



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

D6.4 - Training Programme, Detailed

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:	31 Dec 2020
Actual Delivery Date:	15 Dec 2020
Authors of this Deliverable:	Nicola Mulder, Mamana Mbiyavanga, Saskia Hiltemann, Vera Matser
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:	1 Jan 2019
Duration:	48 months

Table of contents:

1. Executive Summary	3
2. Project objectives	3
3. Detailed report on the deliverable	4
3.1 Background	4
3.2 Description of Work	4
Training Programme	4
Staff Visits	4
Webinar Series	5
Workshops and training events	7
Short training videos	9
Short technical reports	10
Dissemination activities	10
3.3 Future Plans	11
4. Abbreviations	13
5. Delivery and schedule	14
6. Work Packages in CINECA	14



1. Executive Summary

In this deliverable document, we report on the activities in task 6.4 - Training Programme, describe the CINECA training activities in the first 24 months of the project and provide the Training Plan for the next 12-24 months.

As with the first 12 months of the CINECA project, many of the training efforts for the second 12 months also focused on internal learning opportunities and knowledge exchanges. However, all the training and outreach events have been heavily impacted by COVID-19, which removed our ability to hold face-to-face workshops and staff exchanges.

With all the interdependencies between the different CINECA work packages, a staff visit program was put in place to facilitate collaboration amongst the consortium members. We had received a number of requests for staff visits. Unfortunately, owing to the COVID-19 pandemic, only one such staff exchange visit could take place, with the remaining unable to proceed. The staff exchange program has thus switched to a virtual one involving online meetings.

For training interventions targeted at a broader audience, we have set up a [webinar series](#)¹, providing quarterly online learning interventions. These have been able to continue despite COVID-19 restrictions. We ran a total of 6 webinars (3 of these webinars in 2019, and 3 in 2020), with 23 attendees on average, 68% on average of those who registered. Feedback from these webinars is collected and is used to further increase the utility of the webinars going forward. The response rate to the feedback surveys was between 20 and 50% and the average rating for both pace and content has always 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 to facilitate the uptake of CINECA outputs. Eight short videos were produced by work packages on different topics. To increase engagement, the short videos were submitted to ELIXIR's training portal and disseminated via CINECA's various communication channels.

Some face-to-face workshops and training events had been planned for 2020 during the AGM and at international events, such as the IHCC, which was due to take place in Santiago, Chile. Owing to the COVID-19 pandemic, some of these training events had to run virtually (e.g. connecting to AAI workshop), and some had to be postponed (e.g. cohort data curation and harmonisation workshop).

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 dissemination of information of mutual interest through all appropriate channels to the widest potential audience.

¹ <https://www.cineca-project.eu/webinars>



- 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 across continents. The project comprises inter-related work packages addressing different needs. There is a need within the consortium for the training of staff within work packages and for knowledge dissemination across work packages, which are being addressed through staff exchanges. For training outside of the consortium, CINECA products are still being developed; therefore, the short-term focus is on gathering the needs and developing training around current known challenges, such as cohort data harmonisation. A complementary activity is the ongoing webinar series, which serves as both a training and a dissemination initiative. With the COVID-19 pandemic, we have had to adapt our training and dissemination activities significantly over the past 10 months, and the future is uncertain. Therefore, our training plan for the next two years will be tentative and will need to adapt to changing global situations.

3.2 Description of Work

Training Programme

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

Staff Visits

The procedure around staff visits has been set up, and an application form with guidance text has been made available for CINECA partners to request staff visits². Approval is subject to the relevance of the visit and available resources.

To date, one staff visit has taken place and several more were planned but put on hold as shown in tables 1 and 2, respectively.

Table 1: CINECA staff visits completed

Visitor	Host	Topic	Date
---------	------	-------	------

² <https://www.cineca-project.eu/blog/cineca-staff-visit-program>



Alexander Kellmann (UMCG)	Melanie Courtot (EMBL-EBI)	Integrations between MOLGENIS and EMBL-EBI services	01/02/2020 - 29/02/2020
---------------------------	----------------------------	---	-------------------------

As the project has progressed, we expected more intra- and inter-WP staff visits requests. However, with travel currently thwarted, the staff exchange program has been put on hold and interactions are occurring through online meetings. To this end, several cross-work package meetings have been held to discuss interdependencies. Below is a list of some staff visits that were tentatively planned for 2020 and 2021 (Table 2), but were unable to take place. The staff have been interacting online. It is worth noting that there has been a low interest and uptake from the consortium for staff visits and this has been made even harder with the COVID-19 situation. Hopefully, when things return to normal and work packages start to produce more technical outputs, there will be more interest in cross-WP interactions that would require more person-to-person interactions amongst CINECA partner staff. To encourage this, the CINECA training team is undertaking one-on-one online meetings with the individual work packages to gather their training needs and discuss staff exchange options. So far two such meetings have taken place.

Table 2: Planned CINECA staff visits for 2020 and 2021

Visitor (Work Package)	Host/WP	Topic	Date
Kaur Alasoo (WP4)	UMCG	Coordination of RNA-seq workflows between UTartu and UMCG	Q4 2019 / Q1 2020
Andrew Stubbs	EMBL-EBI	Secure Seamless integration with WP3	Q1/Q2 2020
Andrew Stubbs	WP2	Integration of AAI (Milestone M5.1)	Q1 2020
Jorge Izquierdo Ciges	Elixir-Finland	Integration of EGA with CanDIG and ELIXIR AAI	Q4 2020 / Q1 2021
Michal Prochazka	Jonathan Dursi	Integration of CanDIG with ELIXIR AAI	2020

Webinar Series

We set up a CINECA webinar series for delivery of training and dissemination purposes both within CINECA and to a broader audience of stakeholders. The CINECA webinar series aims to discuss ways to address common challenges and share best practices in the field of cohort data analysis, as well as distribute CINECA project results. All CINECA webinars 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](https://www.cineca-project.eu/webinars), and embedded on the CINECA website (<https://www.cineca-project.eu/webinars>).

Six webinars were organised in 2019 and 2020. Two of these webinars describe key challenges that the CINECA project is trying to address, namely federated data sharing and federated data analysis by bringing compute to the data (WP1 and WP4). One webinar was about cohort data harmonisation,



learning from the experiences of the H3Africa project. Two other webinars addressed ELSI issues as one of the key pillars of CINECA.

Table 3: Overview of the webinars delivered in the CINECA webinar series

Date	Title	Speaker
06/06/2019	The CINECA project - Addressing bottlenecks in cohort data analysis	Thomas Keane (EMBL-EBI)
15/11/2019	Connecting to AAI	Mikael Linden (CSC)
12/12/2019	H3ABioNet Experiences in Phenotype Data Harmonisation and Standards Development	Katherine Johnston (UCT)
24/01/2020	Ethical, legal and societal issues in international data sharing	Éloïse Gennet (INSERM), Melanie Goisau (BBMRI-ERIC)
01/10/2020	Status Update Code of Conduct: Teaming up & Talking about it	Michaela Th. Mayrhofer (BBMRI-ERIC)
12/11/2020	Data Gravity in the Life Sciences: Lessons learned from the Human Cell Atlas and other federated data projects	Tony Burdett (EMBL-EBI)

There were between 9 and 40 attendees with an average 68% conversion rate of registrations. All webinars have a feedback form that automatically displays after the end of the webinar and is sent to participants in a follow-up email. Table 4 displays the high-level statistics and feedback collected for the CINECA webinar series.

Table 4: High level statistics and feedback collected from the CINECA webinar series

Name/nmb webinars	Nmb attendees	Conversion	Survey response	Average pace (1-5)	Average content (1-5)
The CINECA project - Addressing bottlenecks in cohort data analysis	36	75%	28%	4.4	4.4
Connecting to AAI	9	75%	22%	5	5
H3ABioNet Experiences in Phenotype Data Harmonisation and Standards Development	12	46%	50%	4.5	4.3
Ethical, legal and societal issues	22	61%	23%	4.2	4



in international data sharing					
Status Update Code of Conduct: Teaming up & Talking about it	20	47%	15%	4.3	4
Data Gravity in the Life Sciences: Lessons learned from the Human Cell Atlas and other federated data projects	40	74%	35%	4.8	4.8
Summary statistics	23	63%	29%	4.5	4.4

Webinar participants are asked what topics they would like to hear about in the future; this year no new requests were received for additional webinars.

Further webinars are being scheduled for early 2021, which are linked to a FAIR (Findable, Accessible, Interoperable, Reusable) hackathon event described below (Table 5). These webinars will address a very important aspect of how CINECA outputs (data and related metadata) should be organised so that they can be more easily accessed, understood, exchanged and reused.

Table 5: Overview of the webinars planned in the CINECA webinar series

Date	Title
Q1 2021	Introduction to FAIR - F, A, I, R principles
Q1 2021	FAIR Software tools
Q1 2021	Making Cohort data FAIR
Q1 2021	Practically Fair
Q1 2021	How to make training FAIR
Q2 2021	Ethics/ELSI considerations
Q2 2021	How CINECA is implementing FAIR

All webinars part of the “How FAIR are you” series

Workshops and training events

There were several workshops and training events planned for 2020; however, only two were delivered as a result of the COVID-19 pandemic and lockdowns. The first was a Hands-on AAI integration workshop planned for the CINECA AGM in March 2020, which had to go virtual as the AGM was converted to a virtual event. The virtual hands-on workshop was aimed at internal CINECA partners planning to integrate their services to the ELIXIR AAI for user authentication and authorisation. Nine CINECA partner institutions were part of the workshop (75% conversion rate),



and participants were given access to virtual machines in the ELIXIR-CZ cloud for practice during the workshop.

The second workshop was a two and a half hour remote hands-on workshop on Advanced AAI integration, which was free and open to external participants too. The training was aimed at 15 attendees, targeting developers and administrators of services that want to integrate some advanced functionality of the ELIXIR AAI for user authentication and authorisation. Participants were expected to have an ELIXIR AAI account, while access to virtual machines in the ELIXIR-CZ cloud was given to participants for practice during the workshop.

We had planned to run a cohort harmonisation workshop at the 3rd International Cohorts Summit for the International HundredK+ Cohorts Consortium (IHCC) in Santiago in May. This summit also went virtual, and since the workshop was designed to be very interactive, it was cancelled as we felt that a virtual event would not achieve the intended impact. Another planned event was a workshop on the Galaxy data analysis platform to take place in Rotterdam in 2020. The workshop also had to be postponed to a later date in 2021 and will be held as a virtual event as discussed in the *Next steps* section.

Table 6: Overview of CINECA training events delivered.

Date	Topic	Trainers
17/03/2020	CINECA AAI Training for Relying Services	Michal Prochazka (Masaryk University), Mikael Linden (CSC)
26/11/2020	ELIXIR AAI online advanced training for Relying Services	Dominik František Bučák (Masaryk University), Mikael Linden (CSC), Michal Procházka (Masaryk University)

Table 7: List of planned upcoming training events

Date	Topic	Organiser	Location
25-27th January 2021	Train-the-trainer	WP6	Remote
15-19th February 2021	The Galaxy Global Course	WP5	Remote
Q1 2021	GA4GH Passports	WP2	Remote
Q1 2021	“How FAIR are you” series - Hackathon	WP6	Remote
3 & 5th March 2021	Using Twitter to promote your project	WP6	Remote
Q1 2021	Session at Mozfest 2021	WP6	Remote
Q2 2021	Beacon set up (technical) and user training	WP1	TBC
Q4 2021	Project at ELIXIR biohackathon	WP6	TBC
Q3/Q4 2021	Data Harmonisation	WP3	TBC, likely



			Africa
Q3/Q4 2022	Clinical/research oriented course on products	WP5	TBC
Q4 2022	End-to-end course covering the range of CINECA products across use cases	All WPs	TBC

Short training videos

A series of short training videos was created to facilitate the uptake of CINECA outputs. 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. Eight short videos were produced by work packages on different topics. 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 accessing training materials, and disseminated via CINECA's various communication channels.

Table 8: Overview of CINECA short training videos (as of 15th December 2020)

Title	Views	Length	Produced by	Date
Data harmonization and enrichment using EMBL-EBI Ontology Tools	146	13:07	WP3	Feb 2020
Solutions for overcoming cohort data integration challenges using ontology: an introduction	67	8:32	WP3	Apr 2020
Annotating data using ontologies	137	9:26	WP3	Apr 2020
CanDIG and ELIXIR AAI interoperability demonstration - ELIXIR user accessing CanDIG service	27	0:48	WP2	Jun 2020
CanDIG and ELIXIR AAI interoperability demonstration - CanDIG user accessing ELIXIR service	23	0:52	WP2	Jun 2020
An introduction to CINECA's Service Catalog - Deliverable 1.1	41	3:20	WP1	Jun 2020
CINECA Discovery Service Catalog - Technical Walkthrough - Standalone Service	29	4:36	WP1	Jun 2020
CINECA Discovery Service Catalog - Technical Walkthrough - Beacon Integration	49	6:38	WP1	Jun 2020
Annotating data using next generation biobanking ontology (NGBO)	17	6:23	WP3	Sept 2020



Some additional training videos are planned by some of the WPs, these will be developed throughout 2021 and 2022 as new products are completed.

Short technical reports

Training materials were also produced in the form of technical reports describing some of the products and activities in work packages. Table 9 lists short technical reports published by CINECA work packages in 2020.

Table 9: Short technical reports published in 2020.

Report title	Date	Author/Editor
Integration of new cohort infrastructures to the ELIXIR AAI	Jan 2020	Michal Prochazka (WP2), Leslie Glass
Sharing patient genomic and biomolecular data across continents	May 2020	Elisa Cirillo (WP4), Leslie Glass
Sustainability and SMEs	Aug 2020	Alex Michie (WP5), Leslie Glass
Uncovering metadata from semi-structured cohort data (CINECA text mining pipeline series)	Sept 2020	Isuru Liyanage (WP3), Leslie Glass
LexMapr - A rule-based text-mining tool for ontology term mapping and classification (CINECA text mining pipeline series)	Oct 2020	Gurinder Gosal (WP3), Leslie Glass
Assigning standard descriptors to free text (CINECA text mining pipeline series)	Nov 2020	Jenny Copara (WP3), Leslie Glass

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 (lecture or poster) by CINECA partners. Table 10 summarises these events for 2020.

Table 10: Events attended by CINECA partners in 2020.

Event	Details	Date	CINECA Lead	Location	People reached
Annual meeting Solve-RD: Presented poster on the analytical Sandbox	Poster	04/03/2020	Lennart Johansson	Virtual	96
EUCAN Datashield beginners workshop	Attendance	24/06/2020	Lennart Johansson	Virtual	90



European Mathematical Genetics Meeting	Lecture	17/04/2020	Kaur Alasoo	Virtual	~200
IHCC summit	Lecture	04/05/2020	Melanie Courtot, Thomas Keane	Virtual	~100
GA4GH connect meeting - Machine readable consent session	Lecture	25/03/2020	Melanie Courtot	Virtual	~500
GA4GH ClinPheno TC	Lecture	19/03/2020	Melanie Courtot	Virtual	~30
GEM Japan workshop	Lecture	28/02/2020	Melanie Courtot	Virtual	~30
16th H3Africa Consortium meeting	Lecture	07/09/2020	Nicola Mulder	Virtual	~350
H3ABioNet Annual meeting	Lecture	03/08/2020	Mamana Mbiyavanga	Virtual	~90
GA4GH 8th plenary – DUO	Lecture	01/10/2020	Melanie Courtot, Lyndon Zass (H3Africa)	Virtual	~500
ECCB conference - ELIXIR::GA4GH: Advancing genomics through expedited data access enabled by standards and ontologies	Lecture	02/09/2020	Melanie Courtot	Virtual	~100
GA4GH 8th plenary - Computable cohort representation	Lecture	28/09/2020	Melanie Courtot	Virtual	~50
H3BioNet seminar - Community standards for cohort data integration	Lecture	21/10/2020	Melanie Courtot	Virtual	~17
ECCB conference - ELIXIR::GA4GH: GA4GH standards impact in H3Africa	Lecture	02/09/2020	Mamana Mbiyavanga	Virtual	~100

3.3 Future Plans

As the CINECA project matures, the focus of training activities will increasingly shift from internal learning interventions to training activities targeted at a wider audience. 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 were cancelled in 2020 and it is unclear when it will be feasible to resume travel, we have had to redesign the training plan. In the first instance, we propose to reallocate travel funds to activities that make the training materials more accessible and to create more videos. We will transcribe the training materials to add subtitles which can then 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. 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.

We will continue training within and outside of CINECA through virtual events in 2021. The upcoming events have been outlined in the tables above under the respective sections (Tables 2, 5 & 7), details for some training events are provided below. If travel resumes during 2021, we can convert some of these to face-to-face events. The first workshop will be on “**How FAIR are you**”, to facilitate the implementation of FAIR principles. This will involve a 7-part webinar series running from January through April 2021 as discussed earlier in the webinar section.

The webinars will be followed by a 2-day virtual hackathon to assess how FAIR materials are. This will consist of 3 streams: 1) cohort data; 2) software; 3) training materials, where participants will bring their data, software and training materials, run them through assessment tools and discuss how to make them more FAIR. The hackathon will be open to anyone, but participation from CINECA and other EUCAN projects will be encouraged.

We are also planning a **Galaxy workshop** for February 2021, a virtual global 5-day Galaxy Training event that will showcase a wide variety of GTN (Galaxy Training Network) [tutorials](#). Because the event will span a wide range of time zones, all training sessions are pre-recorded to allow participants to work through them at their own pace and in their own time. A large community of GTN trainers will be available via online support to answer questions across the time zones.

Finally, we will revisit the **cohort harmonisation workshop** in 2021. We are currently identifying existing cohort data harmonisation efforts through literature searches and our 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. We will need to plan for a virtual event, but hope that by Q3 of 2021 we may be able to run this in person.

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 11 will be available over the next 2 years. Once completed and tested these will be incorporated into the training plan as either webinars, videos or training courses during the quarter following their due date.

Table 11: Upcoming deliverables for which we can plan training.

Tool	WP	Due date
CanDIG and ELIXIR AAI interoperability	WP2	Q2 2020
Query expansion services	WP1	Q4 2020
Cohort minimal metadata model	WP3	Q4 2020
Framework and APIs for executing federated genomics analyses	WP4	Q4 2020
Semantic and harmonisation best practice	WP3	Q4 2021



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 course	All WPs	Late 2022

A final deliverable will be a new end-to-end course covering all CINECA products including federated data analysis, incorporating tools, standards and use cases resulting from the technical work packages and incorporating training on the ELSI standard developed by work package 7.

In addition to these training events, we will continue with staff exchanges internally in the consortium (virtually for now) and will continue to attend international conferences and events to present CINECA outputs and engage with external stakeholders. For 2021 and 2022 some confirmed events include:

Table 12: Training plan to run at upcoming external conferences/events.

Event name	Date	Location
The European Human Genetics Conference	12 - 15 June 2021	Glasgow, United Kingdom
American Society of Human Genetics Meeting	19 - 23 October 2021	Montreal, Canada
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
 ECCB: European Conference on Computational Biology



EGA:	European Genome-phenome Archive
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

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

