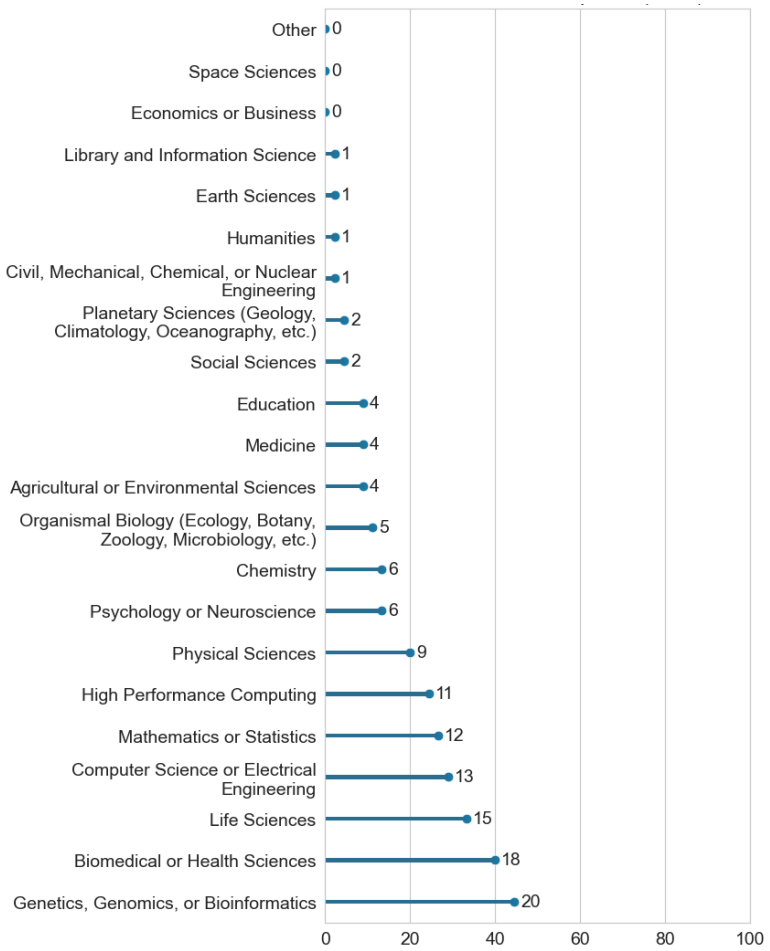

Figure 4 - The relevant fields or disciplines (multiple choice) of the applicants for n = 42, according to the pre-workshop survey.

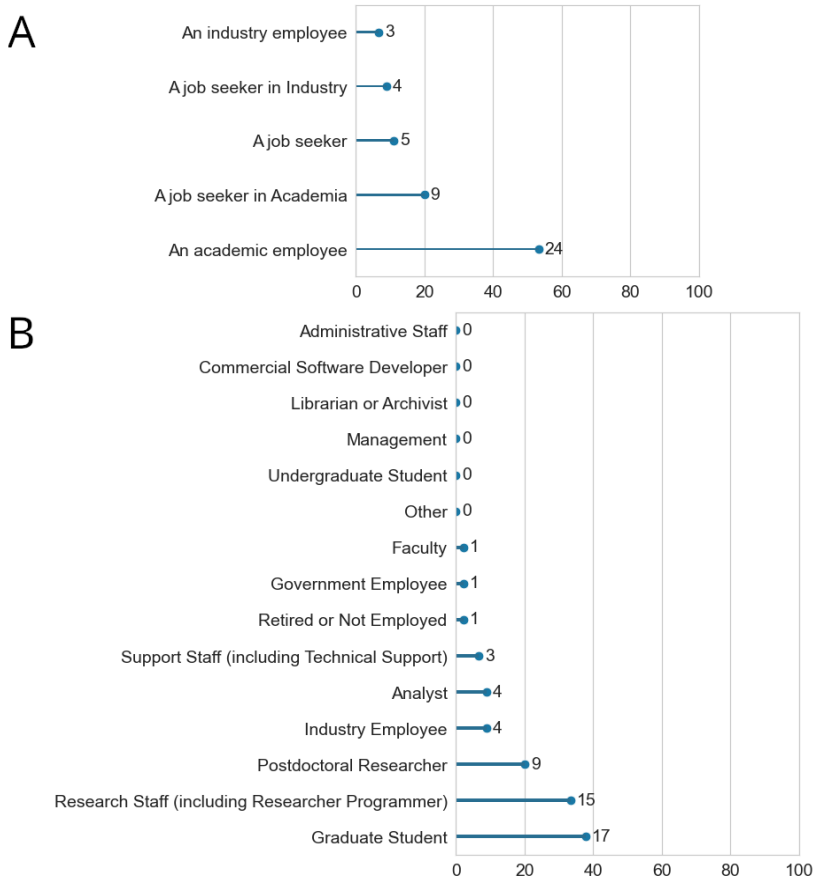

Figure 5- The current job definition of the applicants (single choice) (**A**) for n = 45, and the current occupation/career stage (multiple choice) (**B**) for n = 42, according to the pre-workshop survey.

Additionally, 3 applicants were industry employees and 18 were job seekers (with 9 in Academia, 4 in Industry and 5 either in academia or in industry) (Figure 5-A). Regarding the connections with SMEs, 6 mentioned to be working in an SME, 8 collaborating with SME and 19 aiming to work in an SME (Figure 6).

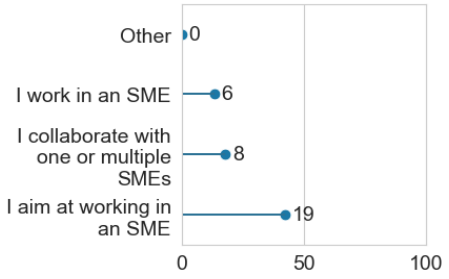

Figure 6 - Connection of the workshop participants to SMEs (n = 33).

Applicants found information about the workshop through various channels, as illustrated in Figure 7, with the majority learning about it via email, social media or receiving direct recommendations from friends or colleagues.

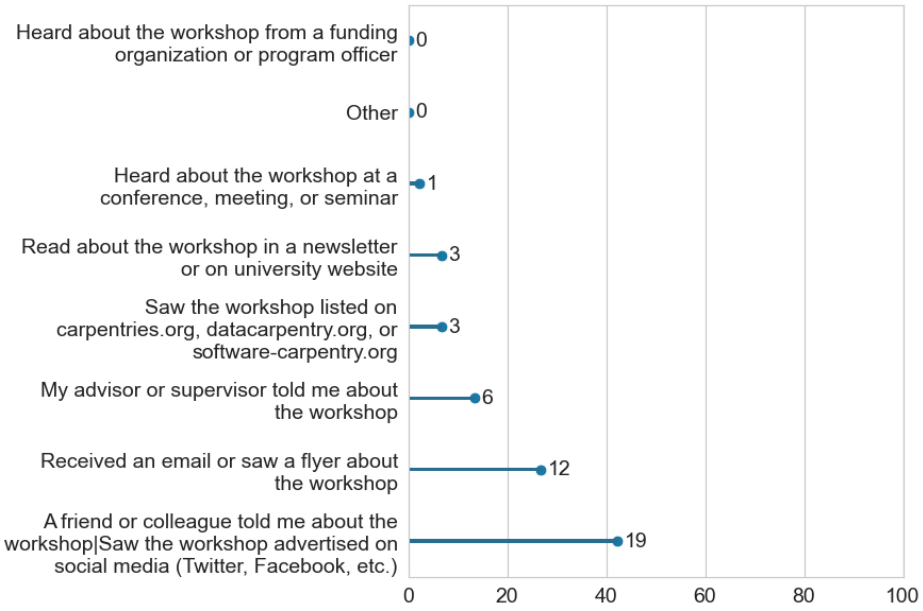

Figure 7 - Answers to the question: “How did you find out about this workshop?” (n = 39).

When asked about the frequency they use certain tools, most participants indicated to use a programming language and the terminal/PowerShell on a daily basis. On the contrary, specialised software with a point-and-click graphical user interface and databases were used only a few times per year. Figure 8 provides a general overview of the skills and interests of the trainees in this workshop, which they might transfer to others with the newly acquired tools from this workshop.

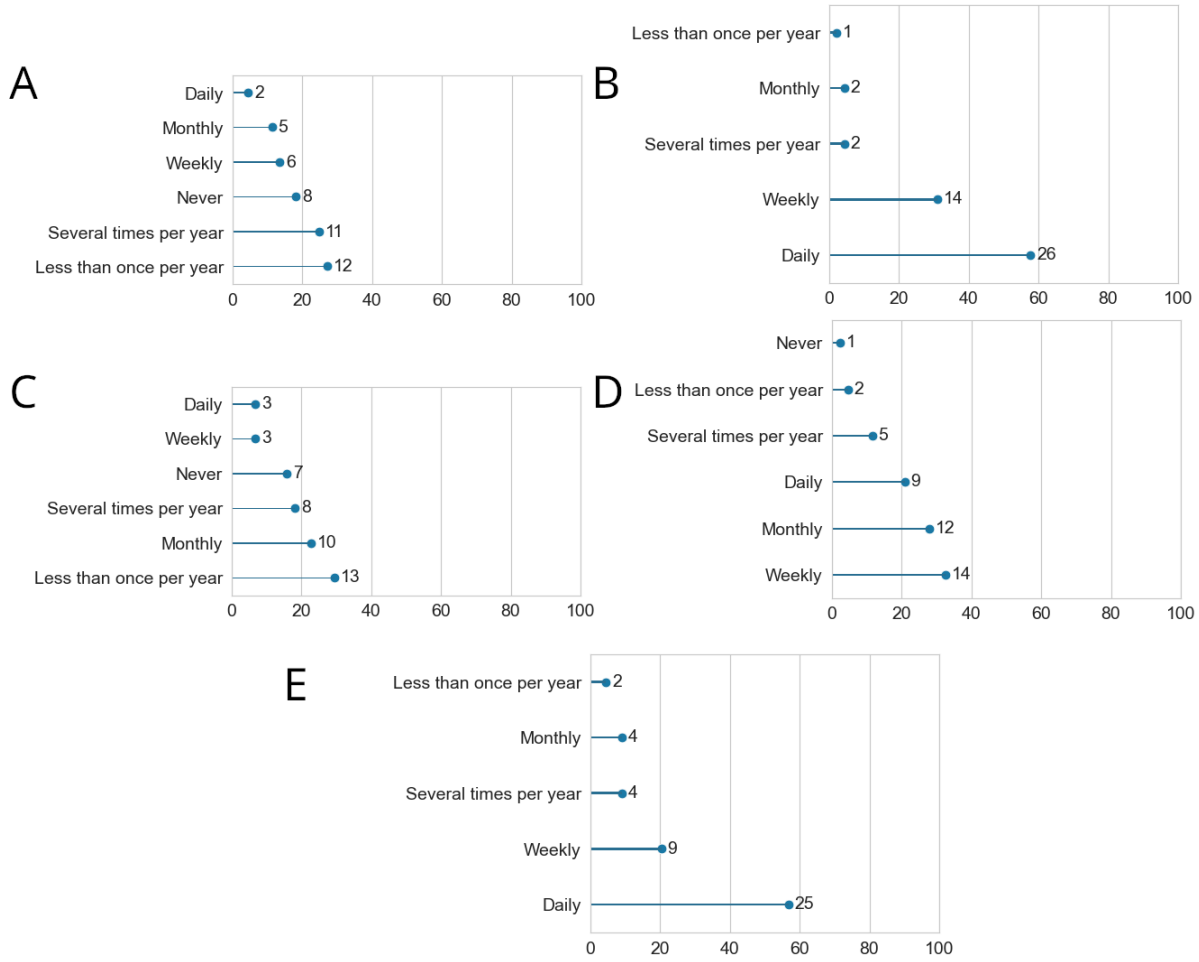

Figure 8 - Answers to the question: "How often do you use any of the following?". This question was optional and trainees could answer to none, some or all the questions (n = 44): **A** - A specialised software with a point-and-click graphical user interface (e.g. SPSS, SAS, ArcGIS, QGIS, Geneious); **B** - Programming languages: R, Python, C++, etc.; **C** - Databases (SQL, Access, etc.); **D** - Version control software (Git, Subversion (SVN), Mercurial, etc.); **E** - Terminal and macOS or PowerShell on Windows.

Most participants were keen on acquiring new skills, with some specifically interested in learning those applicable to their future job and current occupation. Additionally, 22 participants expressed their intention to leverage the acquired skills either to secure a promotion within their current job or to pursue new employment opportunities (Figure 9).

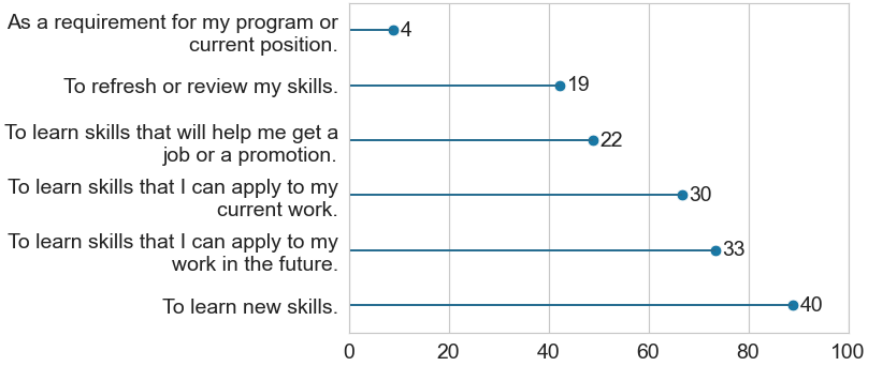

Figure 9 - Answers to the question: “Why are you participating in this workshop?” (n = 45).

Participation

All 47 applicants who submitted their CECAM application were chosen to take part in the workshop. Of these 47 applicants, a total of 45 participants attended the workshop live, although not all of them stayed for the entire duration (Table 3). All 47 received the self-learning materials for them to consult at any time.

Day	Participants	Instructors	Helpers
1	45	2	4
2	42	2	4
3	37	1	4

Table 3 - Number of participants, instructors and helpers per day. The number of participants was obtained from the participant login information captured by Zoom.

Evaluation

After the workshop, 20 participants requested a certificate (with 22 completing the post-workshop survey). In order to receive a certificate, the participant must have filled out the post-workshop survey and provide a link to the repository on GitHub where they carried out the hands-on exercises. The contents and commit history of this repository was used to evaluate the extent to which they successfully carried out the hands-on exercises. Only those who successfully carried out the majority of the hands-on exercises (either during or after the workshop) received the certificate.

In total, 18 participants were issued with a certificate. Two requests were rejected due to insufficient completion of the hands-on exercises.

Impact and outcomes

Daily feedback

At the end of each day, participants were asked for feedback on the following three points:

- Please share one thing that was good about today
- Please share one thing that could be improved about today
- Do you have any other comments?

Over the three days, the workshop received positive feedback, particularly for its clear structure and organization, with 5 participants specifically mentioning the good pacing and

well-prepared materials. The hands-on sessions were a highlight, with 10 participants appreciating the practical exercises and interactive approach.

Some issues with workflow automation topics, including Snakemake and FAIR principles, were mentioned by 2 participants, while Git and repository collaboration were raised once. Time management was noted by 6 participants, with suggestions to extend hands-on sessions and adjust the pacing of certain topics.

Overall, the workshop was well-received, with strong engagement and minor suggestions for improvement in pacing and depth

Post-workshop survey

At the end of the workshop, participants were asked to complete the post-workshop survey consisting of 15 questions, some of which were optional. In total, 22 participants completed this survey.

Regarding the workshop environment and the possibility of interacting with the trainer and helpers, the answers were overall positive (Figure 10 A-B). All participants perceived the instructors as enthusiastic about the workshop and knowledgeable about the material being taught (Figure 10 C-D). All participants could get clear answers to their questions from the instructors (Figure 10-E). Except for two neutral answers, all the participants expressed confidence in their ability to immediately apply what they learned at the workshop (Figure 10-F). Regarding accessibility requirements all participants either answered “No” or did not answer at all. Participants appreciated the instructors' attentiveness to questions, noting that even when questions were not verbalized, they were addressed through useful links, comments, or requests for clarification, demonstrating a strong commitment to the learning experience. The clear explanations and interactive discussions helped participants grasp concepts they hadn’t initially considered, making the sessions even more valuable. Many also felt that the hands-on approach significantly improved their practical skills.

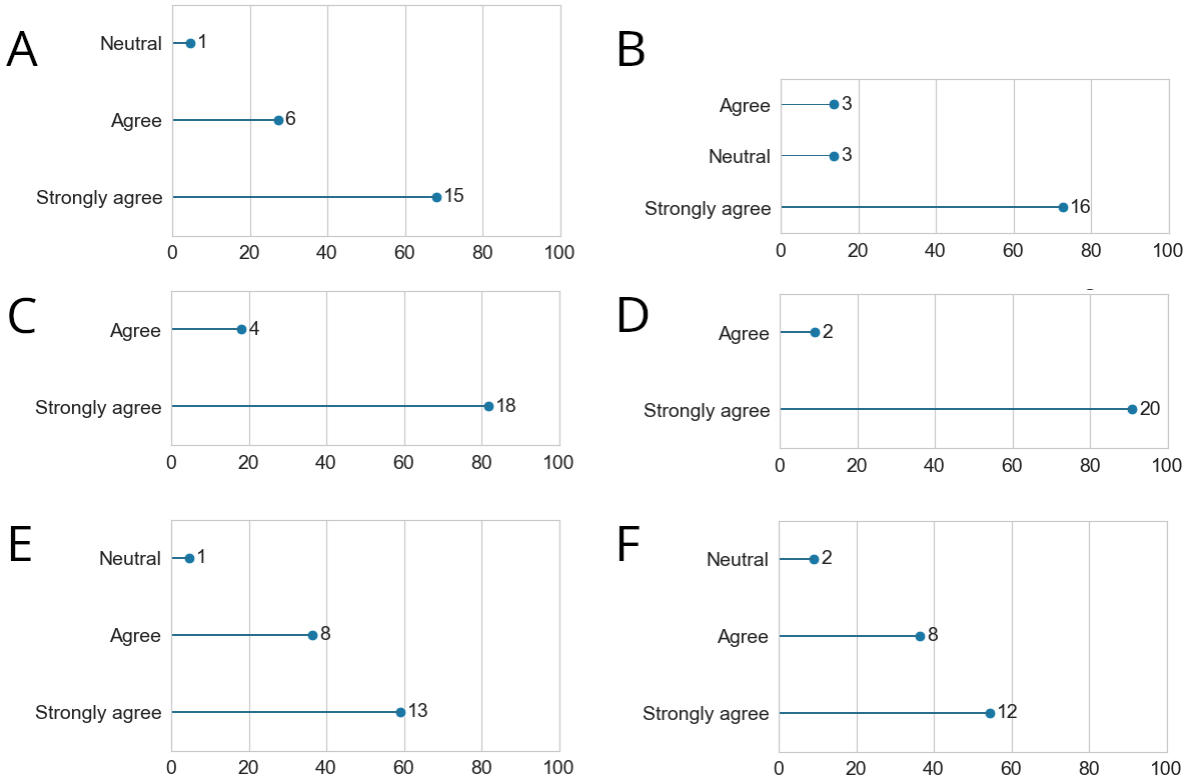

Figure 10 - Rating of participant's agreement with the following statements: **A** - I felt comfortable learning in this workshop environment (n = 22); **B** - I felt comfortable interacting with the instructors (n = 22); **C** - The instructors were enthusiastic about the workshop (n = 22); **D** - The instructors were knowledgeable about the material being taught (n = 22); **E** - I was able to get clear answers to my questions from the instructors (n = 22); **F** - I can immediately apply what I learned at this workshop (n = 22).

Participants were also asked about strengths and ways to improve the workshop.

The major strengths of the workshop have been summarised in the following points:

- Extensive hands-on practice: the guided hands-on exercises were highly valuable, allowing participants to apply concepts immediately.
- Well-prepared and structured materials: the structured approach across multiple days, with topics building on each other, made learning more cohesive and practical.
- Effective use of HedgeDoc: the HedgeDoc document was a key strength, making it easy to follow along, interact with instructors, and consolidate learning.
- Clear and responsive instruction: the instructors provided well-paced, clear explanations and were highly responsive to participants' needs, whether answering questions, sharing additional resources, or adjusting the pace when requested.
- Great balance between theory and practice

- Interactive and inclusive learning environment: the friendly and accommodating atmosphere made the workshop engaging and accessible. The use of screen-sharing tools like Shellshare was particularly appreciated, helping participants without dual screens to follow along easily.

The major areas for improvement of the workshop are summarised below:

- Time management: some sessions felt too fast, while others were too slow. Participants suggested a better balance, more structured pauses, and possibly extending the workshop to allow for more exploration.
- More hands-on practice: a few participants wanted additional time for hands-on exercises, particularly for collaborative work and workflow tools like Snakemake.
- Complex topics: some topics, such as containers, licensing, and automation, were covered too quickly or lacked depth, making them harder to follow for beginners.
- Alignment of teaching styles: participants noted that different instructors had different teaching styles, sometimes making it difficult to follow along. A more structured approach could help ensure clarity.

The feedback on the post-workshop survey was positive and participants were likely to recommend this workshop to a friend or colleague (Figure 11).

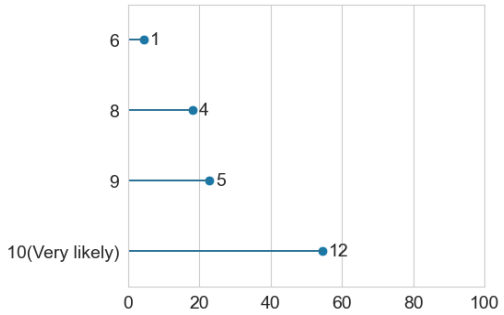

Figure 11 - Answers to the question: How likely would participants recommend this workshop to a friend or colleague? (n = 22). The answers from “0 - Very unlikely” to 4 are not shown as no responses were given in this range.1-3: No, 4-7: Maybe, 8-10: Yes.

Conclusion

The sixth BioNT workshop, “Code & Collaborate: The FAIRytale of Software Development”, was successfully held on February 4th - 6th of 2025, online and cost-free for participants.

The setup for the sixth workshop benefited from the experience gathered during the first five BioNT workshops.

Similar to the latest BioNT workshop, this event saw a good engagement from registration through its entirety. The heightened interest, as evidenced by good participation numbers, can be attributed to more extensive advertising efforts, highlighting the effectiveness of increased partner engagement in promotions. Additionally, the broad appeal of the workshop, catering to researchers, bioinformaticians, data scientists, and professionals across academia and industry, further contributed to its success

One distinguishing factor of this workshop was some of the specific training materials used were taken from the Carpentries Incubator (“Tools and practices for FAIR research software”), where lessons can be considered to be “under development”. The lesson itself was only first piloted in 2024, and had never been given by the instructor. While this presented some challenges, it is a testament to the Carpentries approach to lesson design and instructor training that this was not reflected in any way in the feedback from participants. In addition, the workshop also resulted in modifications and corrections to the lesson that were communicated to the upstream lesson repository (see https://github.com/carpentries-incubator/fair-research-software/pull/162). This also embodies the mutual benefit that can be derived through open lesson design.

One significant challenge for this course was that it is difficult to define a clear set of prerequisite knowledge for such an extended advanced course. In a scenario such as ours, the topics covered were widespread, and a number of participants were interested in only a subset (while being quite experienced in others). This diversity in the learners led to some frustration, particularly on the first day, where it was felt by some that the topics were too introductory and slow-paced, whereas these topics and pace were required by other learners. In an in-person event, this can be mitigated by pairing more experienced and less experienced learners, but this avenue is not open in an anonymous online setting.

The consortium will take the improvements and the individual challenges of the sixth workshop into consideration to further enhance the training provided by BioNT. Overall, the BioNT consortium concludes that the workshop successfully achieved its goals.