

DELIVERABLE 3.1 - TRAINING DELIVERY REPORT



FEBRUARY 11th 2026

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	3.1 - Training delivery report
Work package	3 - Training delivery
Dissemination level	Public
Authors	T. Ninkovic, I. Paredes Cisneros, S. Razick, P. Samarakoon, H. Vega
Delivery date	2026-02-11

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

About this report.....	2
Consortium members.....	2
Table of Contents.....	3
1. Overview of BioNT training activities.....	5
2. Basic curriculum and lessons learnt from its two iterations.....	7
3. Advanced Curriculum - Workshop-specific analysis.....	8
Industry-focused events.....	8
To increase interest in these advanced workshops among industry-linked professionals, we organized two industry-focused events prior to the two last workshops. These industry-focused events served as pre-workshop engagement activities to attract target audiences, particularly SME employees and job seekers, whilst providing opportunities to gather feedback that informed workshop adaptation. Links to the workshop online description and registration.....	8
I. Strategies for training and knowledge exchange with a consulting perspective.....	9
Overview.....	9
Program and Materials.....	9
Support and Certification.....	9
Participation during the workshop.....	9
Table 3: Participation - Strategies for training and knowledge exchange with a consulting perspective.....	9
Updates and Improvements.....	9
Participant Feedback.....	10
Key Achievements.....	10
II. Software Best Practices: Code & Collaborate – The FAIRytale of Software Development.....	11
Overview.....	11
Program and Materials.....	11
Support and Certification.....	11
Participation during the workshop.....	11
Table 4: Participation - Software Best Practices: Code & Collaborate.....	11
Updates and Improvements.....	11
Participant Feedback.....	12
Key Achievements.....	12
III. Applied Machine Learning for Biological Data.....	12
Overview.....	12
Program and Materials.....	13
Support and Certification.....	13
Participation during the workshop.....	13
Table 5: Participation - Applied Machine Learning for Biological Data, Module 1	13
Table 6: Participation - Applied Machine Learning for Biological Data, Module 2	14

Updates and Improvements.....	14
Participant Feedback.....	15
Key Achievements.....	15
IV. System Administration for HPC.....	15
Overview.....	15
Program and Materials.....	15
Support and Certification.....	16
Participation during the workshop.....	16
Table 7: Participation - System Administration for HPC.....	16
Updates and Improvements.....	16
Participant Feedback.....	16
Key Achievements.....	17
4. Impact assessment across BioNT’s training programme.....	17
Survey methodology and response rate.....	17
Tool usage and skill adoption.....	18
Career and professional development outcomes.....	18
Qualitative impact evidence.....	19
5. Summary tables.....	20
Overview of learning objectives for the basic and advanced curricula.....	20
Key statistics of the advanced workshops.....	22
Overview of the study materials’ content and the source of these materials.....	23
Features applied for optimised learning.....	24
Participant feedback.....	25
Cross-programme summary of impact assessment.....	26
6. Conclusion and sustainability considerations.....	27
Sustainability of training materials.....	27

1. Overview of BioNT training activities

BioNT's training programme comprised two complementary curricula designed to address digital skills gaps across the biotechnology and biomedical sectors:

- The basic curriculum targeted professionals, job seekers, and small and medium enterprise (SME) employees requiring foundational competencies in bioinformatics, programming, command-line computing, and FAIR data management principles. This curriculum was delivered twice: Round 1 spanned Q1 2023 to Q3 2024, whilst Round 2 ran from Q4 2024 to Q1 2025, with improvements implemented based on feedback from the initial delivery.
- The advanced curriculum, introduced in the second project year and delivered between Q2 2024 and Q4 2025, focused on equipping data stewards, system administrators, and computational life science professionals with specialised expertise in instructor training methodologies, software best practices, machine learning applications, and system administration for high-performance computing.

To enhance engagement with industry participants, a key component of BioNT's target audience alongside job seekers, the consortium organised two industry-focused events: an "AI in Biomedical Data: Bridging Innovation & Opportunity" seminar (7 May 2025, hybrid format, Oslo) preceding the machine learning workshop, and a "Sysadmins in HPC" meetup (10 October 2025, hybrid format, Barcelona) preceding the system administration workshop. These pre-workshop events successfully attracted professionals from industry, academia, and SMEs, thereby broadening participation beyond traditional academic audiences.

This report presents a comprehensive analysis of lessons learnt from the two iterations of the basic curriculum, followed by detailed workshop-specific analyses for each of the four advanced curriculum workshops and two industry events, and cross-programme impact assessment documenting medium-term outcomes. Detailed analysis of the individual basic curriculum workshops can be found in the [Deliverable 5.2 - second year report](#).

For readers seeking condensed information about workshop delivery, Section 5 provides summary tables consolidating key performance indicators, study materials, improvements implemented, and participant feedback across all training activities. These consolidated data provide insight into the programme's overall impact and effectiveness in reaching and serving BioNT's target communities.

Table 1: Overview of BioNT courses, with dates and key statistics and performance indicators for each of the two curricula (basic and advanced). For the basic curriculum, two rounds were delivered and are reported as Round 1 (R1) and Round 2 (R2).

Workshop title	Date	# Participants/ # applicants	# trainers & helpers	Work in SME Collaborate with SME Aim to work in SME*
----------------	------	---------------------------------	-------------------------	---

B A S I C	A practical introduction to bioinformatics and RNA-seq using Galaxy	R1: 4-8.9.2023 R2: 10-13.9.2024	R1: 73 / 73 R2: 61 / 170	7	R1: 11 11 18 R2: 16 27 69
	From zero to hero with Python	R1: 21-24.11.2023 R2: 17-18.10.2024	R1: 49 / 49 R2: 63 / 107	6	R1: 9 6 13 R2: 12 15 34
	An introduction to high performance computing	R1: 6-8.2.2024 R2: 4-5.12.2024	R1: 45 / 45 R2: 62 / 62	6	R1: 4 9 20 R2: 4 13 28
	Awareness in data management and analysis for industry and research	R1: 5-6.3.2024 R2: 6-7.3.2025	R1: 55 / 55 R2: 38 / 38	7	R1: 8 15 17 R2: 1 10 15
A D V A N C E D	Strategies for training and knowledge exchange with a consulting perspective	25-26.6.2024	24 / 24	7	1 5 9
	Software Best Practices: Code & Collaborate – The FAIRytale of Software Development	4-6.2.2025	47 / 47	7	6 8 19
	Applied Machine Learning for Biological Data	2-6.6.2025	40 / 125	12	3 7 29
	System administration for HPC	21-23.10.2025	30 / 45	8	9 11 11

*SME relationships: Assessed through a multiple-choice question in the registration survey. Due to the multiple-choice format, the total number of responses may exceed the total number of participants.

Table 2: Overview of BioNT industry-focused events

Name of the event	Date	Primary workshop supported	Location	Number of participants
BioNT Industry seminar "AI in Biomedical Data: Bridging Innovation & Opportunity".	07.05.2025	Applied Machine Learning for Biological Data	Hybrid, Oslo, Oslo Science Park, Norway	75, including speakers from Norway, France, and Finland
BioNT Industry meetup "Sysadmins in HPC"	10.10.2025	System administration for HPC	Hybrid, in Barcelona Supercomputing Centre, Spain	60, mostly local talent in industry, academia and SMEs.

2. Basic curriculum and lessons learnt from its two iterations

As shown in Table 1, the basic curriculum included four courses:

1. A practical introduction to bioinformatics and RNA-seq using Galaxy;
2. Introduction to Programming Languages: From zero to hero with Python;
3. Command Line and Cluster Computing: An introduction to high performance computing;
4. Open and FAIR Principles, Data Management: Awareness in data management and analysis for industry and research

The courses in the basic curriculum were delivered twice, which provided a space for implementing lessons learned and improving course delivery, participants engagement and satisfaction. **For the full overview of these courses, please refer to the [Deliverable 5.2 - second year report](#).**

The second iteration (Round 2) of the basic curriculum relative to Round 1 benefited from

1. improved communication and advertising,
2. improved pedagogical methods used in the courses.

These together led to an increased number of applications for the courses. For instance, the RNA-seq workshop in Round 2 received 170 applications for a maximum capacity of roughly 60 seats. Moreover, Round 2 workshops demonstrated a notable increase in applicants identifying as job seekers in the industry and first-time job seekers compared to Round 1 ([Deliverable 5.2 - second year report](#); Table 5, Figure 7). This surge of applicants from diverse categories demonstrated the increased demand for the consortium's offerings. However, this high volume led to challenges in certain instances. As a result, a selection process aligning with the project objectives was introduced to manage participant registration. This selection strategy focused on prioritising applicants from European countries who were either job seekers or currently employed by or affiliated with SMEs. As further detailed in later sections, two workshops in the advanced curriculum received applications well beyond their maximum capacity. Therefore, we applied the process to select the participants, aligning with the goals of the Digital Europe Programme. This selection mechanism allowed the consortium to convert a globally diverse applicant pool into a European-centric participant cohort with a special focus on SMEs and job-seeker status.

Relative to Round 1, data from all the surveys in Round 2 revealed an increased self-reported likelihood of recommending workshops to participants' colleagues or friends ([Deliverable 5.2 - second year report](#); Figure 8).

This upward trend indicates that the pedagogical and logistical improvements implemented were effectively translated into higher satisfaction scores in Round 2. Coupling these improvements with feedback from Round 2 of the basic courses helped the consortium to further refine the design and delivery of the advanced curriculum workshops. The

subsequent sections will detail these specific improvements and present key performance indicators (KPIs) to illustrate the training program's development.

3. Advanced Curriculum - Workshop-specific analysis

The advanced curriculum was delivered between Q2 2024 and Q4 2025, comprising four specialised workshops designed to complement and extend the foundational skills developed in the basic curriculum. Each workshop targeted specific professional competencies required by data stewards, system administrators, and computational life science professionals working in biotechnology and biomedical sectors.

- [Strategies for training and knowledge exchange with a consulting perspective](#)
- [Software Best Practices: Code & Collaborate – The FAIRytale of Software Development](#)
- [Applied Machine Learning for Biological Data](#)
- [System administration for HPC](#)

Industry-focused events

To increase interest in these advanced workshops among industry-linked professionals, we organized two industry-focused events prior to the two last workshops. These industry-focused events served as pre-workshop engagement activities to attract target audiences, particularly SME employees and job seekers, whilst providing opportunities to gather feedback that informed workshop adaptation. Links to the workshop online description and registration

- [Industry seminar - AI in Biomedical Data: Bridging Innovation & Opportunity](#)
 - Pre-workshop event focussing “Applied Machine Learning for Biological Data”
- [Industry meetup - Sysadmins in HPC meet-up](#)
 - Pre-workshop event focussing “System administration for HPC”

The following sections provide detailed analyses of each training activity, documenting delivery methods, participant engagement, iterative improvements, and key achievements that contributed to the programme's overall impact.

I. Strategies for training and knowledge exchange with a consulting perspective

Overview

The "Strategies for training and knowledge exchange with a consulting perspective" workshop was delivered during the project on June 25-26, 2024 and led by ZBMED. In total, 24 participants registered via the CECAM event management platform for the course. All applied participants were accepted for the course as capacity allowed.

Program and Materials

The program was slightly adapted from previous workshops to include a specific focus on consulting, mentoring, and community engagement, needing a higher level of participant interaction. The training materials were sourced from The Carpentries, ELIXIR, and then specialised content was developed for this workshop. All instruction was provided by members of the BioNT consortium, who are experts in diverse skill-sharing, such as training trainers and consulting in bioinformatic-related topics.

Support and Certification

Participants had the opportunity to engage with helpers and instructors via a prestructured collaborative document (HedgeDoc) and Zoom breakout rooms to resolve questions and participate in interactive exercises. Certificates were awarded to those who attended at least one live Zoom session and completed the post-workshop survey. Out of 13 participants who requested a certificate, 13 were granted one (matching the number of post-workshop survey completions).

Participation during the workshop

Table 3: Participation - Strategies for training and knowledge exchange with a consulting perspective

Day	Participants	Instructors	Helpers
1	19	4	4
2	18	4	3

Updates and Improvements

Updates and improvements were made based on the lessons learned in the basic curriculum:

- Specific material creation: Specific training materials were designed to cover unique objectives like consulting, knowledge exchange and community management.
- Pace adjustment: The workshop format was adjusted to accommodate more trainer-participant interaction and discussion.
- Privacy-focused interaction: While the frontal lesson sessions were recorded and participants' questions and comments were handled as usual through HedgeDoc, discussion sessions were not recorded to guarantee the privacy of learners. Even during discussion sessions, the HedgeDoc document was monitored and the usage of camera and microphone was optional, to still allow for full anonymity to the participants that desired so.

Participant Feedback

- The feedback from the workshop was positive ([Deliverable 1.5 - Instructor Training](#); Figure 11). Participants expressed satisfaction with:
 - Instructors and helpers: Described as enthusiastic, knowledgeable, friendly, and accommodating.
 - Learning environment: Participants felt comfortable and praised the respectful communication and the effective use of the collaborative HedgeDoc platform.
 - Applicability: Except for one neutral response, all participants expressed confidence in their ability to immediately apply the skills learned.
- Feedback on the likelihood of recommending the workshop also showed positive responses, with over 76.9% of participants assigning a rating between 8 and 10 ([Deliverable 1.5 - Instructor Training](#); Figure 12).

Key Achievements

A key achievement of this workshop was to balance the focus on soft skills and technical coding elements, which was a gap in standard bio-training. Another achievement was that we managed to achieve the balance between privacy and interactions, which was challenging considering how typical “train-the-trainer” workshops are usually delivered. The workshop was designed (and achieved) to encourage interactions among advanced trainees to not only develop technical skills useful for their professionalisation, but also to self-reflect on the strength and weaknesses of their approaches and attitudes in (peer) education, especially in its applications to digital competencies.

II. Software Best Practices: Code & Collaborate – The FAIRytale of Software Development

Overview

The workshop, titled "Code & Collaborate: The FAIRytale of Software Development," took place from February 4th to 6th, 2025, led by UB. In total, 47 participants registered via the CECAM platform for the workshop and all the applicants were accepted as capacity allowed.

Program and Materials

The program was an intensive three-day virtual event focused on transforming individual coding practices into collaborative development skills. The learning materials were taken from The Carpentries Incubator ("Tools and practices for FAIR research software") and CodeRefinery ([detailed in deliverable 1.7 - software best practices](#)). The trainers were experienced instructors from the BioNT consortium partners who incorporated "real-world" experience into the lessons.

Support and Certification

Participants had the opportunity to use a collaborative HedgeDoc document to resolve issues, ask questions anonymously, and receive real-time assistance from a dedicated team of helpers. Certificates were awarded to those who completed the post-workshop survey and submitted a GitHub repository containing the completed hands-on exercises. Out of 20 participants who requested a certificate, 18 were granted certificates.

Participation during the workshop

Table 4: Participation - Software Best Practices: Code & Collaborate

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

Updates and Improvements

Updates and improvements were made based on the lessons learned during previous workshops:

- Adoption of already existing Materials: This workshop used materials from The Carpentries Incubator that were still "under development," successfully piloting them for a diverse audience.
- Upstream contributions: The workshop resulted in direct modifications and corrections to the lesson materials, which were communicated back to the upstream lesson repository.
- Enhanced advertising: A tailored advertising strategy involving partners and social media (LinkedIn) led to a diverse applicant pool compared to previous courses.

Participant Feedback

- The feedback from the workshop was resoundingly positive ([Deliverable 1.7 - software best practices](#); Figure 10). Participants expressed satisfaction with:
 - Instructors and helpers: Participants rated instructors highly for their enthusiasm, knowledge, and ability to provide clear answers to questions.
 - Learning environment: The environment was described as interactive and inclusive, with tools like Shellshare helping those without dual screens follow along easily.
 - Applicability: 91% of participants reported feeling confident to apply their new skills immediately after the workshop.

Key Achievements

A key achievement of this workshop was the successful engagement of a diverse group of professionals, particularly job seekers and SME-connected individuals. Specifically, the participant group included 19 job seekers in industry and 14 individuals working in or collaborating with SMEs. This allowed for alignment with the project's goal of bridging academic and industrial practices, ensuring the training was relevant to both sectors. Additionally, the workshop attracted a global audience, covering individuals from a broad range of European and non-European countries ([Deliverable 1.7 - software best practices](#); Table 2).

III. Applied Machine Learning for Biological Data

Overview

The "Applied Machine Learning for Biological Data" workshop was delivered during Q3 2025 on May 27th–28th (Module 1) and from June 2nd–6th (Module 2), led by UO. Module 1 served as an intermediate Python course, designed to bridge the gap between the foundational Python skills taught in the basic curriculum and the more advanced requirements of Module 2. This module was offered optionally, welcoming all participants without restriction, for those aiming to enhance their existing Python knowledge.

Module 2 covered advanced Machine Learning (ML) concepts and real-world hands-on sessions, GPU-accelerated Next-Generation Sequence (NGS) data processing theory and practical sessions. Module 2 was mandatory for workshop participants. In total, 125 applicants registered via the CECAM platform for Module 2. Based on the selection criteria, which prioritised applicants working in European countries who were job seekers, SME employees, or aiming to work in SMEs, 40 participants were accepted for Module 2 due to computational resource constraints.

Program and Materials

Module 1 consisted of two half-day sessions, held in the mornings. The theoretical content and short exercises were delivered during these morning sessions. The five-day virtual Module 2 was structured as an intensive training event. These long hands-on sessions (ranging from one to three hours) required participants to work with real-world biological datasets. To enhance interaction and learning, the practical hands-on sessions incorporated coding demonstrations, self-coding and frequent quizzes.

The learning materials were developed from scratch to specifically address real-world biological use-cases and cutting-edge developments in the field (e.g., Accelerated - NGS). Materials were designed to align with the project's mission of bridging academic and industrial practices. The trainers were experienced lecturers from the field, including industry practitioners and experts.

Support and Certification

Participants had the opportunity to use a collaborative HedgeDoc document to resolve issues and ask questions anonymously, supported by an extended set of helpers. The collaborative HedgeDoc document was used for the quizzes in the interactive hands-on sessions. For the advanced module, participants were provided with virtual machines equipped with GPUs. A homework assignment was provided at the end of Module 2, which involved data preparation, model implementation, and evaluation. Certificates were awarded to those who scored above 50 on the homework assignment. Out of 26 participants who requested a certificate, 26 were granted certificates.

Participation during the workshop

Module 1:

Table 5: Participation - Applied Machine Learning for Biological Data, Module 1

Day	Participants	Instructors	Helpers
Session 1	29	1	3
Session 2	28	1	3

Module 2:

Table 6: Participation - Applied Machine Learning for Biological Data, Module 2

Day	Participants	Instructors	Helpers
1	39	3	5
2	34	3	5
3	35	3	5
4	35	3	5
5	30	2	3

Updates and Improvements

Updates and improvements were tailored to maximise the experience of each aspect of the workshop:

- Industry meetup: An industry pre-workshop seminar was held to achieve two main goals: first, to engage the intended target audience and collect feedback to inform the adaptation of Module 2, and second, to promote the upcoming workshop as an important chance to gain new skills. The industry seminar and the BioNT advertising campaign helped attract learners with moderate to advanced general computational skills, who are residing in Europe and affiliated with SMEs ([Deliverable 1.8 - Machine Learning of Biological Data](#); Figure 03, 05, 07).
- Specialised datasets and infrastructure to facilitate the training:
 - Datasets for theoretical sessions: Minimal datasets were used for short exercises during theoretical sessions.
 - Datasets for ML hands-on: For machine learning hands-on sessions, large, real-world datasets were downsampled (a small subset was extracted) to allow learners to complete ML workflows on their personal computers without exceeding resource limits. The link to the source dataset was provided for independent exploration.
 - High-Performance Computing: Sessions on GPU-accelerated NGS processing were supported by providing participants with individual virtual machines equipped with GPUs, enabling them to complete the accelerated genomics hands-on training.
- Real-time program adaptation: Based on daily feedback, hands-on sessions were adjusted during the workshop to include more self-coding time and quizzes to facilitate better understanding.
- Increased outreach to associations of Digital and life science SMEs across Europe.

Participant Feedback

- Participants responded with positive and expressed their satisfaction with ([Deliverable 1.8 - Machine Learning of Biological Data](#); Figure 10):
 - Instructors and helpers: Participants consistently rated instructors as knowledgeable, enthusiastic, and approachable, with ~96% agreeing they received clear answers to questions.
 - Learning environment: 95.9% of participants felt comfortable in the workshop environment, which was described as supportive and conducive to learning.
 - Applicability: 75% of participants expressed confidence in their ability to immediately apply the learned material to their work or research.
- Feedback also showed high recommendation rates, with over 92% of participants assigning a rating between 8 and 10 ([Deliverable 1.8 - Machine Learning of Biological Data](#); Figure 11).

Key Achievements

The workshop successfully achieved its objective of serving the target audience, as defined by the project objectives. This was evident in the highly specific participant demographic, which consisted of individuals residing in Europe who were either currently employed by, affiliated with, or seeking employment with SMEs. (as further detailed in [Deliverable 1.8 - Machine Learning of Biological Data](#); Figure 5).

Another key achievement of this workshop was the significant improvement in participants' self-reported skills across all measured competency areas ([Deliverable 1.8 - Machine Learning of Biological Data](#); Figure 13). This allowed for better integration of Machine Learning approaches into participants' research workflows.

IV. System Administration for HPC

Overview

The System Administration for HPC workshop was delivered during the third year of the project on October 21-23, 2025 and led by HPCNOW. In total, 45 participants were initially selected from the applicants via the CECAM event management platform for the course. Based on the selection criteria (prioritising learners working in European countries who were job seekers, SME employees, or SME collaborators), on a first-come-first-served basis, 30 participants were finally allocated to the course due to the limit on available resources.

Program and Materials

The program was delivered for the first time, so the curriculum was developed from scratch to cover Linux fundamentals, Slurm, containers, automation, and monitoring. The learning materials were taken from materials developed specifically for this workshop (openly

available on Zenodo), and the trainers were invited professionals in the field of High Performance Computing.

Support and Certification

Participants had the opportunity to engage with helpers via a prestructured collaborative document (HedgeDoc) to resolve technical issues and ask questions anonymously. Certificates were awarded to those who fulfilled specific technical criteria on their VMs: creating operational users, installing Warewulf, having a working 'sinfo' command, running monitoring dockers, installing Ansible with correct playbooks, and having a working 'module avail' command. Out of 20 participants who requested a certificate, 20 were granted one.

Participation during the workshop

Table 7: Participation - System Administration for HPC

Day	Participants	Instructors	Helpers
1	33	3	5
2	32	2	5
3	30	3	5

Updates and Improvements

As this was the first delivery of this specific workshop, updates were made during the event itself and based on the project's prior experience:

1. Real-time content adjustment: Based on feedback from Day 1, the content for Days 2 and 3 was adjusted to reduce copy-pasting and allow participants to focus more on instructor explanations, minimising hands-on errors.
2. Infrastructure planning: The setup benefited from experience gathered during the first eight BioNT workshops, particularly regarding audience prerequisites and outreach.
3. Future syllabus refinement: Post-workshop feedback identified a need to balance dense theoretical sessions with more time for hands-on practice in future iterations.

Participant Feedback

- Feedback showed a high recommendation rate with ~95% of respondents rated the workshop 8 or higher ([Deliverable 1.9 - System administration](#); Figure 11).
- The feedback from the workshop was also positive, with exceptionally high satisfaction levels ([Deliverable 1.9 - System administration](#); Figure 10). Participants expressed satisfaction with:

- Instructors and helpers: Participants praised the instructors' enthusiasm, approachability, and clarity. 100% of respondents agreed the instructors were knowledgeable about the material.
- Learning environment: All participants (100%) agreed or strongly agreed that they felt comfortable in the workshop environment.
- Applicability: 73% of participants agreed they could immediately apply the learned material to their work or research.

Key Achievements

A key achievement of this workshop was the successful deployment of personal mini HPC clusters by participants, which significantly boosted their self-reported skills and confidence in system administration. The group of workshop learners consisted of a mix of predominantly experienced technical staff (~60% in HPC/CompSci) and researchers, with a strong connection to SMEs (nearly all participants were employed by, collaborating with, or aiming to work in an SME). This allowed for effective targeting of the desired demographic and alignment with the project's goal to professionalise data management skills. Data collection for this impact assessment remains ongoing to capture additional long-term outcomes as more participants reach the six-month post-training milestone.

4. Impact assessment across BioNT's training programme

To evaluate the medium-term impact of BioNT training activities across both curricula, the consortium conducted an impact assessment survey targeting all past participants approximately six months after workshop completion. The survey assessed tool usage patterns, career and professional development outcomes, and collected qualitative testimonials documenting how participants applied their training in professional contexts.

Survey methodology and response rate

The impact survey was distributed to all BioNT participants who had completed workshops between September 2023 and November 2025, covering both basic and advanced curriculum sessions. The survey received 45 responses, representing approximately 9% of total participants across the programme duration. Response distribution reflected workshop attendance patterns, with Python workshops (31.1%), Data Management workshops (20.0%), and RNA-seq Galaxy workshops (15.6%) contributing the highest response counts. Whilst the response rate presents limitations for generalising findings across the entire participant population, the testimonials and reported outcomes provide valuable qualitative insights into training effectiveness and real-world application.

Tool usage and skill adoption

Survey data demonstrated substantial increases in tool usage frequency following training participation. Before attending workshops, 42.2% of respondents (19 of 45) reported never using the covered tools or resources. Following training, this proportion decreased to 22.2% (10 of 45), representing a 47% reduction in non-usage. Notably, the proportion of participants reporting frequent tool usage (several times per week or daily) increased from 11.1% (5 respondents) before training to 33.3% (15 respondents) after training; a 200% increase in frequent users. Overall, 77.8% of respondents (35 of 45) reported using trained tools at least occasionally in their professional work following workshop completion.

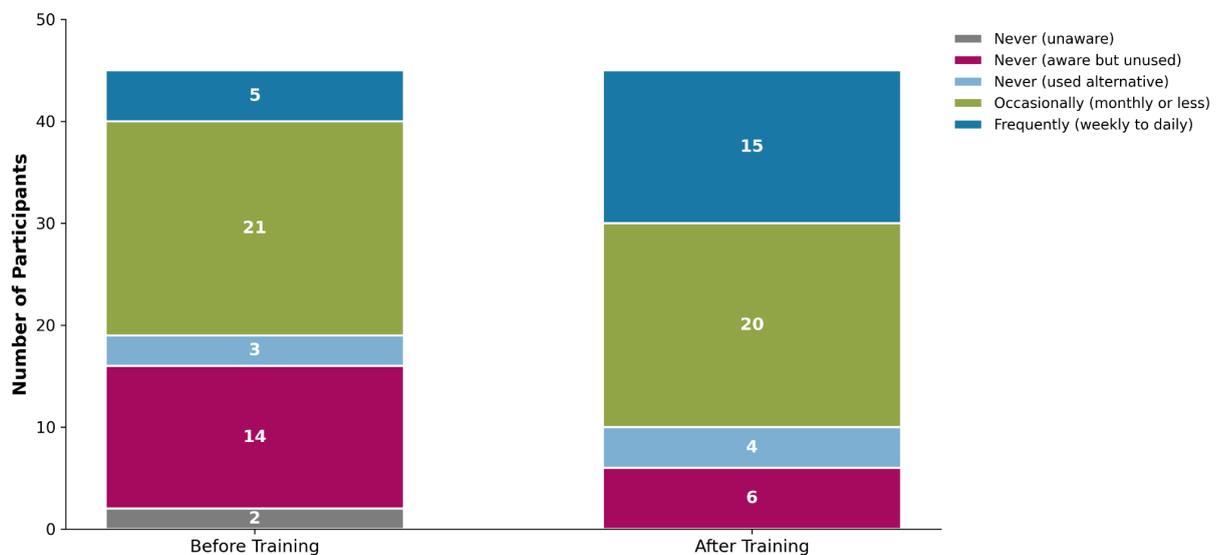


Figure 1: Tool usage frequency before and after training participation (N=45)

Career and professional development outcomes

Participants reported multiple categories of professional impact resulting from BioNT training. The most frequently cited outcome was facilitation of work development or publication (46.7%, 21 respondents), indicating that training directly contributed to research outputs and professional productivity. Career transitioning or employment improvement was reported by 26.7% (12 respondents), demonstrating tangible career advancement outcomes. Additionally, 11.1% (5 respondents) indicated that training assisted with funding applications or grant preparation. Cumulatively, 77.8% of respondents (35 of 45) reported at least one measurable positive professional impact from their training participation.

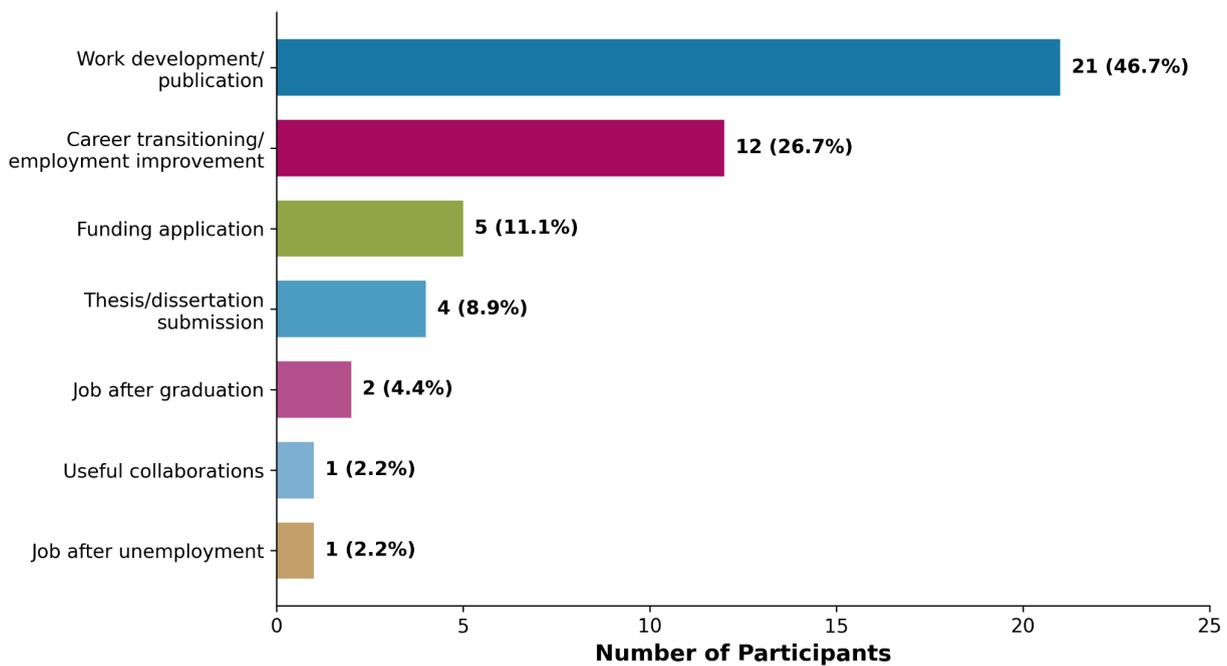


Figure 2: Professional outcomes facilitated by training participation (N=45).

Qualitative impact evidence

The survey collected 25 substantive testimonials providing qualitative evidence of training effectiveness. Representative testimonials included:

"I got to learn a whole lot more about publishing. I used to work in industry. I did not need to publish in my old position. Now I advise people on it, thanks to your course."
- (Data Management workshop participant)

"It has given me a better understanding of the tools to use with my own data and how to approach it."
- (Machine Learning workshop participant)

"It helped to automate some of the tasks and therefore improving my productivity."
- (Python workshop participant)

These testimonials document specific applications of training content in diverse professional contexts, from research publication to workflow automation and data analysis capability development.

5. Summary tables

This section consolidates key information from all BioNT training activities into summary tables, providing a comprehensive reference for workshop characteristics, materials used, improvements implemented, and participant feedback across both curricula.

Overview of learning objectives for the basic and advanced curricula

A practical introduction to bioinformatics and RNA-seq using Galaxy	From zero to hero with Python	An introduction to high performance computing	Awareness in data management and analysis for industry and research
<p>Operate the Galaxy platform to manage and analyse sequencing data, including importing data from external sources and constructing reusable workflows.</p> <p>Perform and interpret quality control of sequencing data, recognising key metrics and common data quality issues.</p> <p>Select, execute, and evaluate RNA-seq read mapping strategies, including explaining the principles and specificity of mapping to a eukaryotic reference genome and assessing mapping quality.</p> <p>Generate and normalise</p>	<p>Use Jupyter Notebooks to edit, manage, and execute Python code in an interactive computing environment.</p> <p>Write simple Python programs that use variables, data types, built-in functions, and external libraries to perform basic calculations and tasks.</p> <p>Apply core Python control structures, including lists, loops, conditionals, and user-defined functions.</p> <p>Manipulate and analyse tabular data using Pandas, including basic dataframe operations.</p>	<p>Identify computational problems suitable for execution on high-performance computing (HPC) clusters, including recognizing the benefits and limitations of parallel execution.</p> <p>Operate effectively in local and remote computing environments, including transferring files and managing files and directories on HPC systems.</p> <p>Submit, monitor, and manage computational jobs on an HPC cluster using a scheduler.</p> <p>Explain the principles of reliable, scalable, and reproducible</p>	<p>Explain the importance of research data management and good research practices in academia and industry..</p> <p>Describe and interpret the European data landscape, including Common European Data Spaces, relevant policies, and regulatory requirements.</p> <p>Identify and explain organisational data management and data governance structures, including enterprise processes, policies, and key roles.</p> <p>Create and manage research data documentation and organisation</p>

<p>gene-level count data to enable valid comparisons between samples.</p> <p>Conduct and interpret differential gene expression analysis, including running DESeq2, analysing outputs, and identifying, annotating, and visualising differentially expressed genes.</p>	<p>Visualise data using Matplotlib to generate informative plots.</p>	<p>scientific workflows, including rule-based and chained execution models.</p> <p>Implement and execute workflow-based analyses using Snakemake, including running workflows on HPC resources.</p>	<p>systems, including Data Management Plans, file and folder structures, and the effective use of electronic lab notebooks with respect to data security and privacy.</p> <p>Apply FAIR principles and reproducible research practices by selecting and using appropriate tools and techniques.</p>
<p>Strategies for training and knowledge exchange with a consulting perspective</p>	<p>Software Best Practices: Code & Collaborate – The FAIRytale of Software Development</p>	<p>Applied Machine Learning for Biological Data</p>	<p>System administration for HPC</p>
<p>Explain fundamental principles of training theory, including adult learning, instructional design, and learner-centred approaches.</p> <p>Apply needs assessment methods to design aligned and inclusive learning experiences, using appropriate teaching strategies, hands-on methods, and formative assessment.</p> <p>Deliver and evaluate effective training sessions, including providing constructive feedback to support learner development and continuous improvement.</p> <p>Manage consultancy activities</p>	<p>Implement collaborative distributed version control workflows, including contributing to and reviewing code in repositories owned by others.</p> <p>Apply effective collaborative software development practices, including structured workflows and systematic code review processes.</p> <p>Design and implement software testing strategies, including local testing and automated testing to ensure code quality and reliability.</p> <p>Develop and manage reproducible computational research environments, including recording</p>	<p>Apply Python-based data manipulation and machine learning techniques (NumPy, Pandas) to preprocess and analyze biological datasets.</p> <p>Select and implement appropriate machine learning approaches (regression, classification, clustering; supervised vs. unsupervised) to address biological research questions.</p> <p>Build, optimize, and evaluate machine learning models using regularization, hyperparameter tuning, suitable performance</p>	<p>Apply core Linux system administration skills in an HPC context</p> <p>Manage shared filesystems, users, services, and packages</p> <p>Set up and administer Slurm workload manager from a system perspective</p> <p>Deploy and support scientific applications in containers</p> <p>Automate configuration using Ansible</p> <p>Monitor and troubleshoot cluster</p>

<p>using structured project management practices, including task tracking, documentation, delegation, and time allocation.</p> <p>Define and apply clear engagement boundaries in consultancy work, including establishing follow-up procedures and distinguishing consultancy activities from research involvement.</p>	<p>and controlling software dependencies and execution contexts.</p> <p>Organise and manage computational software projects across their lifecycle, including project structuring, documentation, and consideration of software licensing.</p>	<p>metrics, and cross-validation.</p> <p>Develop end-to-end machine / deep learning workflows for bioinformatics, including training simple neural networks with PyTorch on genomics data.</p> <p>Ensure scalable and reproducible bioinformatics workflows by using containerization and by comparing CPU and GPU computing strategies for genomic data processing.</p>	<p>health with Prometheus and Grafana</p> <p>Build and manage HPC software environments with EasyBuild and EESSI</p>
--	---	--	---

Key statistics of the advanced workshops

Workshop title	Strategies for training and knowledge exchange with a consulting perspective	Software Best Practices: Code & Collaborate – The FAIRytale of Software Development	Applied Machine Learning for Biological Data	System administration for HPC
Date	June 25-26, 2024	February 4-6, 2025	May 27-28, 2025 (Module 1), June 2-6, 2025 (Module 2)	October 21-23, 2025
Format	Live Online (Virtual Instructor-Led Training)			
Participants	24	47	39	30

Participant information	Large majority (> 65%) Experienced technical staff affiliated with SMEs; all are european residents	Mix of academic, industry professionals and Industry job seekers	All - European residents; Strong industry, SME & job seeker presence (14 SME-affiliated, 18 Job Seekers)	All - European residents & affiliated with SME or Job Seeker's in industry
--------------------------------	---	--	--	--

Overview of the study materials' content and the source of these materials

Workshop title	Strategies for training and knowledge exchange with a consulting perspective	Code & Collaborate: The FAIRytale of Software Development	Applied Machine Learning for Biological Data	System Administration for HPC
Learning topic	Theory of learning, training development, training delivery, consulting, mentoring, community management	FAIR software development principles, version control, code quality, reproducibility, testing, documentation, continuous integration	Data handling (NumPy, Pandas), unsupervised learning (clustering, PCA), classification, regression, deep learning, GPU-powered genomics	Linux administration, Warewulf cluster management, Slurm, Docker, Singularity, Ansible orchestration, Prometheus, Grafana, EasyBuild, EESSI
Source of training material	The Carpentries (Instructor Training), CSCCE materials, ELIXIR material, and custom-developed BioNT content	The Carpentries, CodeRefinery, and Carpentries Incubator ("Tools and practices for FAIR research software")	Custom-developed from scratch specifically for biological datasets and research/clinical use cases	Custom-developed from scratch, tailored to real-world HPC system administration scenarios
Development methods	Adaptation of established open training materials; session-specific design for consulting/mentoring application	Following The Carpentries and CodeRefinery hands-on models; adaptation of curricula with industry-relevant examples	Modular design (Module 1 and 2); tailoring content to address specific gaps learners may have. Cover genomics use cases in hands-on	Development of theoretical modules complemented by extensive hands-on sessions involving the deployment of mini HPC clusters

Accessibility of materials	Openly available on Zenodo; licensed under CC-BY 4.0; recordings facilitated via Zoom	Openly available on GitHub; licensed under CC-BY 4.0	Openly available; licensed under the MIT License	Openly available on Zenodo; licensed under a Creative Commons license
-----------------------------------	---	--	--	---

Features applied for optimised learning

Workshop title	Strategies for training and knowledge exchange with a consulting perspective	Code & Collaborate: The FAIRytale of Software Development	Applied Machine Learning for Biological Data	System Administration for HPC
Key features	Shifted from basic training theory to advanced curriculum; introduced sessions on consulting, mentoring, and community engagement.	Integrated community-piloted lessons from Carpentries Incubator (FAIR research software); focused on shifting from individual coding to collaborative team development.	Divided into optional Module 1 (NumPy/Pandas) and mandatory Module 2 (ML core) to bridge knowledge gaps from basic curriculum.	Shifted from single-node Linux administration to multi-node cluster management using Warewulf and Slurm; introduced automation with Ansible.
Pedagogical features	Transitioned to learner-centred instruction with hands-on methods and constructive alignment; included co-design of course content based on interest.	Applied CodeRefinery hands-on models; emphasised collaborative code reviews and type-along sessions where learners follow instructors in real-time.	Transitioned to specialised material developed for clinical and research genomics use cases; introduced quizzes for knowledge internalisation.	Modified delivery to reduce copy-pasting and increase trainer explanation time; focused on manual construction of a "mini HPC cluster".
Infrastructure features	Used HedgeDoc for anonymous participation and structured Q&A; used Zoom breakout rooms; recommended dual-screen setups for virtual delivery.	Implemented Shellshare for real-time instructor command history visibility; required GitHub account and SSH key setup prior to attendance.	Allocation of 40 specialised Virtual Machines (VMs) equipped with GPUs; used Docker and GPU-powered genomics workflows (Parabricks). Used HedgeDoc for quizzes	Provisioned AWS VMs with virtualisation capabilities; used Prometheus and Grafana for metrics visualisation.

Participant feedback integration	Adjusted break times according to workshop progression; addressed conflict between initial anonymity and later participation expectations in breakouts.	Resulted in modifications to upstream lesson repositories; balanced introductory pacing to accommodate diverse learner experience levels.	Adapted Module 2 based on feedback from pre-workshop seminars; adjusted hands-on sessions to increase self-coding time over code explanations.	Reduced dense theoretical sessions in favor of direct core explanations followed by practicals; addressed AWS region capacity issues by changing instance types.
---	---	---	--	--

Participant feedback

Workshop title		Strategies for training and knowledge exchange with a consulting perspective	Code & Collaborate: The FAIRytale of Software Development	Applied Machine Learning for Biological Data	System Administration for HPC
Recommendation rate (Likely to recommend - 8-10 scores)		77%	95%	92%	95%
Instructors (agree/strongly agree)	Enthusiastic	100%	100%	96%	96%
	Provide clear answers	92%	100%	96%	96%
	Knowledgeable	100%	100%	96%	100%
Immediate applicability (agree/strongly Agree)		84%	91%	75%	73%

Other features that participants liked	Interactive exercises, use of HedgeDoc, balance of practical/theory, anonymous participation.	Hands-on exercises, collaborative code reviews, type-along sessions, use of Shellshare and HedgeDoc.	Hands-on focusing real-world applications, GPU-powered genomics workflows, self-coding sessions, industry meetup integration.	Real-world case examples, deployment of mini HPC clusters on VMs, Ansible/Slurm hands-on, industry meetup.
---	---	--	---	--

Cross-programme summary of impact assessment

Assessment component	Key findings
Response	45 responses (~9% of participants) from 11 workshop sessions
Tool usage	Frequent users increased 200% (from 11.1% to 33.3%); non-usage decreased 47% (from 42.2% to 22.2%)
Top professional outcomes	Work development/publication (46.7%), career transitioning (26.7%), funding applications (11.1%), job acquisition (6.6%)
Overall impact	77.8% reported at least one positive professional outcome
Qualitative evidence	25 testimonials (55.6% response rate) documenting specific applications from research publication to workflow automation and data analysis capability development across diverse professional contexts

6. Conclusion and sustainability considerations

The basic curriculum was delivered in two rounds, yielding valuable insights into the training delivery process. The second iteration benefited from enhanced communication and advertising efforts, leading to a significant increase in applications, particularly among job seekers. This required implementing a selection process, with priority given to European applicants who were either unemployed or associated with SMEs. These key takeaways were subsequently incorporated into the design and delivery of the advanced curriculum.

The BioNT consortium concluded that all four workshops in the advanced curriculum successfully achieved their objectives by reaching the target audience, achieving high participant satisfaction, and enabling effective skill transfer across technical disciplines. The "Strategies for training" workshop demonstrated that interactive consulting training is viable online. The "Code & Collaborate" workshop successfully piloted new FAIR software materials, contributing back to the community, whilst identifying challenges in managing diverse learner experience levels. The "Applied Machine Learning" workshop benefited from a pre-workshop industry meetup that proved highly effective for promotion, attracting learners who reported substantial progression from moderate to strong confidence levels across all measured competencies. The "System Administration for HPC" workshop similarly benefited from industry engagement and delivered highly valued hands-on cluster management training. Across all events, the use of collaborative tools like HedgeDoc fostered a supportive environment, with participants consistently praising the knowledge and enthusiasm of the instructors. Selection processes effectively targeted European job seekers and SME affiliates, ensuring the training reached the intended demographic. Operational lessons highlighted the importance of clear prerequisites to align participant expectations and the value of hybrid events for broader outreach.

Sustainability of training materials

Beyond immediate workshop delivery, BioNT has prioritised the long-term accessibility and sustainability of its training resources. All training materials, including written content, video recordings captured during live Zoom sessions, and participant-shared collaborative notes, are being systematically curated and made available through multiple complementary channels. The [Lhumos platform](#), developed and hosted by EPFL, serves as the primary hub for self-paced learning, providing learners with ongoing access to workshop recordings organised into thematic playlists corresponding to the basic curriculum, advanced curriculum, and community event. These video resources are linked directly to corresponding written training materials, enabling independent learning beyond live workshops.

To maximise accessibility across European linguistic communities, BioNT has implemented a comprehensive translation framework for basic curriculum materials. Using a semi-automated approach combining the DeepL API for context-aware machine translation

with expert manual curation, the consortium has translated training materials from English into German, Spanish, and Italian. Curated translations are progressively made available through the BioNT webpage as quality assurance processes are completed.

The consortium has established sustainable long-term hosting arrangements to ensure materials remain accessible beyond BioNT's funding period. Galaxy Training Network (GTN)-based materials will be hosted, archived, and maintained as part of that project's ongoing efforts, whilst Carpentries-based materials will be permanently hosted on the BioNT Training GitHub organisation space. All lessons will be linked from both the BioNT website and the Lhumos platform, ensuring unified access and synchronisation with live workshop recordings. This approach guarantees that BioNT's training resources will continue to serve learners, trainers, and institutions across Europe, supporting the professionalisation of digital skills in biotechnology and biomedical sectors for years to come.