Common Infrastructure for National Cohorts in Europe, Canada, and Africa
- CINECA -

D6.5 - Training Programme, Annual Report 2021

Work Package:	WP6 – Outreach, training and dissemination
Lead Beneficiary:	European Molecular Biology Laboratory
WP Co-leaders:	Vera Matser (EMBL-EBI) Saskia Hiltemann (EMC)
Contributing Partner(s):	UCT, EMC, EMBL-EBI
Contractual Delivery Date:	December 31st 2021
Actual Delivery Date:	December 16th 2021
Authors of this Deliverable:	Nicola Mulder, Mamana Mbiyavanga, Saskia Hiltemann, Vera Matser, Marta Lloret Llinares
Contributors:	-
Reviewed by:	Cath Brooksbank
Approved by:	Thomas Keane
Dissemination Level:	Public
Grant agreement:	No. 825775 Horizon 2020 (H2020-SC1-BHC-2018-2020)
Type of action:	RIA
Start Date:	January 1st 2021
Duration:	12 months (25-36)

Table of contents:

1. Executive Summary

In this deliverable document, we report on the activities in task 6.4 - Training Programme, describe the CINECA training activities that took place in months 25-36 of the project and provide the Training Plan for the final year.

The CINECA training programme aims to train people within the CINECA consortium as well as external users. Different approaches have been employed, including face-to-face and online courses, hackathons, training videos and staff exchanges. While we waited for CINECA products to be completed, many of the training efforts for the year again focused on internal learning opportunities and knowledge exchanges, but some externally facing events were held to disseminate outputs. All the training and outreach events continued to be heavily impacted by COVID-19, which removed our ability to hold face-to-face workshops and staff exchanges. The staff exchanges were suspended and replaced with virtual meet-ups.

This year we ramped up the webinar series[1], providing nine webinars that also served as online learning interventions. The nine webinars had an average of 29 attendees (the most being 50 for a single webinar), representing on average 57% of those who registered. In addition, feedback from these webinars is collected and is used to further increase the utility of the webinars going forward. The average response rate to the feedback surveys was 25%, and the average rating for both pace and content has consistently been above 4, where 1 is poor, and 5 is excellent.

A series of short training videos (https://www.cineca-project.eu/short-videos) was created last year to facilitate the uptake of CINECA outputs. This year, three new short videos were produced by work packages on different topics, and we started adding English transcriptions to the videos to increase access. To increase engagement, the short videos were submitted to ELIXIR's training portal and disseminated via CINECA's various communication channels.

No face-to-face workshops and training events could take place this year. However, we successfully ran some online training workshops and hackathons. These training interventions included the Galaxy Training Network Smörgåsbord: A Global Galaxy Course, the "How FAIR are you" hackathon and the CINECA GA4GH Beacon training series. CINECA also continued to be represented at relevant conferences and consortium meetings to ensure the dissemination of our outputs.

The ultimate goal of the WP6 training programme is to build a learning pathway that encompasses all the training interventions delivered throughout the project addressing the challenges of federated data sharing and analysis. This deliverable will constitute the final training intervention of the work package that will provide an end-to-end data analysis course. This course will be targeted at end-users and will cover a range of topics from the project outputs and some other topics of interest related to CINECA's goals, including FAIR principles, data harmonisation, Galaxy, ontologies, and data standards and APIs.

---

[1] https://www.cineca-project.eu/webinars

2. Project objectives

This deliverable has contributed to the following objectives:

a) To raise awareness of CINECA and the opportunities and challenges of sharing cohort data on a global scale. Build relationships with related projects to facilitate the dissemination of information of mutual interest through all appropriate channels to the broadest potential audience.

b) To identify and address training needs, both within the consortium itself and more broadly among key stakeholder groups.

c) To develop and deliver training (or other learning interventions, e.g. secondments; communities of practice; hackathons) in areas of highest training needs, integrating the outputs of the technical work packages to deliver the project's training goals.

d) To assess the impact of the project's outreach, training and dissemination activities.

3. Detailed report on the deliverable

3.1. Background

The CINECA project is developing new standards and technical solutions to address existing challenges in responsible human data sharing and analysis across continents. The project comprises inter-related work packages addressing different needs. One of the needs within the consortium is for the training of staff within work packages and for knowledge dissemination across work packages. These are being addressed through staff exchanges and some internal training events. While CINECA products were being developed, the focus was on internal training, but as these are now being released, there is an increasing need to establish training outside the consortium. The current focus is on gathering the requirements and developing training around known challenges, such as cohort data harmonisation. A complementary activity is the ongoing webinar series, which serves as both a training and dissemination initiative. The COVID-19 pandemic has significantly impacted our training and dissemination activities over the last two years, and the future of in-person events is still uncertain. Therefore, our training plan for the previous year will be tentative and will need to adapt to the current uncertainty and ever-changing global situations.

3.2. Description of Work

As mentioned above, the training programme has several parallel activities, including staff exchanges, webinars, short videos, short technical reports and training events. Progress and challenges on these are described in more detail below, along with the tentative plans going forward for the final year of the project.

3.2.1. Staff Visits

The procedure around staff visits was set up early on in the project, and an application form with guidance text has been made available for CINECA partners to request staff visits[2]. To date, one staff visit has taken place before the start of the pandemic. Due to the current travel restrictions, we have moved our focus away from the face-2-face staff visit program favouring virtual alternatives such as internal online training workshops and cross-WP online meet-ups. Since we expect COVID-19 and travel restrictions to remain a barrier to many for the remainder of the project, this focus on virtual alternatives is likely to stay for the project's duration.

3.2.2. Webinar Series

We set up a CINECA webinar series to deliver training and disseminate our activities within CINECA and to a broader audience of stakeholders. The CINECA webinar series aims to discuss common challenges, share best practices regarding cohort data analysis, and distribute CINECA project outputs. All CINECA webinars are open to everyone and include an audience live Q&A session during which attendees can ask questions and make suggestions. All webinars are recorded, made available via the CINECA project YouTube channel, and embedded on the event webpage on the CINECA website (https://www.cineca-project.eu/webinars). Although there were many challenges during the 2021 reporting period, we delivered nine webinars, more than the previous two years, as described in Table 1.

Seven of these webinars were part of the "How FAIR are you" webinar series and hackathon, which consisted of 7 webinars from January to April 2021, culminating in a hands-on workshop (hackathon) at the end of April. This series aimed to facilitate responsible and ethical data and resource sharing and implementation of federated applications for data analysis by implementing FAIR approaches into software, training materials, and cohort data. The first webinar introduced the FAIR concept and discussed applications for health data with a webinar on Open science through FAIR health data networks: dream or reality?. The second webinar on How to make cohort data FAIR explored several informatics techniques, such as ontologies, to make cohort data more FAIR. The next webinar discussed the importance of the FAIR principles for the software tools, presenting the existing initiatives developing tools and raising awareness of the significance of FAIR for software. The fourth webinar discussed the Practical applications of the FAIR data principles, particularly in clinical bioinformatics. The Making training materials FAIR then provided the participants with some tips and suggestions on making training materials more FAIR. The next webinar was on the critical topic of ethical, legal and social considerations in FAIR data sharing. The final webinar by Thomas Keane (EMBL-EBI) took place during the hackathon on the application of FAIR in the CINECA project. This webinar discussed how the CINECA project applies FAIR principles using open standards, such as ontologies and the Global Alliance for Genomics and Health (GA4GH) standards, across the federated network to enable federated data discovery, data access, and analysis.

---

[2] https://www.cineca-project.eu/blog/cineca-staff-visit-program

As part of the EUCAN ELSI Collaboratory (EUCANCan, euCanSHare, CINECA, iReceptor Plus, EUCAN-Connect, ReCoD-ID, and EuCanImage), CINECA hosted a joint webinar about how fostering stakeholder engagement can improve transparency and accountability in international sharing of health data. The two-hour webinar, entitled International Data Sharing: Fostering Engagement, Transparency and Accountability, took place on October 29th and consisted of two parts. The first part included four 15-minute presentations with a Q&A. The second provided a space for researchers in the EUCAN ELSI projects to discuss a wide variety of themes concerning the future of international data sharing.

With CINECA's WP1, we started a CINECA GA4GH Beacon series, consisting of one public webinar and two internal online training events in November. Two online training events will take place next February and will be open to the public with limited places. The public webinar was on the Beacon: a data discovery solution in genomics and health, which took place on November 11th, 2021. The webinar introduced the Beacon v2, its features, addressed security considerations, and illustrated a few use cases.

Table 1: Summary of the webinars delivered in the CINECA webinar series in 2021

Date	Title	Speaker
21/01/2021	Open science through FAIR health data networks: dream or reality?	Kees van Bochove (The Hyve)
17/02/2021	How to make cohort data FAIR	William Hsiao (SFU)
24/02/2021	FAIR principles for the software tools	Carlos Martinez (Netherlands eScience Center)
04/03/2021	Practical applications of the FAIR data principles	Andrew Stubbs (EMC)
18/03/2021	Making training materials FAIR	Sarah Morgan, Anna Swan (EMBL-EBI)
15/04/2021	Ethical, legal and social considerations in FAIR data sharing	Melanie Goisauf (BBMRI)
28/04/2021	Application of FAIR in the CINECA project	Thomas Keane (EMBL-EBI)
29/10/2021	International Data Sharing: Fostering Engagement, Transparency and Accountability	Pilar Nicolas, Alexander Bernier, Thijs Devriendt, Lauren Maxwell, Michaela Mayrhofer (BBMRI-ERIC, CINECA), Bartha Knoppers
11/11/2021	The Beacon: a data discovery solution in genomics and health	Lauren Fromont (CRG)

There were between 13 and 50 attendees with an average 58% conversion rate of registrations. All webinars have a feedback form that is automatically shared with the attendees at the end of the webinar and is sent to participants in a follow-up email. Survey responses ranged from 11% to 44%, with a 25% average. The survey responses suggest that the pace of the webinars was very good (4.5 on average, with 1 being poor and 5 excellent). The attendees also found the webinar content to be very good (4.5 on average). Table 2 illustrates the high-level statistics and feedback collected for all the CINECA webinar series.

Table 2: High-level statistics and feedback collected from the CINECA webinar series in 2021

Webinars	Nmb attendees	Conversion	Survey response	Average pace	Average content	Youtube views
Open science through FAIR health data networks: dream or reality?	50	58.82%	28.00%	4.6	4.6	107
How to make cohort data FAIR	13	52.00%	23.08%	4	4.3	53
FAIR principles for the software tools	23	65.71%	21.74%	4.2	4.4	51
Practical applications of the FAIR data principles	16	48.48%	25.00%	4.25	5	53
Making training materials FAIR	32	59.26%	43.75%	4.46	4.57	73
Ethical, legal and social considerations in FAIR data sharing	19	45.24%	10.53%	5	4.5	96
Application of FAIR in the CINECA project	43	71.67%	20.93%	4.1	4.2	33
International Data Sharing: Fostering Engagement, Transparency and Accountability	44	63.77%	n/a*	n/a*	n/a*	12
The Beacon: a data discovery solution in genomics and health	20	51.28%	n/a*	n/a*	n/a*	8
Average statistics	**29**	**57.36%**	**24.72%**	**4.37**	**4.51**	**45**

*Due to a technical issue, we did not collect feedback for these webinars. Range between 1 and 5 for *Average Pace* and *Average Content.*

3.2.3. Workshops and training events

Notwithstanding all the challenges encountered due to the COVID-19 pandemic and lockdowns, we delivered most of the scheduled training interventions for 2021, with at least two training events each quarter. In addition, seven training events were organised by the CINECA project covering a range of topics relevant to the project, including Galaxy, FAIR principles, GA4GH passports, GA4GH Beacon and science communication. The summary of training interventions delivered in 2021 is given in Table 3.

The first quarter included a very successful GTN Smörgåsbord: A Global Galaxy Course co-organised by CINECA and other partners, running for five days and showcasing a wide variety of Galaxy Training Network (GTN) tutorials. All training sessions, available on the GTN website, were pre-recorded to allow participants from different time zones to work through them at their own pace and time. In addition, a large community of GTN trainers were available via online support to answer questions across the time zones. The training event had over 500 attendees worldwide, with a 97% feedback rate being excellent. Two more training interventions took place on "Using Twitter to promote your project" for beginner and intermediate users, aiming at equipping researchers on communicating

science on social media. These sessions were explicitly delivered for the EUCAN Dissemination and Communication network.

We organised the "How FAIR are you" hackathon in the second quarter, which featured a mix of talks and discussions in plenary sessions and practical implementation in breakout sessions. The talks included experiences in FAIRplus, FAIRsharing, FAIRplus fairification wizard and the FAIR Cookbook. The recordings of this series' webinars and lectures are available on the CINECA website and the project's Youtube channel. Two reports were also published on this series, on the CINECA website ("How FAIR are You" webinar series and hackathon) and on experiences in FAIRplus talk on The Hyve website. In addition, CINECA also organised a training session aimed at CINECA partners on GA4GH passports. The hands-on workshop targeted developers and administrators of services that want to use GA4GH Passports to authenticate users and authorise them to call an API. The 4-hour online training session had 11 participants with a 100% satisfaction level from the feedback survey.

CINECA hosted two online training events in November 2021 as part of the CINECA GA4GH Beacon series, targeted at the current and potential implementers of the GA4GH Beacon. The two online training sessions comprised talks and demonstrations. The first online training session was on Lighting a Beacon: training for (future) implementers. It provided the participants with the tools to get started with the implementation autonomously and identify the resource persons to answer their questions during the implementation process. The second online training in this series was on Beacon v2 reference implementation. It aimed to familiarise participants with Beacon queries using the UI and the API. Initially, these sessions were meant to be open to anyone, though, with the delay of the Beacon version 2 release to 2022, the November training sessions were converted to internal sessions for the CINECA consortium. The open training sessions were rescheduled to 15th and 17th February 2022 and will have a hands-on training component as well as demonstrations.

We had planned to run a cohort harmonisation workshop at the International Cohorts Summit for the International HundredK+ Cohorts Consortium (IHCC). This summit also went virtual, and since the workshop was designed to be very interactive, it was cancelled as we initially felt that a virtual event would not achieve the intended impact. The workshop is planned to take place next year, as discussed in the following sections.

Table 3: CINECA training events delivered in 2021.

Date	Topic	WP	Organisers/Trainers
February 15th-19th	GTN Smörgåsbord: A Global Galaxy Course	WP5	S. Hiltemann (EMC)
June 23rd	Online training on GA4GH Passports	WP2	S. Negru, T. Kataja, M. Linden (ELIXIR-Finland)
April 28th-29th	"How FAIR are you" hackathon	WP6	N. Mulder, M. Mbiyavanga (UCT), V. Matser, M. L. Llinares (EMBL-EBI), S. Hiltemann (EMC)
March 3rd	Using Twitter to promote your project - beginner	WP6	V. Matser, M. Mendonca (EMBL-EBI)
March 5th	Using Twitter to promote your project - intermediate	WP6	V. Matser, M. Mendonca (EMBL-EBI)
November 16th	Lighting a Beacon: training for (future) implementers	WP1	R. Ariosa (CRG)

November 17th	Browsing genomes: a Beacon training for users	WP1	M. Rueda, R. Ariosa (CRG)
November 8th-12th	Project at ELIXIR BioHackathon (FAIR training with Galaxy)	WP5	S. Hiltemann (EMC)

3.2.4. Short training videos

A series of short training videos (see Table 4) were created to facilitate the uptake of CINECA outputs, besides a handful of selected recordings from our training events. These videos can be used as stand-alone training resources to be sent out to participants attending future training courses to support them to develop prerequisite knowledge and as teasers for future webinars. All short training videos produced by the CINECA project are available from the dissemination section of the CINECA website and on the project's Youtube channel. In addition, to increase engagement, the short videos were submitted to ELIXIR's training portal TeSS (https://tess.elixir-europe.org/), which provides a central location for users to access training materials, and disseminated via CINECA's various communication channels. During this reporting period, three short videos were produced on different topics in collaboration with work packages.

WP3 produced the first two videos on the use of ontologies for cohort data harmonisation. The first video, Applying data standards to the harmonisation of COVID 19 datasets from different sources, addresses public health genomics responses to the COVID-19 pandemic and how data standards are being used to harmonise data across jurisdictions drawing examples from the Canadian COVID-19 surveillance and outbreak investigations. The second more general video, Useful ontologies for harmonising cohort data, highlights several well-curated and maintained ontologies useful for annotating cohort data, focusing on the OBO Foundry, a community of practice for interoperable ontology building.

In collaboration with WP2, we published a third short training video on "A common framework for designing portable federated pipelines", which describes a common framework for designing portable federated pipelines based on the joint cohort genotyping WP4 use case. Furthermore, two additional short videos were created during the How FAIR are you Hackathon and subsequently uploaded to YouTube.

Table 4: Overview of CINECA short training videos produced in 2021

Title	Views*	Length	WP*	Topic	Date
Applying data standards to the harmonisation of COVID 19 datasets from different sources	67	7:51	WP3	Data harmonisation, Ontologies	January 2021
Useful ontologies for harmonising cohort data	41	7:01	WP3	Data harmonisation, Ontologies	January 2021
A common framework for designing portable federated pipelines	62	7:32	WP4	Federated data analysis	February 2021
FAIRplus FAIRification wizard	11	26:23	WP6	FAIR	April 2021

| Overview of the FAIRsharing.org | 18 | 32:09 | WP6 | FAIR | May 2021 |

Some of the WPs plan to produce additional training videos throughout the project as new outputs are generated. These will potentially emanate from upcoming deliverables described in Table 10 and additional training materials that would be produced by WPs.

The short training videos have been shown to be one of our most effective strategies to reach out to our different stakeholders. Therefore, we are working to increase the reach of these resources as we make the best use of resources while travel is on hold. We have been working on ensuring that our existing videos and training materials are accessible to non-English speakers. We started creating subtitles for a few pilot videos during this reporting period to gauge response and impact. A translation of these will then follow into other languages such as French and Arabic. We created subtitles for four short videos, including Applying data standards to the harmonisation of COVID 19 datasets from different sources, Useful ontologies for harmonising cohort data, Solutions for overcoming cohort data integration challenges using ontology: an introduction and Annotating data using ontologies videos, all from the WP3, discussing ontologies and data harmonisation.

Table 5: CINECA short training videos transcribed in 2021

Title	Topic	Subtitle language	Date video produced
Applying data standards to the harmonisation of COVID 19 datasets from different sources	Data harmonisation, Ontologies	English	January 2021
Useful ontologies for harmonising cohort data			January 2021
Solutions for overcoming cohort data integration challenges using ontology: an introduction			April 2020
Annotating data using ontologies			April 2020

3.2.5. Short technical reports

Training materials were also produced in the form of technical reports describing some of the project's products and activities in work packages aimed at a more technical audience. In addition, we initiated a new series that discusses how GA4GH standards are being developed and implemented in CINECA and GA4GH standards' uptake by CINECA partners. Four short technical reports were published in the CINECA GA4GH standards series and were featured in some external newsletters such as the GA4GH newsletters. WP4 wrote the first technical report on Implementation of GA4GH standards in CINECA, providing an overview of GA4GH standards and their uses within the CINECA project. The second blog post, Beacon cohorts: A model for cohort discovery in CINECA and beyond, from WP1, focuses on the data discovery part, specifically on the Beacon cohort model, which will help users find information about cohorts of interest. WP2 wrote the third blog post on Passport is the glue between the researcher, data and computing, which discusses how CINECA has implemented GA4GH Passports for the ELIXIR AAI. Finally, WP3 produced a blog post on Powering up data discovery and access using the Data Use Ontology, describing how DUO can facilitate dataset discovery, allowing researchers to request and retrieve only the datasets that match their intended use. In

addition to the CINECA GA4GH standards series, we also published four technical reports. A short technical report from Semantic and harmonisation best practice - D3.2 deliverable by WP3, a short technical report that describes WP4 Joint Variant genotyping use case, a short report on WP6 "How FAIR are You" webinar series and hackathon, and a CINECA project poster summarising the project's progress to date, presented at the 2021 GA4GH 9th Plenary. Table 6 lists the short technical reports published by CINECA work packages in 2021.

Table 6: Summary of short technical reports published in 2021.

Report title	Date	Author/Editor
Implementation of GA4GH standards in CINECA	February 2021	D. Spalding (EMBL-EBI)
Beacon cohorts: A model for cohort discovery in CINECA and beyond	March 2021	L. Fromont (CRG)
Passport is the glue between the researcher, data and computing	April 2021	M. Linden (CSC)
Powering up data discovery and access using the Data Use Ontology	May 2021	M. Courtot (EMBL-EBI)
Semantic and harmonisation best practice - D3.2	June 2021	M. Courtot, I. Liyanage (EMBL-EBI)
"How FAIR are You" webinar series and hackathon	June 2021	N. Mulder, M. Mbiyavanga (UCT)
Joint Variant genotyping use case	July 2021	E. Cirillo (The Hyve), Á. González (CSC), K. Tsukanov (EMBL-EBI)
CINECA project poster	October 2021	M. Mbiyavanga (UCT), L. Glass (EMBL-EBI)
EUCAN Projects General Factsheets	November 2021	EUCAN Cluster Dissemination and Communication group

3.2.6. Dissemination activities

Dissemination of CINECA outputs is also carried out through attendance at conferences and giving lectures. We collected all the events (conferences, workshops, seminars) attended or presented (talks or posters) by CINECA partners.

During this reporting period, engagement with stakeholders has again been limited by the travel bans and lockdowns due to COVID-19. Therefore, all our engagement and outreach occurred through online meetings and events. Despite all the challenges posed by the pandemic, CINECA partners continued with the stakeholder engagement by attending virtual conferences and meetings, as summarised in Table 7. In addition, CINECA also engaged with various stakeholders on topics of common interest, including GA4GH, IHCC, EUCAN Dissemination and Communication, EUCAN Data Catalogue, and EUCAN Harmonisation & federated analysis working groups. For example, with the EUCAN Dissemination and Communication, we have monthly calls and share dissemination efforts in each project. From March, the EUCAN Dissemination and Communication started releasing regular newsletters that include dissemination and outreach communications of all the participating projects, including CINECA, euCanShare, EUCAN-Connect, iReceptor Plus, EUCANCan and ReCoDID.

Table 7: Selected non-CINECA events attended by CINECA partners in 2021.

Event*	Details	Date (2021)	CINECA Lead	People reached
X-Omics/BBMRI-NL Workshop	Lecture	January	J. van der Velde, L. Johansson	+100
TES release on GA4GH	Press release	March	L. Glass	
Personal Genomes 2021	Lecture	April	T. Keane	+100
EOSC-life series on remote training	Lecture	April	S. Hiltemann	+25
ISCB-Africa ASBCB Conference on Bioinformatics	Lecture	June	M. Mbiyavanga, N. Mulder	+500
ELIXIR All Hands	Lecture	June	G. Saunders, S. Scollen, N. Mulder	+150
ELIXIR CONVERGE AGM	Lecture	June	G. Saunders, S. Scollen	+90
ELIXIR UK Health Data Day	Lecture	June	G. Saunders, S. Scollen	+60
Bioinformatics Education Summit 2021	Lecture	June	M. Mbiyavanga, N. Mulder, V. Matser	+90
H3ABioNet/Wellcome Connecting Science NGS workshop	Lecture	June	T. Keane, N. Mulder	423
ISMB/ECCB 2021	Poster	July	N. Naderi, P. Ruch, D. Teodoro	+2000
ISMB/ECCB WEB2021: Making training materials FAIR - experiences, challenges, solutions	Lecture	July	S. Morgan	n/a
ISMB/ECCB 2021	Panel discussion	July	N. Mulder. V. Matser	n/a
Regions 4 PerMed	Lecture	July	Gary Saunders	
European Society of Human Genetics Conference 2021	Poster	August	L. Johansson, M. Swertz	+1000
18th H3Africa meeting	Lecture	August	M. Mbiyavanga, N. Mulder	+500
9th GA4GH Plenary	Lecture, Attendance	September	K. van Bochove, M. Linden, M. Courtot, L. Fromont, J. Rambla, M. Mbiyavanga, N. Mulder, T. Keane	+1000
DPUK Autumn academy	Lecture	September	M. Courtot	
GA4GH Connect	Lecture, Attendance	October	M. Linden, M. Courtot, L. Fromont, J. Rambla, M. Mbiyavanga, N. Mulder, T. Keane	+500
Beilstein Open Symposium	Lecture	October	M. Linden	
SSI Demo - give a try	Lecture	October	M. Linden	+100
BioHackathon 2021 (Barcelona)	Workshop	November	S. Hiltemann	+300

*Event location specified only if the event was not virtual.

3.2.7. Publications

As part of CINECA's efforts to widen and diversify our outreach and dissemination strategies, we have collected peer-reviewed publications by CINECA and CINECA partners that include work done related to the CINECA project (https://www.cineca-project.eu/publications).

3.3. Future Plans

Over the past few years, most training activities revolved around work packages exchanging knowledge within and between each other to advance the project's scientific agenda. However, as the CINECA project has matured and started to complete outputs, the focus of training activities has quickly shifted from internal learning interventions to training activities targeted at a wider audience. As for the past years, we will continue to gauge training needs among CINECA partners and stakeholders and gear training planning towards meeting these needs. Since all face-to-face events since 2020 were either cancelled or moved to a virtual format, and it is unclear when it will be feasible to resume travel, we have constantly had to redesign the training plan. For instance, we are repurposing travel funds to create more videos and activities that make the training materials more accessible. Thus, we have started transcribing the training materials to add subtitles (see **Table 5**) which can be translated into French and other languages (French is the highest priority to increase accessibility in Africa and Canada). This will make the content more accessible to researchers who struggle to follow English, particularly if the lecturer has an accent. In addition, we plan to make a cartoon video of the overall CINECA activities to describe the problems being addressed and how CINECA is developing tools to solve them.

As the project moves to its final stages, we need to expand our training goals within and outside CINECA. Thus, future training events outlined in Table 9 will be geared towards showcasing CINECA outputs for end-users and implementers through focused webinars (Table 8), hands-on sessions (Table 9-11), and demonstration videos (Table 9-10). If travel resumes during 2022, we can convert some of these events to face-to-face events.

[Galaxy](#) is a widely-used data analysis platform, aimed at providing user-friendly access to often complex analysis tools to users without bioinformatics expertise. Galaxy has a strong focus on FAIR data and training principles. Within the CINECA project, analysis pipelines such as those developed within WP5 are made available within the Galaxy platform, enabling their re-use by a large audience and for increased sustainability. Together with other partners, we are planning another session of the **Galaxy workshop** for March 2022. Again, this will be a 5-day global virtual event, spanning a wide range of time zones and showcasing a wide variety of pre-recorded GTN (Galaxy Training Network) [tutorials](#). As in the previous sessions, the learners will be supported by a large community of GTN trainers to answer questions across the time zones.

In addition, we are planning a workshop on **cohort data harmonisation** to be held in the first quarter of 2022. We have identified existing cohort data harmonisation efforts through literature searches and insights from relevant meetings, such as the IHCC and GA4GH. These efforts will be contacted to help develop a curriculum and content for the workshop.

The goal of CINECA is to deliver a paradigm shift of federated research and clinical applications by developing a federated cloud-enabled infrastructure to make population-scale genomic and

biomolecular data accessible. To this end, CINECA has been developing knowledge and implementing tools that enable data federation and interoperability. In addition to the training interventions that have been delivered throughout the project, CINECA will produce an end-to-end data analysis course that will constitute the project's final training deliverable. This course will be implemented as a learning pathway and will cover all the necessary steps for performing a federated data access and analysis. The choice of a learning pathway format is informed both by the pandemic and to aid the sustainability of the training materials. The end-to-end course will make use of the synthetic data produced by the project to facilitate tool developments. The learning pathway will be made available through the EMBL-EBI Competency Hub and/or the EMBL-EBI training programme. The first version of the learning pathway is aimed at end-users and will likely encompass the following modules (currently showing working titles):

- Introduction to data federation
    - What are synthetic cohorts?
- ELSI
    - ELSI "10 things to look out for"
- Find Data
    - Browsing genomes through Beacon
- Access Data
    - DAC access through ENA
    - Galaxy & GA4GH passports
- Harmonise metadata
    - How the CINECA synthetic cohort metadata was harmonised
    - How to check if your metadata is harmonised
- Federated data analysis
    - Federated analysis using Galaxy
- Conclusion

We plan to release the modules as the content becomes ready, which means that not all modules will be released simultaneously. Longer-term, we hope to add additional modules into this structure, such as alternative places to find data or alternative methods for federated analysis, e.g. using Nextflow. The first modules will likely be available from Q2 of 2022.

Table 8: Webinars planned in the CINECA webinar series

Date	Title	Organiser
Q1 2022	Data harmonisation webinars series	WP3

Table 9: List of planned upcoming training events

Date	Topic	Organiser	Location
15 February 2022	Lighting a Beacon: training for (future) implementers	WP1	Virtual
17 February 2022	Browsing genomes: a Beacon training for users	WP1	Virtual
Q1 2022	Join in the Global Train-the-trainer initiative with a CINECA Galaxy classroom	WP6	Virtual

Q1 2022	GTN Smörgåsbord II	WP6	Virtual
Q1 2022	Data Harmonisation	WP3	TBC
Q2 2022	End-to-end federated analysis learning pathway covering the range of CINECA products	All WPs	Online
Q3/Q4 2022	Clinical/research-oriented course on products	WP5	TBC
Q4 2022	Project at ELIXIR biohackathon	WP6	TBC

As mentioned above, it is difficult to plan for future events given the uncertainty around the pandemic; however, based on the CINECA deliverable plan, the tools listed in Table 10 will be made available over the next year. Once completed and tested, these will be incorporated into the training plan as webinars, videos or training courses during the quarter following their due date.

Table 10: Upcoming deliverables for which we can plan training.

Tool	WP	Due date
Text mining processing pipeline for semi-structured data	WP3	Q4 2021
Federated Data Analytics for European Biobanks	WP5	Q4 2021
Federated discovery across registry services	WP1	Q4 2022
Integration of new cohort infrastructures to the ELIXIR AAI	WP2	Q4 2022
Development and application of optimised polygenic risk scores	WP4	Q4 2022
Meta-analysis of splicing and expression QTLs across three cohorts	WP4	Q4 2022
Universal FAIR Data Compliant Federated Biomarker Discovery Service	WP5	Q4 2022
GDPR and FAIR, compliant Diagnostic Services	WP5	Q4 2022
Curation support for care planning	WP5	Q4 2022
End-to-end federated data analysis learning pathway	All WPs	Q4 2022

Finally, we will continue our dissemination activities by attending and presenting at relevant upcoming meetings. Some suggested meetings are listed in Table 11.

Table 11: Outreach plan to run at upcoming external conferences/events.

Event name	Date	Location
International Congress of Human Genetics	13-17 February 2022	Cape Town, South Africa
The European Human Genetics Conference	11-14 June 2022	Vienna, Austria

4. Abbreviations

AAI: Authentication and Authorisation Infrastructure
BBMRI-ERIC: Biobanking and Biomolecular resources Research Infrastructure - European Research Infrastructure Consortium
CRG: Centre for Genomic Regulation
FAIR: Findable, Accessible, Interoperable, Reusable
GA4GH: Global Alliance for Genomics and Health
GDPR: General Data Protection Regulation

H3ABioNet: Pan African Bioinformatics Network
IHCC: International HundredK Cohorts Consortium
QTL: Quantitative Trait Loci
SFU: Simon Fraser University

5. Delivery and schedule

The delivery is on time.

6. Work Packages in CINECA

WP1 - Federated Data Discovery and Querying

WP2 - Interoperable Authentication and Authorisation Infrastructure

WP3 - Cohort Level Meta Data Representation

WP4 - Federated Joint Cohort Analysis

WP5 - Healthcare Interoperability and Clinical Applications

WP6 - Outreach, training and dissemination

WP7 - Ethical and legal governance framework for transnational data-sharing

WP8 - Project Management and coordination

WP9 - Ethics requirements