Deliverable D2.4

Training Final Report

Responsible Partner:	SURF
Status-Version:	v1.0
Date:	14.07.2023
Distribution level (CO, PU):	PU

Project Number:	GA 101017207
Project Title:	DICE: Data infrastructure capacity for EOSC

Title of Deliverable:	Training Final Report
Due Date of Delivery to the EC	30.06.2023
Actual Date of Delivery to the EC	14.07.2023

Work package responsible for the Deliverable:	WP2 - Outreach, stakeholder engagement and service uptake
Editor(s):	Claudia Behnke (SURF)
Contributor(s):	Anna-Lena Flügel (DKRZ) Rene van Horik (DANS) John Favaro (TrustIT)
Reviewer(s):	Debora Testi (CINECA)
Recommended/mandatory readers:	All Partners

Abstract:	The document is the final report on the activities that were carried out during the project in the area of training. It includes reporting on the events (hackathons and datathon) as well as the training materials prepared and packaged as a legacy from the project.
Keyword List:	community, outreach, service, pay-per-click, storage, data, social
Disclaimer	This document reflects only the author's views and neither Agency nor the Commission are responsible for any use that may be made of the information contained therein

Document Description

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
v0.1	13.06.2023	Draft Table of Contents and text contributions	Claudia Behnke
v0.2	06.07.2023	Final text	Claudia Behnke, Anna-Lena Flügel, Rene van Horik, John Favaro
v0.3	12.07.2023	Annotated review returned to authors from internal review	CINECA
v0.4	14.07.2023	Final version addressing reviewer's comments	Claudia Behnke
v1.0	14.07.2023	Final submission to EC	CINECA

Table of Contents

List of Figures

Terms and abbreviations

ASTRON	Astron
BSC	Barcelona Supercomputing Center - Centro Nacional de Supercomputacion
CESNET	CESNET, z. s. p. o.
CINECA	Cineca
CSC	CSC – Tieteen Tietotekniikan Keskus Oy
CyI	The Cyprus Institute
Datacite	DataCite
DDPS	Digital Preservation Service
DKRZ	Deutsches Klimarechenzentrum GmbH
DoA	Description of Action
EC	European Commission
EOSC	European Open Science Cloud
ETHZ	Eidgenössische Technische Hochschule Zürich
EU	European Union
EUDAT ltd	EUDAT ltd
FZJ	Forschungszentrum Juelich Gmbh
GA	Grant Agreement to the project
GRNET	National Infrastructures for research and technology
GWDG	Gesellschaft für Wissenschaftliche Datenverarbeitung mbh Göttingen
INFN	Istituto Nazionale di Fisica Nucleare
IT4I	Vysoka Skola Banska - Technicka Univerzita Ostrava
KER	Key Exploitable Result
KIT	Karlsruhe Institut für Technologie
KNAW-DANS	Koninklijke Nederlandse Akademie van Wetenschappen
KPI	Key Performance Indicator
MPG	Max Planck Gesellschaft zur Foerderung der Wissenschaften e.V.
PID	Persistent Identifier
PPC	Pay Per Click
SIGMA	SIGMA2
SNIC	Uppsala Universitet
SURF	SURFsara BV
TRUST	Trust-IT services
UCL	University College London
ULUND	University of Lund
VA	Virtual Access
WP	Work Package

Executive Summary

This report documents the activities undertaken over the course of the DICE project with regard to training.

Given the importance of training also within the EOSC Future project, a close collaboration with EOSC Future was envisioned from the start of DICE (also because of the larger amount of resources allocated within EOSC Future for the financing of training activities). EOSC Future supported participants from the Western Balkan region with travel grants to the DICE Datathon and to the EUDAT summer school.

Several training-related events were originally planned for execution over the course of the project. These included:

- CompBioMed data platform integration Hackathon
- DICE Digital Preservation Service (DDPS) Hackathon
- LOFAR radio astronomy platform for advanced data discovery and preservation Hackathon
- ICOS community inversion benchmarking tool Hackathon
- Datathon on Sensitive Data

Not all were eventually executed, due to a number of reasons – however, at least one originally unplanned event was added and successfully executed (EUDAT Summer School). Those that were executed were greeted with a high level of satisfaction by the participants.

Furthermore, two originally unplanned Key Exploitable Results were produced in the form of complete training materials that may be used beyond the conclusion of the DICE project.

1   Introduction

The "Data Infrastructure Capacities for EOSC" (DICE) project is a network of computing and data centres and research infrastructures. The project aims to enable a European-wide storage and data management infrastructure for EOSC, providing generic services and building blocks to store, find, and access data consistently and persistently. Specifically, DICE partners are offering 14 state-of-the-art data management services with more than 50 PB storage capacity. The service and resource provisioning is accompanied by enhancing the current service offering to fill the gaps still present to support the entire research data life cycle. DICE provides solutions for increasing data quality and reusability, supporting long-term preservation, managing sensitive data and bridging between data and computing resources.

1.1   About this deliverable

1.1.1   Relationship to previous deliverable

DICE published D2.2 (*Training Programme* [1]) (DICE Consortium, 2021) in M10, which described the objectives of training activities in DICE. We outlined that we planned to organise and facilitate four hackathons and one datathon to address the gaps identified in the Training Programme.

In this deliverable, we will compare the realised training with the planned events and describe the events that were realised and explain how we mitigated deviations of the plan.

1.1.2   Collaboration with the EOSC Future project

From the start, a close collaboration with the EOSC Future project was anticipated since this project has much more resources for developing training material and organising training. EOSC Future focuses on the content of the EOSC portal, and their first training was planned already for September 2021. Besides the collaboration mentioned in (DICE Consortium, 2021), EOSC Future supported participants from the Western Balkan region with travel grants to the DICE Datathon and to the EUDAT summer school.

1.2   Document structure

Section 1 (this section) provides a concise introduction to the project and positions this deliverable within the overall context, particularly the collaboration with EOSC Future and relationships to the previous version of this deliverable.

Section 2 is the heart of the document and describes each event in terms of planning and organisation, and execution. This includes descriptions of the material, reasons for omissions of events, and participant feedback.

Section 3 provides conclusions and **Section 4** provides references.

2   Organised / planned events

The following sections gives an overview of the events that we had planned and those that were realised.

2.1   CompBioMed[1] data platform integration - Hackathon

Date:	Tuesday 21st of June 2022 12:00 till 18:00	Place	CINECA (Bologna/Italy)
Website:	https://www.dice-eosc.eu/news-events/events/data-management-and-publication-dicecompbiomed-hackathon	Participants (in person / remote)	25 registered / 15 joined remote

The event was the first event organised by the task and was held in conjunction with the All Hands Meeting 2022 of CompBioMed[2]. The invited speakers were Narges Zarrabi (DICE/CompBioMed/SURF), Claudio Cacciari (SURF), and Claudia Behnke (DICE/SURF), who guided the participants through Research Data Management focus on the data life cycle and data management landscape – EUDAT B2*Services including Hands-on: B2SHARE & B2FIND. Narges explained the CompBioMed use case from Task 4.2, and Claudio Cacciari explained metadata and data publication by demonstrating the principles of good metadata use with a bag of cookies, much to the participants' liking.

Real-time feedback was also collected from the participants during the hackathon. For example, Figure 1 reflects a question that was asked "What is your current profession / role", yielding responses ranging from data steward to researcher.

---

[1] https://www.compbiomed.eu/

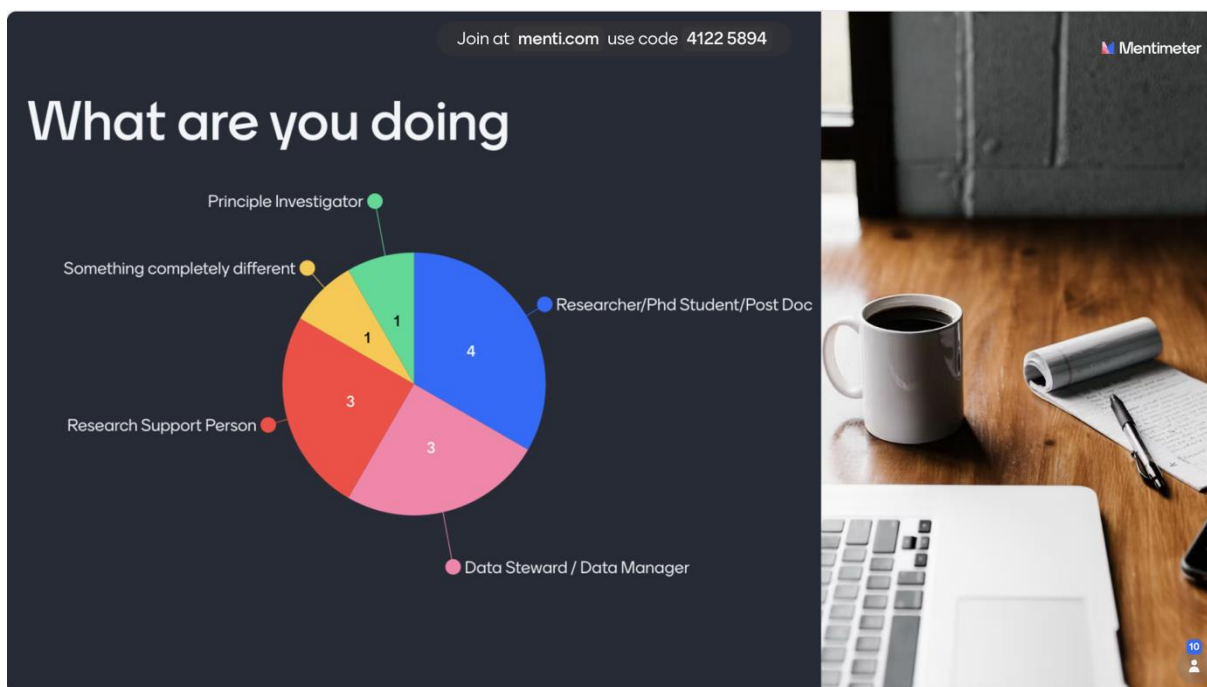[2] https://www.compbiomed.eu/events-2/

Figure 1: Feedback from CompBioMed hackathon participants - profession

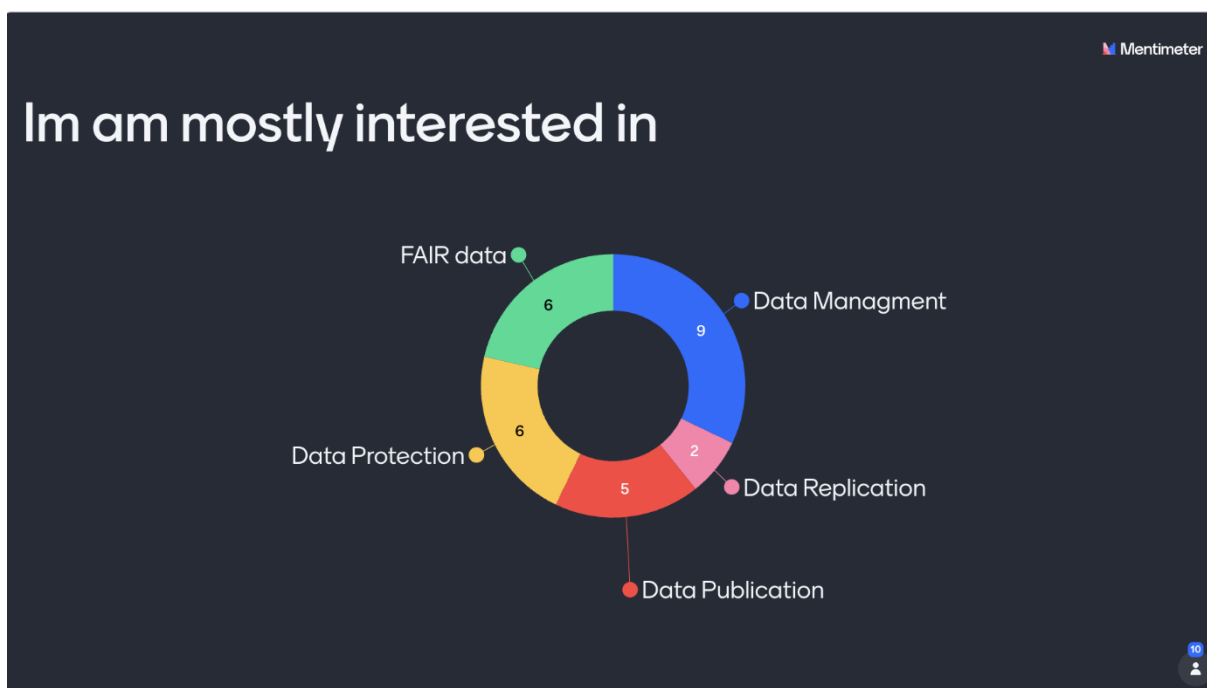Responses to the question "What are your current interests?" are reflected in Figure 2.

Figure 2: Feedback from CompBioMed hackathon participants - interests

Originally the event was planned as an in-person event, but due to travel chaos at the European airports[3], many participants were unable to arrive on time in Bologna/Italy at CINECA. On short notice a remote connection was set-up that allowed the participants to follow the hackathon.

---

[3] https://www.dw.com/en/europe-searches-for-ways-to-ease-summer-travel-chaos-a-62303939

2.2   DICE Digital Preservation Service (DDPS) – Hackathon

Date:	*Event did not happen*	Place	
Website:		Participants (in person/remote)	

One of the tasks of WP4 in the DICE project concerns the development of a digital preservation service, the "DICE Digital Preservation Service" (DDPS). DDPS is intended to transfer research data from a short or medium-term repository (B2SHARE) to a long-term-preservation (LTP) archive (DANS). The activities concerning DDPS are described in Deliverable D4.2 (DICE Consortium, 2022) [3], Chapter 4, "implementation of the LTP policy" and in D4.3 [4].

At the start of the DICE project, the design, development and deployment trajectory of the DDPS was described, and in line with this, an instruction and training plan was set up and included in Deliverable 2.2 (DICE Consortium, 2021). This plan has only been partly implemented, and a considerable amount of the resources for this plan have been re-allocated. An introduction and explanation on the basic components of the service have been presented at the IDCC conference[4]. During the DICE project duration, the DDPS has reached a "proof-of-concept" stage, so outreach concerning implementation and service support was not feasible.

Training resources were now used to assist in the preparation and execution of the Datathon on Sensitive Data (Section 2.5) and to contribute to the EUDAT Summer School (Section 2.6). The training on "data management practices and plan" in the "users track" has been prepared and delivered as part of WP2 of the DICE project.

2.3   LOFAR[5] radio astronomy platform for advanced data discovery and preservation - Hackathon

Date:	*Event did not happen*	Place	
Website:		Participants (in person/remote)	

The LOFAR Observatory operates LOFAR, a unique radio astronomical instrument with stations distributed over Europe, interconnected through a 10 Gbps wide area network and a central processing facility hosted by the Rijks Universiteit Groningen. Following initial processing at the central processing facility, observation data is stored in the Long-Term Archive (LTA) to assess the quality and reduce data volume.

Unfortunately, after several attempts, there was no interest from the community to realise a hackathon. Therefore, this event was cancelled.

---

[4] https://doi.org/10.5281/zenodo.6641722
[5] https://www.astron.nl/telescopes/lofar/

2.4   ICOS[6] community inversion benchmarking tool - Hackathon

Date:	*Event did not happen*	Place	
Website:		Participants (in person/remote)	

The Integrated Carbon Observation System, ICOS provides standardized and open data from more than 140 measurement stations across 13 European countries. The stations observe greenhouse gas concentrations in the atmosphere and carbon fluxes between the atmosphere, the land surface and the oceans. Thus, ICOS is rooted in three domains: Atmosphere, Ecosystem and Ocean.

Unfortunately, the ICOS Hackathon had to be postponed as the workflow was not yet ready for users, but it will be presented in the future ICOS summer schools.

2.5   Datathon on Sensitive Data

Date:	*30.05- 01.06.2023*	Place:	Centrum Wiskunde & Informatica, Science Park 123, Amsterdam
Website:	https://www.dice-eosc.eu/index.php/news-events/events/dice-datathon-sensitive-data	Participants (in person / remote):	30/45

Even though the EUDAT services are not by default well suited for privacy-sensitive data, communities, especially from the health and social science domain, have an increasing demand for storing and managing this data. Within DICE, Task 4.4 enhanced the existing services in such a way that they might be usable for the communities' sensitive data processing. The DICE Datathon on Sensitive Data aimed to present these services to the communities and let them try some hands-on working with the presented use cases, frameworks and tooling.

Together with experts on GDPR and the theoretical basics of sensitive data, DICE T4.4. Partners from CSC Sensitive Data Services, TSD Sensitive Data Services (University of Oslo) and INFN Cloud, as well as EOSC-Future related contributor OpenAIRE (Amnesia Tool), demonstrated a rich portfolio of services for handling all aspects of sensitive data to the participants. Topics included:

1. Sensitive Data, Health Data
2. Data Anonymization and Pseudonimisation
3. Risk Assessment and Mitigating Measures / DPIA, METC, DTIA
4. Partnering and Collaboration
5. Retention Periods / FAIR data / Archiving legislation
6. Transparency and Consent Forms
7. Privacy and Ethics

---

[6] https://www.icos-cp.eu/

8. Technical solutions for handling several sensitive data issues, e.g., access restriction, collaboration, pseudonimisation and discoverability

Communications specialists from Trust-IT Services marketed the event beforehand via event website, DICE and EOSC Future event channels, as well as social media communication on Twitter and Youtube, also during the event. See also Deliverable D2.3 [2] in this respect, to see examples of marketing materials and interviews (DICE Consortium, 2023). All participating partners and EOSC Future WP10 also announced the event on their websites or shared the event info within their professional networks. The event was recorded, and videos are being uploaded for re-use on the DICE/EUDAT YouTube channels[7].

The Datathon was aimed at students and researchers who handle privacy-sensitive data, so their backgrounds and professional roles varied. Most already had some experience or were expected to work with sensitive data, others were beginners and interested in the subject.

Real-time questions during the Datathon revealed professions / roles of the participants (Figure 3) as well as their current interests (Figure 4).

Feedback round at the end of the Datathon showed high satisfaction with the relevance of the topics, level of complexity of the presentations, and trainer availability for questions alike.
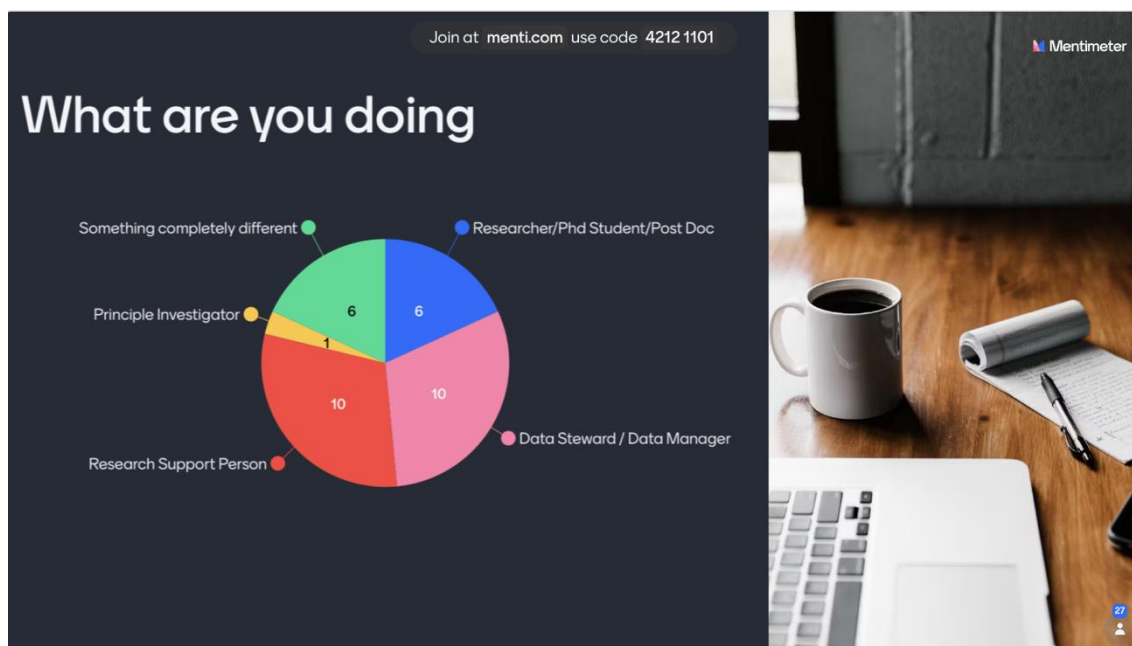

Figure 3: Feedback from Participants of the Datathon — profession

---

[7] https://www.youtube.com/channel/UChCTORr2JvRAihEaILikFrQ?app=desktop

Figure 4: Feedback from Participants of the Datathon — interests

Two partners in this task, T2.2., (SURF and DANS) furthermore contribute to the RDNL course, which recently piloted training on managing personal data (see also section 2.5.1). It will be targeted especially at data supporters[8].

2.5.1   Sensitive Data Online Training (KER)

As mentioned in (DICE Consortium, 2021), Sensitive Data has been an anticipated training topic for DICE from the beginning of the project. Here DICE collaborates with the Dutch initiative RDNL, which has launched material to help researchers to improve their data management. While DICE focused on the creation or enhancement of services that can handle sensitive data (DICE Consortium, 2022) (DICE Consortium, 2021), RDNL is focusing on better understanding the GDPR and how to classify their data.

The DICE Datathon combined hands-on experience in managing sensitive data with the facilities and workflows being developed within the DICE project. Collaborative initiatives and guest presentations from other projects are also planned, offering a rich experience to Datathon participants on all aspects and challenges of sensitive data management. RDNL provided the material for the theoretical part of the Datathon.

Although the DICE project is ending, the material can be reused in other education. To achieve that, DICE has supported RDNL in moving the material to a Moodle course: https://researchdata.nl/diensten/avg-4-data-support/. The Moodle course is free of charge and will be further maintained by RDNL, hence is a perfect example of a Key Exploitable Result.

---

[8] https://researchdata.nl/en/curious-about-new-rdnl-courses-stay-tuned-for-new-material-on-gdpr/

2.6   EUDAT Summer School

Date:	*26.06- 30.06.2023*	Place:	Kajaani/Finland
Website:	https://eudat.eu/summer-schools/eudat-summer-school-2023	Participants (in person/remote):	26/0

The EUDAT Summer School 2023 was organized in collaboration with the DICE project and EOSC Future. Participants were able to apply for travel grants covering courses, meals and accommodation. DICE provided 19 travel grants for participants; EOSC Future added four travel grants for participants from the Western Balkans.
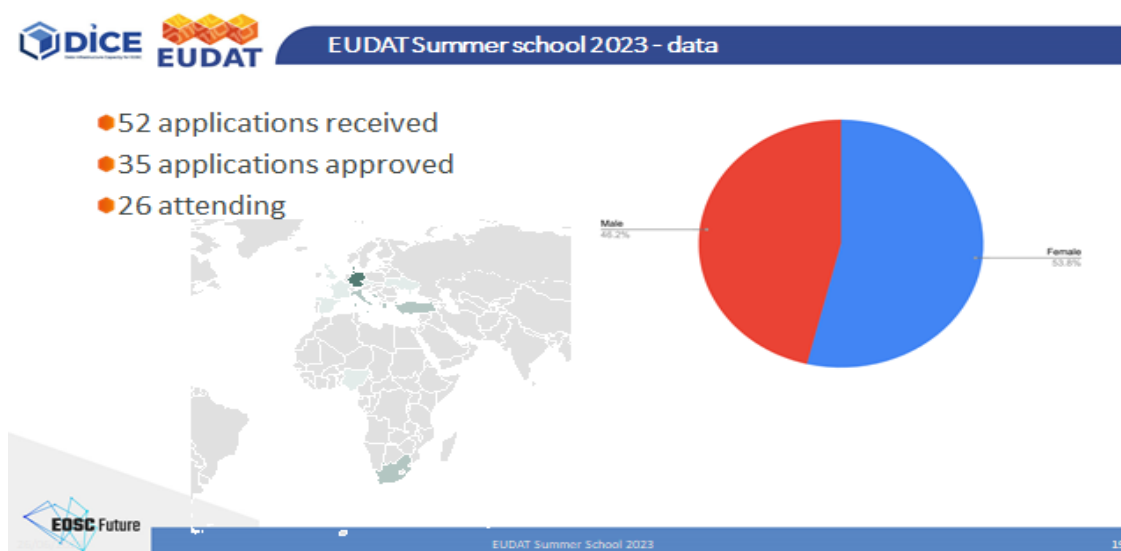

Figure 5: Participant numbers, countries of origin, and gender

The Summer School took place in the beautiful Finnish landscape of Kajaani, home of LUMI[9], the supercomputer of the European High-Performance Computing Joint Undertaking (EuroHPC JU) consortium. As the High-Performance-Computing themed EUDAT Summer School of 2020 unfortunately had had to be cancelled due to the Corona pandemic, EUDAT trainers and service owners were especially delighted to welcome students this year. The 2023 Summer School edition aimed to strengthen the participants' skills required to excel in data management and to enable them to manage and process data throughout the full research-data lifecycle (data discovery, data processing, data analysis, data preservation and data publishing).

For the first time, the Summer School offered two parallel tracks geared towards different audiences: firstly, a Providers Track for research or academic institution IT system administrators, community or data managers looking at providing research data services for their communities, in particular those that are interested in learning how to deploy EUDAT services; secondly, a Users Track, providing early career or Bachelor, Masters, PhD or Postdoc students currently engaged in complex research environments with the latest good practices in research data management, thereby covering the complete research data lifecycle.

---

[9] https://lumi-supercomputer.eu/

The course programme[10] covered EUDAT's complete B2*service suite[11] both for researchers to use the services, as well as providing how-to hands-on sessions to those interested in setting up and re-using the services at their local research institution as administrators. The programme also featured opportunities for socializing like a reception dinner at a historical Finnish manor house, a mixer for trainers, users, and providers to foster an exchange on current projects and to clarify any pending questions, a steamboat ride and, of course, a visit to the LUMI supercomputer at the CSC data centre. Most of the trainers (part of the DICE team) were present on site during the whole Summer School for participants' questions or feedback. Participants' satisfaction was also evaluated at the closing event (see Figure 6 and Figure 7). Especially the hands-on training sessions and the trainers' competencies in their field received positive user feedback.

A post-event survey is also being conducted. The replies received at the time of writing show that the school met the original expectation of about 70% of the trainees with 89% of them satisfied or more with the school organisation overall and it was reported to have been a wonderful opportunity to broaden their knowledge on data management practices and the FAIR data principles.
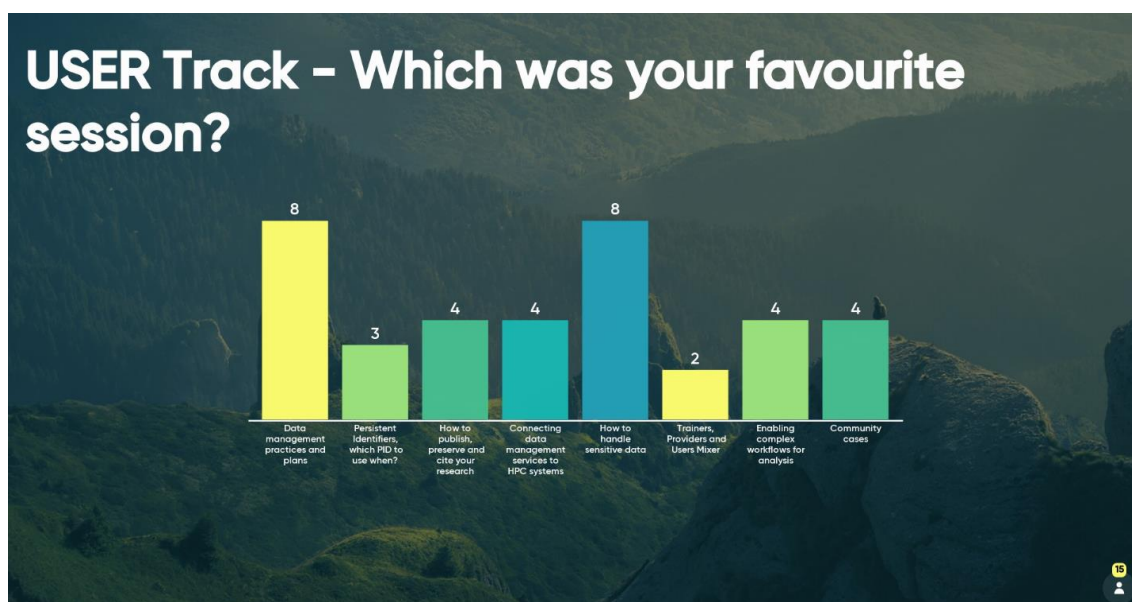

Figure 6: Feedback from the user track

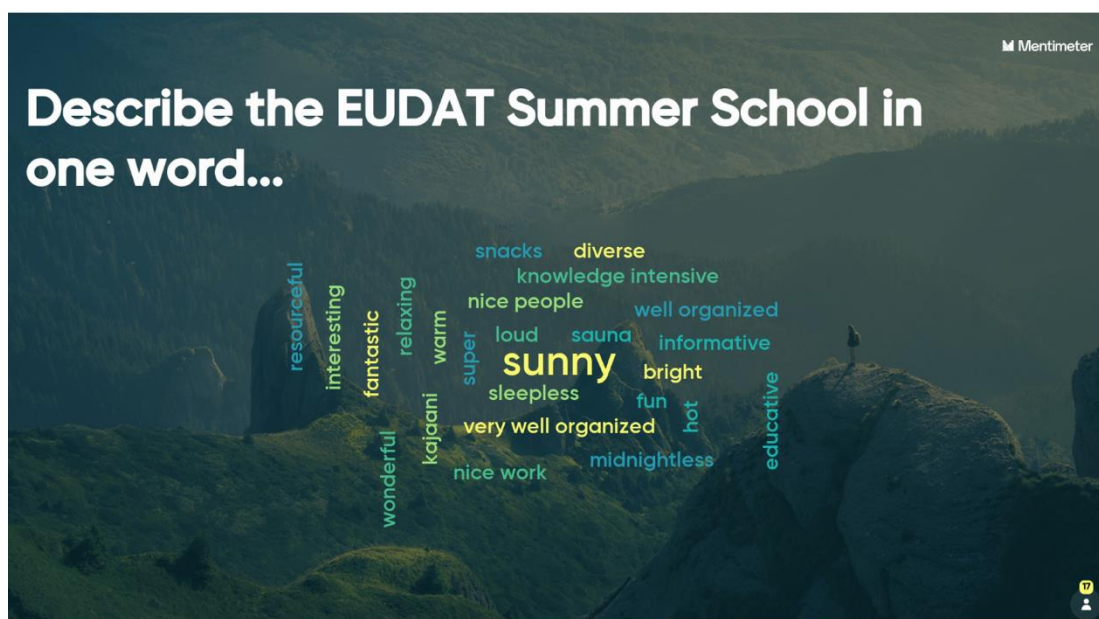---

[10]https://eudat.eu/summer-schools/eudat-summer-school-2023
[11]https://eudat.eu/catalogue

Figure 7: Participant feedback in general

2.6.1 Video Material of the EUDAT Summer School (KER)

The sessions of the EUDAT Summer School were filmed in order to create sustainable material for more trainings of the B2*Services. The equipment and manpower to professionally film the sessions was paid by DICE and the material will be later made available on the EUDAT website. Hence the material is another example of a Key Exploitable Result.

3 Conclusions

As part of the Grant Agreement KPIs were defined at the start of the project to monitor the success of the training activities (as reported in the table below). Even though not all the events were carried out as planned, DICE realized two Hackathons (ComBioMed & EUDAT Summer School), one Datathon (Sensitive Data), and created two KERs. At the same time, the number of expected trainees was reached.

KPI	Target value at M30	Achieved at the end of the project
Training activities to increase uptake of the DICE resources	4 Hackathons User documentation and training material	2 Hackathons & 1 Datathon Training material (i.e. slides, recordings) is being made available after the different events and will be maintained/re-used in the future as part of the DICE KERs
# hackathons participants	120	126 Even if the number of training events was not completely achieved, the overall number of attendees to the hackathons/datathon was higher than planned.

4   References

[1] DICE Consortium. (2021). D2.2 - Training Programme.
https://doi.org/10.23728/B2SHARE.412E4AD7BAE34DC48BA7F8E599A47D33

[2] DICE Consortium. (2023). D2.3 - Final report on Communications and Dissemination.

[3] DICE Consortium. (2022). D4.2 - Pilots for the integration with other services and platforms.
https://doi.org/10.23728/B2SHARE.B27BA476B37740D598A87092E1A527F0

[4] DICE Consortium. (2023). D4.3 - Planning for the integration with other services and platforms.