Collaborative solutions for enhancing inclusiveness of training

Summary report

Authors: Lisanna Paladin, Alex Botzki

Enhancing ELIXIR Training Platform: Workshop Summary Report

Participants

Name	*ORCID id*	*email*	*Contact for future follow-up*
Bruna Piereck	0000-0001-5958-0669	bruna.piereckmoura@vib.be	yes
Elin Kronander	0000-0003-0280-6318	elin.kronander@nbis.se	yes
Luciana Peixoto	0000-0002-3972-3447	luciana.peixoto@biodata.pt	yes
Allegra Via	0000-0002-3398-5462	allegra.via@uniroma1.it	yes
Loredana Le Pera	0000-0002-0076-9878	loredana.lepera@iss.it	
Bérénice Batut	0000-0001-9852-1987	berenice.batut@france-bioinformatique.fr	yes
Owen Appleton	0000-0002-2571-6860	owen.appleton@sib.swiss	yes
Zoi Litou	0000-0002-7442-1137	zlitou@biol.uoa.gr	yes
Lisanna Paladin	0000-0003-0011-9397	lisanna.paladin@embl.de	yes
Alexander Botzki	0000-0001-6691-4233	alexander.botzki@vib.be	yes

Introduction

On the 10th of June, on the occasion of the ELIXIR All Hands Meeting 2024, a workshop was held to address technical challenges and explore open-source solutions for the ELIXIR Training Platform (TrP). This workshop brought together participants from various ELIXIR entities, alongside contributors from diverse backgrounds. The focus was on creating inclusive training materials using a collaborative "call for help" framework.

Participants aimed to leverage collective expertise, particularly in areas such as video editing, subtitle incorporation, multi-modal streaming, and translations.

Participant Demographics

Participants were asked about their involvement in different Platforms, Communities, and Focus Groups within ELIXIR. The responses highlighted the diverse engagement across the network:

- *Representation spanned across multiple entities, including one platform, six communities, three commissioned services, and two focus groups, among others.*
- *Roles included trainers, helpers, coordinators, materials developers, communication managers, and roles specific to ELIXIR, reflecting a wide array of experiences and responsibilities within the network.*

Workshop Motivations and Challenges

Participants shared their motivations for attending and the challenges they hoped to address:

- *Learning how to make training events more inclusive and sharing experiences.*
- *Gaining a better overview of challenges, finding solutions, and understanding trainee challenges to improve course accessibility.*
- *Participating in discussions on available solutions, standardising training materials, and exploring perspectives on standardisation and inclusiveness.*

Table Discussions and Findings

Table 1: Videos from Live Training Events

The discussion on videos from live training events garnered interest primarily from the online participants. The key points covered included the potential and challenges of incorporating multi-modal training, such as podcasts and videos, into ELIXIR's training platform.

Multi-modal Training:

- *Podcasts: It was noted that producing podcasts is time-consuming and requires specialised skills. A central team within ELIXIR could handle this task to manage costs and ensure quality.*
- *Videos: The discussion emphasised the importance of captions for accessibility, particularly for individuals with hearing impairments and non-native speakers. Various approaches were suggested, such as using Zoom for live captions and pre-recorded videos for later captioning via YouTube's auto-generated captions, which then need to be corrected and re-uploaded.*

Challenges and Solutions:

- *GDPR constraints currently prevent the use of AI tools for live captioning within ELIXIR. Alternative tools like Otter are used in other communities, such as the Open Life Sciences (OLS).*
- *Ensuring that videos are FAIR (Findable, Accessible, Interoperable, and Reusable) involves using permanent identifiers and appropriate metadata, possibly leveraging platforms like Zenodo and BioSchemas.*

***Translation**:*

- *There is a need for better models trained specifically for bioinformatics. Strategies include using DEEPL for machine translations and maintaining a glossary of specialist terms to improve translation quality. The OPERAS Research Infrastructure was mentioned as a resource for community translation platforms.*

Table 2: Leveraging LLMs for Training

This table focused on the potential of Large Language Models (LLMs) like ChatGPT to assist in developing course syllabi, creating micro learnings, and formulating learning outcomes using Bloom's taxonomy.

Key Discussions:

- *Course Syllabus Development: Tools like ChatGPT and Gemini can assist in drafting course structures, but their outputs need to be carefully reviewed. Paid versions of these tools offer faster responses and longer memory for previous prompts.*
- *Creating Microlearnings: Participants shared their experiences in using ChatGPT to create short, engaging videos. The importance of prompt engineering was emphasised to obtain specific and useful content.*
- *Formulating Learning Outcomes: While LLMs can help in generating learning outcomes, they require detailed and specific inputs. Trainers need to decide the appropriate Bloom's taxonomy level and learning experiences to achieve desired outcomes.*

Challenges:

- *Limited experience among participants with these tools, but there was significant interest in exploring their potential further.*
- *Licensing issues when creating imaging content were noted as a broader concern beyond just microlearning.*

Table 3: Certifying Training

The discussion on certifying training addressed automatic pipelines for trainer recognition, micro-credentialing, and measuring long-term impact.

Key Points:

- *Certification of Trainers: A potential solution for certifying trainers is through Train-the-Trainer (TtT) courses. However, there is a need for ORCID profiles to include references to training activities, which currently is missing.*
- *Micro-credentialing: The European Commission's definition of micro-credentials could be integrated in partnership with universities. There were discussions about whether organizations should act as certifying bodies or partner with universities for joint courses. A financial model is necessary for long-term sustainability.*
- *Challenges and Solutions: Identifying the best gateway within ELIXIR and beyond for offering certifications and leveraging existing networks like Life Science trainers.*

Table 4: Standardising Formats

The focus was on creating markdown-based formats for training materials and using git-versioned repositories to manage them.

Key Discussions:

- *Learning Platforms: Platforms like Canvas, Blackboard, and Moodle were discussed for their advantages and disadvantages. While they offer accessibility and institutional support, they can be limiting and dependent on institutional choices and budgets.*
- *Git and Markdown: Git was proposed as an alternative to manage versioning and collaboration on training materials, though it requires expertise. Markdown formats, including Jupyter notebooks, rmarkdown, and Quarto, were highlighted for their benefits in creating reproducible and accessible content.*
- *Challenges: Issues with markdown for images and schemes, and making PowerPoint presentations FAIR and git-friendly.*

Recommendations:

- *Promoting the use of git and markdown for creating and maintaining training materials.*
- *Investing in training for using tools like Quarto to make presentations more reproducible and accessible.*

Conclusion

The workshop provided valuable insights into addressing key challenges faced by the ELIXIR Training Platform. The discussions highlighted the importance of adopting open-source tools, enhancing accessibility,

and exploring innovative solutions like LLMs to improve training materials and delivery. Continued collaboration and testing of proposed solutions will be crucial in achieving these goals.

Date: Monday 10th June, 16:00, Sal XI link

Chairs:

Lisanna Paladin
Alexander Botzki

Abstract:

This workshop proposal aims to harness the collective experience of some ELIXIR entities, alongside contributors from diverse backgrounds. The focus is on enhancing the ELIXIR Training Platform (TrP) activities, with a special emphasis on open-source solutions to address technical challenges for the creation of inclusive training material. Embracing a "call for help" framework, we invite participation from all ELIXIR entities, as we intend to pool together the wider possible range of expertise. The Galaxy Community, the Compute Platform and the Generative AI discussion group have expressed interest to join the workshop.
In particular, we aim to explore open-source solutions (possibly with assistance of generative AI) solving urgent issues faced by the Training Platform such as video editing, incorporation of subtitles, multi-modal streaming, translation to different languages, and other accessibility measures.
The workshop will consist of interactions with experts that could already have some solutions in place (e.g. the Galaxy Training Network) providing an opportunity for advertisement of their solution in exchange of the experience that they will contribute. After an introduction of the TrP challenges and aims, attendees will work in group tables, each dedicated to a specific technical issue, enabling guided interactions with the experts' tools and gaining a practical understanding of how to apply them to other training contexts. Participants will be encouraged to share how they have solved similar problems in the entities that they represent. In those cases where we will not be able to identify internal relevant expertise, we will use the workshop to explore how other international communities are addressing the issues.
The outcome of this workshop will be a detailed summary report. This report will comprehensively list and describe the specific challenges discussed during the workshop and solutions collected (tools and/or strategies). Additionally, it will possibly include a draft plan on how to integrate these technical solutions into the existing Training Platform activities. This initiative aligns with ELIXIR's goal of fostering collaboration and knowledge exchange across its diverse entities. By collecting in this document the possible solutions related to the technical challenges of training activities, and by sharing the report publicly on Zenodo, we expect to have an impact well beyond ELIXIR itself, fostering a more effective and accessible learning environment across Life Sciences in Europe.

- *Slide deck listing the challenges we intend to work on*

Time (CET)	Subject
10 mins	Title: Introduction to the workshop Speakers: Alex Botzki and Lisanna Paladin *Slide deck*
Tables in parallel (40 mins)	Table 1: **Videos from live training events** Topics: - Multi-modal training (incorporating streaming) - Editing videos and adding captions easily - Training materials translation Chair: Bérénice Batut *Link to slides or documents*
	Table 2: **Leveraging on LLMs for training** Topics: - From an idea to a draft course syllabus using LLMs - Creating microlearnings with ChatGPT

	- Formulating learning outcomes using Bloom taxonomy using LLMs Chair: Barbara Szomolay, Krzysz Poterlowicz *Link to slides or documents*
	Table 3: **Certifying training** - Automatic pipelines for trainers recognition (ORCID profiles) - (European) micro credentialing integration - Measuring long-term impact Chair: Alex Botzki *Link to slides or documents*
	Table 4: **Standardising formats** - Creating a markdown-based format for training materials - Training materials in git-versioned repositories Chair/Speaker: Geert Van Geest and Alexia Cardona (switch during session) *Link to slides or documents*
40 mins	Title: Joint wrap-up Speaker: Alex Botzki and Lisanna Paladin

Notes (collectively taken)

1. How many ELIXIR entities are represented here?
   Please write the name of the Platform(s) or Community(-ies) you are involved in, including Focus Groups and other activities. If the name is already in the list, add a +1.
   - Training Platform +17
   - Biodiversity Community+1
   - RDM Community+4
   - Galaxy Community+2
   - hCNV Community
   - IDP +1
   - Microbiome Community
   - Node Support Commissioned Services
   - Train-the-Trainer Commissioned Services +2
   - PeoplePulse CoS+2
   - Learning Paths Focus Group +4
   - RI Professionalisation Focus Group
   - Impact+1
   - ELIXIR Communication group +1
   - FAIR Training +2
   - STEERS +4
   - Probably more than I'm not even aware of

2. How would you describe your role in training?
   Please describe in a few words your role(s) in training events or in managing training materials. If your role is already there, add a +1.
   - Trainer +11
   - Training helper
   - Training Coordinator +8
     - ELIXIR+1
     - Institution +3
     - Both +3
   - Training Developer +9
   - Training materials developer (is this the same as above?)
   - ELIXIR Deputy Training Coordinator +4
   - Node Training Coordinator +2
   - Training assistant (within my institution) +1
   - I hang around the training platform people from time to time +2
   - Node Community Manager+1
   - Communication Manager

3. What's your motivation for attending this workshop? What challenges are most important for you?
   If you'd like, describe briefly what's your motivation for attending this workshop in the following section. You can use it also to highlight challenges that you think should be included in the discussion.
   - Learn how to make training events more inclusive +4
   - Share experience +1
   - Get a better overview of the challenges from others, find solutions +4
   - Be more aware of the types of challenges trainees may face and what steps I can take to make the courses I teach more accessible to everyone +1
   - Participate in the discussion of what solutions are available and how to tackle it +2
   - Get effective feedback from 100% attendees?
   - Standardising training materials
   - Learn about different perspectives on standardisation and inclusiveness
   - Watching and seeing if there is anything that can be used in the ELIXIR Almanac

Notes from the table discussion activity

Multiple ELIXIR entities are represented (1 platform, 6 communities, 3 commissioned services, 2 focus groups, 4 groups of other types). Multiple types of roles in training are represented: trainer, helpers, coordinators, materials developers, communication managers and roles specific to ELIXIR.

Table 1, "Videos from live training events", collected the interest of the 4 online participants. All the other table discussions happened in person: Table 2, "Leveraging on LLMs for training", with 8 participants, Table 3, "Certifying training", with 6, and Table 4 "Standardising formats", with 8.

We encouraged people to move between tables if they felt they could contribute to multiple discussions, but this did not happen. The table discussions were summarised and presented in this slide deck.

We selected the challenges to work on to be very practical and requiring solutions based on tools / concrete actions. There might be many more challenges worth to discussCanc.

Table 1:
Multi-modal training
Briefly discussed about podcasts. Podcast production is time consuming and has a cost attached, ELIXIR can help in that.
Then moved to videos. Videos can be captured during live events. Needs caption (brief discussion in the room about captioning, proposal to use AI to address that). In ELIXIR no use of the AI tools due to GDPR issue, Otter used in other communities (OLS). Transcripts need review. GTN was mentioned in this context: ask the trainers to record the videos before, then review captions, names are often wrong. GTN also includes a way to generate videos from slide notes (works for slides). Questions in the room that show interest about this.
Videos can also be captured before the event. Takes more time because requires iterations.
Lead briefly mentioned the other topics they touched upon: How to make videos FAIR? And translation.

Table 2:
Three main topics.
Generative AI requires institutional licence, some examples of tools: chatGPT, Gemini.
Can be useful, but one needs to check the answers, requires specificity. Compared the tools results, paid versions are faster and have a longer memory to remember prompts. Omit cliches and very general expressions. Try a restart, so more to a new thread for testing purposes.
Creating micro-learning with AI: not many have expertise, only some examples of creating short videos on a topic. Licensing issues when generating imaging content.
Learning outcomes: better to start from available materials, specific the taxonomy version and level. AI is not very creative so needs specific input. Could be a useful tool to annotate existing training materials.
General comment: not much experience in these topics, but very interested to test them.

Table 3:

Touched shortly on certification of the trainers, TtT could be a solution. Missing from ORCID: in the section "works" no reference to training (training materials, training events as a trainer - possibly also as a trainee?). Question about the network to be leveraged to make this change (Life Science trainers were mentioned).
Definition by the EC about micro-credentialing. Could be setup in a partnership by the University, should organisation act alternatively to a University? Examples in ELIXIR about joint courses between nodes, so many use cases. All this needs a financial model. Who is the gateway in ELIXIR and beyond?

Table 4:
Canvas and other learning platforms discussed. Many advantages but also disadvantages, dependance on institutional choice and budget.
Git as an "alternative", addresses many problems but needs expertise.
Markdown mentioned as a standard, for documents and presentations. Making presentations becoming easier, especially good advantages to address learning impairments.