BioNT - Network for Training

Deliverable 5.1 | First year report

Co-funded by the European Union

Version 1.0

27 March 2024

Grant Agreement number	101100604 - DIGITAL-2022-TRAINING-02
Action Acronym	BioNT
Action Title	Bio Network for Training
Deliverable number and title	5.1 First year report
Work package number and title	5 Data Analysis & Information
Dissemination level	Public
Authors	R. Alves, S. Di Giorgio, L. Paladin, I. Paredes Cisneros, T. Müller, E. Ortega, approved by all BioNT partners
Delivery date	2024-03-27

Project partners

1. EUROPEAN MOLECULAR BIOLOGY LABORATORY (EMBL)

2. BIOBYTE SOLUTIONS GMBH (BIOBYTE)

3. HPC NOW CONSULTING SL (HPCNOW)

4. UNIVERSITETET I OSLO (UO)

5. UNIVERSITAT DE BARCELONA (UB)

6. INFORMATION CENTRE FOR LIFE SCIENCE (ZBMED)

7. SIMULA CONSULTING (SIMULA)

8. ALBERT-LUDWIGS-UNIVERSITAET FREIBURG (ALU-FR)

9. ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE (EPFL)

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.

First annual report

This document presents the first-year report of the BioNT project. It serves as a comprehensive overview of the project's progress, challenges and achievements in its initial phase. The report focuses on the training workshops first, describing the organising process and then exploring the details of each, the format of delivery and its implications. Subsequently, it provides information on how the activities were advertised. It describes the communities that BioNT collaborated with and provides metrics on the training events. Finally, it describes measures implemented for quality assurance and gives an overview of

the upcoming activities. To support the comprehensive overview provided in this report, readers may find below the project's timeline, encapsulating key milestones achieved during the first year (up to the Annual Meeting), as well as those to come.
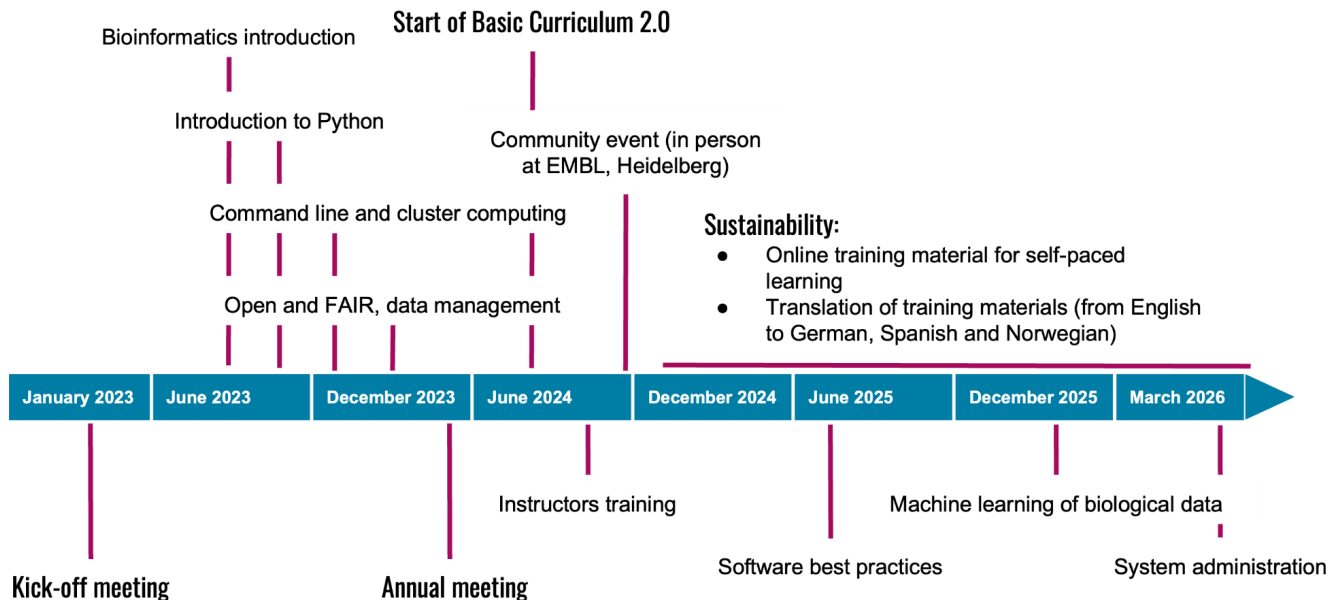

Figure 1: Timeline of BioNT's milestones

Training workshops

Organising the BioNT workshops is a detailed process that requires careful planning and coordination, driven by the consortium's collaborative effort and requiring several months. This section provides an insight into the processes involved in setting up these workshops, from shaping training content to writing the report. These efforts represent a significant part of the consortium's work over the first year, reflecting a commitment to providing valuable and effective learning experiences. The planning starts between four to six months before the training delivery date, and includes these key components:

1. Definition of general and specific training content
2. Training adaptation to BioNT delivery format (if relevant)
3. Definition of the detailed timeline for workshop organisation
4. Definition of the training team: trainers and helpers
5. Setup of workshop registration platform
6. Strategy for workshop advertisement
7. Assessment and acceptance of applications
8. Share finalised course materials (including those newly developed) with trainees
9. Open desk for technical support (if relevant)
10. Dry-run with the training team
11. Workshop delivery
12. Assessment of trainee feedback
13. Issuing of certificates
14. Writing of the deliverable report

These components collectively shape the foundation of the workshop setup process and are presented followingly.

Assessment of training needs for digital skills in life sciences

In the project proposal, the consortium emphasised the importance of crafting specialised training materials tailored to the biotech/biomed sector's workforce needs. Recognizing the critical role of understanding these needs, the consortium collaborated with like-minded entities to conduct a comprehensive Training Needs Survey. In partnership with ELIXIR and its German node, de.NBI, BioNT designed, disseminated, and analysed this survey to assess the interests and preferences of the target audience regarding training content, format, and delivery modalities.

The outcomes of the Training Needs Survey provided valuable insights into the preferences of the target audience across academia, industry, and job-seeking sectors. With responses from 68 participants, including job seekers and small-and-medium enterprise's (SME) employees, the survey identified a strong desire for training in Data Analysis, Software Usage, and Data Management. The survey's findings directly influenced adaptations to the workshop schedule and format and supported the inclusion of interactive online sessions and self-paced e-learning materials. Additionally, the results spurred discussions on refining workshop titles and content to better align with participant interests.

Content development and design

The project built on the consortium's expertise to design and curate the content for each training workshop. This involved sourcing materials from established training resources like the Galaxy Training Network and The Carpentries, while also developing new content tailored specifically to BioNT's trainees.

Bioinformatics introduction

The training workshop titled "A practical introduction to bioinformatics and RNA-seq using Galaxy" was delivered using Galaxy, an open-source platform for accessible, reproducible, and transparent computational research. The European Galaxy server (https://usegalaxy.eu) was used during the workshop, granting participants access to 3,000+ tools, and high-performance computing resources, through a standard web browser, without requiring informatics expertise. To facilitate this workshop, a dedicated job queue and resources were provided via the Training Infrastructure as a Service (TIaaS - https://usegalaxy.eu/tiaas/). This service also provided instructors with a dashboard, enabling them to monitor the participants' progress.

The workshop content consisted of four existing hands-on tutorials accompanied by three topic-introduction slides, already available at the time of the workshop delivery at the [Galaxy Training Network](link): "[Galaxy introduction](link)", "[Quality control](link)", "[Mapping](link)", and "[Reference-based RNA-seq](link)". Two additional hands-on tutorials and one [slide introduction](link) were exclusively designed for and delivered during this BioNT workshop by the consortium: "[Learning about

one gene across biological resources and formats" and "One protein along the UniProt page". Within the Galaxy Community, online training is in high demand, and the training approach developed within the consortium and built upon existing Galaxy infrastructure can serve as a blueprint for how to cater to this demand in the future.

Introduction to programming languages

The training workshop titled "From zero to hero with Python" was based on training materials already available at The Carpentries website. The Carpentries is a nonprofit organisation that teaches software engineering and data science skills to conduct efficient, open, and reproducible research. Their volunteer instructors (4287) have run 4000+ workshops in 65 countries since 2012, with 450+ alone in 2022 (data from the annual report). All their lesson materials are freely reusable under the Creative Commons - Attribution licence (CC BY 4.0) and are stored on The Carpentries website. Additionally, a web-development-focused session was exclusively developed for this workshop.

The first three days of the workshop were based on The Carpentries material "Plotting and Programming in Python", and the fourth day used the newly developed content, "GitLab and JupyterBook" and "Website hosting in GitLab".

Command line and cluster computing

This training workshop, titled "An introduction to high performance computing", also used online training materials available at The Carpentries, mostly for the first two days of the workshop: the first day covered the introduction to the Unix Shell and the second day started with the HPC-introductory tutorials. A lesson program currently in incubation for The Carpentries, the HPC Carpentry, was used for the third day, covering HPC workflow management with Snakemake. This was a customised tutorial designed to be broadly applicable and reflect the wide availability and usage of High Performance Computing (HPC) in academic and industrial environments, particularly relevant when considering the recent accelerated growth of artificial intelligence (AI) solutions that require large or specialised computation.

The trainees were granted access to an HPC resource to be used during the workshop. For this purpose, Magic Castle was used to create a virtual HPC cluster in a public cloud environment. This tool has been used for training purposes for several years by the Digital Research Alliance of Canada, and BioNT has delivered a tutorial on how to use the tool at SC23, The International Conference for High Performance Computing, Networking, Storage, and Analysis (together with the main developers).

This workshop marked the first instance of delivering the complete official lesson program for HPC Carpentry.

Open and FAIR principles, data management

This training workshop, titled "Awareness in Data Management and Analysis for Industry and Research", used online training materials from the FAIRsFAIR Adoption Handbook, and adapted training materials from ZB MED, The Carpentries and Code Refinery. FAIRsFAIR -

[Fostering Fair Data Practices in Europe](#) aims to provide practical solutions for the implementation of the FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles. [ZB MED - Information Centre for Life Sciences](#) is an infrastructure and research centre for information and data in the life sciences. ZB MED aims to ensure the national provision of information and literature in the life sciences for practical applications, teaching and research.

A section on Research Data Management was adapted from an existing slide deck by [ZB MED](#). The slides covering European data spaces were sourced from [NFDI4Health](#) (a slide deck focused on privacy issues, which itself draws from material by ZB MED and Aalto University). The section on good enough practices used content from a lesson in [The Carpentries](#) incubator. The following section on Data Management & Data Governance in the Industry was created specifically for this workshop, drawing from the FAIRsFAIR Adoption Handbook Lesson Plan 16 "Data Management and Governance in Industry and Research" and the [DAMA DMBOK 2ed](#) (Data Management Association - Guide to the Data Management Body of Knowledge 2ed). The training materials about FAIR were generated using the [Library Carpentry: FAIR Data and Software](#) from The Carpentries, albeit in its early phase of development (Pre-Alpha version), necessitating extensive adaptation. The last part of the workshop on reproducible research and how to prepare code to be usable by you and others in the future was derived from the CodeRefinery [lesson](#). The [CodeRefinery](#) provides training and infrastructure for researchers to make their research more reproducible and transparent, furthering the goals of open science and FAIR data management.

All the training materials and slides used for this workshop have been deposited in [Zenodo](#) and have been downloaded more than 180 times already.

Training delivery format

During the project's first year, the BioNT consortium mostly focused on developing a training format that would allow online and free participation from the selected trainees, ensure a comfortable environment for participants, including job seekers and SME employees, and provide a safe space for the trainer–trainee and trainee-trainee interactions, ensuring anonymity of the trainees in all live interactions.

The course registration, and therefore the participant data and communication, was handled through the [EPFL-hosted event management platform managed by CECAM](#). In parallel, EMBL servers were used to collect pre- and post-workshop information through pseudo-anonymised surveys, inspired by those used at The Carpentries. 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 workshop organisers could potentially link participants to their provided feedback in the surveys. This was needed in cases where participants expressed, for example, specific accessibility needs where tailored solutions would need to be discussed with them. 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 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 applicants for the four workshops delivered during the first year were accepted. The communication of the application outcome to all participants, as well as any additional communication, was performed via the CECAM platform.

Inspired by the training format used by CodeRefinery, the consortium opted for a Zoom-based webinar format, using the licence owned by EPFL, to stream the workshop and provide a place for trainer-helper interactions. Trainees were able to join the Zoom room without providing any personal information and with the Q&A, reactions, and chat features disabled. The communication to and interaction with the trainees during the workshop took place exclusively through a collaborative pad (HedgeDoc, hosted at the BIOBYTE servers). The training team, formed by the main trainer and 3-4 helpers, took care of keeping the pad tidy and replying to questions as they came. Several ways to interact with the trainees and assess their learning process were implemented, such as single-choice exercises completed on the pad, individual exercises, guided coding, and daily feedback where relevant.

Dissemination of BioNT's training workshops

The consortium opted for a diverse strategy to advertise BioNT's training workshops, ranging from social media accounts (LinkedIn and X) to word-of-mouth advertising. Several versions of the advertising were created, catering to the different modes of communication. The main communication channels are listed below in the following table. In all cases, potential trainees were referred to the corresponding CECAM-hosted workshop site, where the instructions to register were provided.

Type	Channels	Mode of communication
Training platform	ELIXIR TeSS	Submission of training workshop information
Communication/chat platform	Slack channels (ELIXIR, training hubs, research infrastructures, LifeSciTrainer, OLS, The Carpentries, CodeRefinery, etc.)	Message by consortium member
Social media	BioNT's social media accounts (LinkedIn and X)	Post by the management team, reposted by consortium members and collaborators
Website	Diverse: BioNT, CECAM, de.NBI, NFDI4Microbiota, HPC platform, Galaxy hub etc.	Information submitted by consortium members with connections to those websites
Mailing list	Diverse: ELIXIR training platform, de.NBI training mailing list, CASTIEL 2, The Carpentries, Oslo Bioinformatics, etc.	Information submitted by consortium members with mailing list access

Table 1: Summary of advertising channels used to advertise BioNT's training workshops.

Additionally, advertising these training workshops to the [Digital Skills and Jobs Platform](#) is a requirement of this funding call. Consortium members reached out multiple times to the platform contact points to understand how to do so and filled out the forms available in the [Training Offers](#) tab. Still, unfortunately, the training offers were never made available in the portal.

For the social media channels, tailored images were created with the support of the communications office at SIMULA, to highlight the visual branding of the project. These images are planned to be used continuously for the rest of the project, not only for workshop advertisement but for the dissemination of all BioNT's activities.

Collaborations and synergies

Throughout the past months, very active collaborations have been cultivated between BioNT and other projects and initiatives with similar objectives. These activities benefited the project from various perspectives, from the increased visibility of BioNT as a whole to the targeted advertisement of BioNT's training workshops.

ELIXIR Hub and ELIXIR nodes

One of the first joint projects with the ELIXIR Hub and the ELIXIR node in Germany, de.NBI, was related to the training needs analysis report (BioNT deliverable 2.1). Given the common interest in finding out what the training needs for digital skills in the life sciences sector are, the joint effort focused on designing a survey and assessing the results. This assessment provided useful insights into the training needs and was the base for the training needs analysis report submitted in November 2023. After assessing the initial survey, some improvements were incorporated into the questionnaire and the [updated Training Needs Survey](#) continued to be advertised by BioNT, the ELIXIR Hub and de.NBI.

An additional opportunity to interact and collaborate with the ELIXIR Hub and some of the ELIXIR nodes came from the successful application for funding to [ELIXIR's Knowledge Exchange Scheme](#). Through this scheme, the BioNT consortium was able to have its first in-person meeting in Barcelona, together with members from the ELIXIR Hub and nodes in the same countries of the consortium: Spain, Germany, Norway and Switzerland. The final report from these activities was published by the ELIXIR Hub through their news release and it is currently available on [their website](#).

Based on these collaborations, members from BioNT's management team were invited to discuss further the training needs in the industry sector during ELIXIR's Industry Advisory Committee Meeting, which took place in person (Hinxton, UK, 17th of January 2024). Valuable input was gained from the committee and was summarised in their summary report, [publicly accessible from ELIXIR's website](#).

BioNT's training workshops were advertised through the de.NBI dissemination channels and added to their website. This was particularly beneficial to reach a larger audience within the German community. de.NBI's valuable input about post-workshop training assessment will

also be further incorporated for upcoming training workshops. As an example, the dissemination for the 4th training workshop on their website is [linked here](#).

The Carpentries

All BioNT beneficiaries actively participate in The Carpentries' initiatives. For example, most trainers hold certifications as Carpentries instructors, and some have undergone additional training to become instructors' trainers. This collaboration enables BioNT to integrate and adapt The Carpentries' effective teaching methodologies into the second and third workshops.

BioNT's collaboration with The Carpentries extends also to the organisation of its Community Event, which will be jointly held with the CarpentryConnect 2024. The goal of the joint events is to bring The Carpentries and BioNT communities together to explore and discuss community-led software and data skills training as well as capacity-building initiatives across sectors. This will provide the opportunity to bring together new and experienced community members to share knowledge, network, develop new skills, and develop strategies for training beyond academia while building strong local and regional training communities.

Galaxy Training Network

During the first workshop, the BioNT community leveraged freely available Galaxy accounts for all participants, granting access to specialised computing resources through the training infrastructure service (TIaaS). The learning materials developed during the first workshop are hosted on the Galaxy Training Network (GTN) and thereby expand its collection of self-learning materials outside Galaxy. The GTN facilitated the creation of hands-on and self-learning materials, openly accessible to the entire BioNT community. Additionally, by using Galaxy and TIaaS for upcoming workshops, the Galaxy community could provide an open learning infrastructure for upcoming workshops. The BioNT's basic curriculum workshops delivered during the first year of the project were actively promoted through the Galaxy community hub.

Key performance indicators (KPIs)

In its inaugural year, BioNT successfully conducted four workshops as part of its basic curriculum, with participation from a diverse group of individuals: A total of 223 trainees in these workshops, including 47 SME employees and 59 job seekers. It is worth noting that while the total number of trainees reached 223, some individuals may have attended multiple workshops, contributing to the overall count.

Following each workshop, participants were invited to complete a post-workshop survey to provide feedback about their experience during the training activities. Among the attendees, 111 completed the post-workshop survey, indicating their active engagement and commitment to the workshop's objectives. These individuals are considered successful participants who actively contributed to the workshop experience and its overall

effectiveness. Moreover, the post-workshop survey showed that more than 90% of participants would recommend the workshop to friends and colleagues. This high recommendation rate shows the perception of the workshops and their significant contribution to participants' professional development.

Workshop	Bioinformatics Introduction	Introduction to programming languages	Command line and cluster computing	Open and FAIR principles, data management
Applications received	60	50	45	68
Participants who completed the post-workshop survey	28	21	19	44
Participants who received a certificate	16	17	15	28
SME employees who benefit from the workshop	13	9	4	8
Job seekers who benefit from the workshop	21	17	11	10
Survey participants who would recommend the course to friends and colleagues	96%	100%	90%	95%

Table 2: KPIs for the four BioNT workshops delivered during the first year, as part of the basic curriculum. The last metric was extracted from the post-workshop survey and it is related to the question "How likely would participants recommend this workshop to a friend or colleague?" (rated on a scale from 0 to 10, where 0 represents very unlikely and 10 indicates very likely, considering responses from 7 to 10 only)

Regarding data collection, deliberate decisions were made to exclude questions about the specific SME where participants were working, which implies that no data can be derived about the number of SMEs involved. This strategic choice aimed to keep the survey focused on information directly relevant to workshop organisers, especially in light of the potential personal nature of employer/employment status data.

Quality assurance

Ensuring the quality of the workshops is essential for BioNT, and this is achieved by adapting the training workshops according to the trainee's feedback and the results of the Training Needs Survey.

Participants in the specific workshops had multiple occasions to express their needs and provide feedback. During registration, they were asked to fill out a pseudo-anonymised workshop survey collecting relevant data for the instructors, such as their background

knowledge about the topics or the operating system they would be using during the workshops. Unless interacting with a remote system or a platform was part of the workshop content, to prepare for the workshop they were asked to install the needed tools locally in their machines, instructions were provided as well as dedicated meeting times to receive support in doing so. The goal of this approach is to provide all the tools to keep testing the skills they acquired during the workshop even after it and integrate them into their day-to-day activities.

Feedback plays a crucial role in the quality assurance process. Participants were invited to provide daily feedback for each workshop, with the opportunity to share one positive aspect of the day, one area for improvement, and any additional comments they may have. This daily feedback facilitates necessary improvements and adaptations in real-time. For example, adjustments included adapting the pace of the workshop or addressing accessibility issues such as screen size limitations or microphone problems. Additionally, for the last workshop, issues emerged with the collaborative document used for communication due to the high number of participants, highlighting the need to provide clear instructions on its usage for the following days. This was followed by a more detailed explanation that was positively received the following day.

At the end of each workshop, participants were asked to complete a post-workshop survey to gather additional insights and feedback for future improvements. This process was essential for evaluating the overall quality of the workshop, including aspects such as the environment, materials, and instructor knowledge. The feedback collected was carefully analysed by the instructors to identify any comments related to the structure or format of the workshop and to determine potential areas for improvement. For example, for the second workshop, a recommendation for using a two-screen setup was provided on the CECAM webpage under the "Recommendation" section. Many participants mentioned that this recommendation was not effectively communicated. Therefore, for the subsequent workshop, this information was included in all communications and emphasised at the beginning of the workshop.

This feedback will also help refine the delivery of the second round of the basic curriculum, ensuring that future sessions are adjusted to better suit the evolving needs and expectations of participants. After the first two workshops were delivered, some of the results of the Training Needs Survey were analysed. Participants in the Academia category expressed a preference for evening training sessions, organised as half-day sessions. In contrast, participants from the Industry category preferred full-day training sessions condensed into fewer total days. To cater to the specific needs of SMEs and industry professionals, the BioNT consortium decided to continue testing the delivery of training workshops at various times of the day, guided by the presented survey results. This resulted in the shortening of the last two workshops to two full days.

In conclusion, ensuring the quality of the workshops was an essential aspect of BioNT's training activities. This objective was accomplished by considering participants' feedback and examining the results of the different surveys.

Planning of upcoming activities in BioNT

Self-paced training materials

BioNT aims to provide full accessibility to the training resources in English, including written materials, video recordings, trainees' shared notes, etc. Future activities involve the curation and publication of the training resources generated during the online training sessions. This is currently being planned and organised for the workshops in the basic curriculum.

Translations of training materials

BioNT's sustainability activities involve translating the training material provided during the courses to different languages spoken in the European Union. Following the capacities of the consortium members, materials will be translated into Norwegian, Spanish and German.

Upcoming online workshops

The learning materials, existing and newly developed, for the delivery of the basic curriculum, will be used again for the second delivery, starting in 2024 and possibly extending to 2025. The feedback collected during the first delivery proved to be extremely helpful in refining the training materials and the training delivery and therefore the consortium expects similar or even better results regarding the trainee's perception. Given that during the past year, BioNT's visibility has increased, it is also expected to reach a larger audience. Parallely, the delivery of the training workshops in the advanced curriculum is scheduled to begin in June 2024 and extend to the beginning of 2025, as per the timeline presented in the proposal.

Community Event

The planning of BioNT's Community Event started in the first year of the project. Given the shared interests and goals of the two initiatives, and to enhance visibility and appeal within the training community, BioNT collaborated with The Carpentries community: BioNT's Community Event will take place jointly with the CarpentryConnect. The joint event will be hosted by the EMBL in Heidelberg, Germany, from the 12th to the 14th of November 2024.

The CarpentryConnect and BioNT community event - Heidelberg 2024 aims to be the key community-building and networking event for The Carpentries and BioNT's community in Europe, with worldwide participation. Under the theme of "Community-led Training Beyond Academia," the event underscores its focus on bringing together both new and experienced community members to exchange knowledge, network, cultivate new skills, and strategize for training beyond academia while fostering robust local and regional training communities.

The event's provisional program is available here and it includes four types of sessions: Breakout sessions, skill-up or training sessions, workshops and lightning talks/posters followed by a day and a half dedicated to the mini-hackathon and curriculum development session.