

## D7.4: Report on Training Workshops

Author(s)	Angus Whyte, Ellen Leenarts, Jerry de Vries, Rahul Thorat, Eileen Kuehn, Gergely Sipos, Valentino Cavalli, Vasso Kalaitzi, Kevin Ashley
Status	Final
Version	V 1.2
Date	2018-11-30

### Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

### Abstract:

This report documents the delivery and outcomes of the four main WP7 training events held between November 2017 and August 2018. Three were consultative workshops targeting trainers in the Research Infrastructure community and coordinators of research data management services in Institutions. The fourth was the GridKa School 2018, which offered the opportunity to test the applicability of the Skills Framework developed in WP7 to a broad data science training curriculum.

The consultative workshops identified priorities resulting from discussion of three main challenges: (1) infrastructure for training; (2) assisting research performing organisations to develop the competences and capabilities for open data science; and (3) policy coordination on skills and reward mechanisms to stimulate open research data practices.

The report then describes the relevance of GridKa School training to the skills gaps previously identified in WP7. This leads to an assessment of the topics that trainees were interested in learning about, and how these relate to the EOSCpilot Skills Framework. Finally, the report draws conclusions from the two kinds of event organised in WP7; discussion of opportunities and challenges to delivering the skills improvements needed, and delivery of training to do that.

Document identifier: EOSCpilot -WP7-D7.4	
Deliverable lead	UEDIN-DCC
Related work package	WP7
Author(s)	Angus Whyte, Jerry de Vries, Ellen Leenarts, Rahul Thorat, Eileen Kuehn, Valentino Cavalli, Kevin Ashley
Contributor(s)	Gergely Sipos, Vasso Kalaitzi,
Due date	30/11/2018
Actual submission date	30/11/2018
Reviewed by	Helene Brinken (SUB-UGOE) ; Mark Thorley (UKRI)
Approved by	Brian Matthews (UKRI)
Start date of Project	01/01/2017
Duration	24 months

### Versioning and contribution history

Version	Date	Authors	Notes
0.1	07/06/18	Rahul Thorat, Ellen Leenarts (DANS), Valentino Cavalli, Vasso Kalaitzi (LIBER), Gergely Sipos (EGI) Angus Whyte (DCC)	Draft compilation of workshop 1 and 2 agendas, notes and outcomes
0.2	02/11/18	Angus Whyte (DCC)	Report structure
0.3	02/11/18	Eileen Kuehn (KIT), Valentino Cavalli (LIBER)	GridKA School outcomes integrated into report
0.4	06/11/18	Angus Whyte, Jerry de Vries, Ellen Leenarts	Further editing of sections 3 and 4, introduction, executive summary
0.5	9/11/18	Angus Whyte, Jerry de Vries, Ellen Leenarts, Eileen Kuehn, Valentino Cavalli	Further editing of sections 3 and 4
0.6	13/11/18	Angus Whyte	Further editing all sections and draft conclusions
0.7	29/11/18	Angus Whyte, Kevin Ashley, Eileen Keuhn, Ellen Leenarts	response to review comments, final proofread
0.8	29/11/18	Kevin Ashley	Formatting and final corrections

The European Open Science Cloud for Research pilot project (EOSCpilot) is funded by the European Commission, DG Research & Innovation under contract no. 739563

**Copyright notice:** This work is licensed under the Creative Commons CC-BY 4.0 license. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0>.

**Disclaimer:** The content of the document herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the document is believed to be accurate, the author(s) or any other participant in the EOSCpilot Consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the EOSCpilot Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the EOSCpilot Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

## TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	6
2. INTRODUCTION	6
2.1. Overall objectives in the project context	6
2.2. Purpose of this document	6
2.3. Relationship to other project outputs	7
2.4. Structure of this document	8
3. Consultative workshops to clarify needs and expectations	9
3.1. Framework for the discussion	9
Consultative Workshops	10
3.2. Consulting research infrastructures and e-infrastructures	11
3.2.1. Part 1 Scene Setting at EOSC Stakeholders Forum [W-1A]	11
3.2.2. Part 2 Options and Feedback from EUDAT2018 [W-1B]	13
3.3. Consulting trainers from research institution RDM services (IDCC) [W-2]	16
3.4. Conclusions: skills needs and expectations of EOSC	19
4. Delivering training for data science and stewardship	22
4.1. Background on GridKa School	22
4.2. Method	22
4.3. GridKa School as a testbed for the EOSCpilot skills framework	23
4.4. Assessing the applicability of the EOSCpilot Skills Framework	25
4.4.1 Applicability of the Framework for organisers to plan a training event	25
4.4.3 Applicability of the Framework in terms of trainees' learning objectives	25
5. Overall Conclusions	30
Annex A. GLOSSARY	32
Annex B. Requirements Workshop Report	33
Notes from EOSCpilot WP7 Skills Requirements Workshop Mercure Hotel, Amsterdam Airport, 17 May 2017	33
Attendees	33
Background	33
Skills landscape- presentations	33
Introduction- EOSCpilot skills implications and impacts	33
Afternoon Session: Data stewardship – what is it, where are the gaps, how to improve skills and collaboration?	36
Introduction:	36
What data stewardship competencies are relevant to EOSCpilot?	36
Where the gaps are in our competence model and in skills development activity	37
Conclusions: How should EOSCpilot address the gaps	38

Annex C. Mentimeter results from consultative workshops	39
Workshop W-1B EUDAT 2018 Porto	39
Workshop W-2, IDCC 2018 Barcelona	41
ANNEX D Complementary Data and Statistics of GridKa School 2018	45
1) Questionnaire templates	45
Participant questionnaires	45
2) Skill gaps mapped to programme topics	46
Skills for integration	46
Skills for service management	47
Skills for effective use	48
Skills for assessment	48
Additional contributions	48
3) Statistics regarding professional groups	48
4) Statistics regarding service roles	49
5) Learning objectives and their coverage	51
6) Overall effectiveness of GridKa School	53

## LIST OF FIGURES

Figure 1 - Participant views on feasibility of methods to make training materials FAIR	14
Figure 2 - What should be EOSCs priorities for skills development? ('EUDAT workshop 1-B')	14
Figure 3 - What should be EOSCs priorities for skills development? ('IDCC workshop 2')	17
Figure 4 - What forms of train-the-trainer support would you expect from EOSC? (IDCC workshop 2)	17
Figure 5 - How feasible is it to make training materials FAIR using the following methods? (IDCC workshop 2)	18
Figure 6 - Mean ratings (1-10) of relevance for plenary lectures and tutorials for GridKa School starting from 2014.	25
Figure 7 - Learning objectives that are of interest to at least 20% of GridKa School attendees.	26
Figure 8 - Effectiveness of GridKa School in enhancing competences per service role.	27

## LIST OF TABLES

Table 1 - Count of lectures and tutorials from GridKa School agenda that were mapped to EOSCpilot competence gaps	24
---	----

## 1. EXECUTIVE SUMMARY

This report documents the delivery and outcomes of a WP7 requirements workshop held in May 2017, and of the four subsequent WP7 training events held between November 2017 and August 2018. Three of these events were consultative workshops involving trainers and coordinators of research data management (RDM) services. The fourth was the GridKa School 2018, an event with an established reputation for data science and data management training. WP7 involvement in organisation of this event gave the opportunity to test the applicability in this context of the Skills Framework.

The consultative workshops targeted trainers in the Research Infrastructure community, and coordinators of research data management services in Institutions. The report identifies priorities from their discussion of three main challenges 1) infrastructure for training; 2) assisting research performing organisations to develop the competences and capabilities for open data science; and 3) coordination of national-level skills policies, strategies and reward mechanisms to stimulate open research data practices.

The report then describes the relevance of GridKa School training to the skills gaps previously identified in EOSCpilot D7.2, and the lesson learned from applying the framework in an existing context. This leads to an assessment of the topics that trainees were interested in learning about, and how these relate to the EOSCpilot Skills Framework. This was carried out by co-organising the delivery of the curriculum, including consultation with its trainees on the expected learning objectives and how well these were met.

Finally, the report draws conclusions on how the workshops have informed the work of the project to address challenges to the delivery of skills improvements needed in the EOSC environment.

## 2. INTRODUCTION

### 2.1. Overall objectives in the project context

The workshop and training activity in WP7 was carried out with four main purposes aligned to the project objectives.

- Consult the training and education community about the scope of the skills in data stewardship that need to be improved for the research community to benefit from the EOSC environment, and the appropriate forms of skills development activity.
- Identify community expectations of EOSC as a harmonized training delivery infrastructure offering services to develop skills.
- Deliver and evaluate pilot training activity intended to address skills gaps found through landscape analysis and engagement with other EOSCpilot WPs.
- Assess trainee's expectations of the relevance to their learning objectives of data stewardship topics drawn from the skills gaps and more broadly from the EOSCpilot Skills Framework.

### 2.2. Purpose of this document

The purpose of the report is to document the workshop and training events organised by EOSCpilot WP7, and report on their outcomes with respect to the aims above. The report describes the delivery and outcomes of a requirements workshop (Amsterdam, May 2017) involving the WP partners, and four subsequent events held between November 2017 and August 2018. Three were consultative workshops, and these were held at the EOSC Stakeholders Forum (Brussels, November 2017) [W-1A], the EUDAT2018 conference (Porto, January 22-25) [W-1B], and the International Digital Curation Conference IDCC18 (Barcelona, February 19-22) [W-2]. These workshops targeted trainers in the Research Infrastructure community, and coordinators of research data management services in Institutions. The report identifies priorities drawn from the participants' discussion of the three main challenges the workpackage addresses, as follows:

- How can EOSC support research training providers to contribute to international level training infrastructure?
- How can EOSC assist research performing organisations to develop the competences and capabilities

- for open data science?
- How can EOSC coordinate national-level policies, strategies and reward mechanisms to stimulate open research data practices?

The report then describes work to deliver training addressing skills gaps, and to assess the overlaps between the scope of that training and the EOSCpilot Skills Framework. This was carried out through co-organisation of a major data science and data management training event, the GridKa School 2018, and through consultation with its trainees and tutors on the expected learning objectives and how well these were met.

It is important to note that the report covers the events that were scheduled in the Description of Work and organised by EOSCpilot. This excludes a number of events:

- 1) Events that EOSCpilot participated in but were organised by others, and are already reflected in the D7.3 report e.g. the EDISON project (at the 3rd EDISON Data Champions Conference, Warsaw June 2017, and at the DI4R Conference, Brussels, December 2017), OpenAIRE (September 2017), and CODATA Summer School (July 2018);
- 2) Events that apply the EOSCpilot Skills framework reported in D7.3 in conjunction with user organisations or stakeholders. Outcomes of these will be reported in D7.5 together with responses to an online consultation. The events in this category are: a workshop on skills for open science (with 4TU Research Data Service at the Technical University of Delft in September 2018); and forthcoming webinars on the EOSCpilot Skills Framework (with the LIBER Working Group on Digital Skills in Libraries), and on ethics in the EOSC environment (with WP3 partners).

### 2.3. Relationship to other project outputs

The requirements workshop took place as the skills work-package was taking shape around its first task T7.1 - Internal and External liaison and gap analysis. The task had a broad focus on identifying the data handling skills necessary for 'open data science', a shorthand expression for the union of two sets of skills; for performing science openly, and for performing science with the aid of data-intensive techniques. The task scope included identifying skills requirements emerging from the first of three phases of Science Demonstrators (EOSCpilot WP4), and from relevant initiatives to develop skills through training and other means, especially among Research Infrastructures, e-Infrastructures, research producing organisations, and H2020 projects. Work on competence frameworks was a particular focus, as this stage of the work was to inform the first WP7 deliverable, D7.1 Landscape Analysis and Competence Framework. The requirements workshop was therefore a key step towards collating the work done and scoping the D7.1 report delivered six weeks later at the end of June 2017.

The training workshops which took place in the following months (July 2017 - February 2018) were designed as consultative events. The initial requirements workshop had identified the need for this, considering the extent of existing training resources and the need for further evidence of gaps and priorities. The first of these consultative events was W-1A, and discussion of priorities both there and at the subsequent DI4R conference (December 2017) contributed to the second WP7 deliverable, the D7.2 *Interim report and Catalogue of EOSC Skills Training and Educational materials*.

The consultation was continued through to two half-day workshops at conferences early in 2018; the first being the EUDAT2018 conference (Porto, January 22-25), and the second the International Digital Curation Conference IDCC18 (Barcelona, February 19-22). The follow-up discussion of skills gaps and training support priorities in these fed into D7.3 *Skills and Capability Framework*.

The conclusions drawn from this report will also contribute to D7.5 *Final Report on EOSC strategy for sustainable development of skills and capabilities*.

## 2.4. Structure of this document

Section 3 describes the design, delivery and outcomes of the consultative workshops W-1A, W-1B and W-2.

Section 4 describes the scope of the GridKa School 2018, how its curriculum was shaped to address skills gaps identified in the EOSCpilot project, and the overlaps between this curriculum and the EOSCpilot Skills Framework.

Section 5 provides the report conclusions.

Annexes contain a glossary of terms used, a report from the Requirements Workshop, consultative workshop participants' responses to questions on the challenges discussed, and GridKa School trainees' responses on evaluation questionnaires.

### 3. CONSULTATIVE WORKSHOPS TO CLARIFY NEEDS AND EXPECTATIONS

#### 3.1. Framework for the discussion

##### Requirements workshop

A need for workshops was identified at the outset in the description of work for EOSCpilot WP7, where the focus of the training was described in general terms as ‘open data science’. This was envisaged as combining the principles and methods associated with both data science and open science, in a manner to be clarified during the project. A requirements workshop, held early in the project, was a means towards this end.

The requirements workshop was held in May 2017 (see notes in Annex A), involving project partners and input from external experts. The discussion resulted in a sharper focus on data stewardship in the open science and data science context. The discussion drew on various sources presented in the workshop:

- EOSCpilot High-Level Expert Group first report <sup>1</sup>
- competence frameworks for data science, <sup>2</sup> and ‘data information literacy’ <sup>3</sup>
- interim results from the OSPP Skills Working Group survey of researchers on open science skills <sup>4</sup>
- training developed in CORBEL and Rltrain <sup>5</sup>
- engagement with the first Science Demonstrators <sup>6</sup>

The Science Demonstrators were short projects (EOSCpilot WP4 sub- projects) proposed by research groups and institutions. Their role in the project was to pilot newly integrated services and resources for specific scientific use cases. By the time of the workshop it was already clear that the Demonstrators involved advanced skills and domain knowledge to manage complex datasets and software requirements. The workshop concluded that ‘data stewardship’ in the EOSC context would need to be scoped broadly enough to encompass skills associated with a variety of roles. A subset of the roles and competences defined in the EDISON project was highlighted, including domain research, data science engineering, and data analytics, as well as the data management role more usually associated with data stewardship.

Science Demonstrators in EOSCpilot were to be funded in three tranches, had a short duration, and were funded because they already employed the pre-requisite technical skills and domain knowledge. It became clear at this stage that it would be difficult in the short term to derive training needs for EOSC from the Demonstrators, or to deliver training to those projects according to the WP7 schedule. There were three main implications for the work-package:

1. Evidence of skills gaps from Science Demonstrators would need to be monitored as their outputs emerged, and fed into iterations of the Skills Framework.
2. The training community in research infrastructures and institutions would need to be consulted on their priorities for an EOSC training strategy, including the infrastructure to deliver events and materials.
3. Training relevant to EOSC services and Science Demonstrator outputs would need to be delivered when the nature of these became clearer later in the project.

---

<sup>1</sup> First report of High Level Expert Group on the EOSC: <https://ec.europa.eu/digital-single-market/en/news/first-report-high-level-expert-group-european-open-science-cloud>

<sup>2</sup> Demchenko, Y. Belloum, A. and Witkowski, T, (2016.) EDISON Data Science Competence Framework v.0.7 EDISON Project, <http://edison-project.eu/data-science-competence-framework-cf-ds>

<sup>3</sup> Data Information Literacy <http://blogs.lib.purdue.edu/dil/what-is-data-information-literacy/>

<sup>4</sup> Final Report of the ERA Working Group on Skills, available at: <https://ec.europa.eu/research/openscience/index.cfm>

<sup>5</sup> See Rltrain and CORBEL workshop description at: [http://rltrain.eu/news/-/asset\\_publisher/iZ6BmkMaEhsa/content/rltrain-and-corbel-workshops](http://rltrain.eu/news/-/asset_publisher/iZ6BmkMaEhsa/content/rltrain-and-corbel-workshops)

<sup>6</sup> See EOSCpilot website at: <https://eoscipilot.eu/science-demonstrator-topics>

The Skills and Capability framework in D7.3 contains the work done to respond to the first point. The consultative workshops in this section represent the outcomes on the second point, and the GridKa School results described in section 4 address the third point.

### Consultative Workshops

The three events described further in this report are:

1. 'Building Data Stewardship Expertise in Europe: How can we fill the gaps?' EOSCpilot Skills Session. Stakeholders Forum Brussels. 28 Nov. 2017 (referred to as W-1A)  
<https://eoscpilot.eu/content/building-data-stewardship-expertise-europe-how-can-we-fill-gaps>
2. 'The EOSC as a 'skills commons' providing FAIR training for FAIR data stewardship' co-located event at EUDAT Conference "Putting the EOSC vision into practice", Porto. 25 Jan. 2018 (referred to as W-1B')  
[https://eudat.eu/eosc\\_as\\_a\\_skills\\_commons\\_providing\\_fair\\_training\\_for\\_fair\\_data\\_stewardship](https://eudat.eu/eosc_as_a_skills_commons_providing_fair_training_for_fair_data_stewardship)
3. 'The EOSC as a 'skills commons' for developing research data stewardship skills at scale' workshop at IDCC 18 "Beyond FAIR - from principles to practice to global join up", Barcelona 19 Feb. 2018 (referred to as W-2) <http://www.dcc.ac.uk/events/idcc18/workshops#workshop2>

Workshops W-1A and W-1B intentionally had the same structure and discussion questions, but involved different audiences. The results of these were used to inform W-2, which was an elaboration of the format either dealing with questions whose answers were still not clear, or moving some topics forward based on results from earlier work.

### Methods used

Our approach for gathering feedback during the workshops was to ask structured questions on identified discussion themes, gathering feedback in small group discussions and live polling using an online tool. Because of the limited available time at W-1A (during the stakeholders forum) we did not use live polling there.

### Discussion themes

For the consultative workshops we selected three challenges for discussion. The main issues were derived from the Description of Work for the work-package, then more specific questions were elaborated from successive workshop discussions as it became clearer which issues were salient to the audience. The main issues that recurred across these discussions were as follows:-

*Challenge 1: How can EOSC support research training providers to contribute to international level training infrastructure?*

Questions

- Stakeholders' event : Should it harvest information into a central catalogue about training events and materials from EOSC participating organisations?, As a central catalogue was highly desired by the audience we asked the following question at the EUDAT and IDCC workshop sessions: Should FAIR principles be extended to training resources, and if so what kinds of resources are worth the effort to make reusable?
- Should EOSC collate materials across infrastructures into an EOSC wide portal?
- Should EOSC perform quality assurance and/or certification of providers, and/or badging of content?
- Should EOSC monitor what is being provided and attempt to fill gaps either in the content or mode

of delivery?

*Challenge 2: How can EOSC assist research performing organisations to develop the competences and capabilities for open data science?*

Questions

- How can EOSC assist institutions to plan the skills required to deliver their strategies and services for implementing FAIR principles, open science and data science?
- Should EOSC broker the supply and demand for disciplinary-focused training across institutions and research infrastructures?
- What are the biggest gaps in cross-disciplinary skills for data stewardship?

*Challenge 3: How can EOSC coordinate national-level policies, strategies and reward mechanisms to stimulate open research data practices?*

Questions

- What information could EOSC collect and publish to inform national-level policies, strategies and mechanisms?
- What can EOSC do to encourage and amplify efforts of funding bodies, institutions and other stakeholders to recognise researchers' skills for data stewardship and open research practices?
- What can EOSC do to nurture the career structures and rewards for professional support staff who contribute to open research practices?

## 3.2. Consulting research infrastructures and e-infrastructures

These groups were targeted in W-1A and W-1B.

### 3.2.1. Part 1 Scene Setting at EOSC Stakeholders Forum [W-1A]

#### Agenda

The EOSC Stakeholders Forum 2017 targeted research infrastructures, research e-Infrastructures and research communities. It provided an opportunity to set the scene for the later workshops by clarifying the issues and options needing further discussion. The Skills session (W-1A) was a one-hour parallel session.<sup>7</sup> It aimed to bring together professionals associated with the field of research data management (RDM) and open science, particularly those working in the infrastructures, to identify creative solutions for future EOSC training infrastructure and actions needed to make them happen. The session started with three 5-minute introductory lightning talks. After these talks, the participants were invited to join a table addressing one of the three main challenges. Participants could then move around the tables of their choice. Facilitators at each table suggested topics and summarised the issues and options discussed.

#### Speakers

The session at the stakeholder forum was chaired by Kevin Ashley (EOSCpilot, DCC). The speakers of the three 5-minute introductory lightning talks were:

- Robin Rice (Head of Research and Data Support Section, University of Edinburgh, UK); presented MANTRA, UoE Training matrix, Training and Awareness raising workshops.
- Kevin Ashley (EOSCpilot, DCC); presented the EOSC pilot infrastructures, and
- Gareth O'Neill (President, European Council of Doctoral Candidates and Junior Researchers, Eurodoc, the Netherlands); presented gaps in data management planning for the researchers.

#### Responses to challenges

<sup>7</sup> Agenda for the session: <https://eoscipilot.eu/content/building-data-stewardship-expertise-europe-how-can-we-fill-gaps>

***First challenge: How can EOSC support research training providers to contribute to international level training infrastructure?***

During the discussion a ‘two-tier’ model for training and skill development for EOSC was suggested:

- Tier-1: Core skills and training topics should be covered with high-quality curated materials and support for self-paced or instructor-led courses. These enable EOSC to fulfil its baseline commitments in federating services and enabling researchers to carry out open science. To make sure every skill/topic area is covered new courses might be needed to fill gaps.
- Tier-2: Domain specific and specialised skills and training topics. The view expressed was that EOSC does not need to control quality and ensure full coverage in this area. This will be more ‘donated’ than curated content.

Both tiers could be filled by EOSC harvesting existing content from providers and by new content being delivered in response to demand, possibly commissioned by EOSC from a range of providers.

***Second challenge: How can EOSC assist research performing organisations to develop the competences and capabilities for open data science?***

During the discussion several actions were identified for EOSC. First, EOSC should match skills demand and skills gaps to the provision of training. A more centralised catalogue of training would be useful here in order to be able to assess current provision. Second, EOSC should offer a (possibly domain-dependent) checklist defining minimum standards for training. These combined would result in a grid of domains at various levels, ‘vertical’ discipline-focused training and ‘horizontal’ more generic training.

Gaps acknowledged at the workshop were the lack of sufficient metadata and contextual information for researchers to understand the data created by other researchers. On top of this, critical thinking skills are needed to understand how other people’s data may be reused. Furthermore, the EOSC needs to provide guidance to researchers on the FAIR data principles in relation to the EOSC services.

Some expressed the view that EOSC should support a ‘train the trainer’ approach, but probably not provide training itself.

***Third challenge: How can EOSC coordinate national-level policies, strategies and reward mechanisms to stimulate open research data practices?***

To improve open research data practices institutional data policies need to be in place, and these should communicate the need for researchers and others who support them to acquire skills. Skills need to include those required to use standardised data management plan (DMP) templates for the various disciplines, and to manage data according to the DMP.

It was suggested that EOSC could provide an EOSC repository for open research data policies for infrastructure development, research collaboration, data sharing and skills development. Participants also saw the need for a governance framework (strategic body, executive body, steering body) to improve reward mechanisms. Researchers should be rewarded for data management, with emphasis on recognising efforts to publish their data. Efforts in EOSCpilot WP3 to establish an ‘open science monitoring framework’ were seen as a potential route towards establishing such a rewards mechanism, beyond its basic goal of checking that RDM is happening. It was seen as important that all stakeholders who may play a role in EOSC are given the opportunity to apply to become members of the stakeholder panel.

### 3.2.2. Part 2 Options and Feedback from EUDAT2018 [W-1B]

#### Agenda

W-1B was a full morning workshop promoting dialogue between those with a cross-disciplinary skills remit and those with a more domain-specific focus. Focussing on issues around providing a ‘skills commons’ by offering FAIR training materials that fulfil needs for domain-focused examples, suitable for the variety of organisations and roles concerned. Trainers and others interested in skills development were invited to discuss the three challenges and other key factors that help or hinder development of individual competences and capabilities for their users and their organisations. The audience consisted of trainers and developers of EOSC services.

The workshop was co-chaired by Kevin Ashley (EOSCpilot, DCC) and Marjan Grootveld (DANS). The agenda started with five short presentations and afterwards three breakout groups were created who took forward discussion begun at W-1A.<sup>8</sup> To move the discussion forward a live poll was held to gather audience responses to the challenges. Then the group broke into three groups, to discuss the issues raised.

#### Speakers

- Kevin Ashley (EOSCpilot, DCC), chaired the event and gave an introduction.
- Three presenters looked at training and skills work in three EOSC projects:
- Angus Whyte (DCC) spoke about EOSCpilot: its definition for data stewardship; a formalisation of roles and responsibilities to ensure research objects are managed in accordance with FAIR principles; using a landscape of skills resources as the basis for identifying skills gaps e.g. data policy requirements, tools, domain standards and workflow for cloud resource utilisation.
- Gergely Sipos (EGI) spoke about EOSC-hub; Integrating and managing services for the European Open science Cloud. Information on four aspects of the EOSC-hub were presented: data and tools services, federation services (e.g. AAI), process and policies (e.g. security regulations), federated operations (e.g. lightweight certification of providers). The scope of the EOSC-hub is on 50 already available services, generic topics (FITSM) and data management planning.
- Ellen Leenarts (DANS), presented training by OpenAIRE Advance; In the current OpenAIRE project there are two work packages related to training National Open Access Desks to assist researchers in the transition to open data by default and open science practices: First a task force for RDM, Legal and Policies related training (WP3) and second a multiplier approach of the Open Science Helpdesk (WP4).
- Angus Whyte (DCC), presented the EOSCpilot Skills Framework; Mapping competencies to service capabilities for data-intensive research.
- Ellen Leenarts (DANS), presented FAIR training - applying FAIR principles to training resources.

#### Responses to Mentimeter

Mentimeter was used for live polling of attendees. 54 conference attendees registered for this co-located session. At the start and at the end of the session there were about 20 attendees, in the middle of the time around 35. 20 to 25 people responded to the questions in the live polling. Full results can be found in the Annex; they are summarised here:

Examples to build capabilities to use EOSC: participants scored nearly even scores for data stewardship, application skills and job advertisement information.

On how to stimulate researchers for RDM, the consensus was divided between carrot and stick based activities such as making a DMP mandatory (stick) or provide additional rewards for good data management (carrot).

With regard to the question ‘What are the most feasible methods for making training materials FAIR ‘Adding identifiers and standard metadata to ensure findability’ and ‘Add non-restrictive licenses to encourage reuse’

---

<sup>8</sup> [https://eudat.eu/eosc\\_as\\_a\\_skills\\_commons\\_providing\\_fair\\_training\\_for\\_fair\\_data\\_stewardship](https://eudat.eu/eosc_as_a_skills_commons_providing_fair_training_for_fair_data_stewardship)

were seen as most feasible by the 14 respondents to this question. The average is shown in coloured circles whereas the distribution is shown as background colours. The participants saw a less obvious case, perhaps depending on the research domain, for using standard formats and vocabularies (see figure 1 below).

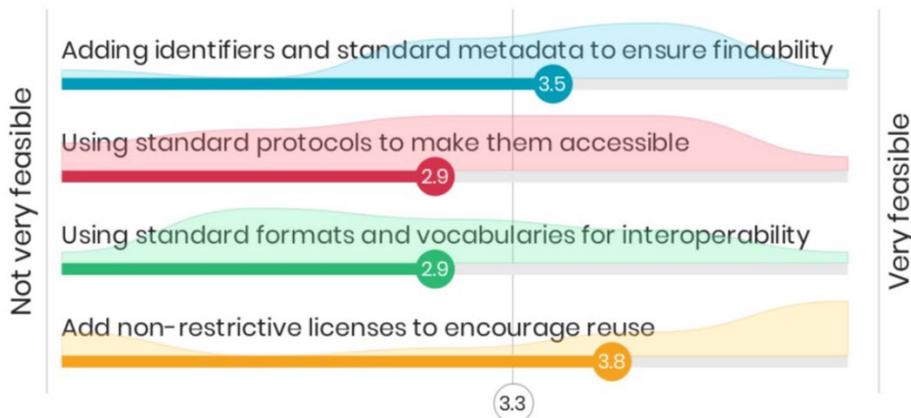


Figure 1 - Participant views on feasibility of methods to make training materials FAIR

Asked about EOSC’s priorities for skills development, the top three priorities chosen by the 22 respondents were: ‘Support train-the-trainer approaches’, ‘Catalogue existing provision of training (based on skills framework)’ and ‘catalogue provision of other skills resources, e.g. guides and handbooks’.



Figure 2 - What should be EOSCs priorities for skills development? (‘EUDAT workshop 1-B’)

**Responses to challenges**

After the Mentimeter poll three breakout groups were created who took forward the discussion started at the Stakeholder forum. The results of the Mentimeter poll were taken into account and were the start of the discussion.

***First challenge: How can EOSC support research training providers to contribute to international level training infrastructure?***

There was a strong consensus that the EOSC should adopt or endorse output from RDA working groups and other projects rather than reinvent the wheel. EOSC could have a central coordinating role in “harvesting” and making existing information related to Open Science, Data Science, RDM, etc. more discoverable. This may be a labour-intensive task and the cost- effectiveness would need to be assessed against the alternative of infrastructures and organisations doing this themselves.

Some participants of the workshop envisioned a marketplace as proposed within EOSC-hub as a means to link services with training resources, trainers and training events. EOSC could offer some added value for example by organising ‘trainers/experts for hire’ across the various projects and infrastructures. This could encompass a train-the-trainer approach.

EOSC could keep track of and visualise the popularity of training resources. This might indicate the value of the resources. Benchmarking the quality of RDM training by acquiring feedback from trainees or from peer-review was also seen as highly desirable. By coordinating information and feedback on training there is potential for EOSC to develop quality measures. Discussion participants recognised that any scheme for certifying quality could be difficult to establish, especially given that this could only be implemented gradually, meaning that absence of certification would not imply lack of quality.

***Second challenge: How can EOSC assist research performing organisations to develop the competences and capabilities for open data science?***

The EOSC should improve opportunities for RIs and e-Infrastructures to offer relevant materials and expertise to university research data services, in order to complement and enrich the cross-disciplinary training they offer, and fill gaps in disciplinary-focused materials. EOSC must enhance the findability of skills resources and expertise across the research communities and infrastructures. EOSC could encourage proactive outreach by the RIs towards institutional RDM services, helping both sides to achieve their goals of broadening access to research communities and stimulating cross-disciplinary research.

A core set of high quality and up-to-date EOSC training resources would support the train-the-trainer approach. Train-the-trainer is an effective strategy only so far as the subject material is up-to-date and relevant, and for generalist trainers this can be challenging in topics where the scope and content quickly change, such as domain standards. Train-the-trainer is also only effective for class-based training (physical or virtual) and not to other forms of skills acquisition such as self-paced learning.

Suitable topics for a core tier of training resources were thought to include: Citation, persistent identifiers, data protection, dealing with commercial sensitivities around data access, licensing, long-term preservation including migration and software or service dependency issues.

e-Research Centres have a role in coordinating local cross-institutional support, for example helping to build individual institutions’ capacity to enlist ‘data champions’ who can address their needs for disciplinary-focused training. Large-scale and long-term research of cross-disciplinary collaborations occupy a middle ground between data-intensive domains and the ‘long-tail’ of others. Collaboration partners have diverse practices and standards, so they have a strong need for mutual learning.

The EC was looked on to take a role in promoting standards. The revision of the European Charter for Researchers and Code of Conduct for Recruitment suggested in the OSPP report on Rewards <sup>9</sup> was

---

<sup>9</sup> [https://ec.europa.eu/research/openscience/index.cfm?pg=rewards\\_wg](https://ec.europa.eu/research/openscience/index.cfm?pg=rewards_wg)

mentioned as a possible way to help promote the use of standard sets of competences and contribute to their development.

***Third challenge: How can EOSC coordinate national-level skills policies, strategies and reward mechanisms to stimulate open research data practices?***

National-level policies and funding are essential to be a good researcher and to use EOSC in the optimal way. A lot of funding is expected from the countries, and some policies will necessarily be built up on a national level. NOADs were seen as important as national ambassadors/intermediaries by some participants.<sup>10</sup> The e-infrastructures and RIs often have national representatives and it was suggested that EOSC can better communicate with these e-infrastructures and RIs than with member states. National funding agencies can help with giving mandates, policies and guidance on the Why and (generic) How of Open Science and research data management.

### **3.3. Consulting trainers from research institution RDM services (IDCC) [W-2]**

#### **Agenda**

W-2 was organised to test a number of assumptions WP7 was working on. About 40 people participated this workshop. The workshop started with three presentations and as in the EUDAT workshop W-1B there was interactive polling with all participants, breakout sessions with three challenges and final conclusion/discussion with a question “How can the EOSC help RIs/libraries to close the stewardship skills gaps?”<sup>11</sup>

#### **Speakers**

- Angus Whyte (DCC), gave a presentation with special attention to the landscape of research infrastructures, e-infrastructure projects and the skills and competencies framework.
- Ellen Verbakel (4TU Delft), presented the data steward program at the 4TU Delft.
- Ellen Leenarts (DANS) presented the training components of EOSC-hub and OpenAIRE-Advance and on the topic of improving fairness of training materials.

#### **Responses to mentimeter**

30 people responded to the Mentimeter poll. The majority (62%) were from the area of research libraries, about one third (28%) from service providers and the rest (10%) from research centres.

The majority of the attendees (50%) were interested in EOSC for data skills and career development across the research sectors. Figure 3 shows the top EOSC priorities for skills development for the attendees. In descending order these were; ‘Support train-the-trainer approaches’ (15 votes), ‘Fill skills gaps by developing and/or commissioning training’ (14 votes) and ‘Define minimum standards of training’ (13 votes). In this audience, the quality measurement of training provided was seen as more important compared to the participants of W-1B.

<sup>10</sup> A NOADs is a ‘National Open Access Desk’ of the OpenAIRE network, see: <https://www.openaire.eu/what-is-the-openaire-network-noads?highlight=WyJub2FkcyJd>

<sup>11</sup> <http://www.dcc.ac.uk/events/idcc18/workshops#workshop2>

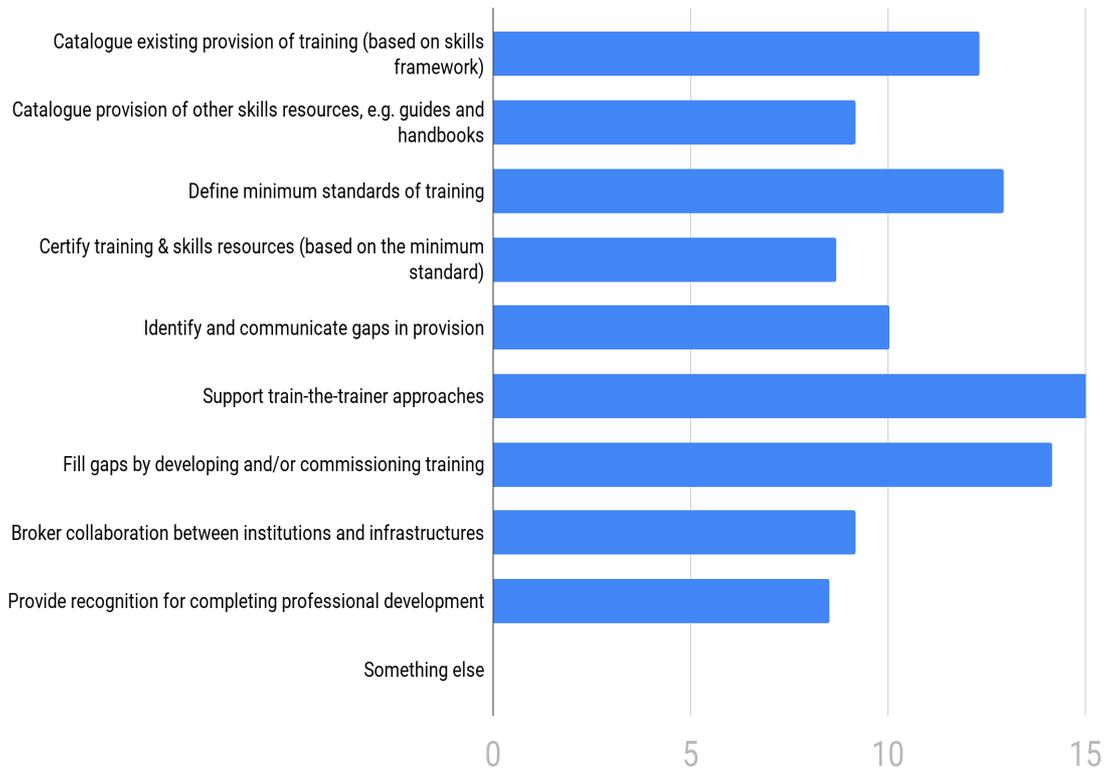


Figure 3 - What should be EOSCs priorities for skills development? (IDCC workshop 2')

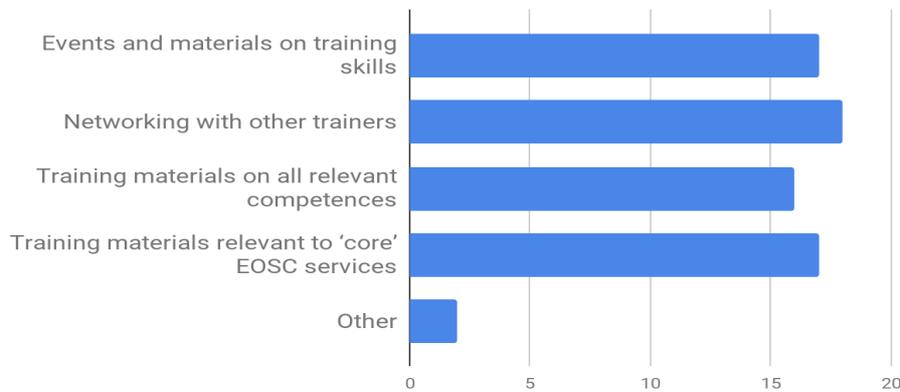


Figure 4 - What forms of train-the-trainer support would you expect from EOSC? (IDCC workshop 2)

There were no major differences in enthusiasm for different forms of train-the-trainer support – there was a demand for all of them. In the answers to the follow-up question on ‘What forms of skills resources would you expect EOSC to support’ it was clear that the attendees preferred the support of training courses and online webinars as skills resources. The role profiles and job description had lowest preference by the attendees.

Figure 5 shows results on the question on making training resources FAIR. On a scale from 1 to 5 the participants could choose how feasible it is to make training resources FAIR. The average is shown in coloured circles whereas the distribution is shown as background colours. As at W-1B attendees suggested that adding persistent identifiers and standard metadata and adding non-restrictive licenses are most feasible actions for making training materials more FAIR.

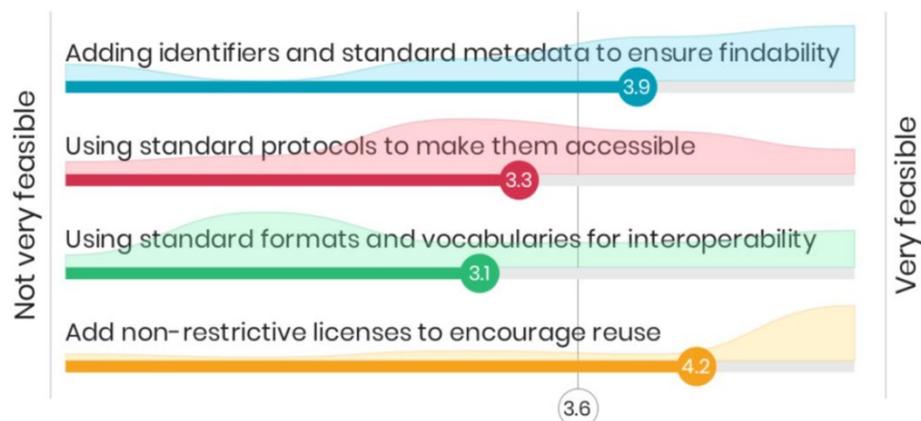


Figure 5 - How feasible is it to make training materials FAIR using the following methods? (IDCC workshop 2)

### Responses to challenges

The results of the Mentimeter poll were used as input for the discussion rounds. As at W-1B, participants were divided into three discussion groups, each group discussing one challenge.

#### ***First challenge: How can EOSC support research training providers to contribute to international level training infrastructure?***

Participants welcomed the idea of extending the FAIR principles to training resources. EOSC should be leading by example; suggested actions included distributing use cases on how to make your research FAIR, use cases per domain, encountered difficulties, and found solutions. The extension of FAIR principles to training resources could become part of the Open Science movement, similar to open education resources. In addition, repository managers/services should have support for FAIR training resources, possibly including standardised formats.

There is a need to perform quality assurance, e.g. through the certification of providers, or badging of content in a central catalogue of training materials and events, but the participants in the workshop considered certification very time consuming. They liked the idea of peer review of training resources by knowledgeable groups of educators. They said that users of materials could add comments: who they are, what was their goal/context, did it work (etc.) as this is all very valuable information for others. It was also suggested that training materials could be based on other training materials.

There was some discussion of the potential for a federated approach to add value. Participants thought it desirable for EOSC to identify gaps, but there was no consensus on whether to fill gaps through any form of centralised activity. Information could be collected to visualise gaps according to a framework of desirable skills, and identify demand through metrics such as the number of downloads.

#### ***Second challenge: How can EOSC assist research performing organisations to develop the competences and capabilities for open data science?***

The discussion covered issues and opportunities for institutions and other EOSC stakeholders to collaborate. The issues raised included maximising the use of learning/training resources that are already there, e.g. cataloguing them to enhance the visibility and accessibility of training content. This would depend on effectively using a list of competencies and skills to encourage researchers to learn and get going, e.g. to identify core skills or learning materials. The competences for open science and data science might usefully be considered as a continuum e.g. in skills frameworks, but neither term has settled meanings. These

meanings vary according to the disciplinary focus, so it can also be useful to treat open science and data science as separate specialisms when providing training.

A second issue was how to get people started, e.g. effective train-the-trainer and engagement approaches. The EOSC should foster networking among trainers to share their expertise and strategies and the EOSC should enable researchers to self-assess and recognize their own skills and competences.

A third issue was how to take advantage of the availability of DMPs to use these as training resources. There was support for giving further consideration to making DMPs FAIR to encourage peer review, and also machine-actionable so that text and data mining can be applied to enable comparison within domains, and enable researchers to judge for themselves how close they are to the practices of their peers.

***Third challenge: How can EOSC coordinate national-level skills policies, strategies and reward mechanisms to stimulate open research data practices?***

During this discussion two main questions were addressed and discussed. The first question was about amplifying stakeholders efforts: how to raise awareness and visibility and to increase feasibility. Many people feel the sense of urgency around RDM. It is needed to make sure that DMPs are followed up and monitored. Mandating DMPs has the capability to stimulate people; people feel more engaged. Case studies are useful to provide good examples of practice. Policies from funders and publishers are a more straightforward motivator than rewards, but badging systems seemed to be working at some institutions.

The second question was about how EOSC could reward people through better incentives and career structures. Participants saw it as essential to reward RDM rather than rely only on citations of outputs. Some ideas of possible rewards were expressed: identifying creditable activities to include in a CV, providing opportunities for scholarships, offering career development grants and people exchange programmes Most of these combine rewards with a mechanism for further skills development.

Participants saw a need to improve career prospects for data professionals: the data steward role was thought to offer potential to improve the reputation and legitimacy of data librarian roles. This emphasises the different roles that data professionals can take at different points in the research lifecycle.

A third question was the tangible contribution of EOSC to helping professionalise the role of trainers. Certification of training was one option thought likely to help.

### **3.4. Conclusions: skills needs and expectations of EOSC**

The workshops achieved their aim of validating aspects of our approach, showing where change was necessary and indicating the feasibility of certain actions we intended to propose.

The workshops increased the visibility of data stewardship as an important topic for participants in the host events. They drew attention to EOSC projects that are dealing with data skills, not only EOSCpilot but also others like EOSC-hub and OpenAIRE Advance. Generally, the workshops were very interactive due to the discussion groups and live polling. As some of the workshop organisers have a role in EOSC-hub and OpenAIRE Advance the workshops offered opportunities to inform work in these projects.

From the discussions of the two workshops we can extract for each challenge certain recurring themes that inform the ongoing consultation on EOSCpilot training and education strategy and recommendations. The actions suggested in response to each challenge contain a degree of overlap, suggesting some actions deliver multiple benefits by addressing more than one challenge.

***First challenge: How can EOSC support research training providers to contribute to international level training infrastructure?***

Participants of our workshops expect an infrastructure for training in EOSC that does the following:

- Provides a comprehensive catalogue harvesting metadata and information on training resources from organisations already providing training portals, based on common metadata for learning resources, on a model similar to that applied in the ELIXIR TeSS system<sup>12</sup>.
- Adopts a 2-tier approach to delivery of training resources, as envisaged in the EOSCpilot Skills Framework, Tier-1 : Core skills and training topics, Tier-2 : Domain specific and specialised skills and training topics.
- Extends the FAIR principles as far as possible to training resources, by providing assessment criteria.
- Reuses existing materials wherever possible, including RDA outputs.
- Monitors the provision of training and possibly attempt to fill gaps (there was not a clear consensus on this point).
- Develops some form of benchmarking of training materials according to their compatibility with FAIR and other EOSC principles, e.g. through certification, open badge mechanisms, peer review, or popularity metrics such as downloads. Further investigation is needed of the feasibility and desirability of these approaches and this is beyond the remit of EOSCpilot to conclude.
- A marketplace of services and training - e.g. in EOSC-hub.

***Second challenge: How can EOSC assist research performing organisations to develop the competences and capabilities for open data science?***

- Provide guidance or training on how to apply metadata and contextual information to data and other outputs (such as software) to implement FAIR principles.
- Implement support measures for train-the-trainer approaches and networking among trainers e.g. the new Community of Practice by OpenAIRE Advance, to share their expertise and strategies.
- Promote sharing of good practices through case studies.
- Promote standards for describing learning resources and exchanging information on training.
- Further explore scenarios for findability of training: 1) from user perspective (as a researcher you want to find the most relevant courses), and 2) a ‘machine-interoperable perspective’ (so training can be found by IT services looking for it).
- Offer information on training across Europe via a ‘marketplace’ of IT and soft services.
- Promote the EOSC as a mechanism for matching providers of training and learning resources to organisations with skills gap.
- Identify rules of participation for organisations to provide training that is ‘compatible’ and ‘compliant’ with EOSC policies (e.g. if you want your training to be in the “catalogue” it is compatible if you provide basic metadata, compliant if you take additional steps to make it FAIR).
- Making DMPs FAIR, and also machine-actionable, to better enable organisations to facilitate their peer-review, and benefit from their usefulness as learning resources.

***Third challenge: How can EOSC coordinate national-level policies, strategies and reward mechanisms to stimulate open research data practices?***

- Work that is currently undertaken to standardise domain based DMPs is also highly valued given their potential as resources for training.
- EOSC can coordinate efforts on skills by involving stakeholders for open data research policy and implementing cross e-infrastructures and RIs communication, potentially requiring a stakeholder group to coordinate skills development and encourage good practice by developing an EOSC

<sup>12</sup> <https://tess.elixir-europe.org/>

repository for best practices, e.g., open research data policy, infrastructural development, research collaboration, data sharing, skills development and rewards.

- EOSC can encourage an incentives-based approach to open practice, so that institutions reward data and application producers similarly to paper authors, reward DMPs as well as citations, and provide career prospects for data professionals.
- The OpenAIRE NOADs and other national intermediaries may be utilised further to help researchers.
- A global perspective is needed for skills development, just as it is for research. EOSC should therefore include skills-related issues in its global engagement with similar infrastructure initiatives.
- By creating an EOSC website that provides consistent information on skills related to EOSC services, instead of separate project websites for the EOSC projects as EOSpilot, EOSC-hub and OpenAIRE Advance.

Complementary to the ideas that were shared in the discussion groups with regard to the identified challenges were the responses to the live polling question on what the priority for EOSC in skill development should be:

1. Supporting train-the-trainer approaches:

Top of the list in W-1B and W-2.

2. Catalogue existing provisioning of training:

Second priority in the W-1B workshop and fourth in W-2.

3. Define minimum standards for training.

4. Fill gaps by developing/commissioning training:

Feedback on this point had less consensus, with strong views for and against.

In D7.5 the work package will come back to these priorities for EOSC in skill development.

## 4. DELIVERING TRAINING FOR DATA SCIENCE AND STEWARDSHIP

This section describes how the work package addressed the following two objectives in the context of the GridKa School 2018.

- Deliver and evaluate pilot training activity intended to address skills gaps found through landscape analysis and engagement with other EOSCpilot WPs.
- Assess trainees' expectations of the relevance to their learning objectives of data stewardship topics drawn from the skills gaps and more broadly from the EOSCpilot Skills Framework.

### 4.1. Background on GridKa School

KIT, a partner of WP7, organises the Gridka School, a summer school held annually and traditionally covering topics about data management and analytics, Cloud Computing, and modern programming, with special focus on applicability for distributed systems and Big Data. The GridKa School is well-established and one of Europe's largest summer schools in the context of Scientific Computing. It has evolved from an initial event to introduce Grid technology to the High Energy Physics community to encompass many learning objectives and topics targeting a large set of groups. Ernst et al.<sup>13</sup> describe the target groups of the GridKa School as including:

- "Graduate and postgraduate students from various science disciplines with an engineering or technical Background.
- Scientists and researchers of different scientific communities, which have interest in deployment and intensive usage of cloud and big data technologies and related tools.
- Industry partners with strong interest in collaboration with scientific communities in the field of IT."

The programme of the GridKa School especially targets new technologies, tools and trends. It has a track record of attracting researchers from domains such as High Energy Physics, Astro-particle Physics, Environmental Science, Bioinformatics, and Informatics - data-intensive domains that were well represented in the EOSCpilot Science Demonstrators. This made the audience a suitable target for training on the skills gaps identified in EOSCpilot, and for testing the relevance of the broader range of topics in the Skills Framework. GridKa School participants range from graduate and PhD students, to researchers who expected to be among EOSC clients, whether as EOSC end-users, EOSC system managers, or EOSC suppliers.

### 4.2. Method

The GridKa School 2018 was organised by Eileen Kühn (EOSCpilot, KIT), with the intention of providing training that would address gaps identified in EOSCpilot. This lead role also gave the opportunity to test the applicability of the Skills Framework to planning an intensive training event. The event was held on August 27 - 31, 2018, at the Centre for Advanced Technological and Environmental Training, Campus North, KIT in Karlsruhe, Germany. The direct involvement of WP7 in the organisation of GridKa School enabled us to shape the schedule of the school, and survey the trainees. This offered a basis for assessing the match between the Skills Framework topics and those the audience for this event considered relevant to their learning objectives.

It should be noted that the school programme and the EOSCpilot Skills Framework were not expected to overlap completely. The school covers Parallel/Concurrent programming and other topics which are currently beyond the scope of the EOSCpilot Skills Framework. Tutors of the GridKa School for example commented that topics such as "concurrency programming", or "Software development for computational sciences" are missing from the current version, which explicitly focuses on stewardship. Section 2.1.1 of D7.2 provides the

---

<sup>13</sup> Ersnt, M. et al (2015) 'GridKa school - Teaching information technologies since 2003' [2015 IEEE Global Engineering Education Conference \(EDUCON\)](https://doi.org/10.1109%2FEDUCON.2015.7096003) pp. 395 - 402 <http://dx.doi.org/10.1109%2FEDUCON.2015.7096003>

motivation for the focus on stewardship which is identified as a priority in the EOSC Declaration and other sources as described in the D7.2 report.<sup>14</sup>

The GridKa School focuses more broadly on hands-on topical training to ensure effective use of today's technologies by users. Therefore, the tutors at GridKa School are selected based on their expert-level knowledge of topics, while teaching usually is not their real core business. This contrasts with the workshops described in Section 3, whose participants were training coordinators and tutors closely involved in professional development of researchers and support staff.

To evaluate the applicability of the Skills Framework, we wanted to ensure that the programme addressed the skill needs we identified in D7.2, and updated in D7.3.<sup>15</sup> The relevance of the topics listed in the Skills Framework was tested indirectly, by applying it to scope the programme content and analysing the extent to which participants viewed the topics, and the subsequent training they received, as relevant to their learning objectives.

The assessment was conducted through electronic surveys with the participants, and interpretation of the results was informed by observant participation in the organisation of the event. We circulated surveys for each day of the event to all participants. The surveys were sent the following day via email<sup>16</sup> to ensure both comparability and recall.

The template questions for the survey to measure the applicability of the EOSCpilot Skills Framework for planning a training event as well as the effectiveness for learners can be found in Section 1 of Annex D.

### 4.3. GridKa School as a testbed for the EOSCpilot skills framework

#### 4.3.1 Agenda

The GridKa School is a one-week summer school. During the school week, each day has four plenary lectures scheduled in sequential order in the mornings. After a one-hour lunch break, participants reconvene to join for one of up to six parallel hands-on tutorials in the afternoons. Furthermore, the school includes several social events (Tarte Flambee event on Tuesday, Evening Lecture on Wednesday, School Dinner on Thursday) to foster collaboration, networking, and discussions between trainees, speakers and tutors. Participants can choose to join the school for a single day or the whole week.

As a result of the direct involvement in the organisation of GridKa School, we were able to align the discussion/decision on GridKa School topics and learning material with the skills gaps that were identified in the EOSCpilot Skills Framework development. This alignment not only provides benefits to extend the EOSC training catalogue - we also expected improved relevance, coverage, and topicality of GridKa School 2018. The full programme of lectures, tutorials, and social program is available online.<sup>17</sup>

Table 1 provides the mapping of the agenda items (count of lectures and tutorials) to the skill gaps.<sup>18</sup> The full mapping including the titles of lectures and tutorials is provided in Section 2 of Annex D. The Table shows

<sup>14</sup> Kuehn, E. and Streit, A. (2017) D7.2 Interim report and catalogue of EOSC skills training and educational materials. <https://eoscipilot.eu/content/d72-interim-report-and-catalogue-eosc-skills-training-and-educational-materials>

<sup>15</sup> Whyte, A. and Ashley, K. (2018) D7.3 Skills and Capability Framework. <https://www.eoscipilot.eu/content/d73-skills-and-capability-framework>

<sup>16</sup> Due to preferences by participants following the GDPR we had to exclude some participants from the list of recipients as we were not allowed to send them information via email.

<sup>17</sup> <http://indico.scc.kit.edu/event/427/timetable/#all.detailed>

<sup>18</sup> During the course of programme planning the identified gaps from D7.2 and D7.3 have been merged for consistency and to better visualize the resulting mapping of lectures and tutorials to the listed categories.

that we scheduled lectures to lay theoretical foundation for most of the needs except for *Monitoring of federated data and compute services*. We considered the theoretical foundations for this topic to better be included in a tutorial. We supplemented the theoretical competences with hands-on tutorials where possible. For example, we did not plan a specific tutorial for supporting *Domain knowledge and domain-specific methods* as specific knowledge should be better covered in domain-specific events. Instead, we selected some specific challenges from differing domains to be presented as a lecture. Overall, the match between competence gaps and learning elements was high.

In summary, the GridKa School 2018 had 17 plenary lectures, 1 evening lecture and 17 tutorials. A total of 39 speakers and tutors were engaged in delivering the course. 142 participants (not counting HNSciCloud and HDF AAI participants) from 17 countries attended the event.

**Table 1 - Count of lectures and tutorials from GridKa School agenda that were mapped to EOSCpilot competence gaps**

Category	Competence topic	Lectures	Tutorials
Skills for integration	System management and administration including container technologies	9	6
	(Long-term) data management, curation, preservation and provenance	4	2
	Federated authorization and authentication including data security in the cloud	1	1
	Domain knowledge and software development to adapt workflows to cloud workflows	8	9
Skills for service management	System management and administration of virtual infrastructure	6	3
	Federated authorization and authentication including data security in the cloud	1	1
	Monitoring of federated data and compute services		1
Skills for effective use	Decision-making for when using the cloud is appropriate	2	
	Data security in the cloud	1	1
	Data management, curation, including provenance to enable reusability and long-term preservation	2	3
	Domain knowledge and domain-specific scientific methods	2	
Skills for assessment	Research strategy, ethics and policy in the age of open science	4	
Other topics not EOSCpilot gaps		2	1

## 4.4. Assessing the applicability of the EOSCpilot Skills Framework

### 4.4.1 Applicability of the Framework for organisers to plan a training event

To assess the overall perceived relevance of the programme to participants we asked them to rate the relevance of lectures and tutorials<sup>19</sup>. The results of the perceived relevance of lectures and tutorials are shown in Figure 6 comparing GridKa School 2018 with ratings for previous years. The figure shows the overall mean of the ratings of each lecture/tutorial respectively with regard to the complete programme. This means, the analysis not only contains invited talks that were selected based on the Framework but also lectures and tutorials that were considered relevant by the local organisers.

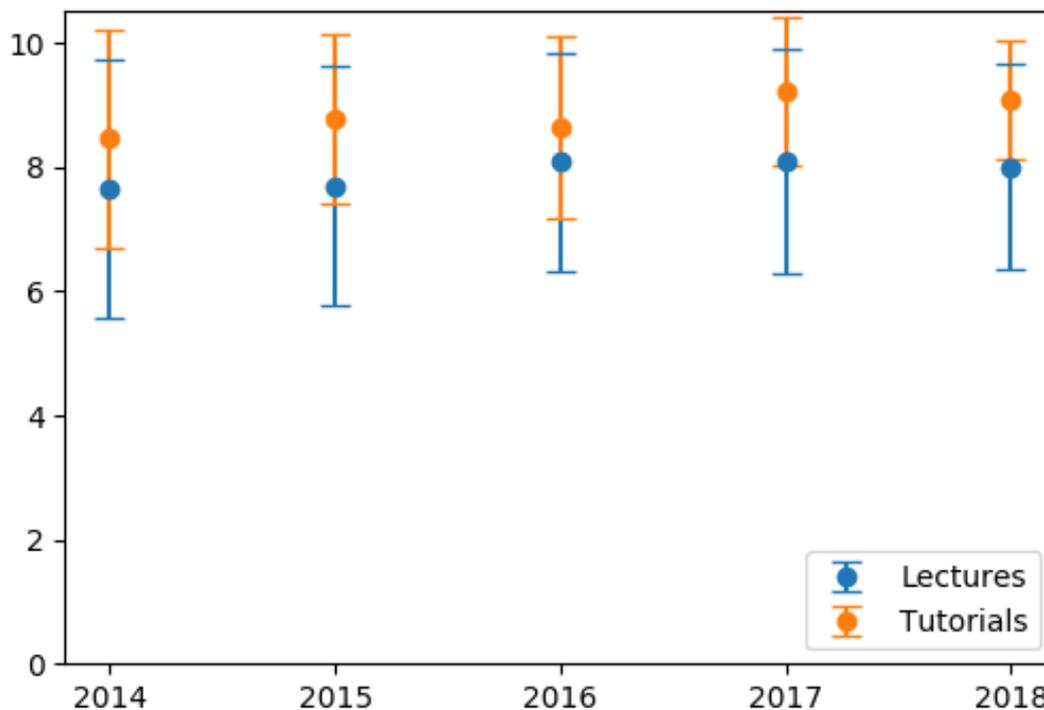


Figure 6 - Mean ratings (1-10) of relevance for plenary lectures and tutorials for GridKa School starting from 2014.

(The analysis for GridKa School 2018 includes all plenary lectures and tutorials, not only the invited contributions addressing the skills gaps.)

The mean rating of lectures and tutorials for GridKa School 2018 was considered very high by the participants. While this is not necessarily a consequence of using the EOSCpilot Skill Framework as a basis for identifying topics, it does suggest that topics were a good match to participants' interests. The following section assesses this in more detail.

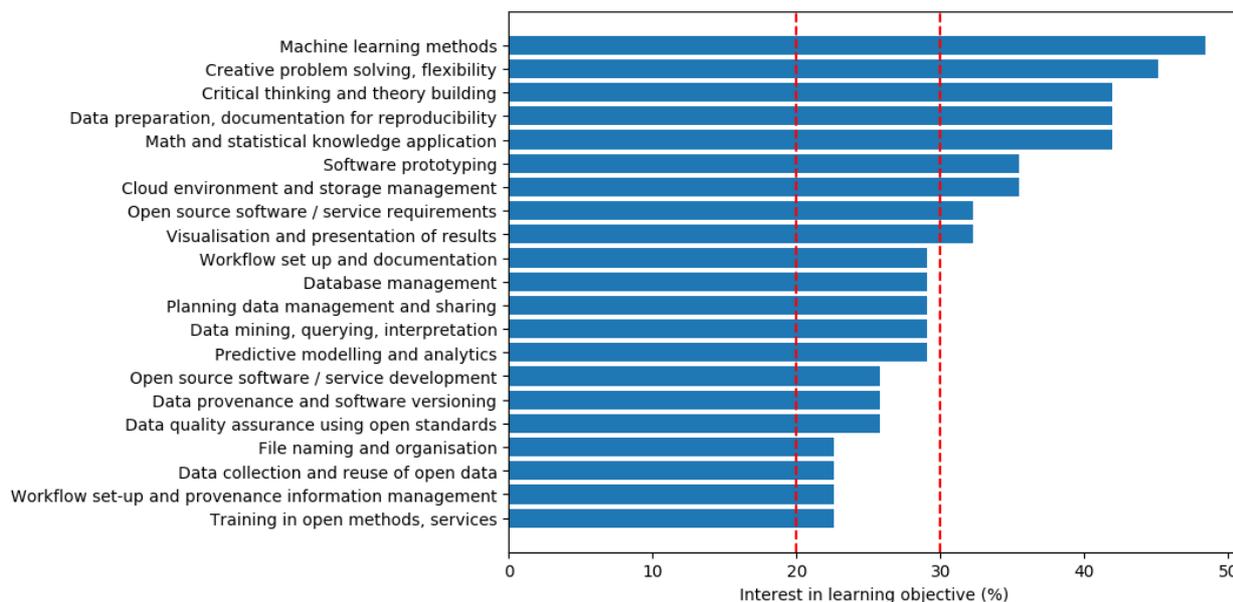
### 4.4.3 Applicability of the Framework in terms of trainees' learning objectives

One of the main use cases of the EOSCpilot Skills Framework is helping trainees to identify learning resources based on their learning objectives. This requires users to be able to specify their learning objectives. We therefore asked trainees to indicate which specific skills/competences of those listed in the Framework they were interested in as learning objectives. As the programme of GridKa School has been built with a focus on expected skills gaps (compare Section 4.3.1), the question is whether trainees are successful in identifying their learning objectives and assessing the effectiveness of provided trainings. As expected, participants of GridKa School 2018 show a large overlap with the expected EOSC System Users (compare Annex D, Figures

<sup>19</sup> We also asked about the quality of lectures and tutorials. However, quality is more related to presentation itself and, therefore, less important to assess the outcome of planning events with the help of the Framework.

D.1-D.4) and relevant domains for the EOSC. We therefore also expected a large overlap in learning objectives.

Figure 7 lists the learning objectives that were identified by at least 20% of participants. This illustrates the topics in the Skills Framework that have greatest appeal to the GridKa School audience. A full list of identified learning objectives is provided in Annex D, Figure D.5. Most learning objectives that interested at least 30% of the GridKa School participants relate to current trends, e.g. *Machine learning methods*, and specific technical competencies, e.g. *Data preparation, documentation for reproducibility, Cloud environment and storage management, or Software prototyping*. Reducing the threshold to at least 20% adds competencies for information and data management, e.g. *Database management, Planning data management and sharing, or Data quality assurance using open standards*.



**Figure 7 - Learning objectives that are of interest to at least 20% of GridKa School attendees.**

(The vertical lines mark the thresholds of 20% and 30%. Learning objectives that are of interest for at least 30% of attendees are mainly in the context of data science.)

Some of the topics identified in the Skills Framework did not map well to the participants' identified learning objectives. The interest in competences such as *Authentication and authorisation management* or *Access control and management* is of interest to less than 20% of participants, *Ethical, legal and data policy compliance* is of interest only for less than 5% of participants or *Developing a profile of open research* that is not even considered relevant by a single participant. This raises importance for supporting the planning of training events based on identified skills gaps to raise the awareness of topics that are critically important for EOSC, but might not be considered relevant learning objectives by trainees themselves. Indeed, feedback on plenary lectures and tutorials showed (compare Figure 6) that these topics were considered relevant and, therefore, there is a need to raise awareness about these topics in the first place. As such, the Skills Framework may help to raise awareness that these topics represent desirable skills and promote a focus on improving these skills explicitly.

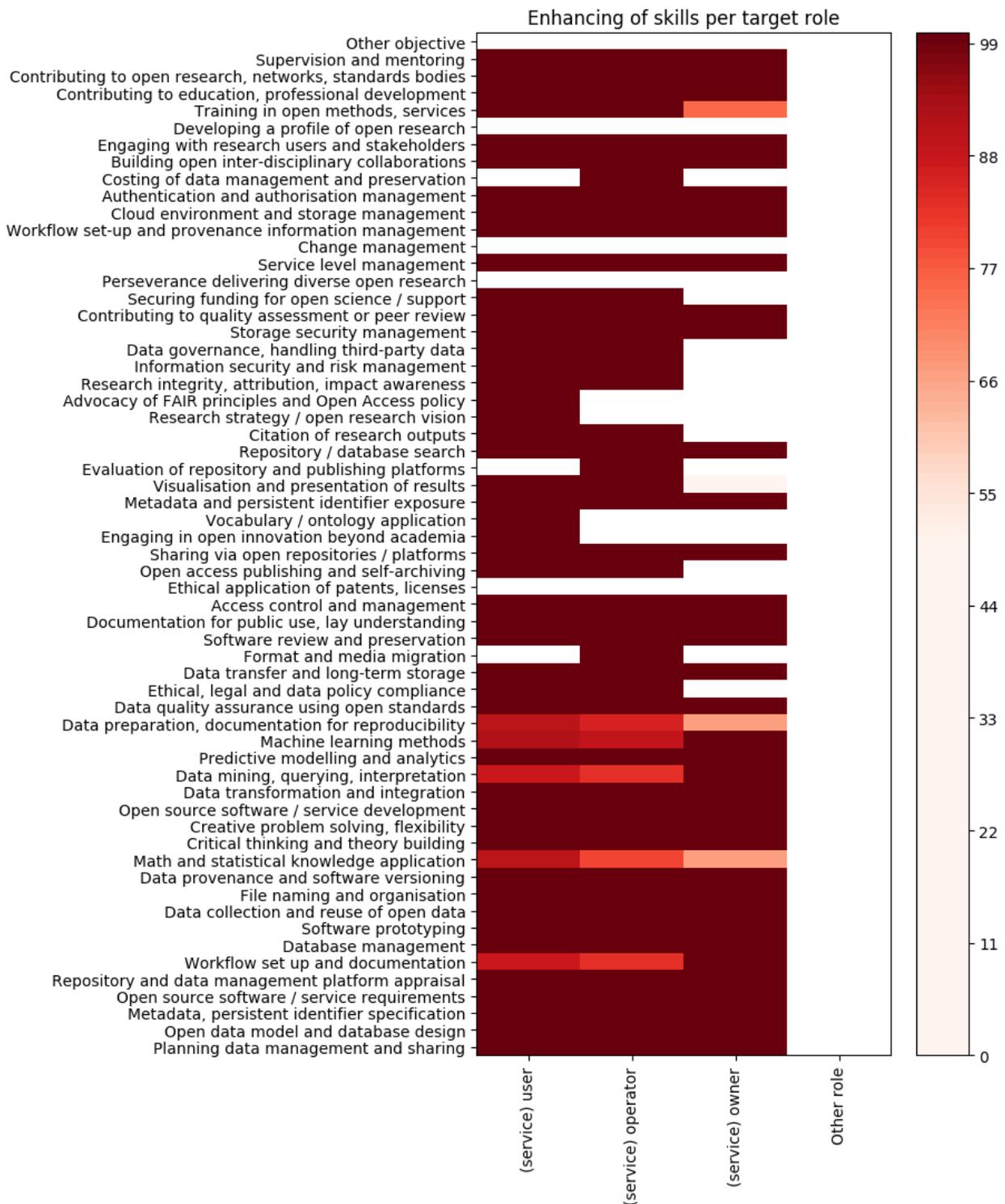


Figure 8 - Effectiveness of GridKa School in enhancing competences per service role.

(White areas indicate that a given competence is not considered relevant by a given service role. The range of colours indicate the effectiveness of GridKa School in addressing the competence topics listed.)

To assess the degree to which the GridKa School provides a means of *enhancing* particular set of skills from the EOSC Skills Framework, we analysed the participants’ responses on effectiveness of GridKa School in enhancing learning objectives by service roles. 88% percent of participants perceived the school as effective in addressing their selected learning objectives. Figure 8 shows the percentage of positive responses per learning objective versus a specified service role. The white areas show where a competence is not

considered a learning objective for a given service role. The same analysis per professional role is provided in Annex D, Figure D.6.

Figure 8 indicates that (service) users have the broadest topical range of improvement. Higher-level roles such as service owners have mainly overlapping topics with the two other roles, i.e. service operator and service user. Notably, people that identified themselves with higher-level roles usually identified themselves as belonging to these base roles as well.

However, the uniform coverage in the base role indicate that people could not differentiate clearly between the learning objectives. Only experienced participants with a broad range of roles, i.e. people that are experienced in diverse and maybe distinct fields, could isolate areas of improvements. This suggests that the level of experience is coupled to the effective direct use of the Framework by potential trainees. The uniform coverage of effectiveness is further coupled to the design of the questionnaire: To reduce complexity of the questionnaire, participants did not have to assess each selected learning objective with regard to the effectiveness in addressing it, but the overall effectiveness. Some of the competences, e.g. *Data preparation, documentation for reproducibility* and *Machine learning methods* were specifically provided at GridKa School 2018 in selected tutorials. However, tutorials only have a limited number of seats so that not all of the participants were able to attend specific tutorials addressing their learning objectives. This is also supported by the freeform feedback we collected (compare Annex D, Section 5).

Other freeform comments suggest that the primary issue identified by participants was the availability of places for their choice of tutorials, and selection of the most relevant topics for hands-on tutorials. Some of those that were over-subscribed were on topics listed in the Skills Framework. In other cases, competing learning objectives were addressed by courses running in parallel. This suggests the Framework helped to identify some topics that this audience finds highly relevant. However, it also underlines the potential to make training content more accessible to those unable to attend in person.

#### 4.4.4 Conclusions on GridKa School

The GridKa School provided training viewed as highly relevant to an audience of data scientists and researchers in data-intensive domains who are expected to be prominent among EOSC end users. Course content was tailored towards topics previously identified in EOSCpilot as skills gaps.

The planning of the agenda of GridKa School 2018 by selecting specific topics matching the identified gaps worked well. Aligning the programme with the skills gaps helped the organisers to select specific topics and people to invite for giving lectures and providing tutorials. It helped to put a focus to specific topics on the one hand and ensure a balanced programme on the other hand. We intend to confirm this positive experience of utilising the EOSCpilot Skills Framework in event planning by repeating the exercise in future.

There is some overlap between the topics in the EOSCpilot Skills Frameworks and those that GridKa School trainees identified as their own learning objectives. Of 59 topics listed, 21 were identified by at least 20% of trainees, of which 7 matched a higher threshold of 30% of the trainees. Those of most relevance were:

- Software prototyping
- Math and statistical knowledge application
- Machine learning methods
- Data preparation and documentation for reproducibility
- Critical thinking and theory building
- Creative problem solving and flexibility
- Cloud environment and storage management

These were the topics in the Skills Framework that attracted the broadest interest from the GridKa School attendees. To the extent that these attendees are representative of target EOSC users with an interest in data science, and the Skills Framework topics are representative of stewardship topics, the trainees preferences for the above topics suggests these are areas of strong cross-over between the distinct areas of data science and stewardship.

A small number of topics that were previously identified in EOSCpilot as gaps (e.g. from Science Demonstrators and consultations) were identified as learning objectives by very few GridKa School attendees. These included ethical and legal compliance, data preservation topics, and some aspects of service management. This does not mean that these skills are not relevant, but that trainees assessment of their skills needs are not the only criterion for skills acquisition.

## 5. OVERALL CONCLUSIONS

EOSCpilot WP7 has conducted workshops that both provide training (GridKa School) and consult with training providers. The consultative workshops have asked about skills development challenges and responses to those that should be priorities for the EOSC and its stakeholders. These challenges were initially drawn from the WP7 plan in the EOSCpilot Description of Work, and elaborated on through successive workshops, partly based on the desk research on the skills landscape (D7.1) and partly on the ‘training as a service’ approach and skills framework drafted in D7.2 and D7.3 respectively. The framework was tested through application at an established summer school, which broadly confirmed the utility of the framework. It also showed that self-selection by trainees of topics for skills is not the only criterion to be used; they appreciated the value of training on topics such as ethical and legal compliance that they would not have selected themselves.

Trainees at the GridKA school indicated their competence had been enhanced on more of these topics than they had initially seen as learning objectives. This indicates the importance of the educator’s role in highlighting topics that trainees might not otherwise choose to learn about through more self-directed methods.

Discussion themes from three EOSCpilot consultative workshops have been described earlier in this report, and, together with lessons from the summer school, allow us to draw the following conclusions:

- 1) Successive drafts of a Skills and Capability framework for stewardship have been presented to trainers and stakeholders in skills development, with overall positive comments and suggestions of particular skills to emphasise. The number of participants across the three workshops is relatively small (60-70) however, and there is a need to broaden out the consultation with the publication of the Skills and Capability Framework in D7.3.
- 2) There was a broad consensus that EOSC should provide a training catalogue and registry, aggregating metadata about training and learning resources from training portals offered by organisations in the EOSC ecosystem.
- 3) The Framework is helpful in the planning of broadly-based training events, both for the organisers and attendees. It helps to identify learning objectives for both groups and to understand how well the training meets those objectives.
- 4) The choice of skills to acquire should not be left to trainees alone; guidance from supervisors and event organisers, informed by the framework, will ensure that they acquire skills of value that they might not identify themselves.
- 5) There was general support for the idea that EOSC should fulfil the role of a ‘marketplace’ of skills, aiming to broker supply and demand for skills training and exchange between research infrastructures and institutions to help ensure training and development is effectively targeted.
- 6) There was support for a distinction between different types or ‘tiers’ of data skills; those that may be considered generic (cross-project or cross-domain) or ‘thematic’ (project-specific or domain-themed). However, there were some objections to the suggestion that EOSC should have a monitoring role to assess training and determine where there are gaps that need filled. There was some support for the idea that EOSC should commission training and learning resources to fill gaps in generic skills and capabilities. However, some felt that domain-specific services and infrastructures should be left to address the domain-specific skills gaps.
- 7) There was also support for EOSC offering certification of training and learning resources, which must meet a minimal set of criteria to ensure these resources are findable, accessible, interoperable and reusable. These criteria should be consistent with FAIR principles and good practices recognised by the Open Educational Resource community, e.g. open licensing of content. There might be a similar distinction to be made between learning resources that are ‘compatible’ and ‘compliant’, similar to

those applied for the governance of services in EOSC and similarly implying a body capable of defining and applying these criteria. However, the utility of this is dependent on the acceptance of such criteria for EOSC services as a whole.

- 8) There was some mention of relevant standards in terms of a general need for them, but little discussion of specific standards. The main example was Open Badges, as these were believed to be popular with trainees and a means of tracking impact of training. There is a need for further work in EOSC to determine which standards should be further supported, and how standards serving different but complementary purposes could be used together, e.g. learning resource metadata, learning records and contribution to research outputs (e.g. CRediT).

The Final Report of WP7 (D7.5) will report on further consultation on these points with the broader communities of interest on skills development. This includes a workshop organised in collaboration with TU Delft in September 2018 to explore whether the Framework can usefully be mapped to researcher career stages.

## ANNEX A. GLOSSARY

Term	Explanation
<b>Capability</b>	<i>Competence</i> applied at a research team or organisational level, with a defined level of expertise and responsibility, to perform a service role or work in the EOsc environment
<b>Competence</b>	An element (topic) of theory or practice e.g. ‘workflow set-up and management’, combined with an <i>expertise level</i> to indicate whether someone has an awareness of the area, or an ability to do it, or expert knowledge of it.
<b>EOsc End-user</b>	An EOsc System User consuming EOsc Resource(s) by means of EOsc Service(s) (e.g. the EOsc Portal) to accomplish a task. She can be a Researcher, a Research Administrator, or a Third-party Service Provider.
<b>EOsc Supplier</b>	An EOsc System User when taking care of the provisioning of an EOsc Service Component(s) enacting an EOsc Service by the EOsc System. It is further specialised in the following sub-roles: EOsc Service Component Supplier, Data (Service) Supplier, and EOsc Service Developer.
<b>EOsc System Manager</b>	An EOsc System User when taking care of the management of the EOsc System. It is further specialised in the following sub-roles: EOsc System Owner, EOsc System Top Manager, and EOsc Service Provider.
<b>EOsc System User</b>	The role played by every actor (human or machine) exploiting the EOsc System according to the EOsc Policy. An EOsc System User might be further specialised in roles including EOsc End-users, EOsc Suppliers, and EOsc System Managers.
<b>Expertise level</b>	A description of level of <i>competence</i> ; e.g. comprehension, ability to apply, or expert knowledge of the given competence, i.e. be able to evaluate its application or synthesise new ways of applying the relevant knowledge
<b>Professional group</b>	A person's domain of responsibility defined by a set of <i>competences</i> , e.g. domain research, data science/ analytics, data management, data service engineering
<b>Service</b>	A <i>service</i> described in the EOsc service portfolio, or service catalogue, that offers value or reduces risk to a <i>professional group</i> or other designated community
<b>Service role</b>	Role in the application of a <i>service</i> i.e. service operator, or user
<b>Skill</b>	The application of a <i>competence</i> or <i>capability</i> to a specific context, e.g. deployment of a service. Skills may be specified in the form of a <i>skills user story</i>

## ANNEX B. REQUIREMENTS WORKSHOP REPORT

### Notes from EOSCpilot WP7 Skills Requirements Workshop Mercure Hotel, Amsterdam Airport, 17 May 2017

#### Attendees

Femmy Admiraal (1), Themis Athanassiadou (2), Kevin Ashley (3), Cath Brooksbank (4) Elly Dijk (1), Magdalena Getler (3) Marjan Grootveld (1), Sarah Jones (3), Eileen Kuehn (5), Gareth O'Neill (6) Simone Sacchi (7), Vasso Kalaitzi (7), Brian Mathews (8), Gergely Sipos (2) , Sarah Steele (8), Jerry de Vries (1), Angus Whyte (3)

(1) DANS (2) EGI (3) DCC, UEDIN/ UGLAS (4) EBI (5) KIT (6) EURODOC (7) LIBER (8) STFC

#### Background

This first internal EOSCpilot Skills Requirements workshop aimed to contribute to WP impacts in the following areas:

- facilitating the sharing of training and educational material, tools, applications and data for high-level education and professional training.
- development of a model for sharing of expertise in Data Stewardship and Data Management.

The morning session was dedicated to short presentations offering external (to WP7) perspectives on skills development and gaps, plus results of initial landscape analysis and requirements gathering. In the afternoon session participants discussed the implications for the actions to be taken in WP7, and the approach to capturing and analysing these requirements, to enable further gathering of these throughout the pilot and the identification of appropriate responses through training or other means.

#### Skills landscape- presentations

##### *Introduction- EOSCpilot skills implications and impacts*

Kevin Ashley, DCC

Kevin began by placing the skills and capability aims of the EOSCpilot in the context of its broader objectives. He highlighted these as joining up existing research and e-infrastructure, defining policy and governance for this loose federation, and proving it can work through the science demonstrators. With the notable exception of cluster projects, the norm is that separate infrastructures provide training and/or define the skills needed to use their services.

The problem EOSCpilot addresses is not so much a lack of capacity in skills provision, but insufficient integration of that provision across infrastructures and uncertainty that the right skills are being developed for open data-intensive science, or that delivery is scalable to meet increased demand. Postgraduate education is contributing some of the skills needed through research methods training. Some infrastructures are good at defining what skills are required to use and interoperate with them. Training provision is also important for all RIs as this is now one of the metrics on which they are assessed. What is needed is greater attention to skills required to interoperate between infrastructures, make these work with local institution-level provision, and support researchers to use a cohesive set of services.

So the focus of the WP is on what specific skills people and organisations need, and how these should be delivered/acquired/assessed. Training is not necessarily the most efficient way of acquiring some of these skills, compared with on-the-job learning. The WP is addressing the requirements alongside other initiatives that it is important to engage with. These include the Belmont forum WP on data curation skills

in climate sciences, the GO-FAIR proposed coordination node on stewardship skills, the FOSTER+ project's continuing work on open science skills, relevant groups within the Research Data Alliance, and the various RI cluster projects.

### *Developing skills across ELIXIR and the Research Infrastructure landscape*

Cath Broosbank, Head of Training EMBL-EBI (remote presentation)

Cath related her experience of skills development in EMBL-EBI. The European Bioinformatics Institute is Europe's home for biological data services, a major international trusted provider for the life sciences, and hosts the ELIXIR technical hub. Training is part of the organisation's mission, involving all service teams and research groups. In 2016 165+ people were involved in its development and delivery to researchers, including Cath's team of 17. Delivery modes include online and face-to-face training, which is both off-site and on-site, and a train-the-trainer programme.

Competence frameworks are central to the approach taken at EBI, which involves defining skills requirements in competency profiles that each list the indicative knowledge, skills and behaviours for the competence areas within its scope. These are used in global and pan-European initiatives like ELIXIR to develop training programmes, and are also a resource for other course providers. A similar approach is taken in the RItrain project, which focuses on 'Driving leadership for research infrastructures', and the CORBEL cluster project, which has applied the EDISON framework.

Training needs are identified through surveys and workshops, competences defined, and gaps with existing training analysed. The training is developed and delivered to draw on available expertise and adapt to the needs of busy individuals. This has led to a move away from full-time study, lectures and exam-based assessment to learning that can be embedded in everyday work, through participation in communities of practice, sharing experience and using 'bite-sized' learning that is short, very modular and flexible, which may be used on an individual or team basis. RItrain for example has employed webinars and staff exchanges as the basis for professional development.

The materials are openly licensed, and has been re-used a lot in university courses. Visibility of the materials is enhanced by making it available to aggregators like the ELIXIR TeSS system, and through GOBLET.

### *Perspectives on professional development in data stewardship, current activity of the Education & Skills WG under the Steering Group on Human Resources Management (SGHRM)*

Gareth O'Neill, President European Council of Doctoral Candidates and Junior Researchers

Gareth gave an overview of the scope of the SGHRM Education & Skills Working Group, which he became involved in through their work with the European Commission on researcher mobility, and his role in representing early career researchers through EURODOC. The SGHRM has recently adopted a focus on the Digital Agenda, with a brief to inform the work of the Open Science Policy Platform through 8 working groups considering a broad range of open practices.

On that basis the Education & Skills Working Group working group conducted an online survey targeting researchers across Europe, on their Open Science practices. The survey closed only one day before the workshop so only interim results were available. Gareth highlighted themes from analysis of responses to questions most relevant to the workshop. The survey had a sizeable response from researchers across a range of maturity and seniority levels, a mostly even disciplinary spread (with the exception of Humanities), and equal gender balance. He drew a number of preliminary conclusions, including improved provision in the following areas:-

- awareness of Open Data initiatives is low
- training courses on data management are required
- clear guidelines for data management are required

- use of data management plans is low
- support for data management is inadequate
- data management needs to be supported through rewards and funding

### *Improving data stewardship skills and their recognition, EOSC and related initiatives*

Marjan Grootveld, DANS

What existing provision is there for data stewardship skills development to support data-intensive research, especially across domains? What approaches/ examples illustrate how this could be scaled up? What are the main challenges and opportunities to develop cross-domain expertise?

Marjan addressed these questions based on DANS experience in skills support including community workshops, webinars, and summer schools; and spanning EUDAT, OpenAIRE, CESSDA, CLARIAH, DARIAH, EHRI, and PARTHENOS. DANS also participates in RDA Working/ Interest Groups and, at a national level, has provided training through Essentials 4 Data Support, and RDM consultancy for funders and universities.

On the issue of existing provision, the e-infrastructures have strong offerings e.g. PRACE, EUDAT, EGI, and there are also useful examples from the Research Infrastructures captured in the emerging EOSC skills catalogue. Certification of existing training is immature; DANS has found little demand for training certificates, although there is interest in Open Badges.

The scale and cross-domain emphasis of EOSC is ambitious and challenging to address in skills terms, considering that in many domains the RIs still have only a few early adopters. However there is some existing collaboration between e-Infrastructures training activities to build upon. For example EUDAT is aiming to hold a workshop for trainers at DI4R17. Skills swap or exchanges for trainers themselves might be further considered.

On the issue of scaling up, Marjan emphasized that the Human Factor is key. Communities of Practice around certain skills or certain services need to be nurtured if such communities are to be enabled to live beyond individual projects.

One approach to scaling up and across domains is to use researchers who are expert at science communication (e.g. like Mauro Martino) with the vision to see beyond single domains and articulate research issues to a broad audience.

On a more fundamental level, trainers should force themselves to make their materials more reusable to audiences beyond the specific groups they are first designed for.

### *Science Demonstrator skills development context*

What gaps have been highlighted by Science Demonstrators in the skills needed by their target user communities, to achieve the planned impact? What relevant work is already fostering data science skills for cross-domain collaboration, through Research Infrastructures, e-infrastructures, related H2020/ other projects?

**High Energy Physics, Photon-Neutron**, Eileen Kuehn, KIT: Eileen talked through the response from HEP and Photon-Neutron demonstrator contacts to the WP7 skills questionnaire in WP4's first reporting. Although it had not been feasible to discuss their responses directly with them, local contacts in the HEP community had been able to help interpret the requirements.

**Textcrowd** (Elly Dijk, DANS) Elly described the Textcrowd contacts' responses to the skills questionnaire.

**Pan-Cancer, EOSCpilot Services and Integration** (Gergely Sipos, EGI) Gergely talked through the response from Pan-Cancer demonstrator contacts to the WP7 skills questions in WP4's first reporting, and the current status of WP5 and WP6. The main skill gaps identified so far are in the areas of engineering, and data management/ stewardship rather than data science or analytics. In engineering terms, Pan-Cancer requires a very complex and demanding cloud computation and storage framework (e.g. 1Pb storage) and the provisioning of this is a challenge for WP5. Operating large scale cloud based analysis systems requires a broad set of technical skills that are not commonly possessed by end-users – (data security, networking,

Linux administration, distributed architectures, operational management). The relevant skills are covered by existing resources; a forthcoming EBI cloud deployment course, and MOOCS provided by edX. Some contextualization of these to the EOSCpilot demonstrators may be useful, focusing on:

- Decision making for when using the cloud is appropriate
- Data security in the cloud
- Management of virtual infrastructure

- The main data management skills issue identified so far is understanding what the SD setups do to deliver trustworthy results for researchers.

- Another generic issue is the sustainability of training for the SDs; the challenge of ensuring content on the key skills and capabilities are kept up to date, when the SDs have a short lifespan.

### *Training infrastructure in RIs and e-Infra's (Sarah Jones, DCC)*

What tools and services could facilitate hands-on training with demonstrator applications? Where are relevant collections of learning materials? How may information on relevant training events or materials be aggregated from Research Infrastructures, eInfrastructures and external sources. Sarah presented a draft of a layered model proposed by Angus Whyte and Gergely Sipos. This was expanded and updated as a result of the discussion, and forms section 4 of the D7.1 report.

### *Introduction to discussion of key competence areas (Angus Whyte, DCC)*

Overview of data stewardship competencies, and the relationship between data management and other EDISON competence areas (data science engineers, analytics experts and researchers). Angus talked through the main points of a draft framework. This was expanded and revised as a result of the discussion, and forms sections 6 and 7 of the D7.1 report.

## **Afternoon Session: Data stewardship – what is it, where are the gaps, how to improve skills and collaboration?**

### **Introduction:**

Magdalena Getler, DCC

To introduce the discussion Magdalena Getler (DCC) presented a brief summary of key points identified in several influential reports on data stewardship skills needs; from CLIR<sup>20</sup> and the Belmont Forum.<sup>21</sup> The key points of the CLIR report, based on interviews with early career researchers in the social sciences included their lack of formal training in long-term data curation, or of consideration of this as a priority. The report recommended the integration of data specialists into research teams, and highlighted the challenge of inter- and trans-disciplinary research. The latter was the focus for the Belmont Forum e-Infrastructures and Data Management group survey of researchers and data managers in global change research. Largest challenge in data use included data complexity lack of data standards and exchange standards, finding relevant existing data, data management and storage. The main skills areas needing most improvement were data processing and analysis, programming, and data management.

### **What data stewardship competencies are relevant to EOSCpilot?**

The first part of the discussion focused on three questions about skills in the EOSC Science Demonstrator context that earlier presentations had introduced;

<sup>20</sup> Jahnke L, Asher A, Keralis SDC (2012) The problem of data. Technical report, Council on Library and Information Resources <https://www.clir.org/pubs/reports/pub154/>

<sup>21</sup> Belmont Forum (2016): <http://bfe-inf.org/document/skills-gap-analysis>

1. What data stewardship competencies have we found to be well covered in current skills development activity?
2. What data stewardship competencies are needed to integrate Science Demonstrator applications into EOSC, deploy them as services, and apply them in their intended research contexts (using results from Science Demonstrator liaison)?
3. What do the SD applications expect from people who better fit the EDISON descriptions of data scientists, engineer, analytics experts, and other who may need additional skills or awareness (professionals or citizen scientists)?

Key points and conclusions were:

- WP7 should be able to support EDISON and other initiatives to get further consensus on key competences for data stewardship, and especially on how these relate skills embedded in research teams, to those associated with data librarianship.
- It was suggested that feasibility of embedded roles depends to some extent on disciplinary cultures, e.g. the size of project teams and extent of standardisation. The competence framework should aim to accommodate differences in context, as the split between embedded and centralised data management roles and the level of responsibility that each can take on will depend, for example, on the sustainability of data management roles at various level of organisation (team, department, research centre, institution).
- The competence framework also needs to be at a high enough level for its development to be feasible within EOSCpilot.
- The reported needs from the 4 Demonstrators reporting so far are broader than data management aspects of stewardship, also reaching into data processing, analysis etc. Many of the demonstrator needs are more about engineering and deploying services in the cloud.
- It is not clear that Demonstrators have identified the skill requirements that are needed for the services to have an impact with users, or to address long-term data stewardship.
- WP7 should differentiate between what users / researchers need to know, e.g. about applying FAIR principles using the Science Demonstrators, and what the relevant service providers and intermediaries would need to do, to be able to operate the demonstrator applications as services in EOSC.

### Where the gaps are in our competence model and in skills development activity

1. What changes to the draft competencies are needed to reflect the current skills development activity we have profiled?
2. What changes are needed to meet skills gaps identified in earlier presentations?
3. How should competence framework reflect disciplinary differences in embedding of skills within the research team, and in the use of 'external' services from the organisation or from third parties?

Key issues discussed here were:

- EOSCpilot should not duplicate the detailed competence analyses already done, e.g. by EDISON, the RDA Interest Group on ETDH, and Purdue University Libraries but offer a high level synthesis, as well as some general principles and dimensions relevant to WP7. The D7.1 report should offer a menu of competences for designing training on the use of the demonstrators to achieve their impact, and their relevance/ utility for that purpose can be validated later in the project.
- CORBEL's application of the EDISON competence framework may be an exemplar that D7.1 can generalize from.
- For EOSC, the scope of stewardship should include at least awareness level competences that overlap with data science, research and engineering e.g. on moving data or code to the cloud, managing software repositories, and on reproducibility.
- Ease of use and presentation: Mozilla Foundation Web Literacy framework offers a useful example of how to present the competence framework.

- The SFIA framework of IT competences is worth considering as an exemplar of how to include responsibility level as one of the dimensions proposed for EOSCpilot.
- In task 7.3 the competence framework will be extended to include capabilities. These should map to the organizational level responsibilities in the competence framework, i.e. for stewardship beyond the individual research project level (regardless of whether that means a department, research centre, institution, or third-party service provider).

### Conclusions: How should EOSCpilot address the gaps

1. Identify appropriate training delivery, whether through EOSCpilot workshops or third-party activity/ materials
2. Identify skills strategy and policy issues for stakeholder engagement
3. Identify an approach for certification of training e.g. to pick up from EDISON
4. Make recommendations for institutions and funders on policy changes to improve skills capacity management and the career development of data stewards and related professions

The target needs to be realistic. The WP cannot possibly meet all skills gaps for all Demonstrators, or target all domains. The main issue is feasibility of structuring training delivery (or skills acquisition) to fit with the short life cycle of Science Demonstrators. Designing training to fit the specific needs of individual Demonstrators would inevitably mean content is skewed towards those running earlier in the project, but only deliverable after they have run. For the same reason it may be more feasible to focus provision on the outcomes from Science Demonstrators, e.g. illustrating how these help researchers apply FAIR principles, or benefit wider use of the infrastructures, rather than the skills needed to build those services and integrate them into EOSC.

The first training workshop is scheduled for November 2017 at the multi-stakeholder event. This should partly focus on engaging stakeholders who would have an interest in the competence framework as a reference resource for skills development, and those who would directly apply it in their organisations. The event could also engage further with the ERA working groups on Skills and the closely related Rewards group. Meanwhile we need to follow up on their surveys.

The training part of the WP should target people responsible for skills development within RIs and eInfras. Approaches to on-the-job learning found useful in CORBEL and elsewhere, i.e. bite-sized learning, skills exchanges should be considered.

The Training Infrastructure to be described needs to highlight the networks of people and facilities available for delivering training, as well as systems and processes.

Further investigation of TeSS is needed to establish whether it may be recommended as a cross-domain solution for harvesting training events/ materials, and whether these may be tagged with competences from EOSCpilot or other frameworks.

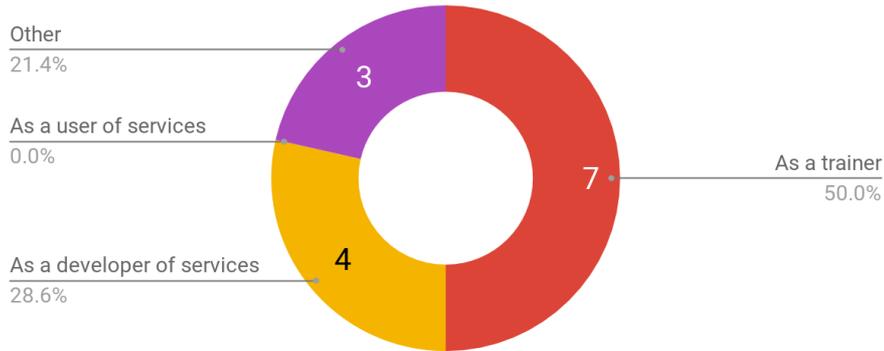
Formal certification may be desirable for training that focuses on deployment or application of EOSC services, however it may also be a barrier to building up networks of trainers knowledgeable on data stewardship topics relevant to EOSC. Informal more light-weight approaches to certification e.g. Open Badges, are worth further consideration as an incentive for participation in training. Some evidence is available that this approach works as an incentive to apply data skills towards data sharing.

CC-BY licensing should apply to training content.

### ANNEX C. MENTIMETER RESULTS FROM CONSULTATIVE WORKSHOPS

#### Workshop W-1B EUDAT 2018 Porto

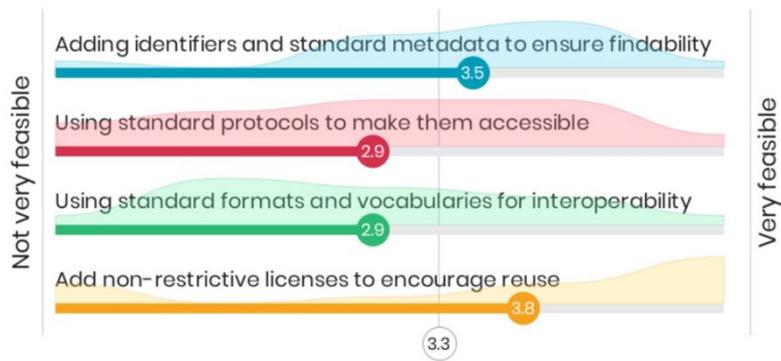
How would you contribute to EOSC?



14

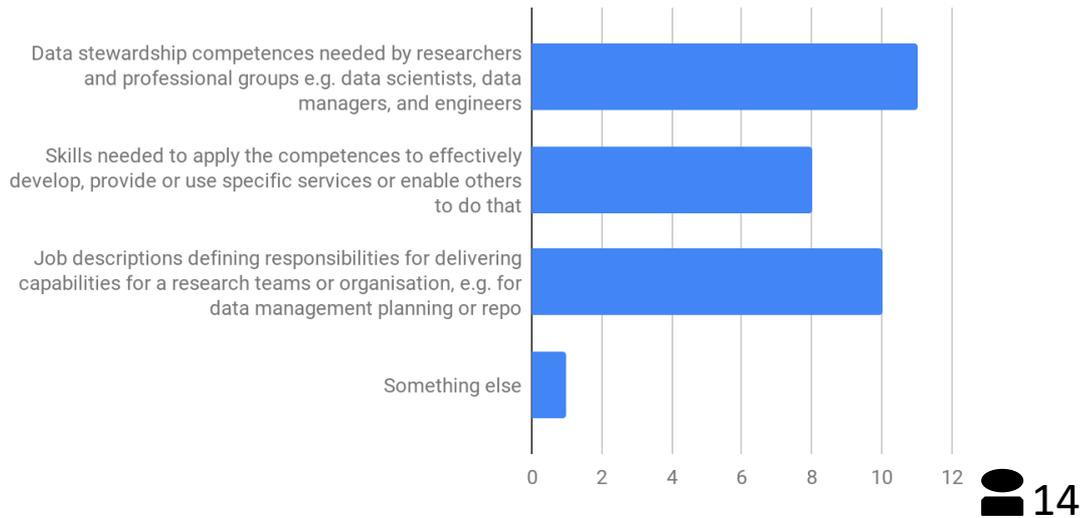
How feasible is it to make training materials FAIR using the following methods?

Mentimeter



14

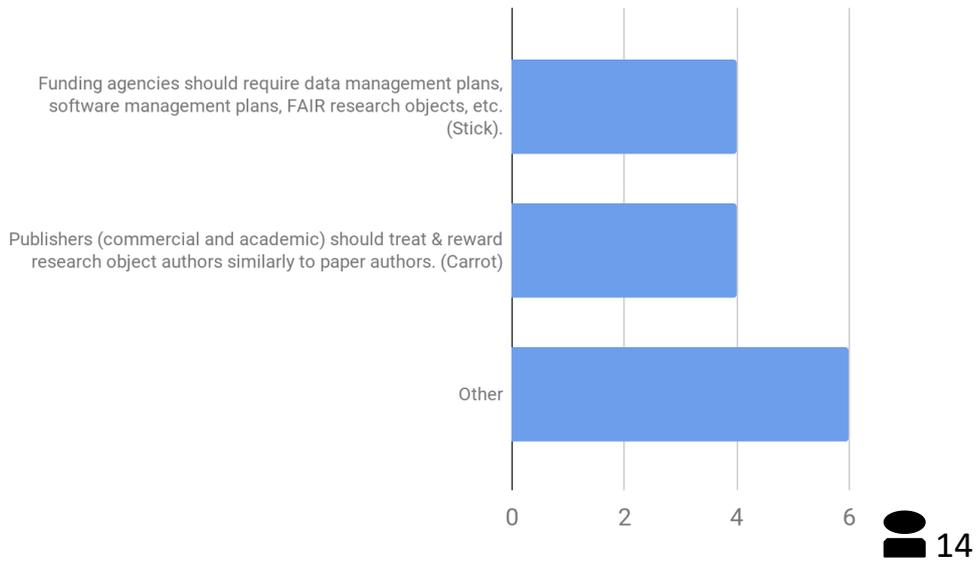
### What kind of examples would be useful to help you build capabilities to work with the EOSC? (choose all that apply)



### What should be EOSCs priorities for skills development?

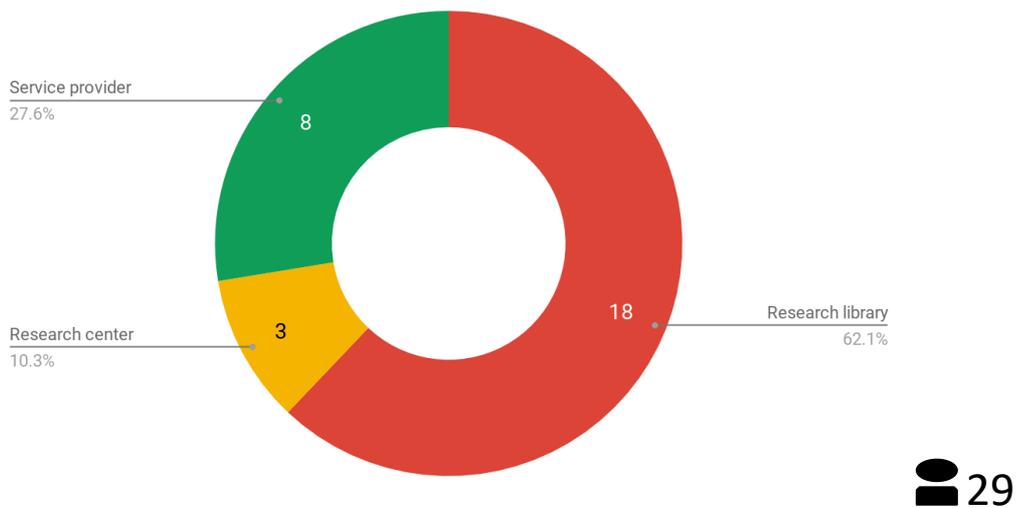


How could/should Open Science practices be best encouraged within scientific communities? With a stick or with carrots?

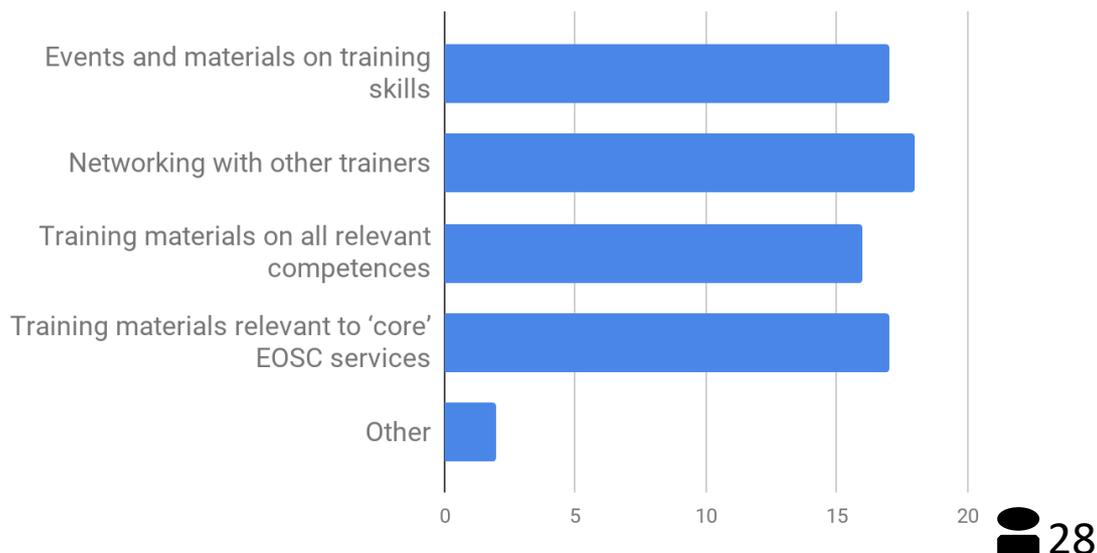


**Workshop W-2, IDCC 2018 Barcelona**

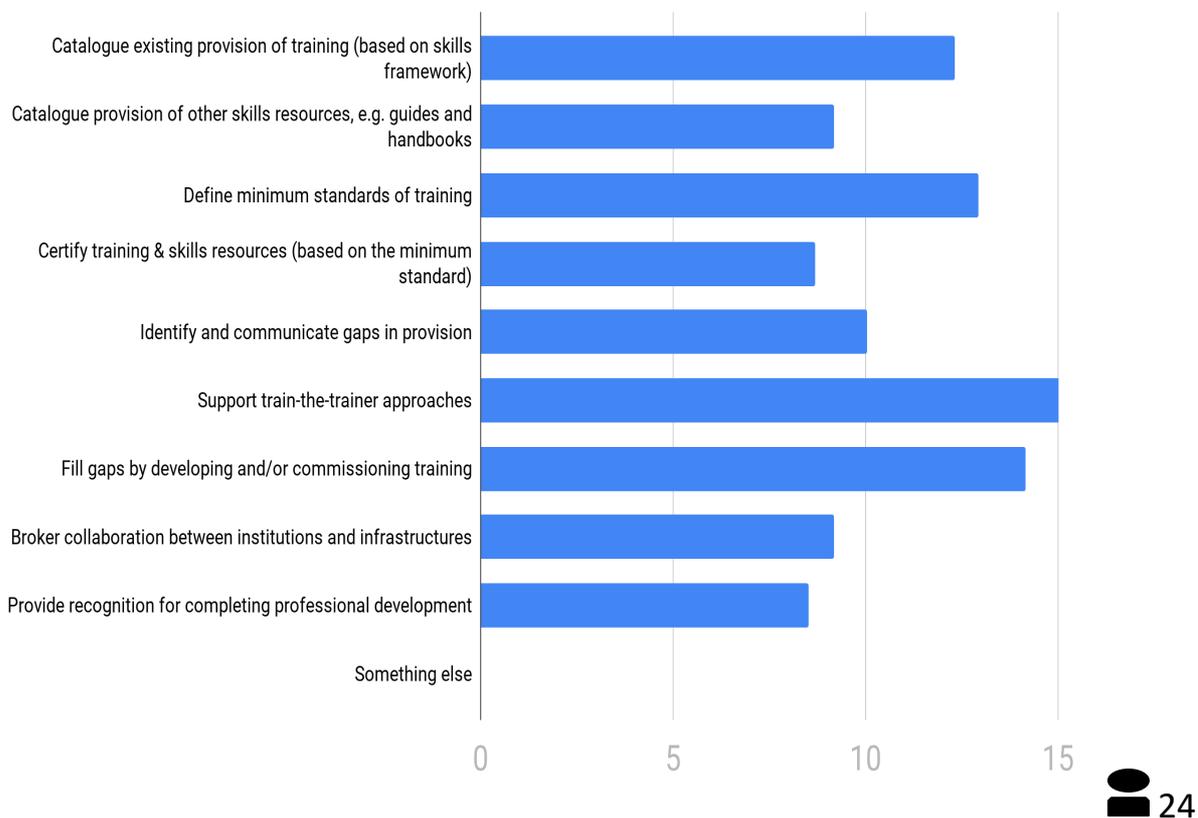
Where are you based?



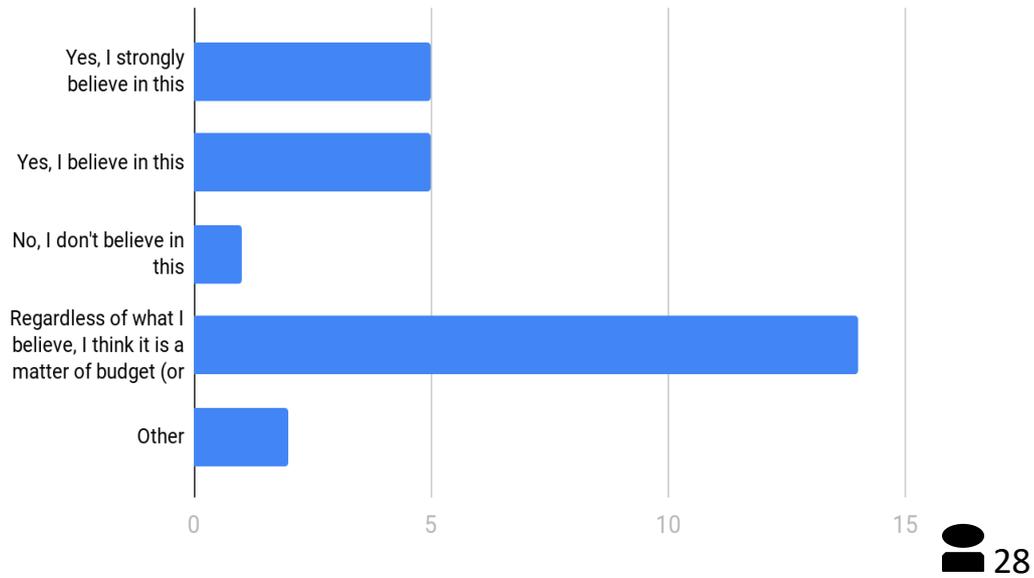
### What forms of train-the-trainer support would you expect from EOSC?



### What should be EOSCs priorities for skills development?

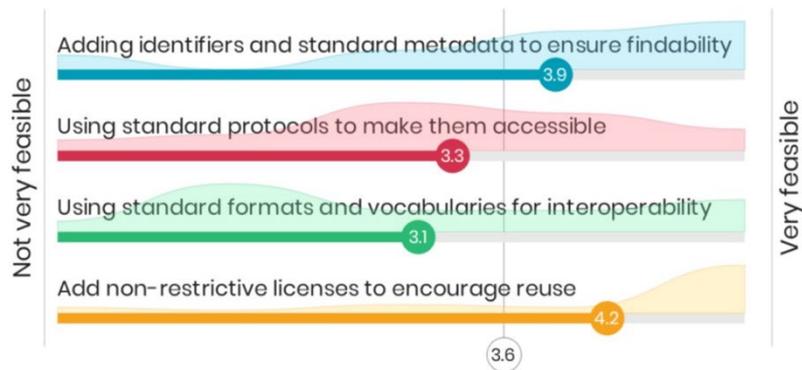


### Should every research group have a data steward?



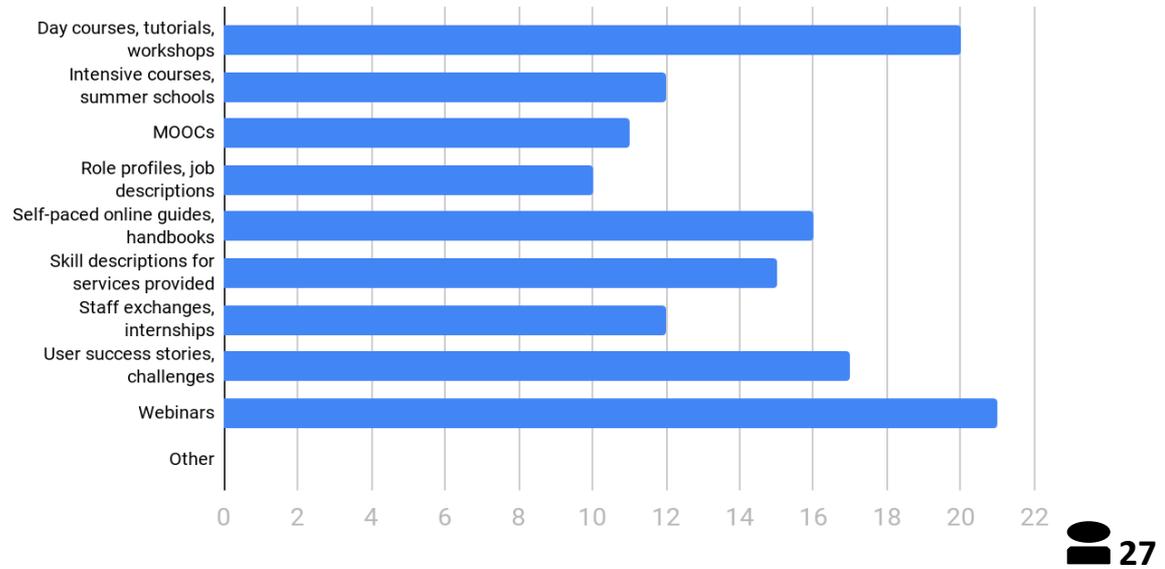
### How feasible is it to make training materials FAIR using the following methods?

Mentimeter

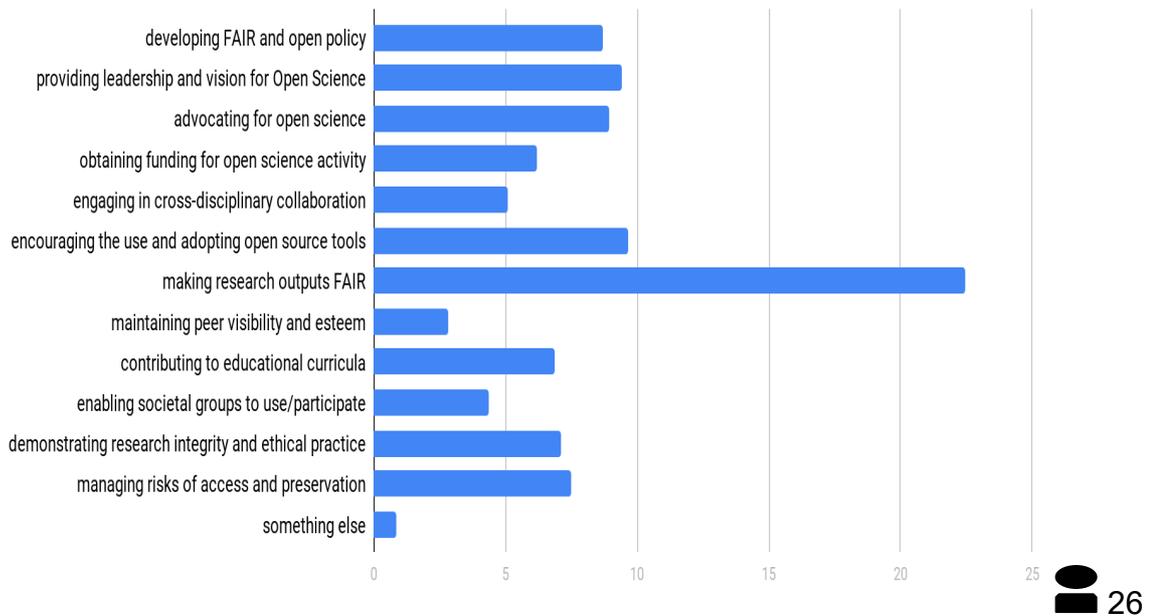


25

### What forms of skills resource would you expect EOSC to support?



### What should institutions reward researchers and support professionals for?



## ANNEX D COMPLEMENTARY DATA AND STATISTICS OF GRIDKA SCHOOL 2018

### 1) Questionnaire templates

#### Participant questionnaires

- Questions for each plenary lecture
  - Please rate the quality and relevance of the plenary lecture
    - Array with 10 point choice (1 – very poor, 10 – excellent)
- Questions for parallel hands-on tutorials
  - Please select the tutorial you attended. Based on this, you can evaluate the quality and relevance of contents as well as the difficulty level you expected from the course and experienced during the course.
    - Which tutorial did you attend
      - List of tutorials with comment
    - What was your level of knowledge before attending the tutorial?
      - List (No previous knowledge; Beginner; Intermediate; Expert) with comment
    - Please rate the difficulty of the tutorial you expected from available information by tutors.
      - List (Beginner; Intermediate; Expert; None) with comment
    - Please rate the difficulty level you experienced during the tutorial.
      - List (Beginner; Intermediate; Expert; None) with comment
    - Please rate the quality and relevance of the parallel session you attended
      - Array with 10 point choice (1 – very poor, 10 – excellent)
- Questions about learning objectives
  - Please tell us about your learning objectives. To improve GridKa School even more in the following years, we evaluate your learning objectives with regards to the Skills and Capability Framework that is developed by EOSCpilot.
    - Which professional roles do you belong to?
      - Multiple choice (Domain Researcher; Data Scientist / Analyst; Data Manager; Data Service Engineer) and 'Other'
    - Which of the target roles match your learning objective?
      - Multiple choice ((service) user; (service) operator; (service) owner) and 'Other'
    - When you registered for GridKa School, which of the skills/competences match your learning objective? Are there any skills missing in the list?
      - Multiple choice (Planning data management and sharing; Open data model and database design; Metadata, persistent identifier specification; Open source software / service requirements; Repository and data management platform appraisal; Workflow set up and documentation; Database management; Software prototyping; Data collection and reuse of open data; File naming and organisation; Data provenance and software versioning; Math and statistical knowledge application; Critical thinking and theory building; Creative problem solving, flexibility; Open source software / service development; Data transformation and integration; Data mining, querying, interpretation; Predictive modelling and analytics; Machine learning methods; Data preparation, documentation for reproducibility; Data quality assurance using open standards; Ethical, legal and data policy compliance; Data transfer and long-term storage; Format and media migration; Software review and preservation; Documentation for public use, lay understanding; Access control and management; Ethical



- Tutorial: Introduction to using HTCondor to run distributed compute Jobs and Workflows on Servers, Clusters, Grids, or Clouds
- Tutorial: Hacking Hands-On
- (Long-term) data management, curation, preservation and provenance
  - Lecture: Leading the digital transformation in chemicals
  - Lecture: Reliability in Supercomputing – Trends and Challenges
  - Lecture: Building global bridges and connections: The Research Data Alliance
  - Lecture: Open Science for Big Data
  - Tutorial: Scalable and reproducible workflows with Pachyderm
  - Tutorial: Databases for large-scale science
- Federated authorisation and authentication including data security in the cloud
  - Lecture: Cloud Security in the Context of Fog and Edge Computing
  - Tutorial: Hacking Hands-On
- Domain knowledge and software development to adapt workflows to cloud workflows.
  - Lecture: Using the Computer as a Wet Lab: From Biomolecular Simulations to Simulating Tissue
  - Lecture: VICE: From desktop to cloud or HPC – Virtualized Research Environments
  - Lecture: COBalD – Boosting Scientific Workflows with Opportunistic Resources
  - Lecture: Six Key Challenge Areas driving innovation in Distributed High Throughput Computing
  - Lecture: GPU Programming 101
  - Lecture: Quantum Computing and IBM Q: An Introduction
  - Lecture: Fixed-Point and Adaptive Precision Methods for Numerics at Exascale
  - Lecture: Boosting productivity of researches integrating IT services into an interactive analysis platform
  - Tutorial: Introduction to Go
  - Tutorial: Introduction to Python
  - Tutorial: Julia: high performance programming the easy way
  - Tutorial: Collaborative Software Development
  - Tutorial: Introduction to using HTCondor to run distributed compute Jobs and Workflows on Servers, Clusters, Grids, or Clouds
  - Tutorial: Scalable Scientific Analysis in Python using Pandas and Dask
  - Tutorial: Productive GPU Programming with OpenACC
  - Tutorial: Parallel programming with OpenMP and MPI
  - Tutorial: Quantum Computing

### Skills for service management

- System management and administration of virtual infrastructure
  - Lecture: Reliability in Supercomputing – Trends and Challenges
  - Lecture: Cloud Security in the Context of Fog and Edge Computing
  - Lecture: ViCE: From desktop to cloud or HPC – Virtualized Research Environments
  - Lecture: COBalD – Boosting Scientific Workflows with Opportunistic Resources
  - Lecture: HNSciCloud overview
  - Lecture: GÉANT Network Overview and Evolution Plans
  - Tutorial: HNSciCloud Tutorials by Contractors
  - Tutorial: Introduction to using HTCondor to run distributed compute Jobs and Workflows on Servers, Clusters, Grids, or Clouds
  - Tutorial: Hacking Hands-On
- Federated authorisation and authentication including data security in the cloud
  - Lecture: Cloud Security in the Context of Fog and Edge Computing

- Tutorial: Hacking Hands-On
- Monitoring of federated data and compute services
  - Tutorial: Behind the scenes perspective: into the abyss of profiling for performance

### Skills for effective use

- Decision-making for when using the cloud is appropriate
  - Lecture: ViCE: From desktop to cloud or HPC – Virtualized Research Environments
  - Lecture: Boosting productivity of researches integrating IT services into an interactive analysis platform
- Data security in the cloud
  - Lecture: Cloud Security in the Context of Fog and Edge Computing
  - Tutorial: Hacking Hands-On
- Data management, curation, including provenance to enable reusability and long-term preservation
  - Lecture: Open Science for Big Data
  - Lecture: Thirteen modern ways to fool the masses with performance results on parallel computers
  - Tutorial: Scalable and reproducible workflows with Pachyderm
  - Tutorial: Databases for large-scale science
  - Tutorial: Introduction to the SciPy stack and Jupyter Notebooks
- Domain knowledge and domain-specific scientific methods
  - Lecture: Using the Computer as a WetLab: From Biomolecular Simulations to Simulating Tissue
  - Lecture: Fixed-Point and Adaptive Precision Methods for Numerics at Exascale

### Skills for assessment

- Research strategy, ethics and policy in the age of open science
  - Lecture: Open Science for Big Data
  - Lecture: Beyond Algorithms. Even Better Systems by Understanding Human Special Features
  - Lecture: Building global bridges and connections: The Research Data Alliance
  - Lecture: Thirteen modern ways to fool the masses with performance results on parallel computers

### Additional contributions

- Lecture: Cognitive Computing – (The new AI)
- Lecture: Exploring Big Data with Visual Analytics
- Tutorial: Machine Learning with Neural Networks

## 3) Statistics regarding professional groups

Figure D.1 represents the distribution of the professional groups identified in the skills framework amongst the participants. Many trainees felt they belong to one or more of the groups described in the model (see Figure D.2): Domain researcher, Data Scientist/Analyst, Data manager and Data Service Engineer, whilst a few chose “other” (User Support, Scientific Computing Support, Scientific computing engineer, system administrator, DevOps, Student)

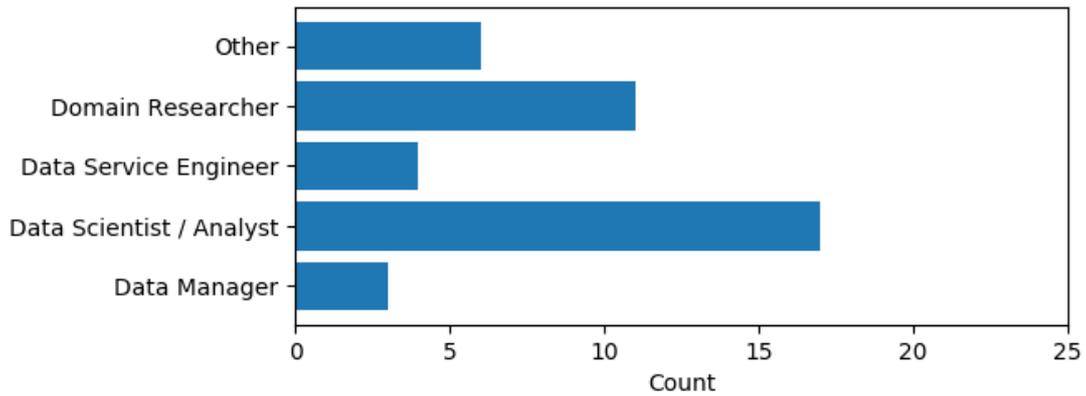


Figure D.1: Number of participants per professional role based on the responses of 31 participants.

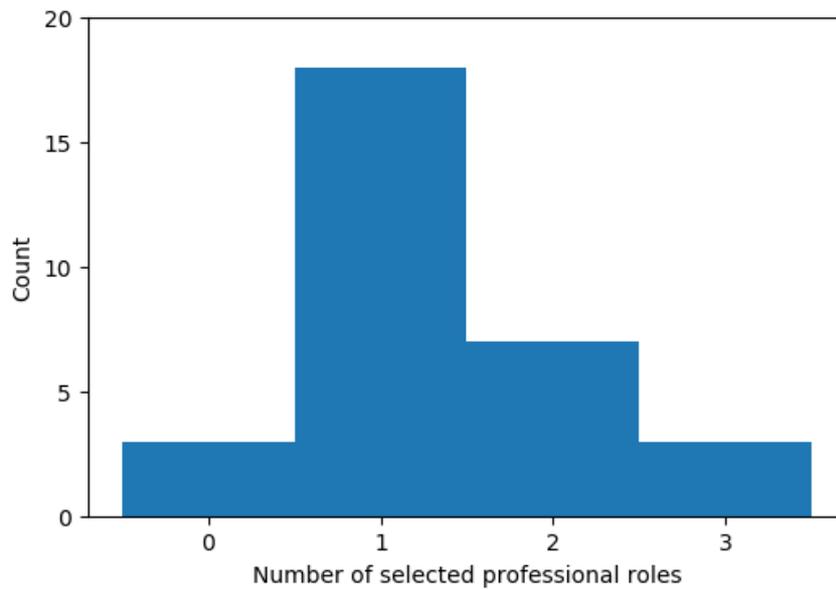


Figure D.2: Distribution of number of selected professional roles per participant

#### 4) Statistics regarding service roles

Figure D.3 represents the distribution of the services roles identified in the skills framework amongst the participant. Many trainees felt they belong to one or more of the roles described in the model (see Figure D.4): (service) user, (service) operator, or (service) owner, whilst a few chose “other” (developer / devops).

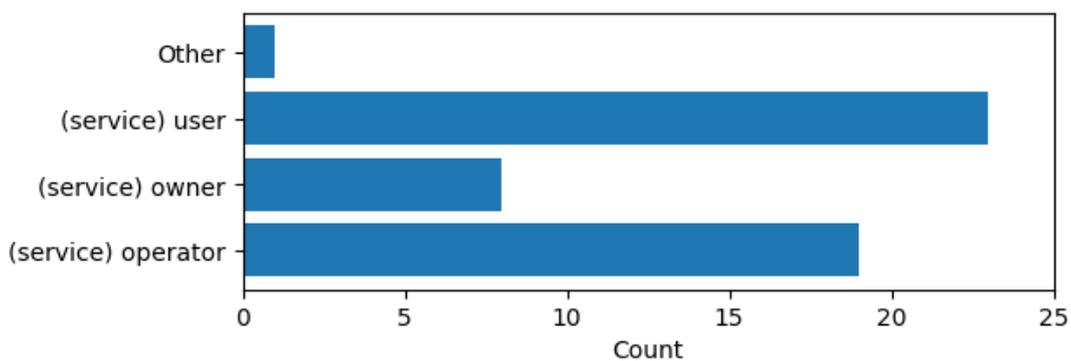


Figure D.3: Number of participants per service role based on the responses of 31 participants.

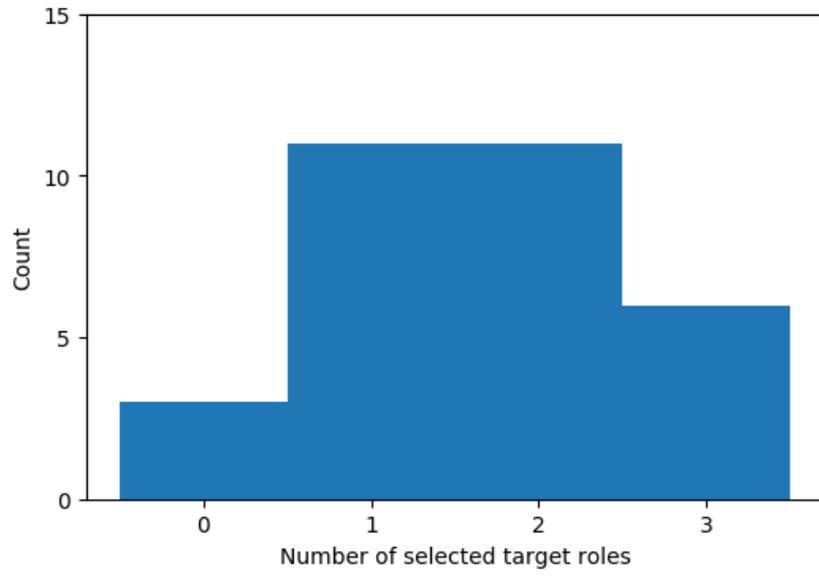


Figure D.4: Distribution of number of selected service roles per participant

### 5) Learning objectives and their coverage

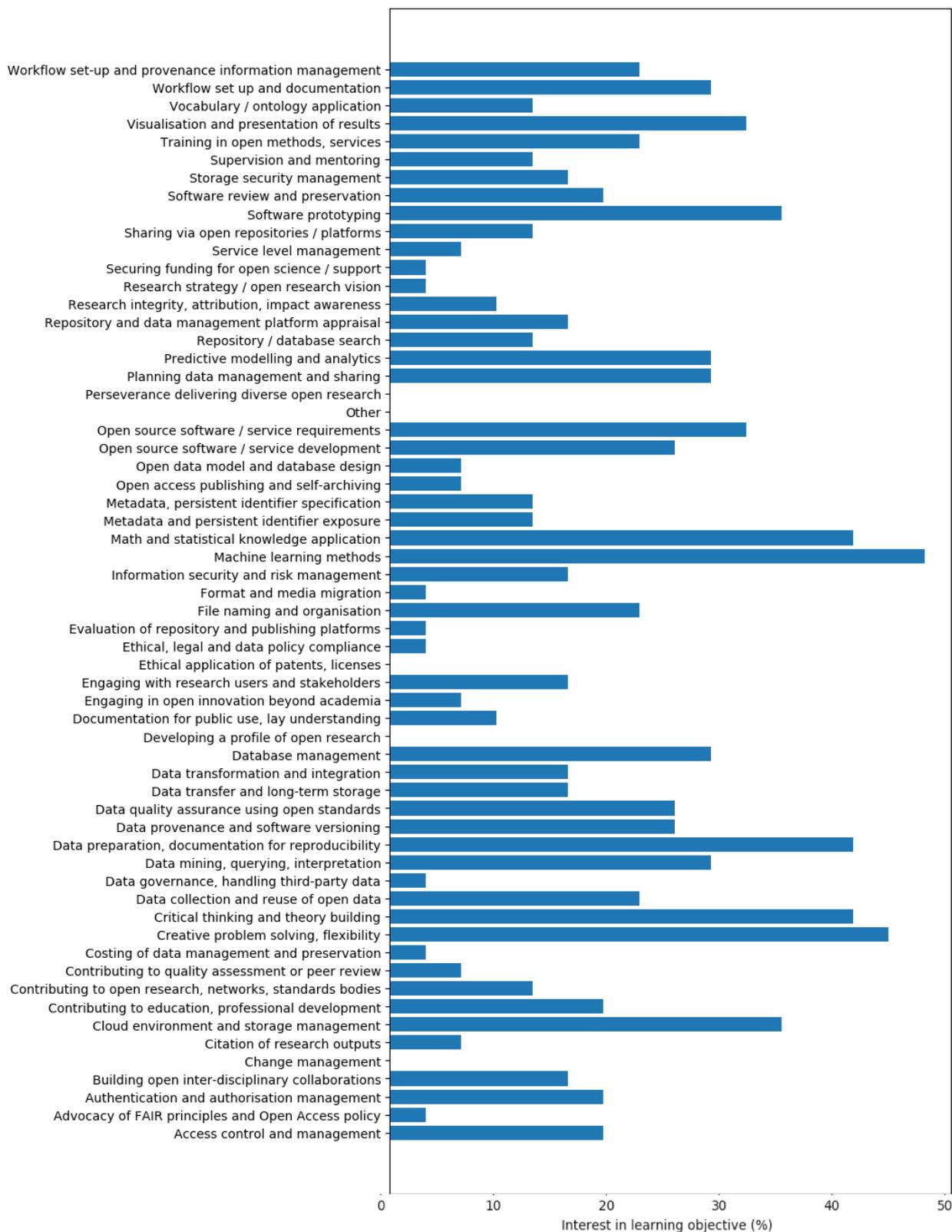


Figure D.5: Percentage of people selecting a specific learning objective.

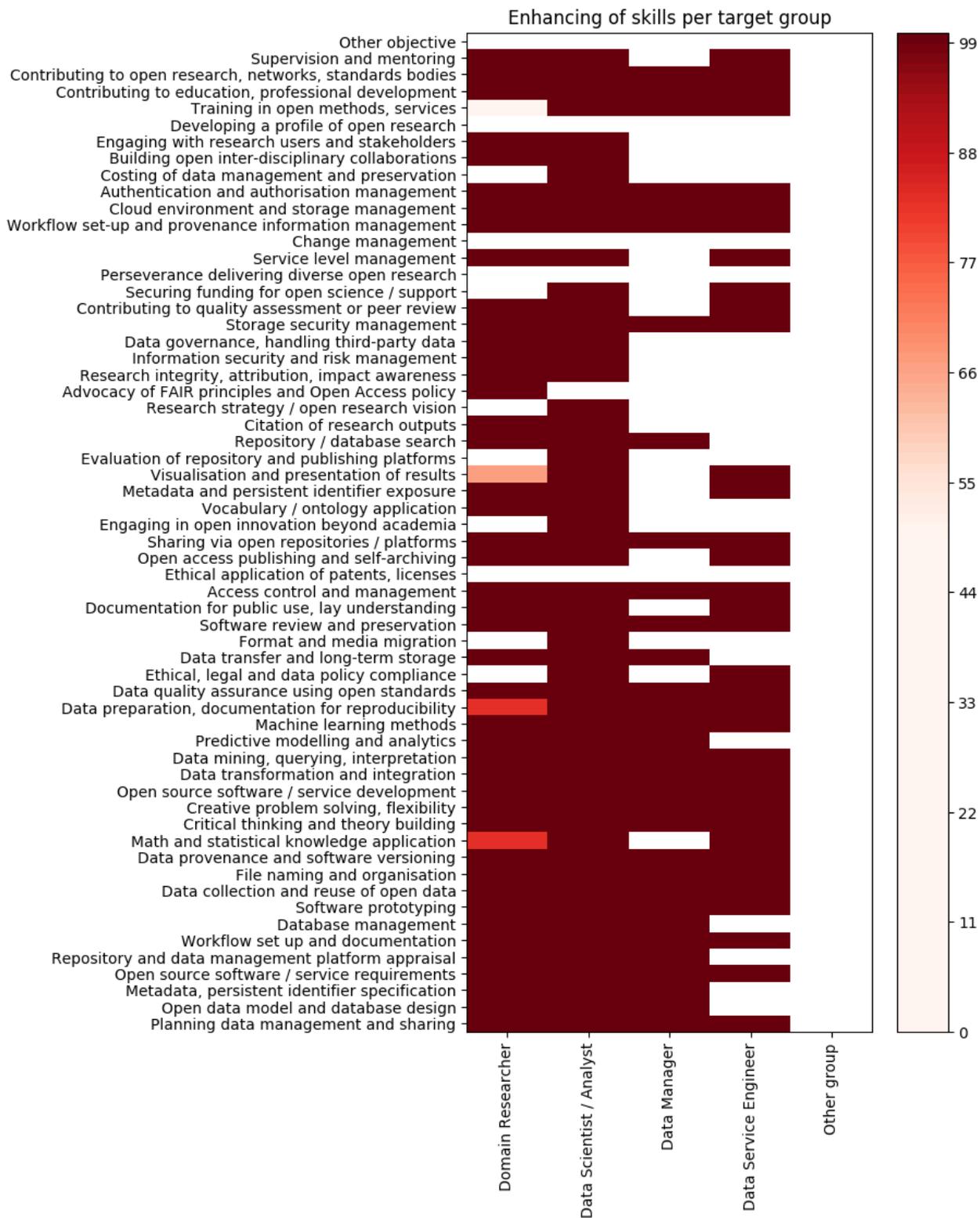


Figure D.6 : Effectiveness of GridKa School in enhancing skills per target group

### 6) Overall effectiveness of GridKa School

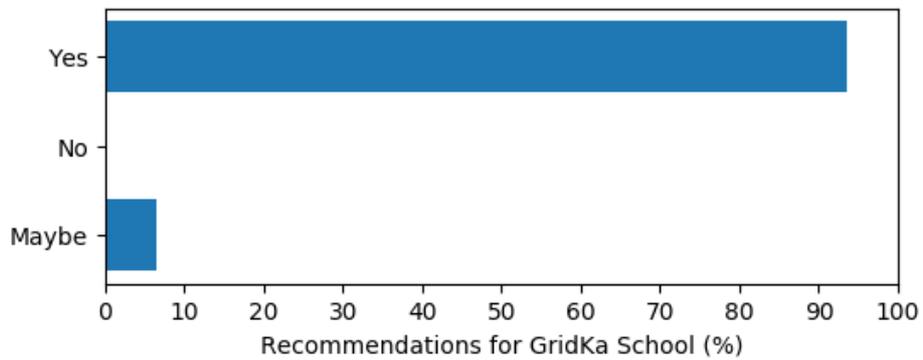


Figure D.7: Percentage of recommendations for future GridKa Schools

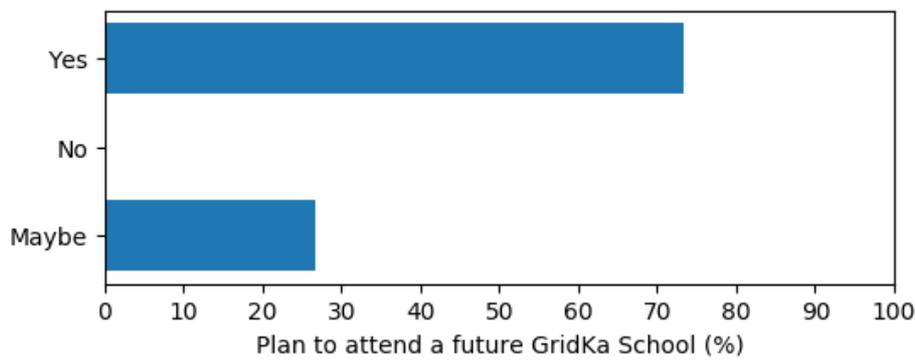


Figure D.8: Percentage of participants planning to attend a future GridKa School

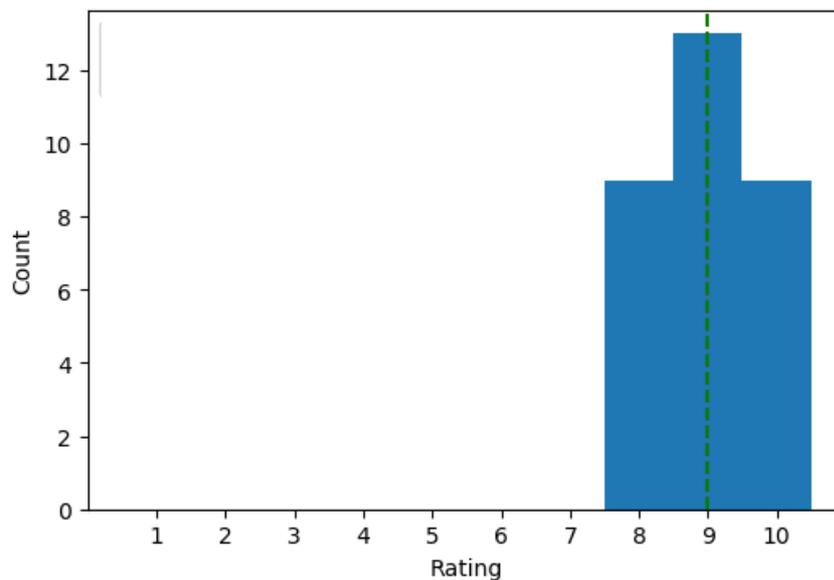


Figure D.9: Overall rating (1-10) of GridKa School

Some quotes from the survey responses:

- It has been very difficult to choose among the broad variety of very interesting hands-on sessions. It'd take twice the time to attend to all the topics I was interested in.

Nonetheless, it was an great school and I really enjoyed it! I'm already looking forward to coming back next year :)

- There has been a huge amount of very interesting hands-on sessions and it was very difficult to choose. One would have needed at least twice the time to cover all of the interesting topics. All in all I really enjoyed the school but it was way too short! :)
- I think there are two points which can be improved.  
1/ Some tutorials are fully booked right after the registration starts. This is not the first time (at least it was the same last year), and it is always frustrating to register and having several tutorial sessions already full.  
2/ I think the lunch break is too short. I would increase it to 15 mins at least (at least a break of 1h15 would be needed, maybe 1h30 would be best). If we start the afternoon session at 1:30 PM, then it is still possible to end it at 6:30. Attendees will have more time to eat, to relax, and to discuss with each others and with the tutors (discussion is, I think, an important part of such a school).
- It was an interesting mix of plenary talks, given by the different backgrounds of the speakers. However the whole school was a bit biased towards HEP.  
It was hard to choose from the set of offered tutorials, as a lot of them sounded very interesting. It'd be nice if all the material still missing in Indico, could be uploaded :)  
The FTU/SCC location was a bit far-off from the city-center.  
Big thanks to Eileen, the exception handlers, and the whole GridKa School Team!
- This was indeed a productive week!  
A suggestion about parallel tutorial sessions. It may be nice to have video recordings of tutorial sessions so that nobody misses anything.