



Project Title	Fostering FAIR Data Practices in Europe
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

D7.1 FAIR in European Higher Education

Work Package	WP7
Lead Author (Org)	Lennart Stoy (EUA)
Contributing Author(s) (Org)	Bregt Saenen (EUA), Claudia Engelhardt (UGOE), Joy Davidson (DCC), Birgit Schmidt (UGOE), Vinciane Gaillard (EUA)
Due Date	29.02.2020
Date	24.02.2020
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.3629683

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

FAIRsFAIR Work Package 7 “FAIR Data Science and Professionalisation” aims to develop resources and build communities that support the uptake, if not mainstreaming, of RDM and FAIR practice within higher education curricula. In order to achieve these objectives, the present report aims to build a foundation for the identification of existing practices and needs of higher education institutions. It builds on a web-based questionnaire with 90 responses and two focus groups with a total of 50 participants implemented between September and November 2019. The research covered several dimensions of research data management at HEIs relevant for the implementation of FAIRsFAIR WP7, as well as WP3 “FAIR Data Policy Practice” and WP6 “FAIR Competence Centre”. These dimensions included: Institutional research data management policies; support services for research data management; competence development of students and graduates; universities and EOSC; and FAIRsFAIR support for universities

Versioning and contribution history

Version	Date	Authors	Notes
0.1	29.11.2019	Lennart Stoy (EUA)	Document structure created
0.2	02.12.2019	Lennart Stoy (EUA)	Initial data and charts from survey added
0.3	09.12.2019	Bregt Saenen (EUA)	Section 3.5, 3.6 added
0.4	13.12.2019	Lennart Stoy (EUA)	Sections 1, 2 added
0.5	03.01.2020	Lennart Stoy (EUA)	Section 3.1 added
0.6	20.01.2020	Lennart Stoy (EUA)	Section 3 - Focus group inputs added
0.7	29.01.2020	Claudia Engelhardt (UGOE), Joy Davidson (DCC)	Sections 3.2, 3.3, 3.4 added
0.8	03.02.2020	Lennart Stoy (EUA), Bregt Saenen (EUA), Claudia Engelhardt (UGOE), Birgit Schmidt (UGOE), Vinciane Gaillard (EUA)	Completed draft for internal review
1.0	24.02.2020	Lennart Stoy (EUA), Bregt Saenen (EUA), Claudia Engelhardt (UGOE), Joy Davidson (DCC)	Revision following comments from internal review; final edits before submission to EC

Disclaimer

FAIRsFAIR has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no. 831558. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

Abbreviations and Acronyms

DMP	Data management plan
EDSF	EDISON Data Science Framework
EOSC	European Open Science Cloud
EUA	European University Association
FAIR	Findable, Accessible, Interoperable, Reusable
FAIR4S	EOSCPilot framework of FAIR data stewardship skills for science and scholarship, and draft recommendations on FAIR training
FAIRsFAIR	Fostering a FAIR data culture in Europe (H2020 project)
HEI	Higher Education Institution
HLAC	High Level Advisory Committee
IPR	Intellectual property rights
OS-CAM	Open Science Career Assessment Matrix
OSPP	Open Science Policy Platform
RDM	Research data management
SSH	Social sciences and humanities
UC3M	University Carlos III of Madrid
WP	Work package

Table of contents

Executive Summary	5
1. Introduction	7
2. Methodology	12
2.1 Survey and questionnaire	12
2.2 Focus groups	14
2.3 Data availability	14
3. Results	15
3.1 Sample description and awareness of FAIR	15
3.2 Policies and support services	18
3.4 Competences and programmes	27
3.5 Universities and the European Open Science Cloud	38
3.6 Potential for FAIRsFAIR support	41
4. Summary of main findings	43
4.1 Addressing FAIR in higher education	43
4.2 Making research data policies and support services ready for FAIR	44
4.3 Universities and the European Open Science Cloud	45
Annexes	46
Annex 1A - Survey questionnaire	46
Annex 1B - EDSF and FAIR4S in the survey questionnaire	59
Annex 2 - Focus Group in Madrid	63
Annex 3 - Focus Group in Amsterdam	65
Annex 4 - Data for Figures 14a-c	67

Executive Summary

FAIRsFAIR Work Package 7 “FAIR Data Science and Professionalisation” implemented a landscape analysis of Higher Education Institutions (HEIs) during late 2019. The objective of the analysis was to assess the status quo of research data management and the FAIR principles at HEIs. Both a **web-based questionnaire** with 90 responses and **two focus groups** with a total of 50 participants were conducted between September and November 2019.

The research covered several dimensions of research data management at HEIs relevant for the implementation of FAIRsFAIR WP7, as well as WP3 “FAIR Data Policy Practice” and WP6 “FAIR Competence Centre”. These dimensions included:

- Institutional research data management policies
- Support services for research data management
- Competence development of students and graduates
- Universities and EOSC
- FAIRsFAIR support for universities

Awareness of the FAIR principles

Awareness of the FAIR principles is perceived as **relatively high only among professional support staff and university leadership**. Researchers and students are seen as least aware of the FAIR principles. Immediate actions may focus on teaching fundamentals of research data management (RDM) and FAIR data that more specialised competencies can be built upon.

Referring to reproducibility, research integrity and transparency of research was also seen as potentially effective in raising awareness of the FAIR principles. RDM and data stewardship should be understood as an important part of the research process.

FAIR and RDM in higher education programmes

Universities express a **clear interest to advance data stewardship competencies and skills for students, researchers and professional staff**, although not every role or discipline has the same requirement. Many HEIs have a strategy towards the development of digital literacy and skills, at different levels within the institution and express the intention to develop such a strategy in the future. This may prove an opportunity to ‘mainstream’ more specific research data-related skills and competencies throughout a HEIs portfolio of educational offers.

Where available, the **provision of RDM-related teaching is lowest at bachelor level, low to moderate at master level and strongest at doctoral level**. There are only slight differences between domains, with life, natural and engineering sciences usually more advanced than social sciences and humanities (SSH) fields or multidisciplinary/generic offers.

In terms of coverage of specific skills groups, there are broad trends across educational levels and disciplines. Data analytics and statistics as well as compliance with legal and regulatory frameworks (data protection, privacy, intellectual property rights (IPR) and ethical issues) are most commonly addressed. Nevertheless, respondents expressed a **strong need to strengthen the teaching of**

different research data-related competences at all three levels. This is most urgent for doctoral education, where almost 90% of HEIs reported a need to support data management skills and competencies and roughly 80% report a need to support data analytics and data science competences.

Universities indicated a **strong need for practical guidance** on how to follow the FAIR principles in different domains and disciplines. Universities would moreover welcome train-the-trainers actions, sharing of good practices, and a domain-specific competence framework and model courses/outputs from FAIRsFAIR.

Making research data policies and support services ready for FAIR

Many participating HEIs reported the existence of institutional research data policies (60%) and an even larger number (77%) indicated that they offered research data support services to researchers and staff.

Overall, **HEIs tend to encourage rather than mandate many RDM activities**. While over half of respondents indicated that they indeed mandate sharing of data and data deposit, explicit **reference to FAIR in RDM policies is not yet common practice**. Nonetheless, aspects of FAIR are often included implicitly. As FAIR data is an emerging concept in many areas, it may be sensible for HEIs to retain implicit reference in policies while encouraging and supporting researchers to make data as FAIR as possible and emphasising how the local policy aligns with funders' requirements.

The results show a gap between those primarily involved with formulating policies and those primarily involved in supporting adherence to them. It is recommended that both the **development and implementation of a policy should be a collaborative effort**. While the responsibility for the support and implementation of policies may be assigned to a more specialised service (e.g. the library), the formulation of policies takes place through a broader consultation of internal stakeholders.

As the development and implementation of RDM policies and support services is an emerging area requiring new or additional competencies from those responsible, **existing staff will need to be trained and/or new staff to be hired**. Sufficient resources must be made available by institutions but possibly also by other organisations such as research funders.

Universities in the European Open Science Cloud

Universities report a **low awareness of the European Open Science Cloud (EOSC)** among different groups within an organisation, with the exception of university leadership and research data support. Responding institutions note a **lack of use cases for EOSC and limited institutional capacity** as a main barrier to engage with EOSC.

On the other hand, universities do see **possible benefits of EOSC**, in particular as a system **facilitating collaborative research and increasing the visibility of research** carried out within the institution. Possible ways to ensure engagement of HEIs with EOSC could be to include EOSC onboarding support for institutions in the further EOSC implementation strategy.

1. Introduction

FAIRsFAIR – Fostering Fair Data Practices in Europe – aims to support the emergence of a FAIR data culture (i.e., one where research data are findable, accessible, interoperable and reusable) in the context of the European Open Science Cloud (EOSC).

The vision of EOSC is to be “a virtual commons where science producers and science consumers come together for more insights, new ideas and more innovation”¹ by the means of federating existing and new infrastructures and making available research data and ancillary services for the management, analysis and curation of said data.

Creating EOSC is a joint effort between scientific communities, repositories, research infrastructures, institutions and numerous other stakeholders. For researchers, EOSC will offer a platform for seamless data discovery, sharing, re-use and analysis. Indeed, the success of EOSC depends upon the use of EOSC services by researchers. For the sharing of research data, this means the wider adoption of research data management (RDM) practices and the FAIR principles² – as increased production and use of FAIR data is the cornerstone of EOSC.

However, current research data management practices and implementation of the FAIR principles vary across disciplines and domains. As a result, the provision of training and competence development of researchers has been identified as a main challenge for EOSC:

- The **2017 EOSC Declaration** lists *Data culture and FAIR data* as its first domain of priorities, further detailing that “European science must be grounded in a common culture of data stewardship” and that “[o]nly a considerable cultural change will enable long-term reuse for science and for innovation of data created by research activities”. It specifies that “necessary skills and education in research data management, data stewardship and data science should be provided throughout the EU as part of higher education, the training system and on-the-job best practice in the industry.”³
- The final report and recommendations of the European Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC), **Prompting an EOSC in Practice**, highlights the need to “[b]uild a workforce able to execute the vision of the EOSC by ensuring data stewards, data and infrastructure technologists and scientific data experts who are trained and supported adequately”.⁴

¹ European Commission (2019). *European Open Science Cloud (EOSC) Strategic Implementation Plan*. Luxembourg: Publications Office of the European Union. Retrieved 21 February 2020, from: <https://doi.org/10.2777/202370>.

² Wilkinson, M., Dumontier, M., Aalbersberg, I., *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018, pp. 1-9. Retrieved 21 February 2020, from: <https://doi.org/10.1038/sdata.2016.18>.

³ Retrieved 21 February 2020, from: https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf.

⁴ European Commission (2018). *Prompting an EOSC in practice. Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC), 2018*. Luxembourg: Publications Office of the European Union. Retrieved 21 February 2020, from: <http://doi.org/10.2777/112658>.

- The **European Open Science Policy Platform (OSPP)** has, in the context of EOSC, recommended to “[d]evelop Open Science and data skills among all the key stakeholder categories” adding that “[t]o reap the benefits of the EOSC in support of open science, skill development in the area of Information Technology (IT) and data literacy should be supported at all levels, from the primary school up to policy makers”.⁵ In further reports, OSPP has also called on stakeholders to “[f]oster Open Science literacy as essential to European competitiveness at the global level, together with other digital and information competencies”.⁶ These Open Science skills have been further detailed in a dedicated Working Group report, including a section on “Skills Related to Data Management and Open Data”, distinguishing between “researchers and technicians that are at the ‘core’ of data engineering” and “researchers from other disciplines that consider technical skills such as data management to be an addition to their primary research field”.⁷
- The final report and action plan from the European Commission expert group on FAIR data, better known under its title **Turning FAIR into Reality** (hereafter abbreviated as *TFiR*), included more specific recommendations. *TFiR* suggested a “concerted effort [...] to coordinate and accelerate the pedagogy for professional data roles”, including a curriculum framework supporting FAIR data science. More efforts should also be made to train data scientists and data stewards. This could be achieved, for instance, by integrating “[d]ata skills, including an appropriate foundational level in data science and data stewardship, in undergraduate and postgraduate training across disciplines”.⁸
- On a broader level, the **New Skills Agenda for Europe** embraced a “focus on digital skills”.⁹ Under the subsequent **Digital Education Action Plan**, Open Science skills “from data management to legal aspects, including technical skills, such as data stewardship, data protection, scholarly communication and dissemination” have been included as a priority line of action.¹⁰

⁵ Open Science Policy Platform (May 2017). *Report on the governance and financial schemes for the European Open Science Cloud*. Retrieved 21 February 2020 from: https://ec.europa.eu/research/openscience/pdf/ospp_euro_open_science_cloud_report.pdf.

⁶ European Commission (2018). *OSPP-REC. Open Science Policy Platform Recommendations*. Luxembourg: Publications Office of the European Union. Retrieved 21 February 2020, from: <http://doi.org/doi:10.2777/958647>.

⁷ European Commission (2017). *Providing researchers with the skills and competencies they need to practise Open Science. Open Science Skills Working Group Report*. Luxembourg: Publications Office of the European Union. Retrieved 21 February 2020, from: <http://doi.org/10.2777/121253>.

⁸ European Commission (2018). *Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data*. Luxembourg: Publications Office of the European Union. Retrieved 21 February 2020, from: <http://doi.org/10.2777/1524>.

⁹ European Commission (2016). *A new skills agenda for Europe. Working together to strengthen human capital, employability and competitiveness*, COM(2016)381 final, 10 June. Retrieved 21 February 2020, from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016DC0381>.

¹⁰ Retrieved 21 February 2020, from: https://ec.europa.eu/education/education-in-the-eu/european-education-area/digital-education-action-plan-action-5-open-science-skills_en.

- The **Commission Recommendation on access to and preservation of scientific information** also calls EU member states to “set and implement clear policies [...] for the necessary skills and competences of researchers and personnel of academic institutions regarding scientific information”.¹¹

There are further factors besides the development of EOSC that may drive the need for researchers and other professional staff to be skilled in data management.

- The future European Framework Programme for Research & Innovation, **Horizon Europe**, scheduled to start in January 2021, is likely to include requirements for RDM following the FAIR principles and in the form of Data Management Plans (DMPs).¹²
- The **Directive on open data and the re-use of public sector information**, better known as Open Data or PSI Directive, stipulates that “Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available (‘open access policies’), following the principle of ‘open by default’ and compatible with the FAIR principles”.¹³ The directive will be implemented into national law by July 2021.
- **Funding organisations** are increasingly requiring data management planning compatible with the FAIR principles and conducting concerted efforts to harmonise DMP requirements in line with FAIR.¹⁴ This is also evidenced by the intention of research funding organisations to include FAIR compliance in future review of their funding policies.¹⁵
- **Publishers** are also increasingly developing common requirements for data sharing alongside publications.¹⁶

¹¹ European Commission (2018). *Commission recommendation on access to and preservation of scientific information*, (EU)2018/790, 25 April. Retrieved 21 February 2020, from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018H0790>.

¹² European Parliament (2019). *Legislative resolution on the proposal for a regulation of the European Parliament and of the Council establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination (COM(2018)0435 – C8-0252/2018 – 2018/0224(COD))*, P8_TA(2019)0395, 17 April. Retrieved 21 February 2020, from: http://www.europarl.europa.eu/doceo/document/TA-8-2019-0395_EN.html.

¹³ European Parliament & Council of the European Union (2019). *Directive on open data and the re-use of public sector information (recast)*, (EU)2019/1024, 20 June. Retrieved 21 February 2020, from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024>.

¹⁴ Science Europe (2018). *Practical Guide to the International Alignment of Research Data Management*. Brussels: Science Europe. Retrieved 21 February 2020, from: https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf.

¹⁵ Fosci, M., Richens, E., & Johnson, R. (2019). *Insights into European research funder Open policies and practices*. Apeldoorn: SPARC Europe. Retrieved 21 February 2020, from: <http://doi.org/10.5281/zenodo.3401278>.

¹⁶ See e.g. Hrynaszkiewicz, I., Simons, N., Hussain, A., & Goudie, S. (2019). Developing a research data policy framework for all journals and publishers. *Preprint*. Retrieved 21 February 2020, from: <https://doi.org/10.6084/m9.figshare.8223365.v1>; McQuilton, P., et al. (2020). *FAIRsharing Collaboration with DataCite and Publishers: Data Repository Selection, Criteria That Matter*. Retrieved 21 February 2020, from:

Accordingly, there is a clear need for better training of researchers and students about Open Science and, more closely related to EOSC, to foster the development of research data management skills and competencies. Skills and competencies related to research data and Open Science also relate to broader discussions of digital/data skills and competences at institutional and European level.

Taken together, these drivers mean that RDM, in close alignment with the FAIR principles, will need to become more standard practice, akin to a whole cultural change in the way research is conducted. This has wide ranging implications for higher education institutions and requires institutional change beyond training and skills provision, e.g. in the form of institutional policies and support services.¹⁷ The institutional capacity to create an environment that facilitates “FAIR” research practice will require investment in the development and retention of skilled human resources.¹⁸

Within FAIRsFAIR, Work Package 7 “FAIR Data Science and Professionalisation” aims to develop resources and build communities that support the uptake of RDM and FAIR practices and elements within higher education curricula. It builds directly on the *TFiR* recommendations that “[d]ata skills, including an appropriate foundational level in data science and data stewardship, should be included in undergraduate and postgraduate training across disciplines”. Specifically, it seeks to:

- Develop a **Competence Framework for FAIR** to support professionalisation, recognition and uptake of FAIR data science and data stewardship
- Develop **model courses, curricula, and training** documents to support uptake in university curricula design
- Train and support universities in **using the Competence Framework and model curricula** to speed up adoption

In order to achieve these objectives, the present report aims to build a foundation for the identification of existing practices and needs of higher education institutions. While there is a relative abundance of studies and reports on research data management policies and support provision, the area of training provision within university curricula has, to the knowledge of the authors, received little attention.

<https://doi.org/10.17605/OSF.IO/N9QJ7> and STM Research Data (2020). *Share – Link – Cite. Introducing the STM 2020 Research Data Year*. Retrieved 21 February 2020, from: <https://www.stm-researchdata.org/>.

¹⁷ See e.g. European University Association (2017). *Towards Open Access to Research Data. Aims and recommendations for university leaders and National Rectors’ Conferences on Research Data Management and Text and Data Mining*. Brussels: EUA. Retrieved 21 February 2020, from: <https://eua.eu/resources/publications/414:towards-open-access-to-research-data.html> and League of European Research Universities (2013). *LERU Roadmap for research data*. Leuven: LERU. Retrieved 21 February 2020, from: <https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf>.

¹⁸ See e.g. Rat für Informationsinfrastrukturen (2019). *Digital competencies – urgently needed! Recommendations on career and training prospects for the scientific labour market*. Göttingen: RfII. Retrieved 21 February 2020, from: <http://www.rfii.de/download/digital-competencies-urgently-needed-october-2019/>.

Due to the involvement of the European University Association (EUA) and the means to disseminate to the more than 800 university members of the association, activities in this report moreover addressed HEI-related questions for other Work Packages (WP) of FAIRsFAIR. The development of the survey instrument and the analysis of results have been done in collaboration with WP3 *FAIR Data Policy and Practice* and WP6 *FAIR Competence Centre*.

The report is structured in the following sections:

Section 2 explains the methodology used for data collection.

Section 3 presents the findings of the survey and focus groups and is divided into six sub-sections corresponding with the topics covered in the survey.

- The types of respondents and the sample are explained in section 3.1;
- The types and implementation of research data support policies are addressed in section 3.2;
- Section 3.3 informs about support services for research data management;
- Section 3.4 reports how FAIR and RDM skills and competences are addressed at responding universities and their respective programmes and courses;
- Section 3.5 addresses a series of questions related to universities views on EOSC;
- Section 3.6 details the suggested action areas for FAIRsFAIR as identified by responding institutions.

Finally, Section 4 summarises the results and presents the key conclusions derived from the analysis. It is structured in three sub-section corresponding to the main aspects investigated in this deliverable.

- Section 4.1 summarises the main findings relating to RDM in higher education programmes;
- Section 4.2 reports on the results in the area of RDM policies and support services;
- Section 4.3 addresses the results in relation to universities and EOSC.

2. Methodology

In order to collect quantitative and qualitative information, two different methods were combined to gather the data that this report draws its conclusions from. First, an international survey of universities was run in order to collect quantifiable information and examples from universities. This survey was conducted between 19 September and 15 November 2019. In total, the survey gathered 90 responses from institutions in 24 European countries. Second, two ‘focus groups’ were organised to share, substantiate and frame the findings of the survey on the topics “Universities, research data management and the FAIR principles” and “Teaching (FAIR) Data Management and Stewardship”, respectively.

2.1 Survey and questionnaire

The questionnaire-based approach followed the practice of EUA in surveying member universities, as for instance in the [2017-2018 EUA Open Access Survey](#). Partners in WP7 developed a survey questionnaire throughout summer 2019 that addressed the different dimensions mentioned above. Other FAIRsFAIR WPs, notably WP3, were involved in the development of the questionnaire as well.

The questionnaire was structured around the following sections containing a total of 31 structured or open-ended questions. The full questionnaire is attached in Annex 1A and an overview of the content is provided in Table 1 below:

Heading	Content
1) General information	<ul style="list-style-type: none"> • Background and contact details of the respondent (Questions 1-4) • Profile of the institution (Questions 5-6)
2) Current research data management policies, support services and competence development	<ul style="list-style-type: none"> • Awareness of the FAIR principles (Question 7)
2a) Policies	<ul style="list-style-type: none"> • Existence, main elements and responsibility for research data management policies (Questions 8-11)
2b) Support Services	<ul style="list-style-type: none"> • Existence, main elements and responsibility for research data management support services (Questions 12-16) • Open-ended comments (Question 17)

2c) RDM competences of students	<ul style="list-style-type: none"> ● Existence of global approach or strategy to address data literacy and skills (Questions 18) ● Teaching of data-related competences at Bachelor, Master and Doctoral level (Questions 19-21) ● Use of existing resources or frameworks in the context of data-related competences (Questions 22-23) ● Open-ended comments (Question 24)
3) Thinking ahead: universities, research data competences and EOSC	
3a) Future competences	<ul style="list-style-type: none"> ● Perceived need to strengthen teaching of competences at Bachelor, Master and Doctoral level (Question 25) ● Open-ended comments (Question 26)
3b) Universities and EOSC	<ul style="list-style-type: none"> ● Awareness, perceived benefits and barriers and difficulties for institutions to engage with EOSC (Question 27-29)
3c) Supporting universities	<ul style="list-style-type: none"> ● Possible areas and priorities of FAIRsFAIR to support universities (Questions 30-31)

Table 1. Questionnaire structure

The survey was designed to collect responses that align with existing frameworks on data science and data stewardship competences, in particular the EDISON Data Science Framework (EDSF)¹⁹ and the EOSC FAIR4S framework²⁰ for research data management and stewardship (see Annex 1B). Several questions on the main elements of research data policies were adapted from previous EUA surveys to reflect the changing landscape of EOSC and FAIR.

Dissemination

The survey was launched on 19 September 2019 with an invitation to universities registered in the EUA contact database. Specifically, university rectors and vice-rectors responsible for research activities were targeted. Contacting leadership is usual practice at EUA, as it reduces multiple responses from the same institution – although usually, responding to the survey is assigned to a central service or unit responsible for the data collection across an institution. The initial invitation was sent to 1378 recipients on 19 September 2019. A second mailing to the same 1378 contacts was done on 30 October 2019.

Additional dissemination of the survey was carried out via EUA and FAIRsFAIR’s media channels, social media, websites, and newsletters as well as those of project partners. The survey was also sent to relevant contacts and mailing lists such as, e.g., at CODATA, The Guild of European

¹⁹ Retrieved 21 February 2020, from: <http://edison-project.net/>.

²⁰ Retrieved 21 February 2020, from: <https://www.eoscpilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities>.

Research-Intensive Universities, or Science Europe. Other EOSC-funded projects were informed about the data collection through the EOSC-5b Landscape Task Force.

2.2 Focus groups

As indicated in the FAIRsFAIR description of work, the survey was to be complemented by two 'focus groups' in order to collect qualitative insights and data from stakeholders.

The two focus groups were organised along the main themes of the survey, respectively:

- *Universities, Research Data Management and the FAIR Principles*, hosted by University Carlos III of Madrid, 30 October 2019, 26 participants (for the programme see Annex 2).
- *Teaching (FAIR) Data Management and Stewardship*, hosted by the University of Amsterdam, 19 November 2019, 25 participants (for the programme see Annex 3).

Both focus groups applied an interactive approach including presentations, plenary and panel discussions. The themes and discussion questions were developed by EUA in collaboration with the local host and FAIRsFAIR partners.

Main target groups of the two focus groups were, mainly, university leadership or staff with responsibility for either the development or implementation of RDM policies or roles in developing teaching and training for RDM competencies. While those were prioritised, it did not exclude participants with other relevant backgrounds. Participants stemmed from university libraries, information services, data management services, graduate schools and training programmes, research administration and infrastructure staff. Several national or regional RDM initiatives and programmes were also represented.

The implementation of the first focus group on the topic "Universities, Research Data Management and the FAIR Principles" at UC3M was coordinated with WP3. The second focus group with the topic "Teaching Data Management and the FAIR Data Principles" was organised with the intention to support the launch and implementation on Task 7.3 on the development of a FAIR data competence framework since a meeting of the EDISON user community took place on 20 November 2019.

The results of the focus group discussions have been introduced in the present report as additional input where relevant. Brief summaries of the focus groups were produced and published on the FAIRsFAIR website.²¹

2.3 Data availability

The quantitative data collected through the survey will be made available through Zenodo at <https://doi.org/zenodo.3629687>.

²¹ Retrieved 21 February 2020, from:

<https://www.fairsfair.eu/articles-publications/outcomes-fairsfair-focus-group-universidad-carlos-iii-de-madrid> and <https://www.fairsfair.eu/articles-publications/fairsfair-focus-group-university-amsterdam>.

3. Results

3.1 Sample description and awareness of FAIR

In total, the survey gathered 90 responses from institutions in 24 countries within Europe and one response each from an institution in the USA and India, presumably due to the global dissemination of the survey through channels such as CODATA. The distribution is shown in Figure 1. The United Kingdom (n=11), Spain (n=9), and the Netherlands (n=8) are the top three countries measured by the number of responding institutions. Germany, Italy, and Sweden count six responding institutions each. The pattern of responses is similar to that seen for the *FAIR Data Policy and Practice* survey in which the majority of responses came from countries that currently demonstrate a relatively high engagement with Open Science and the FAIR principles.²²

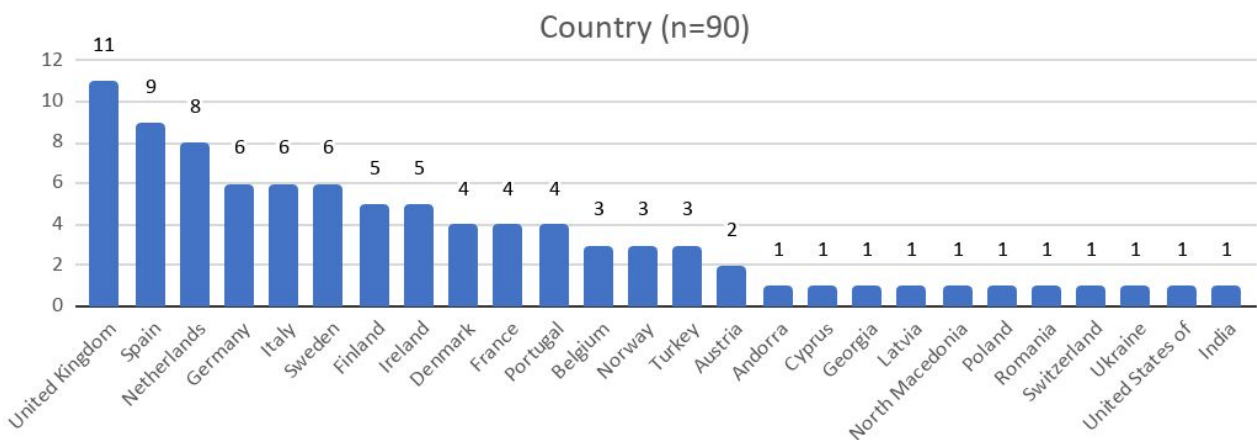


Figure 1. Countries of responding institutions

Respondents were asked to signal the type of higher education institutions, i.e. whether they represented a comprehensive university (i.e. covering all or most academic disciplines) or a specialised institution (e.g. medical science, music and arts schools). Results are reported in Figure 2 below. The majority of respondents (approx. 63%) indeed stem from comprehensive universities. Technical universities represent roughly 18% of the sample. Thirdly, specialised institutions comprise around 9% of the responding institutions. Only 7% of responding institutions are universities of applied sciences. Several open/distance learning institutions also completed the survey

²² Davidson, J., & Whyte, A. (2019). *FAIRsFAIR Policy and Practice Survey 2019 data for D3.1_D3.2_D6.1 (Version 1.0)* [Data set]. Retrieved 21 February 2020, from: <https://doi.org/10.5281/zenodo.3550529>.

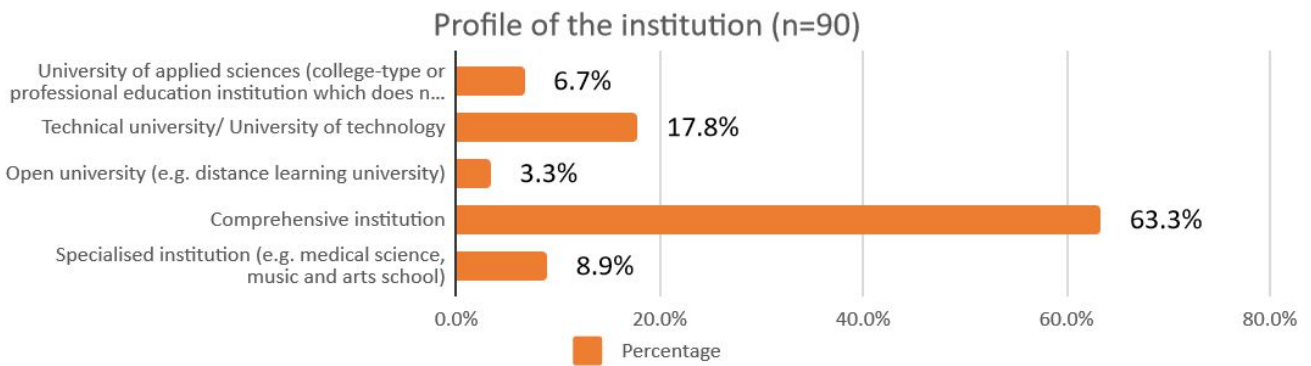


Figure 2. Profile of responding institutions

The majority (53%) of responding institutions can also be characterised as relatively large institutions with more than 1,000 researchers and doctoral candidates measured in full-time equivalents (see Figure 3 below). Universities with 500-1,000 and 100-499 researchers, respectively, each represent 21% to 20% of the sample.

What is the total number of researchers (FTE), including doctoral candidates, working at your institution? (n=90)

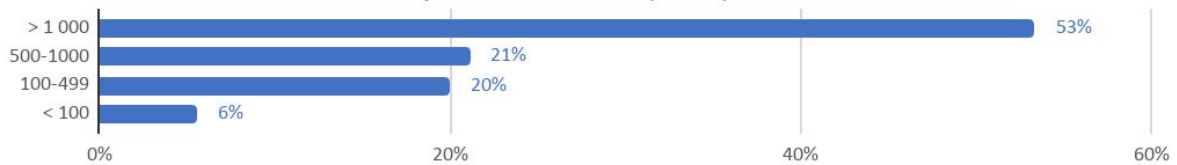


Figure 3. Number of researchers at responding institutions

Lastly, respondents were also asked to rate the awareness of the FAIR principles among different stakeholder groups within their institutions (Figure 4). Here, a clear majority answered that the awareness is highest among professional and support staff, as well as among the institutional leadership. Researchers and early-stage researchers are perceived to have lower awareness. Students are generally seen as least aware of the FAIR principles.

How would you rate the awareness of the FAIR principles within your institution?

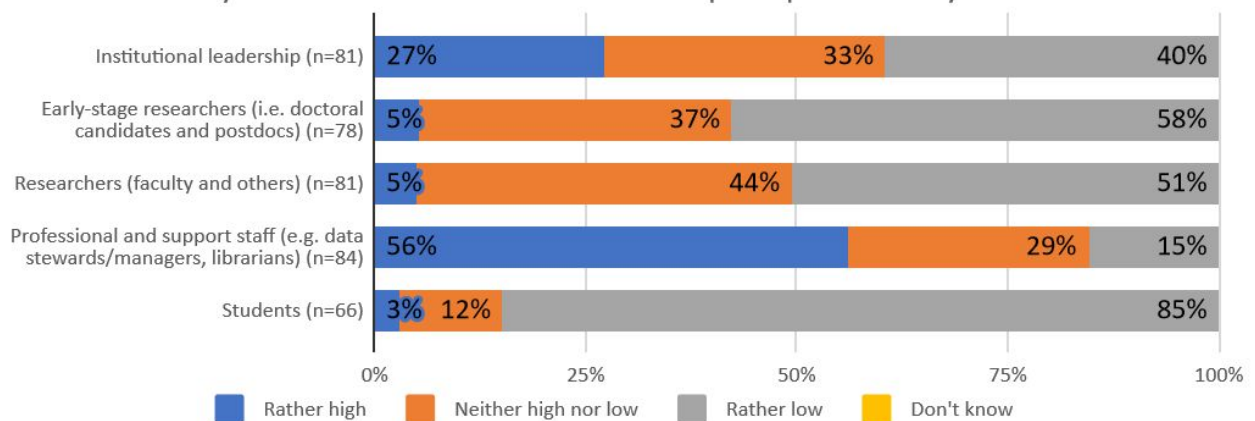


Figure 4. Awareness of FAIR principles within institution

The profile of responding institutions suggests preliminary conclusions. Firstly, there is a **need to increase outreach** across all EU member states and associated countries. In particular, EU13 countries are underrepresented in the sample. Moreover, as only a small number of responses was received from universities of applied sciences, there remains a need to identify their awareness and current levels of activity in relation to RDM and FAIR.

Finally, the awareness of researchers and early-stage researchers, and students of the FAIR principles is generally perceived as lower than that of professional and support staff, and institutional leadership. This highlights a **need for more awareness-raising among researchers** and their respective communities.

Input from Focus Group

The need for more awareness-raising has also been raised during the two focus groups. Participants noted, for instance, that Open Access publishing dominates awareness of Open Science practices and that **awareness does not automatically translate into practice**, e.g. due to constraints of researchers as a result of high workload. Several participants remarked that awareness is lower outside of the libraries and the research office, and specific disciplines with more existing standards and practices, such as life sciences. On the upside, it was also highlighted that **many researchers already follow certain aspects of FAIR** without being aware of it.

Focus group participants also warned that new terminology can add to confusion, including misunderstanding the relationship between open and FAIR data. Researchers, for instance, were seen to be more interested in issues of reproducibility and research integrity. **Linking FAIR data and RDM with references to reproducibility, research integrity and transparency of research** may be useful to raise awareness.

Participants emphasised that **focussing on early career researchers and training** would be a most important measure to raise awareness of the FAIR principles. Furthermore, any measure should be clearly supported and owned by the university leadership.

3.2 Policies and support services

A full section of the survey was devoted to the status of RDM policies and support services at the responding institutions. As depicted in Figure 5, the majority of HEIs respondents to this survey (60%) do have a policy on research data management, which is twice that reported in the EUA survey of 2017-2018 suggesting that those already engaged with research data are more likely to respond.²³ It should be noted however that the majority of responding institutions are based in countries currently deemed to be quite engaged with the Open Science agenda. Also, as noted above, more than half of survey respondents (53%) come from institutions that employ more than 1,000 researchers (including doctoral candidates) suggesting that the majority of survey respondents represent research-intensive organisations.

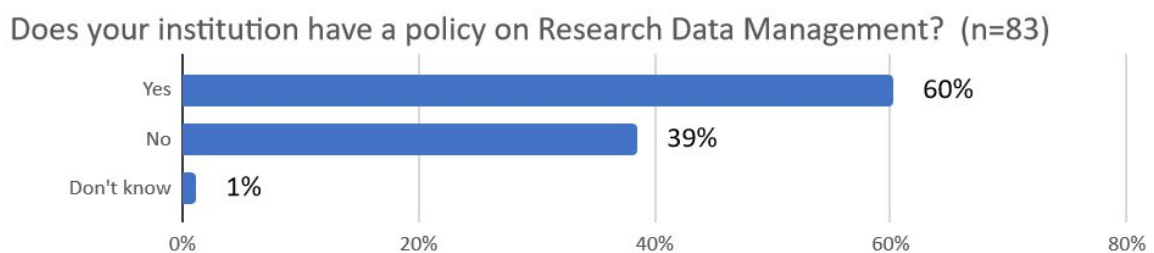


Figure 5. Existence of RDM policy

Policy development and provision of support

For those respondents who indicated that they do have a policy on RDM, the survey then aimed to identify which stakeholder groups were involved in *developing the policy*. A follow-up question asked who was responsible for the *implementation of support services* and policies.

Figure 6 compares which stakeholders were involved in policy *development* versus those involved in the *implementation* of policies and support. Figure 6 shows that the majority of respondents indicated that academic leaders/management and research administration were primarily involved in developing RDM policies (as indicated by 68% and 60% of respondents respectively). Just under half of those responding to this question stated that IT was primarily involved in developing the policy (46%). Around a third of respondents stated that other professional support roles were involved in developing the RDM policy including Legal (36%), Library (30%), and Data Protection (28%). Among the five respondents selecting 'other' roles being involved in developing the RDM policy, data management/data stewards were specifically mentioned three times.

²³ The 2017-2018 EUA Open Access Survey Results found that, out of 319 responding universities, 21% already had a policy and 38% were in the process of developing one. While the findings of the given report indicate an increase of the share of universities with RDM policies, this result may also be attributed to the different samples of each survey.

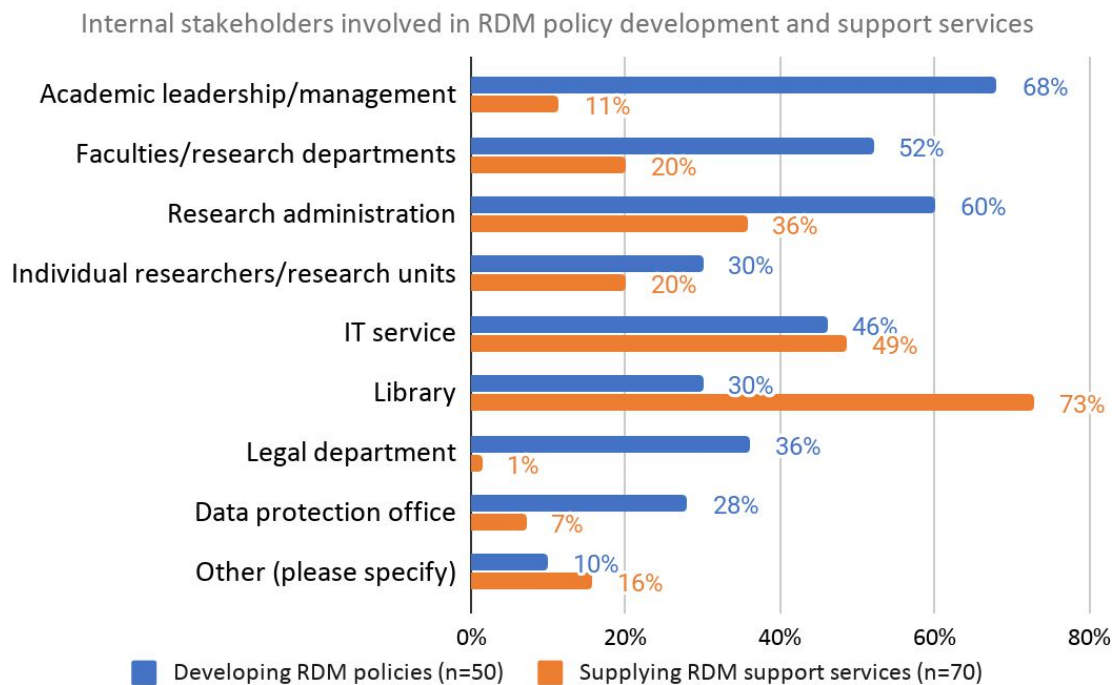


Figure 6. Stakeholder involvement in developing RDM policies and providing support

As noted above, professional and support staff were deemed those with the highest level of awareness about the FAIR principles, followed by institutional leadership at a significant distance. Figure 6 shows that while the research administration unit is heavily involved with developing policies, other professional support roles (Library, Legal, Data Protection) appear to be less involved. Given their expertise in several areas relevant to FAIR (e.g., data repositories, licensing, and access restrictions), it is advisable to involve a broader range of support staff in developing RDM policies.

As can also be seen in Figure 6, 73% of respondents indicate that the Library is primarily involved in *delivering RDM support services* for the institution compared with only 30% stating that the Library was *involved in the development of the policy*. As the key support provider, Libraries should be more actively involved in the development of RDM policies to ensure that supporting adherence with the policy is feasible with the financial and human resources available.

As noted above, just over a third of respondents to the survey question about support service provision indicated that Research Administration were involved (36%). As the policies of funding bodies have a key influence on researchers' behaviour,²⁴ research administrators may be the first point of contact for researchers looking for advice when preparing new grant applications. It is therefore advisable that research administrators are aware of what funders expect with regard to RDM and are able to either provide guidance directly or to signpost researchers to relevant local sources of guidance and support (e.g., Library), or to external guidance and support services.

²⁴ Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., & Whyte, A. (2019). *D3.1 FAIR Policy Landscape Analysis (Version v1.0_draft)*. Retrieved 21 February 2020, from: <https://zenodo.org/record/3558173>.

Faculties and research departments were identified as being primarily involved in developing RDM policies by more than half of respondents to the question on policy development (52%). This is good practice and helps to identify any disciplinary challenges as well as opportunities to leverage local support. However, when answering the subsequent question about support service provision, only 20% of respondents indicated that faculties and research departments were involved in providing such services.

Just under one third of respondents indicated that individual researchers were involved in developing the RDM policy. As researchers will be those tasked to comply with the policy, it is recommended that a cohort of researchers – ideally representing different domains and career stages – be actively involved in drafting and testing policies for feasibility and to highlight where additional support may be needed to support compliance. The survey results indicate that the Data Protection and Legal teams are currently under-utilised in providing support with less than 10% of respondents stating that they were involved in (7% and 1% respectively).

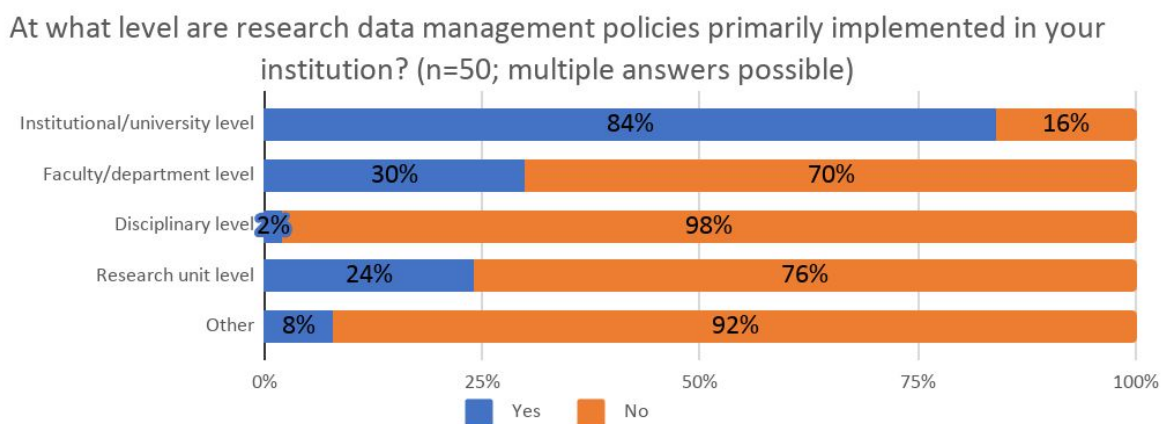


Figure 7. Level of implementation of research data policy

The survey next asked respondents to describe the level at which the RDM policy is implemented. Most respondents indicated that implementation occurs at the institutional/university level (84%) followed by the faculty/department (30%) and research unit level (24%) (Figure 7). Considering faculties/research departments' relatively high level of involvement in developing the policies (compare Figure 6 above), it may be advisable to consider involving this stakeholder group more actively when implementing the RDM policy, especially considering the preference for distributed/embedded data stewards. Disciplinary level policy implementation was very rare (2%).

The design and implementation of institutional research data policies was discussed during the Research Data Policies and the FAIR data principles focus group held at *University Carlos III de Madrid* in October 2019. There was broad agreement among focus group participants that the effective implementation of RDM policies requires a **collaborative approach with all relevant operational units** working together to develop and provide support. There was also general agreement that policy implementation might include a pilot phase where institutional processes and operational workflows are defined, tested and reviewed to ensure feasibility prior to policy ratification and full implementation.

Participants at the focus group also discussed **practical challenges** encountered during the implementation of their institutional policies. The key barriers to successful implementation were

deemed to be a lack of time on the part of researchers for RDM and the great disparity among researchers' current levels of awareness about the FAIR principles – either very high or very low and not much in between. This disparity can make developing appropriate support and guidance a challenge.

In addition, participants highlighted that researchers' attitudes towards data sharing tend to be discipline-specific, suggesting that bottom-up actions and support are important. The **role of faculty or department level data stewards** was seen as potentially transformative in overcoming this barrier. Accordingly, there is potential for faculties/departments to fill the gap between generic and domain-specific guidance and support at the local level.²⁵ Indeed, some HEIs are introducing discipline-specific support such as TU Delft where each faculty has a “dedicated Data Steward to answer questions, provide advice, and help develop appropriate solutions for research data management and sharing”.²⁶ However, this approach can be resource-intensive and, as such, may not be feasible for all HEIs. Work such as that being piloted by the Swedish National Data Service²⁷ to provide pooled discipline-specific support could be useful models for other countries. In addition, there remains a need to provide HEIs with access to a shared set of domain-specific guidance and resources to better support RDM practice.

Policy content

For those respondents who stated that they had a policy on RDM, the survey aimed to identify the main elements addressed by these policies. Respondents were asked to indicate whether a number of elements were covered by the institutional RDM policies and, where relevant, whether these elements are mandatory or optional (Figure 8). Most HEI RDM policies already include provisions on Open Access, data management planning, and data storage – either as optional or mandatory elements (96%, 96% and 94% respectively). The elements most often included in the RDM policy as mandatory aspects were data storage (55%), data management planning (54%), and data protection (47%).

Provision for FAIR is specifically covered in just over half of the policies of respondents (27% as a mandatory element and 33% as an optional element). For reference, more than three quarters of researchers responding to the *State of Open Data 2019* survey had never heard of the FAIR principles (54%) or had heard of but were not familiar with them (28%).²⁸ Acknowledging that FAIR as a concept is still rather unknown to researchers, it may indeed be sensible for HEIs to avoid requiring ‘full’ FAIR compliance and, at this stage, encourage and support researchers to make data as FAIR as possible. This could be achieved by providing practical guidance and support that references FAIR and indicates how the local policy and support aligns with funders' requirements for FAIR data.

²⁵ Research carried out by FAIRsFAIR WP3 also found that guidance at the institutional level tends to be generic and may not be sufficient to meet researchers' domain-specific challenges.

²⁶ Retrieved 21 February 2020, from:

<https://www.tudelft.nl/en/library/current-topics/research-data-management/r/data-stewardship/>.

²⁷ Retrieved 21 February 2020, from: <https://snd.gu.se/en/about-us>

²⁸ Digital Science, et al. (2019). *The State of Open Data Report 2019*. Retrieved 21 February 2020, from: <https://doi.org/10.6084/m9.figshare.9980783.v2>.

The findings of policy characterisation carried out as part of Work Package 3 revealed that the data policies of HEIs most often suggest data sharing rather than require it.²⁹ Moreover, an open consultation carried out in Work Packages 3 and 6 found that the policies of funding bodies have the greatest influence over researchers' behaviour.³⁰ As such, HEIs should seek to align with funders' requirements in order to support compliance and to avoid ambiguity or conflicts between institutional and funder requirements. It is worth emphasising that, supporting researchers to comply with external mandates is an important course of action even for institutions without their own RDM policies in place.

What are the main elements of your institutional research data management policy?

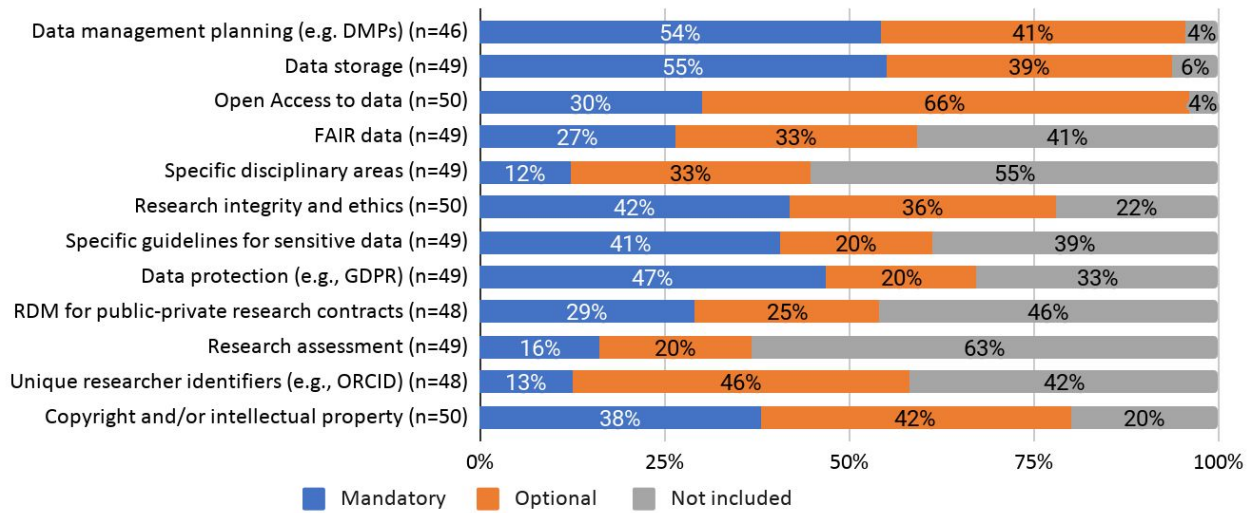


Figure 8. Main elements of RDM policies

The elements least covered in the RDM policies include research assessment and specific disciplinary areas (63% and 55%, respectively, do not address them). As career incentives and rewards are a key motivation for researchers, it may be advisable for HEIs to consider how their RDM policy might be integrated with their approach to career assessment and promotion. Indeed, this is part of a broader discussion between HEIs, research funding organisations and other actors on how to expand the range of academic activities being incentivised and rewarded. For example, there has been a marked increase in the number of HEIs signing up to the Declaration on Research Assessment (DORA) declaration³¹ in recent years, as well as institutions modelling their approach on the Leiden Manifesto³² and other community-generated initiatives focused on more responsible research evaluation practices.

²⁹ Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., & Whyte, A. (2019). *D3.1 FAIR Policy Landscape Analysis (Version v1.0_draft)*. Retrieved 21 February 2020, from: <https://zenodo.org/record/3558173>.

³⁰ Davidson, J., & Whyte, A. (2019). *FAIRsFAIR Policy and Practice Survey 2019 data for D3.1_D3.2_D6.1 (Version 1.0)*. Retrieved 21 February 2020, from: <http://doi.org/10.5281/zenodo.3550529>.

³¹ Retrieved 21 February 2020, from: <https://sfedora.org/>.

³² Retrieved 21 February 2020, from: <http://www.leidenmanifesto.org/>.

As funding body policies have started to introduce expectations on HEIs to become signatories of DORA or equivalent (such as the Wellcome Trust³³), the number of HEIs becoming signatories will no doubt continue to increase. However, a recent EUA study demonstrated that publications remain key indicator for research assessment with less than half of responding HEIs considering other research outputs – such as data – to be important or very important³⁴. This suggests that additional efforts are needed to ensure that research assessments encompass a broader range of activities and outputs, including research data.³⁵

Types of support provided

Besides the existence of policies for RDM and FAIR data, the survey also inquired about the provision of support services for RDM. Both can occur separately, e.g. in cases in which support services are established but a policy has not yet been adopted. The vast majority of survey respondents have research data support services in place – regardless of whether their HEI has a policy on RDM or not (see Figure 9).³⁶

Does your institution currently have any research data support services in place? (n=82)

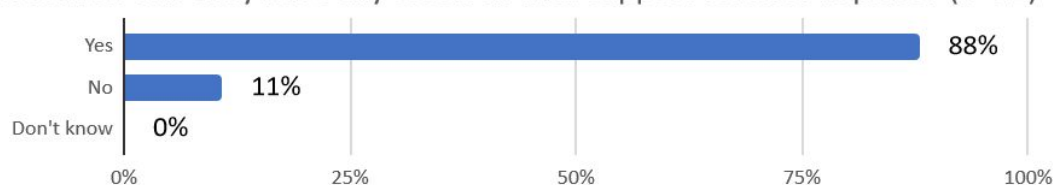


Figure 9. Research data support services are provided.

The share of institutions with support services largely mirror findings an open consultation carried out as part of WP3 and WP6 which revealed that the vast majority (77% of HEIs responding to the question) provide in-house support for FAIR to some degree. The *2017-2018 EUA Open Access Survey Results* confirm this distribution, with e.g. approximately 80% of universities providing support through the library or another department.

For those respondents indicating that they have research data support services in place, the survey next aimed to identify what types of support are currently being made available to researchers interested in making their data FAIR. As depicted in Figure 10, more than three quarters of respondents to this question indicated that they provide training for researchers (80%) and have a dedicated section of the institutional website for RDM (80%). These findings indicate a positive trend compared to the sample of the *2017-2018 EUA Open Access Survey*, in which only around 50% of universities were providing training and only around 40% maintained institutional webpages on RDM.

³³ Retrieved 21 February 2020, from:

<https://wellcome.ac.uk/news/wellcome-updates-open-access-policy-align-coalition-s>.

³⁴ European University Association (2019). *2019 Open Science survey results on research assessment*. Retrieved 21 February 2020, from: <http://doi.org/10.5281/zenodo.3435325>.

³⁵ It should be noted that research assessment may not necessarily be addressed in an institution’s research data policy, but in separate documents and guidelines.

³⁶ Support services can predate the formalisation of data policies. In cases where there is no formal institutional RDM policy (yet), HEIs may still aim to provide support to researchers for adhering with funders’, publishers’ or legal requirements.

When it comes to specific aspects of FAIR practice, about two thirds of the HEI respondents to this question stated that they have provisions for supporting compliance with legal and ethical requirements and the FAIR principles (64%) as well as for publishing FAIR outputs on their own or recommended repositories (59%). In addition, about two thirds of HEIs responding to this question support their researchers in developing an open research strategy and vision (59%).

What type of support does your institution provide to researchers interested in making research data FAIR? (n=69)

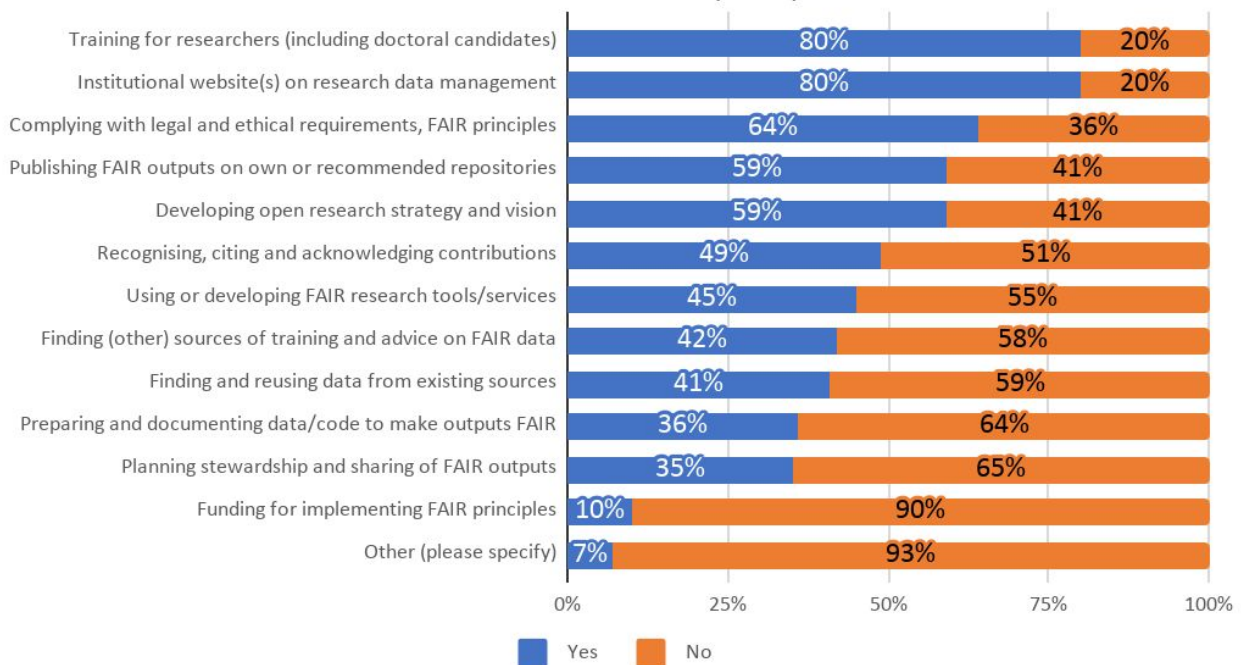


Figure 10. Types of support provided for making data FAIR

About half of the HEIs responding to the question indicated that they provide support for recognising, citing and acknowledging contributions (49%) and for using or developing FAIR research tools/services (45%). Finding sources of training and advice on FAIR data was cited as a service offering by slightly fewer of those HEIs responding to the question (42%) as was finding and reusing data from existing sources (41%). The latter suggests that there is a gap in support provision that needs to be filled if we are to realise a FAIR data culture where FAIR data is reused.

When it comes to supporting the preparation and documentation of data/code to make these outputs FAIR, just over one third of HEIs responding to this question stated that they currently have support provisions in place (36%). A similar number indicated that they currently provide support for planning stewardship and sharing of FAIR outputs (35%). These findings suggest that there is a need for improved support provision to realise increased production and stewardship of FAIR data outputs. Participants of the RDM policies and the FAIR principles focus group also highlighted that there is a need for support specifically relating to the interoperability and reusability of data.

Among HEIs responding to this question, there were very few that provide funding for implementing FAIR principles (10%). This reflects the findings of policy characterisation carried out in Work Package 3 which found that only one out of the 11 HEI policies reviewed clearly stated that costs

associated with RDM and FAIR data are supported by the institution when justified. As suggested in the recent FAIR Data Landscape Analysis³⁷, establishing a balanced approach to supporting the costs associated with making and keeping data FAIR over time is something that requires cooperation between HEIs, funding bodies and publishers.

There were five respondents who provided ‘other’ types of support that are being developed. One responding institution is developing a Digital Badge in the Responsible Conduct of Research, a significant portion of which is dedicated to FAIR data management. Providing badges for successful completion of training is a good way to incentivise participation especially if the badges are used in some way to inform annual performance reviews and/or career assessment. Another HEI is developing its training around the creation of a DMP. Such practical and workflow-oriented training is a good idea and can lead to better learning outcomes.

Professionalisation of support

Has your institution established specific research data support roles (e.g. data stewards, research data managers)? (n=70)

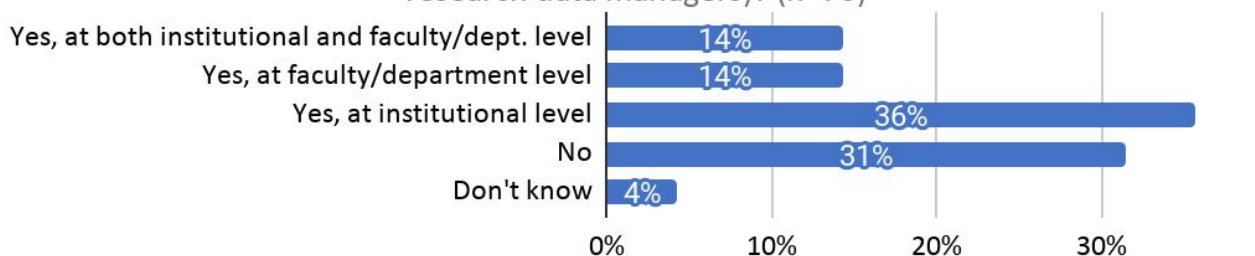


Figure 11. Research data support roles established by institutions

The final question of this section asked whether HEIs have established specific research data support roles, and if so, at what level these roles operated. Respondents were able to select multiple options. As can be seen in Figure 11, the majority of respondents to this question indicated that they have introduced specific data management support roles at the institutional level (36%). A smaller number have established such roles at the faculty/department level (14%). In a small number of cases, there is support at both institutional and faculty/department levels (14%). Lastly, at 31% of responding institutions, specific RDM support roles have yet to be established.

Further inputs from focus groups

Participants of the focus group highlighted that **coordination and assignment of responsibilities** among the many professional support services and operational units is crucial. This includes IT services, legal departments, technology transfer offices, research support and administration, libraries, research integrity and ethics offices, as well as data protection officers. Given the broad range of services involved, participants felt that a centralised approach to dissemination and communication would be most effective to ensure that key issues are amplified and to avoid sending mixed messages to researchers.

In addition, participants emphasised the **benefits of employing decentralised data stewards or data managers** who provide disciplinary knowledge as essential for close collaboration with

³⁷ Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., & Whyte, A. (2019). *D3.1 FAIR Policy Landscape Analysis (Version v1.0_draft)*. Retrieved 21 February 2020, from: <https://zenodo.org/record/3558173>.

researchers. These may be located within a specific department/faculty or research groups (as large and well-funded research groups often have a dedicated data manager).

A challenge regularly encountered is the **funding for data management positions**, as staff are often hired through externally funded research projects. Focus group participants stressed that the expertise associated with project-funded support roles can be lost when the project ends and the funding runs out. As such, participants felt that there is a need to define funding strategies to sustain such positions beyond the life of the project. Focus group participants also emphasised the need for skilled IT staff to work closely with research groups (i.e., software engineers).

Finally, establishing RDM support services implies a **reformulation of tasks and resource allocation within specific professional support services** of the university. Participants emphasised that, in many institutions, RDM can be considered a new responsibility. Existing staff may need to be trained and new staff may need to be hired – both of which will require adequate resource allocation in terms of time and money. New career paths will need to be defined with the increasing need for professional data stewards.

All these factors imply a need for institutional strategies for securing, embedding and sustaining RDM skills through new hires, changes to existing roles, re-skilling, professional development, and training. Focus group participants also emphasised that both intra-institutional and inter-institutional collaborative networks were useful to support the development and implementation of RDM policies and support services.

3.4 Competences and programmes

Supporting skills and competence development within higher education programmes is a main objective of FAIRsFAIR WP7. The survey and focus groups were specifically designed to assess the state of skills and competence provision within universities as well as the perceived needs to foster the teaching of RDM-related skills within higher education.

First, the survey asked whether the institution has an approach or strategy to address data literacy and RDM skills for students. These results are reported in Figure 12. In total, 38% of the respondents who replied to this question stated that their institution had such an approach. This is divided between 8% where strategies are in place on institutional, faculty or department levels. Another 21% indicated that one existed on the institutional level. 9% reported one at faculty or department level. 31% reported one at faculty or department level.

Does your institution have an approach or a strategy to address data literacy and skills of students?

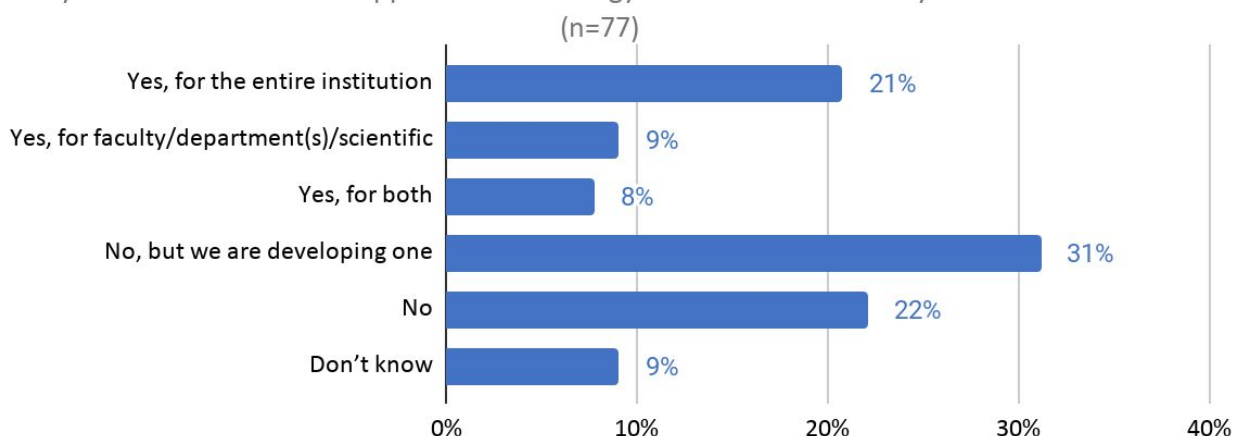


Figure 12. Institutional strategy for data literacy and skills

Overall, 53% of respondents stated that their organisation did not have such a strategy. While this makes the proportion of institutions without a relevant strategy considerably larger than the one of those who do have one, there is an approach or strategy in development at 31% of institutions, so this ratio is likely to change in the future. As a result, with 69% institutions either already having or currently developing an approach to tackle data literacy and skills, the responses indicate that the majority of HEIs are increasing their efforts in this regard.

Coverage of broad competencies by discipline and educational level

The next question was designed to capture the extent to which data-related skills and RDM skills are already being addressed by responding universities. For this purpose, respondents were first shown the broad skills groups defined by EDSF, namely “Data analytics and statistical techniques”, “Data software engineering”, “Data Management” and “Applying general data science methods”.

Subsequently, respondents were asked to indicate whether *any one of those options* were addressed at each the bachelor, master and doctoral level in five broad sets of disciplines

(Humanities and Social Sciences, Life Sciences, Natural Sciences, Engineering Sciences and Generic/multidisciplinary).³⁸ A five-tiered scale from “never” to “always” was used (see Annex 1A).

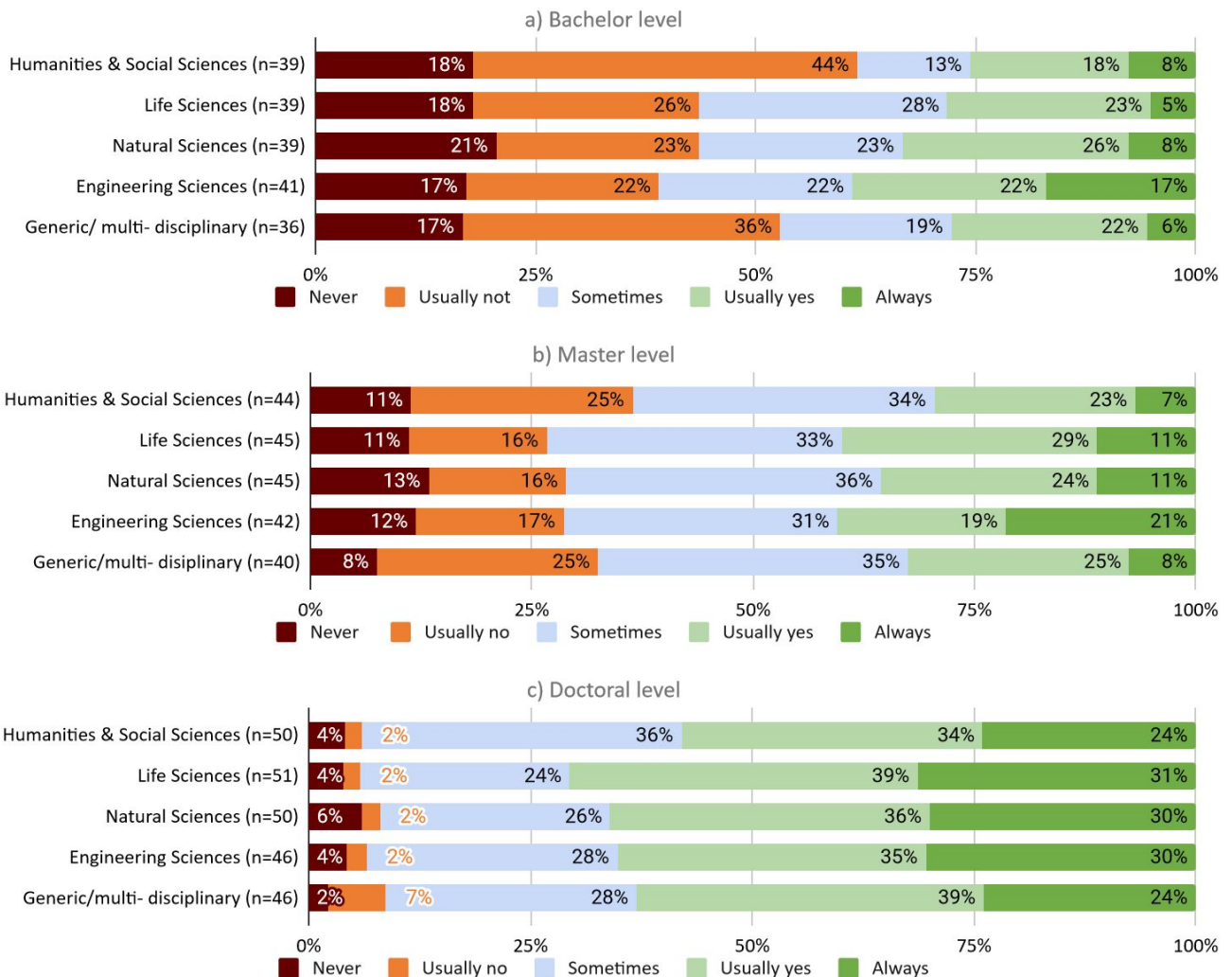


Figure 13a-c. Are data science and management skills addressed? (per educational level and domain)

The results are displayed in Figures 13a-c. The absolute number of responses ranged, depending on the discipline, from 36 to 41 for the bachelor, from 40 to 45 for the master and from 46 to 51 for the doctoral level. There was also a considerable number of responses for the options “Don’t know” and ‘Not applicable” which are not included in the charts but are reported in Table 2.

The results show that research data-related competences are being addressed with increasing frequency from the bachelor to the doctoral level. At the bachelor level (Fig. 13a), the number of HEIs teaching these skills “usually not” or “never” is slightly higher than that of those which “usually” or “always” do – except for the Engineering Sciences where the numbers are even with 39% each. The lowest rate of inclusion of RDM competences was reported for the humanities and social

³⁸ The exact wording of the question was: “To your best knowledge, in which broad set of disciplines and at what level is your institution addressing one or more of the above-mentioned research data-related competences.”

sciences – with 26% of institutions responding “usually yes” or “always” vs. 62% responding “usually not” or never” – and at the generic or multidisciplinary level with 28% positive and 53% negative answers.

At the master level (Fig. 13b), the relation of HEIs that are “usually not” or “never” addressing relevant skills versus those that “usually” or “always” do is shifting slightly towards the latter for the life sciences (27% vs. 40%), the natural sciences (29% vs. 35%) and the engineering sciences (29% vs. 40% resp.). The relation remains even for the generic level (33% vs. 33%) and remains tilted a little towards the negative side for the social sciences and humanities (36% vs. 30%). A rather large number of respondents indicated that their institution “sometimes” addressed research data-related skills, reporting between 31% and 36% depending on the subject area.

At the doctoral level (Fig. 13c), only 6%-9% of HEIs, depending on the domain, reported to “never” or “usually not” include research data-related content in teaching, while 24%-31% of respondents indicated that they “always” did. In addition, 34%-39% responded that such topics were “usually” addressed in doctoral teaching in all disciplinary areas. Together, the latter two add up to proportions between 58% (for the SSH) and 70% (for the life sciences). Furthermore, the rates of respondents who said RDM competences were sometimes part of teaching were also quite high, ranging from 24% for the life sciences to 36% for the social sciences and humanities.

		Never	Usually no	Sometimes	Usually yes	Always	n (valid)	Don't know	N/A	n
Bachelor	Humanities & Social Sciences	7	17	5	7	3	39	11	7	57
	Life Sciences	7	10	11	9	2	39	12	5	56
	Natural Sciences	8	9	9	10	3	39	12	6	57
	Engineering Sciences	7	9	9	9	7	41	8	7	56
	Generic/ multi- disciplinary	6	13	7	8	2	36	13	5	54
Master	Humanities & Social Sciences	5	11	15	10	3	44	8	6	58
	Life Sciences	5	7	15	13	5	45	9	4	58
	Natural Sciences	6	7	16	11	5	45	8	4	57
	Engineering Sciences	5	7	13	8	9	42	6	8	56
	Generic/multi- disciplinary	3	10	14	10	3	40	11	4	55
Doctoral	Humanities & Social Sciences	2	1	18	17	12	50	7	6	63
	Life Sciences	2	1	12	20	16	51	8	4	63
	Natural Sciences	3	1	13	18	15	50	8	4	62
	Engineering Sciences	2	1	13	16	14	46	7	8	61
	Generic/multi- disciplinary	1	3	13	18	11	46	10	4	60
		Never	Usually no	Sometimes	Usually yes	Always	n (valid)	Don't know	N/A	n

Table 2. Are data science and management skills addressed? (absolute figures)

Coverage of specific competences per educational level

While the initial questions of the survey aimed to develop a ‘broad’ picture of data-related skills at different levels and disciplines, the survey also aimed to obtain data about more specific data-related and research data management skills and competences. Respondents were presented with a list of nine specific research data-related topics and asked to indicate to what extent those were currently being delivered at their institution at bachelor, master and doctoral level, again on a

scale from “never” to “always”.³⁹ The results are illustrated in detail in Figures 14a-c.⁴⁰ There was also a considerable number of responses for the options “Don’t know” and “Not applicable” which are not included in the charts but are reported in Annex 4.

In general, the results show that , currently, teaching of the topics in questions is lowest at the bachelor and highest at the doctoral level, which seems to be in accordance with a similar tendency in the findings of the previous question.

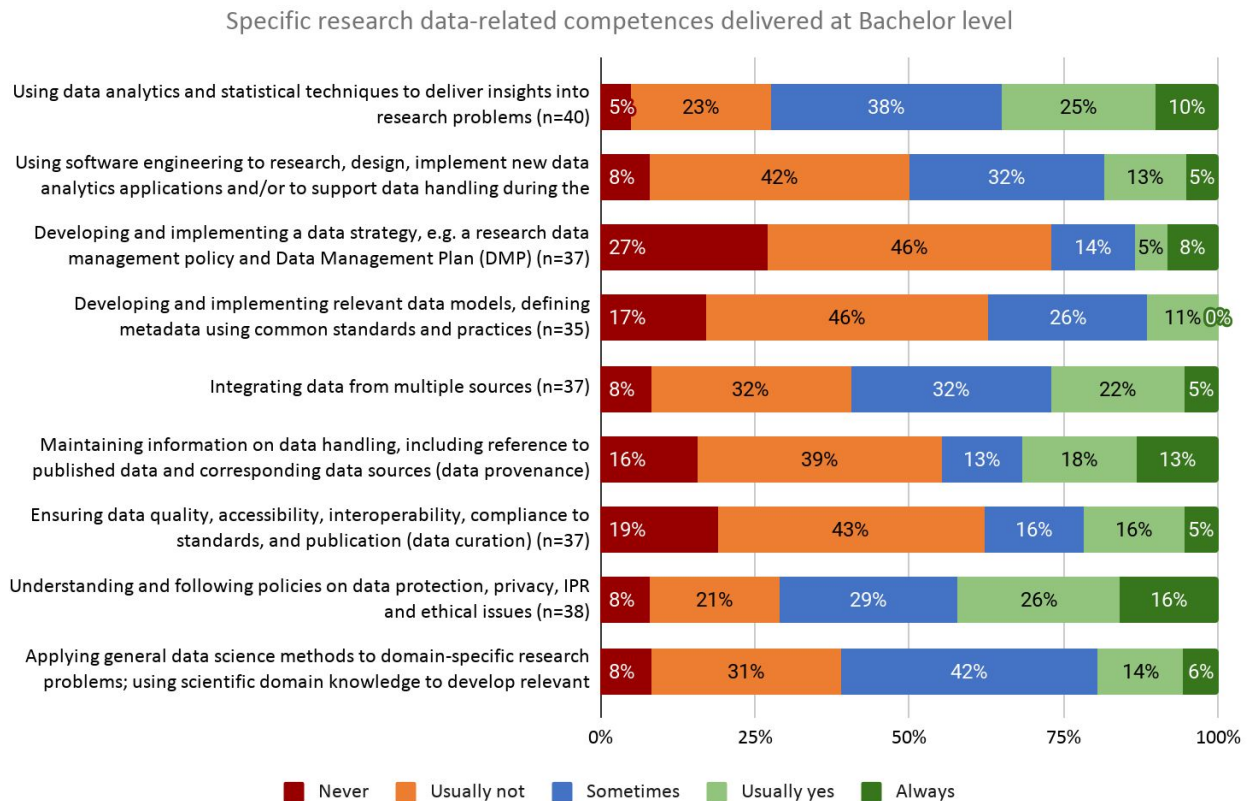


Figure 14a. Specific competences at Bachelor level

As seen in Figure 14a, none of the topics achieved very high results for the options “always” (0%-16%) and “usually yes” (5%–26%) at Bachelor level. The two competences rated highest in this respect are “Understanding and following policies on data protection, privacy, IPR and ethical issues” and “Using data analytics and statistical techniques to deliver insights into research problems”. In all cases except for those two, the options “never” (5%-27%) and “usually not” (21%-46%) were selected more frequently than their counterparts “always” and “usually yes”.

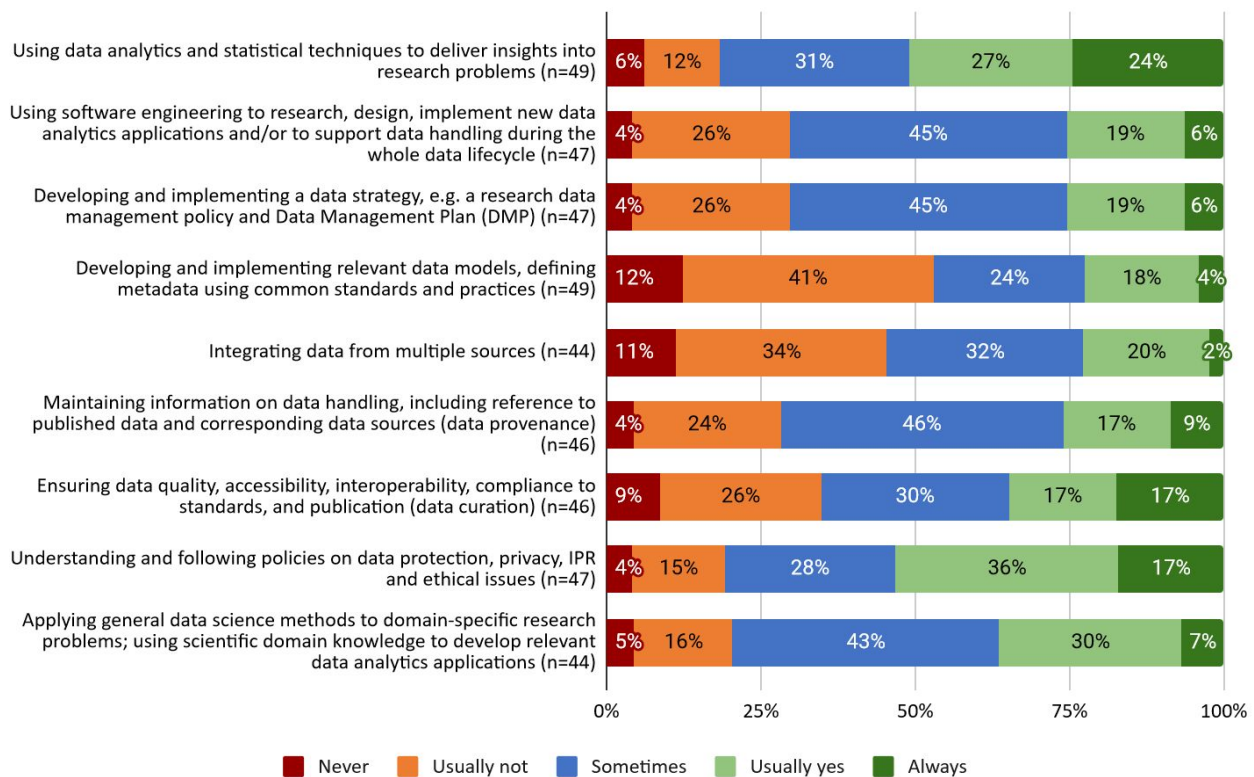
At master level (see Figure 14b), “always” (2%-24%) and “usually yes” (17%-36%) received slightly more, “never” (4%-12%) and “usually not” (12%-41%) on the whole less responses compared to bachelor level.

³⁹ The exact wording of the question was: “To your best knowledge, what are the more specific research data-related competences delivered by your institution at bachelor, master, and doctoral level?”

⁴⁰ Please note that the charts exclude “don’t know” and “not applicable” responses. For the bachelor level, between 11 and 13 respondents (depending on the individual item) said they could not answer this question. The respective numbers for the master level are 7 to 10, for the doctoral level 3 to 8.

Nevertheless, there are still only three skills groups with a higher score for “always” and “usually yes” combined than for the sum of “never” and “usually not”: “Understanding and following policies on data protection, privacy, IPR and ethical issues” (53%), “Using data analytics and statistical techniques to deliver insight into research problems” (51%) and “Applying general data science methods to domain-specific research problems; using scientific domain knowledge to develop relevant data analytics applications” (37%). Those three are also the most popular topics in our sample at master level. There is an almost even number of positive and negative responses in terms of “Ensuring data quality, accessibility, interoperability, compliance to standards and publication (data curation)” (33%-34%).

Specific research data-related competences delivered at Master level


Figure 14b. Specific competencies at Master level

Other topics are mostly not part of RDM-related competencies at the HEI respondents. Least likely to be addressed at master level are “Integrating data from multiple sources” and “Developing and implementing relevant data models, defining metadata using common standards and practices”. Between 24% and 46% of respondents selected “sometimes”, depending on the item.

While there are rather gradual differences between bachelor and master level, the picture changes when looking at the results for the doctoral level, which are illustrated in Figure 14c. The numbers of responses indicating that a particular competence was “never” or “usually not” addressed in teaching drop to a very low level of 0%-6% and 2%-15%, respectively. The proportions of responses stating that related content was “always” or “usually” delivered in this context are higher, ranging from 7%-33% and 27%-43% respectively. The two most popular are “Understanding and following policies on data protection, privacy, IPR and ethical issues” (“always” and “usually yes” at 73% of

responses) and “Using data analytics and statistical techniques to deliver insights into research problems” (“always” and “usually yes” at 71% of responses).

With regard to the remaining skills groups, percentages for “always” and “usually yes” added together lie between 38% and 52%, the – rather large – proportions of responses for “sometimes” between 36% and 48%. None of the topics can be clearly identified as the least popular on doctoral level. However, it is noteworthy in the context of the FAIR principles that “Ensuring data quality, accessibility, interoperability, compliance to standards, and publications (data curation)” does receive almost the lowest attention in current education at doctoral level.

Specific research data-related competences delivered at doctoral level

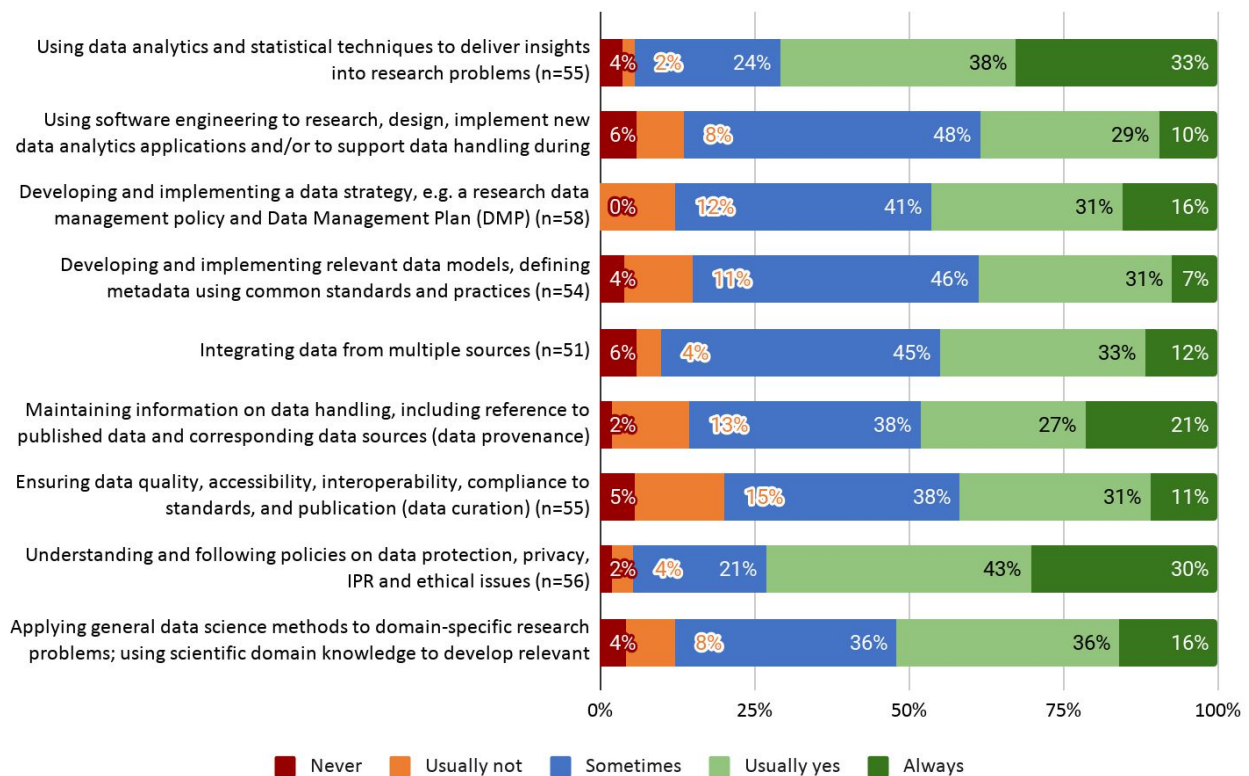


Figure 14c. Specific competences at doctoral level

An open-ended question in the survey collected additional details of specific courses and programmes that address the competences above. Respondents were asked “Can you provide an example of a curriculum or course that includes one or more of the data-related competences mentioned above?”. Overall, 41 institutions answered this question. From further analysis we can conclude:

- At least 11 institutions mentioned research data management training at the doctoral level, usually through research training and methods courses.
- 13 universities entered details about courses at master level; 5 at bachelor level.
- 9 universities moreover responded that they offered more generic courses such as trainings provided by the university library, Software Carpentry⁴¹ and training courses on data science.

⁴¹ Retrieved 21 February 2020, from: <https://software-carpentry.org/>.

In terms of areas or disciplines mentioned, several topics could be identified:

- a high number of programmes is related to data science, business analytics and information and communication systems (8x);
- one programme addresses cyber security;
- two programmes or courses relate to medical science;
- two programmes or courses are in the area of biology or biostatistics;
- two programmes focus on the 'data librarian';
- three courses address information literacy or competency and open science.

Use of existing frameworks, resources and tools

The survey also asked respondents if their institutions were using existing frameworks for research data management, data science or open science when developing training activities or curricula. A number of international frameworks were provided as answer options.

Results displayed in Figure 15 show that most institutions do not make use of existing frameworks (49%) or instead make use of (national) frameworks and guidelines not listed in the answer options (40%). Regarding the latter, several respondents listed specific examples in the comments, including the Vitae Researcher Development Framework (RDF), the Irish National Framework on the Transition to an Open Research Environment (NORF), the Swedish National Data Service (SND), etc.

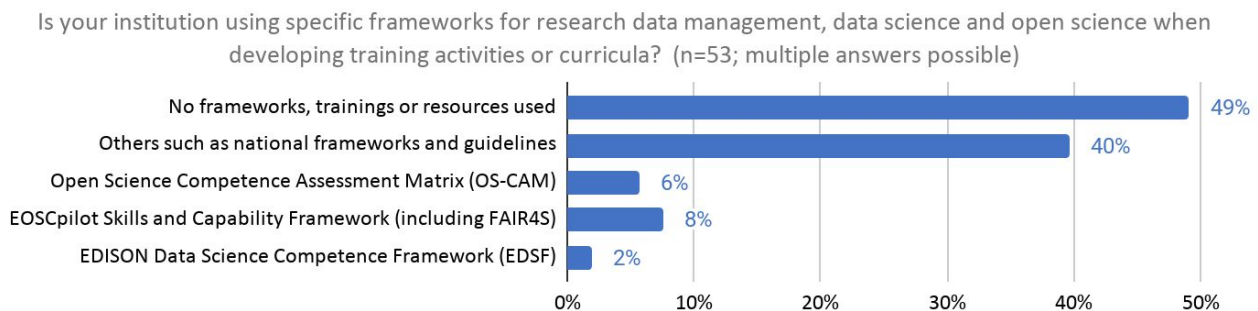


Figure 15. Use of frameworks for RDM, data science and open science for training or curricula development

In comparison, only a small minority makes use of the existing frameworks that were listed as answer options. This is the case for the EOSCpilot Skills and Capability Frameworks (including FAIR4S) (8%), the Open Science Career Assessment Matrix (OS-CAM) (6%)⁴² and the EDISON Data Science Competence Framework (EDSF).

⁴² The framework was mistakenly labelled Open Science *Competence* Assessment Matrix in the survey. This original wording is retained in Figure 15.

Is your institution using resources, trainings or other services from other projects or initiatives? (n=64; multiple answers possible)

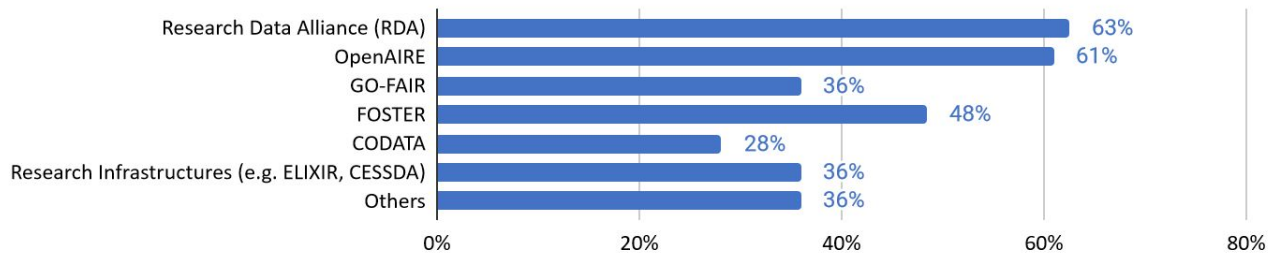


Figure 16. Use of existing resources, trainings or other services from other projects or initiatives

The survey then asked respondents if their institutions are using resources, trainings or other services from relevant projects or initiatives. Figure 16 shows that most institutions make use of RDA (63%) and OpenAIRE (61%) resources or trainings, followed closely by FOSTER (48%), GO-FAIR (36%), research infrastructures (e.g. ELIXIR, CESSDA) (36%) and CODATA (28%). In addition, 36% of respondents indicated that their institutions use other resources, training or other services not included as an answer option. These refer mostly to national resources.

Needs to strengthen teaching at different educational levels

After having collected information about the current state regarding the integration of the FAIR principles in teaching at the HEIs in the sample, the final question aimed to capture the respondents' assessment regarding future needs by asking *“Does your institution believe there is a need to strengthen the teaching of specific research data-related competences at the bachelor [master/doctoral] level?”* in four groups of competences: “Data analytics and statistical techniques”, “Data software engineering”, “Data Management” and “Applying general data science methods” derived from EDSF. Answers could be given on a three-tiered scale: low need; neither high nor low need; high need. The results are visualised in Figure 17a-c. All data, including responses for the options “Don’t know” and ‘Not applicable” are reported in Table 3.

The results indicate that the respondents see such a need at all three levels with increasing urgency towards the doctoral one. But even at the bachelor level (see Figure 17a) the percentage of institutions that see a high need for strengthening FAIR-related teaching comprises between 34% (regarding “Data software engineering”) and 54% (in terms of “Data analytics and statistical techniques”). Another 27% to 35% see a moderate need, leaving only a proportion of 19%-39% of responding institutions that merely see a low need.

At master level (see Figure 17b), the percentages of HEIs indicating that they see only a low need to strengthen teaching regarding the topics in questions shrink to 6%-14% while the ones stating a high need increase to 47%-66% and the ones indicating a neither high nor low need stay at about 26%-39%. The most urgent need to be represented to a greater extent in teaching at master level is expressed with regard to “Data analytics and statistical techniques” and “Data management”.

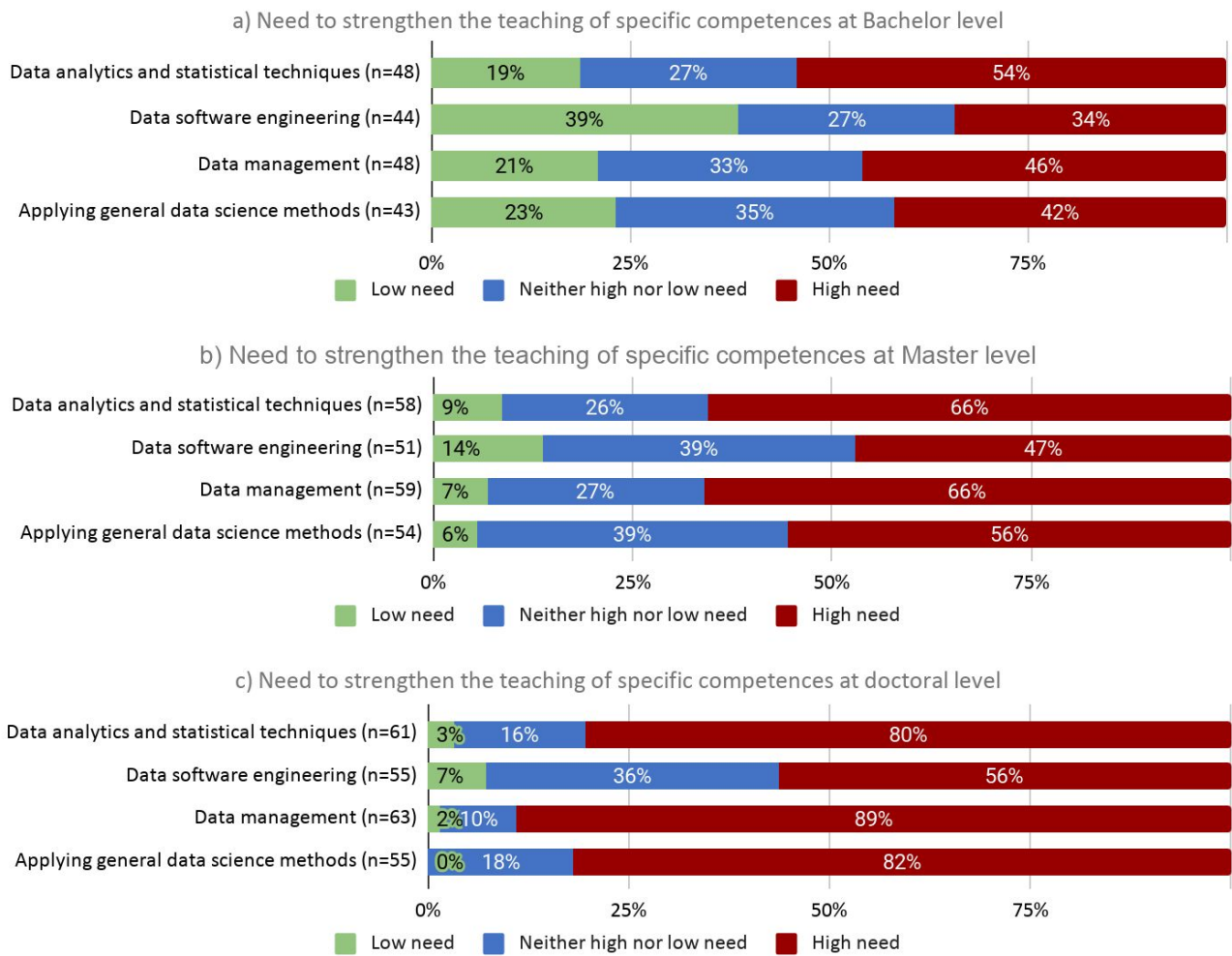


Figure 17a-c. Reported needs to strengthen specific competences by educational level

	Low need	Neither high nor low need	High need	n (valid)	Don't know	N/A	n	
Bachelor	Data analytics and statistical techniques	9	13	26	48	7	5	60
	Data software engineering	17	12	15	44	10	5	59
	Data management	10	16	22	48	7	4	59
	Applying general data science methods	10	15	18	43	11	5	59
Master	Data analytics and statistical techniques	5	15	38	58	3	0	61
	Data software engineering	7	20	24	51	8	2	61
	Data management	4	16	39	59	3	0	62
	Applying general data science methods	3	21	30	54	7	0	61
Doctoral	Data analytics and statistical techniques	2	10	49	61	2	0	63
	Data software engineering	4	20	31	55	6	1	62
	Data management	1	6	56	63	1	0	64
	Applying general data science methods	0	10	45	55	7	0	62
	Low need	Neither high nor low need	High need	n (valid)	Don't know	N/A	n	

Table 3. Reported needs to strengthen specific competences by educational level (absolute figures)

At doctoral level (see Figure 17c), the need to strengthen related teaching is regarded as high by 80%-89% of respondents in terms of “Data analytics and statistical techniques”, “Applying general

data science methods” and “Data management” and by 56% with respect to “Data software engineering”, indicating a very strong need for action on a broad range of data-related topics.

Input from Focus Group

The focus group held at the *University of Amsterdam* in November 2019 addressed the teaching and acquisition of RDM-related competences. Participants discussed how to address **RDM at Bachelor and Master level**. They agreed that, at this stage, there is a lesser need for a full ‘mastery’ of RDM, but rather of building awareness and the application of basic RDM principles. This corresponds with the rather low degree to which research data-related competences are currently being addressed at the institutions that responded to the survey. Especially the discovery and re-use of data was perceived as important; data production, curation and sharing were of less importance. It was also stressed that most graduates at this level will not pursue an academic career, and therefore the relevance for the (non-academic) labour market should be considered. This may however differ depending on the discipline in question.

A problem noted here was that **established curricula change at irregular intervals** – and adding additional content is always a challenge. It is also important that training material is maintained and regularly revised to be relevant; this can be a problem particularly for online collections of training material.

A parallel discussion was held on **RDM training in doctoral education**. It was highlighted that doctoral training is important for cultural change in academia. Moreover, RDM is related to professional and transferable skills, good scientific practice, awareness of ethics, integrity and legal questions, reproducibility and accountability, and collaborative research. Furthermore, knowledge of RDM could be seen as a competitive advantage, e.g. to benefit from funding opportunities. However, some focus group participants shared their concern that addressing RDM and FAIR might be “too late” in some disciplines if left to doctoral education.

Participants also raised a set of problems and challenges. A major issue is a lack of incentives and rewards for career advancement that stem from RDM, e.g. data publishing, within academia but also outside of it. Doctoral candidates also depend on support from the research group or supervisor/principal investigator, who should encourage RDM practices. Participants argued that it is sometimes difficult to engage and reach the senior research community on this topic. For some individual doctoral candidates, RDM may seem like an overwhelming task, in particular when no services or good practices are available for an early-stage researcher. Among others, participants also recommended a “train the supervisors” programme or mandatory module/interactions between doctoral candidates and supervisors on RDM.

A further topic for discussion was whether and how to address **generic data stewardship and FAIR data competences**. It is worth pointing out that the notion of “generic” competences was thoroughly problematised at the focus group. Often, ‘generic’ is used as a synonym of discipline-agnostic. Even though this group of skills is often thought of as being directed at a more homogenous group than the discipline-specific ones, the target audience is actually very diverse, consisting of – for example – students and researchers at different career levels, but also staff members of libraries, data centres and other infrastructure organisations. Another interpretation of

generic skills would be that they represent a baseline, common ground or fundamentals that more specialised competences can be built upon.

Participants also touched upon the role of **soft skills or transferable skills** such as communication and organisational skills. These are deemed highly relevant for data stewardship, but it can be difficult to fit them in a dedicated FAIR skills curriculum as a stand-alone topic area. One way to tackle this could be to integrate these in lessons and exercises on FAIR skills or letting students gather practical experience in partnerships with other actors (e.g. industry, infrastructure institutions). However, the question whether soft skills should be part of a data steward curriculum or if they should rather be part of research training in general alongside topics such as research ethics and integrity unlikely has simple answers.

Good practices mentioned by participants include a format of “coffee and science” sessions, organised e.g. by the library, online courses, the FOSTER project⁴³ and material and resources produced by, for instance, ELIXIR⁴⁴. Generally, participants preferred training with practical exercises and real-life examples (e.g. “bring your own example/problem” methods). Lectures should be complemented by small-group tutorials where possible. It was also noted that training should be offered in a modular/tailored structure for different levels of competence.

Finally, another discussion revolved around **discipline-specific competences**. As mentioned above, one main challenge is the different level of development of RDM practices and standards across disciplines. Participants agreed that within certain fields there is little to no awareness of RDM. Therefore, even standardisation of data is difficult in certain fields if there is limited awareness of what data standards or metadata are. These fields may benefit from taking generic resources and making them discipline-specific or from transfers of good practices. Humanities were mentioned as one field lacking metadata standards or where project-specific metadata is common, thereby preventing interoperability. It was noted that different data types also have their own standards across disciplines (e.g. imaging data, genetic data). Therefore, data type-specific training may be just as important as discipline-specific training.

The **findability of training material and resources** was also discussed. Participants were supportive of the suggestion to develop mechanisms to find the information, training and support that is available, e.g. in a “catalogue of catalogues” of trainings under the umbrella of the EOSC Portal. This could be facilitated if the training material itself was FAIR, i.e. equipped with the necessary metadata. It should also have reusable licensing material. CODATA was mentioned as one initiative which produces relevant resources that should be more applied/integrated at the university level. Data stewards, for instance embedded in faculties/departments, could then play an important role in disseminating these materials to researchers, essentially scanning for the appropriate training and communicating them to researchers to follow appropriate courses.

⁴³ Retrieved 21 February 2020, from: <https://www.fosteropenscience.eu/>.

⁴⁴ Retrieved 21 February 2020, from: <https://elixir-europe.org/platforms/training>.

3.5 Universities and the European Open Science Cloud

This section looks at the awareness of and views on the European Open Science Cloud (EOSC) in European universities. Having focused on the state of research data management in previous sections, the survey report now turns to the link between the institutional and European level in this field.

The survey asked respondents how they would rate the awareness of EOSC within their institution. Specifically, the question asked them to rank the awareness of the following groups from ‘Rather high’ to ‘Rather low’: institutional leadership, early-stage researchers (i.e. doctoral candidates and postdocs), researchers (faculty and others), professional and support staff (e.g. data stewards/managers and librarians) and students.

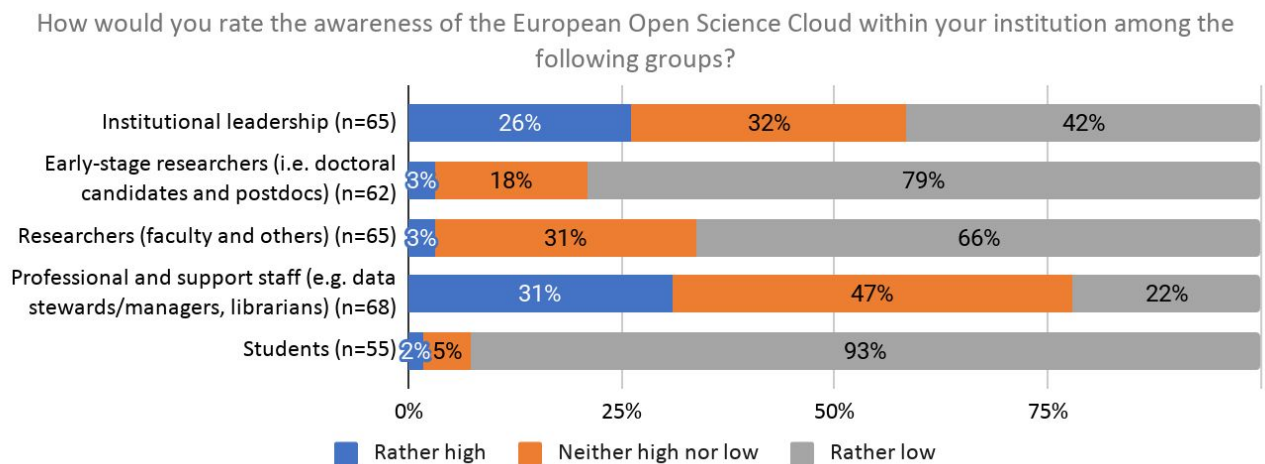


Figure 18. Awareness of EOSC across internal stakeholders

The survey results show a stark difference between the level of awareness of institutional leadership and professional and support staff compared to that of researchers, early-stage researchers and students (see Figure 18). The highest levels of awareness are seen for professional and support staff (31%) and institutional leadership (26%), while the lowest levels can be observed for students (93%), early-stage researchers (79%) and researchers (66%). However, this divide is less severe than it seems at first sight. It should be noted that respondents also indicate a low level of awareness under institutional leadership (42%) and by professional and support staff (22%).

The survey asked respondents where they see potential benefits created by the future EOSC for their institutions. The survey results identify three leading benefits (Figure 19). Respondents mainly see the potential benefits of EOSC in terms of creating opportunities for data-driven research collaborations (72%), increasing the visibility of research data created within their institution and easier access to services not available within their institution (65% each). 39% of respondents indicate the strengthening of their institutional capacity for data-driven science as a potential benefit of EOSC.

Where do you see potential benefits created by the future EOSC for your institution? (n=69, up to three choices)

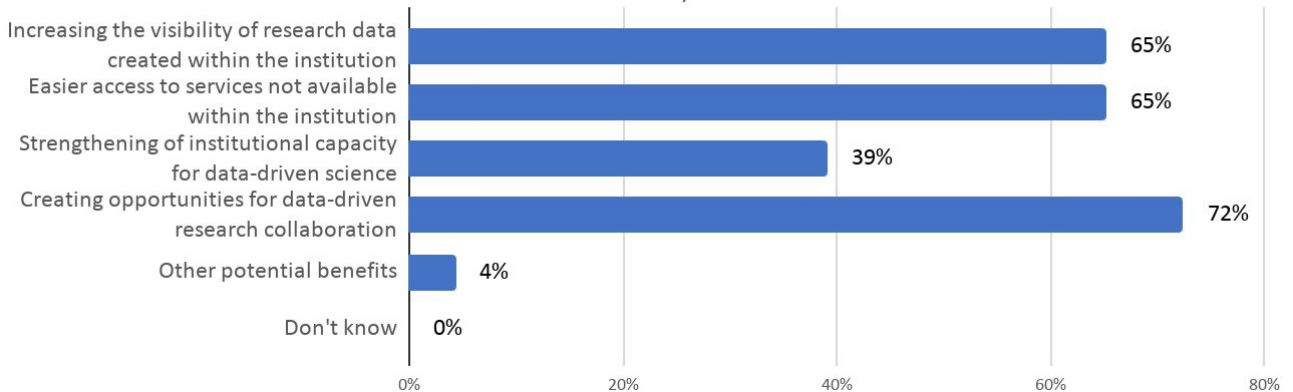


Figure 19. Perceived potential benefits of EOSC

The survey then asked respondents where they see barriers and difficulties for their institution to engage with the future EOSC. Results identify several leading barriers and difficulties (Figure 20). Respondents see the largest difficulties in their limited awareness of EOSC and its potential benefits (75%), lack of clear use cases (48%), limited institutional capacity (e.g. skilled staff and support structure; 43%), lack of incentivising policies or rewards from external actors (39%) and lack of interest or resistance from researchers (35%).

Where do you see barriers and difficulties for your institution to engage with the future EOSC? (n=69; up to four choices)

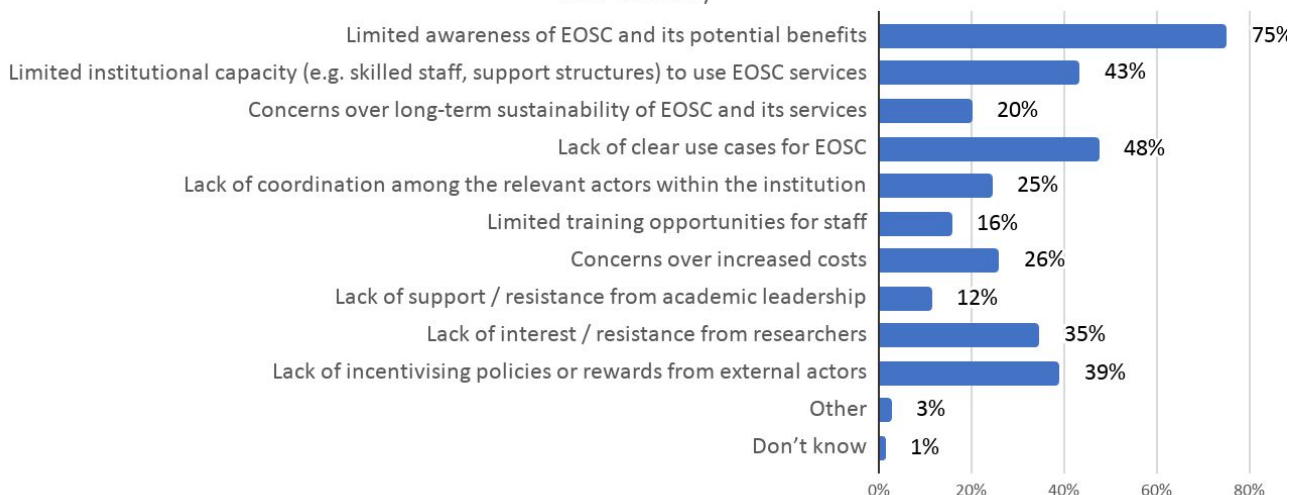


Figure 20. Perceived barriers and difficulties to engage with EOSC

Lesser but still prevalent barriers and difficulties, including concerns over increased costs (26%), lack of coordination among the relevant actors within the institution (25%), concerns over long-term sustainability (20%), limited training opportunities for staff (16%) and lack of support or resistance from academic leadership (12%).

Input from Focus Groups

The matter of university perception of EOSC was also briefly addressed during the Focus Groups. Beyond confirming a low level of awareness, participants emphasised that universities are hosts to (e-)infrastructures and research data services that could be made available through EOSC. Institutions therefore require clear rules of participation for services and entities in EOSC.

Participants raised various other points, such as uncertainty about costs related to linking institutional services to EOSC and legal questions, especially related to sensitive data. Moreover, where funders require RDM, this should be incentivised and rewarded, and funded appropriately.

Focus groups also addressed the ‘multi-level’ nature of EOSC. EOSC relates to individual research organisations and universities and even their subdivisions (e.g. core facilities, faculties/departments, libraries or repositories); it relates to national policies for research data; it also has a global dimension as science is often collaborative, and EOSC is an opportunity to facilitate such interactions. Thus, the implementation of EOSC needs to leave space for the balance between needs at different levels. In some areas, for example, data standards are not yet ‘FAIR’ – and might require more time until they reach FAIR maturity.

The discussions also touched upon open questions about the involvement of the private sector in EOSC and the potentially global scope of EOSC. Several participants raised concerns about commercial companies gaining ‘exclusive profit’ from open data. Similarly, the idea of commercial tech technology transfer is perceived as, sometimes, at odds with the notion of open data – which may lead to conflicts between different university units and services.

3.6 Potential for FAIRsFAIR support

This section looks at how FAIRsFAIR can support universities to develop and implement Findable, Accessible, Interoperable and Reusable (FAIR) research data management policies and teach FAIR research data management competences to Bachelor and Master students and early-stage researchers (i.e. doctoral candidates and postdocs).

Support for developing and implementing FAIR research data management policies

The survey asked respondents how FAIRsFAIR can support their institution to develop and implement FAIR research data management policies. The survey results identify three leading ways for the project to provide support in this regard (Figure 21). Respondents mainly see opportunities for FAIRsFAIR to train professional and support staff in the area of research data management (64%), to improve the availability of tools and resources to inform universities about research data management (49%) and to train researchers in FAIR research data management in specific domains (46%).

How can FAIRsFAIR support your institution in developing and implementing FAIR data and research data management policies? (n=70; up to three choices)

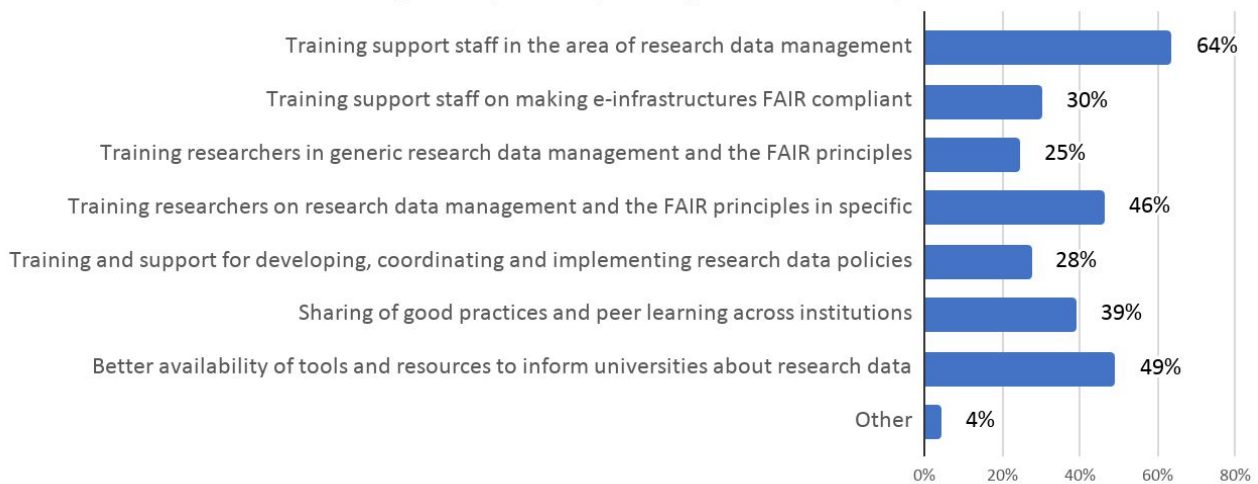


Figure 21. FAIRsFAIR support for developing and implementing FAIR and RDM policies

Other support opportunities are also recognised by respondents, albeit to a lesser extent. These include sharing of good practices and peer learning across institutions (39%), training professional and support staff to make e-infrastructures FAIR compliant (30%), training and support for developing, coordinating and implementing research data management policies (28%) and training researchers in generic FAIR research data management (25%).

Support for teaching FAIR research data management competences

The survey asked respondents how FAIRsFAIR can support their institution in teaching FAIR research data competences to Bachelor and Master students and early-stage researchers (i.e. doctoral candidates and postdocs). The survey results identify three leading ways for the project to provide support in this regard (Figure 22). Respondents mainly see opportunities for FAIRsFAIR to train the trainers (55%), provide domain-specific model courses and curricula (51%) and share good practices and peer learning across institutions (49%).

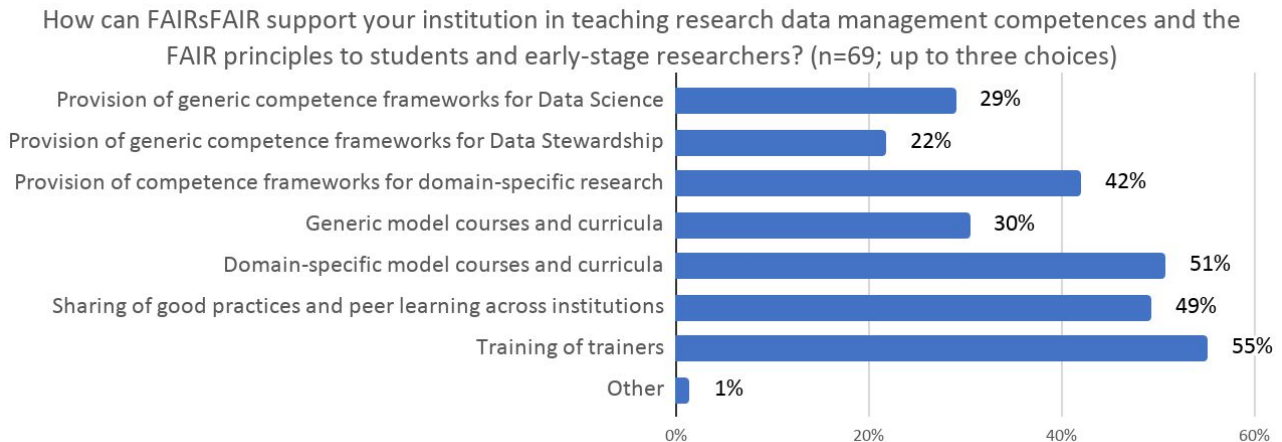


Figure 22. FAIRsFAIR support for teaching FAIR RDM competences

Other support opportunities are also identified by respondents, albeit to a lesser extent. These include providing competence frameworks for domain-specific research data management (42%), providing generic model courses and curricula (30%), providing generic competence frameworks for data science (29%) and providing generic competence frameworks for data stewardship (22%).

Input from Focus Groups

The focus groups were an opportunity to further qualify the recommendations collected through the survey. Participants voiced strong support for the role of FAIRsFAIR in collecting, curating and communicating good practices. Moreover, ‘practical’ examples and resources on how to make research data FAIR – within the context of a given researcher’s discipline – were deemed as an important course of action. Such measures would help reduce the gap between theoretical and practical knowledge. This recommendation stems from the observation that “most topics and trainings are still very general, which may not always be very useful for researchers”.

In relation to education and training, participants pointed out that different levels of “FAIRness” of data and compliance with the FAIR principles are possible. Training, resources, or the way how FAIR is addressed could be tailored according to this “level of difficulty”.

4. Summary of main findings

This report analyses and discusses a wide range of topics relevant to RDM and FAIR data at universities. As organisations with diverse missions for research, teaching, and more, implementing the FAIR principles represents a complex process that touches upon different aspects of the operation of an HEI. This report aims to present insights into three main dimensions, namely a) the integration of FAIR and RDM into higher education activities, b) the status of RDM policies and support services and their relation to the FAIR principles, and c) university perspectives on EOSC. This section presents the main findings and synthesises possible recommendations on how to develop RDM and FAIR competencies and capacities within the higher education sector.

4.1 Addressing FAIR in higher education

In order to promote cultural change within research organisations such as universities and among researchers, a **minimum level of awareness of RDM and FAIR** should be established. This could be achieved through more targeted communication measures and training based on more specific needs for different roles and disciplines. It is important to emphasise the benefits of RDM – and, by extension, FAIR data – for individual researchers, science and scholarship in general. As part of this, to reach early-stage researchers in particular, **FAIR and RDM should be advocated in a larger context of meeting open science and research integrity goals**. Practicing data stewardship should be understood as an intrinsic aspect of the research process and communicated as such to avoid the impression of RDM being a mere compliance activity.

The survey and focus group results reveal that universities express a **clear interest to advance research data-related skills and competences of students, researchers and professional staff**. Here, different levels of competence needs should be considered as not every role or discipline has the same requirement. Many HEIs already have a strategy towards the development of digital literacy and skills, sometimes at different levels within the institution. Even more express the intention to develop such a strategy in the future, demonstrating that HEIs see a need to develop the digital skills and competencies of their students. The intention to develop digital skills in a broader sense may also **prove an opportunity to attach more specific research data-related skills and competencies throughout an HEIs educational portfolio**.

Where available, the **teaching of research data-related skills and competencies is lowest at bachelor level, moderate at master level and strongest at doctoral level**. There are only slight differences between domains, most distinct at the bachelor level, with life, natural and engineering sciences being usually slightly more advanced than SSH disciplines or multidisciplinary/generic offers across the skills groups used in this study. This points towards a higher need in SSH contexts to develop RDM practices and standards, also in order to address them in educational programmes.

In terms of coverage of specific skills groups, there are broad trends across educational level and discipline. “Data analytics and statistical techniques” and “Following policies on data protection, privacy, IPR and ethical issues” are the two skills groups addressed most frequently on doctoral level, as well as on bachelor and master level. Encouragingly, the numbers of HEIs indicating that a particular topic, based on the EDSF framework, was “never” or “usually not” covered are very low at the doctoral level. Nevertheless the **findings indicate an opportunity to increase the coverage of data analytics and RDM-topics in teaching at all levels**.

To conclude, there is clearly a strong need to strengthen the teaching of specific research data-related competences at all three levels, but most urgently on the doctoral one indicated by 89% of responding HEIs with respect to “Data management”, by 82% in terms of “Applying general data science methods, 80% with regard to “Data analytics and statistical techniques” and 56% relating to “Data software engineering”.

Finally, an expressed need of responding institutions is for accessible guidance for researchers and students to follow and, subsequently, apply in practice. **Practical guidance on the application of the FAIR principles in different domains and disciplines, and related skills and competences**, would be useful to overcome this status quo.⁴⁵ Main suggestions from universities moreover include train-the-trainer actions, sharing of good practices, and a domain-specific competence framework and model courses/outputs.

4.2 Making research data policies and support services ready for FAIR

The design and implementation of institutional research data policies and support services for researchers are crucial enablers for advancing FAIR research data management practices at universities. The report shows a comprehensive picture of the content and status of both policies and support services at the responding institutions. **Sixty percent of participating HEIs reported having institutional research data policies** and an even larger number indicated that they offered research data support services to researchers and staff. However, as many respondents to the survey come from institutions in countries where engagement with Open Science is quite high, this finding may not be representative of the entire higher education sector in Europe.

In responding HEIs, there appears to be a gap between those primarily involved with formulating policies and those primarily involved in supporting adherence to them. It is recommended that both the **development and implementation of a policy should be a collaborative effort**. While the responsibility for the support and implementation of policies may be assigned to one or more specialised services (e.g. the library, data competence centre), the formulation of policies should take place through a broader consultation of internal stakeholders. Failure to engage all relevant operational units in policy development may result in policies that are not fit for purpose. It is therefore recommended that **all relevant units tasked with supporting the policies be involved with their development and pilot testing**. Moreover, most support services are implemented by central units, though inputs collected through the focus groups tend to support the **establishment of decentralised support**, e.g. through ‘embedded’ data stewards within faculties and departments where resources allow.

In terms of content, HEI **policies tend to encourage rather than mandate many RDM activities**. And while over half of respondents indicated that they indeed mandate sharing of data and data deposit, **reference to FAIR in RDM policies is not yet common practice**. In-house support for FAIR is available to varying degrees in many HEIs. However, as awareness of FAIR remains relatively low among researchers, focussing on aligning guidance and support for FAIR data, an

⁴⁵ This could be reflected in the development of the Competence Framework in FAIRsFAIR Task 7.3 and the development of further training material in Task 7.4. Such a step would allow the exploitation of synergies with activities in Work Package 6 and Work Package 3 of FAIRsFAIR - and other EOSC-related projects.

incremental approach – increasing FAIR requirements over time without causing unnecessary confusion for researchers, e.g. in areas without existing standards – may be most effective.

Importantly, as the policies of funders tend to carry most weight, **HEIs should seek to align with funders' requirements** in order to avoid ambiguity or conflicts between institutional and funder requirements. For institutions without a formal policy in place yet, this could mean seeking to support compliance with legal and funders' requirements before developing additional mandates.

To deliver RDM support services, **existing staff may need to be trained and/or new staff hired**. Sufficient time and/or financial resources must be made available to enable this by institutions, possibly with the support of research funders. In addition, clear career paths for data stewards and other professional RDM staff should be established within universities.

4.3 Universities and the European Open Science Cloud

Universities report a **low level of awareness of EOSC** among different groups within an organisation, with the notable exception of university leadership and research data support services. Universities also note a **lack of use cases for EOSC** and indicate that a main barrier to engage with EOSC is **limited institutional capacity**. On the positive side, universities do anticipate possible benefits of EOSC, in particular as an infrastructure that **facilitates collaborative research and increases visibility of research carried out within the institution**. Possible ways to ensure engagement of HEIs with EOSC could be to include EOSC capacity building and/or onboarding support for institutions (e.g. e-infrastructure staff, data stewards and managers) in the further EOSC implementation strategy.

Despite acknowledging that publisher research data requirements have relatively strong influence on researchers' RDM practices, some participants were also concerned that private companies could control access to publicly-funded research data – a well-known problem in the field of scholarly publishing for publications. The engagement of industry in EOSC should therefore be carefully designed in the Rules of Participation to develop EOSC as a **publicly-owned infrastructure** that has open principles at its core.

Annexes

Annex 1A - Survey questionnaire

- A ⊗ shown in a multiple choice question denotes an exclusive response.
- An asterisk (*) indicates a question to which an answer was required.

EUA-FAIRsFAIR Survey on Competences and Policies for Research Data Management and FAIR Data

In the context of the FAIRsFAIR project, the European University Association ([EUA](#)) is conducting a survey on different dimensions of research data management within higher education institutions. The survey covers specifically:

- the state and nature of institutional policies and support services for research data management;
- the relevance of the FAIR principles for research data;
- the teaching of research data competences to students, researchers and staff;
- awareness of and views on the European Open Science Cloud (EOSC).

In doing so, the survey mainly addresses topics and issues that have not been studied in the [2017-2018 EUA Open Access Survey](#). Main target group of the survey is university leadership and management staff responsible for the development and implementation of research data management. EUA especially encourages universities to participate that have not yet developed institutional approaches to research data.

Purpose

Results from the survey will inform the content and structure of activities of FAIRsFAIR and EUA aimed at supporting universities to foster a FAIR research and culture, such as:

- The development of a competence framework and training resources for FAIR competences
- The design and implementation of capacity development workshops and training opportunities for researchers and professional staff within universities working on research data management and curriculum design
- Support for policy and decision-making in the fields of research data and the European Open Science Cloud.

Structure

This survey is structured in three parts which address the following areas:

1. General information about your institution and your personal contact details.
2. The current state of (FAIR) research data management policies, support services and competence integration at your institution.

3. Your institution's views on future needs for research data competences, the European Open Science Cloud and possible lines of action of the FAIRsFAIR project.

Deadline

The deadline to submit your response is **15 November 2019**.

Guidelines for filling out the survey

- The survey should be filled out by the people or departments closely involved in developing and implementing FAIR data policies, support services and competence integration at your institution. **However, please note that only one response per institution should be submitted.**
- To facilitate collaboration between people and departments within your institution, [please find here a printable PDF version](#) of the survey. In case you need to consult with others within your institution, we suggest that you review the PDF version before you fill in the survey online. Please note **only the online version can be used to submit your final answers.**
- The survey saves answers per page as you click the “**Next**” button and move to the following page. You can exit the survey if you wish and re-enter by copying the link you have received in the **same device and browser** from which you first accessed it. The pages you have filled in up to that point will be saved. Please note that you will also be able to go back and make changes to your answers before submitting them.
- Your participation in this research is voluntary, and you may decline to participate without risk. While it is useful to be complete in your responses to the survey, you are free to withdraw from the study at any time. The data from any questions that were answered before exiting the survey will be recorded.
- Once you have reached the end and submitted, you will be given a full overview over your responses that you can print out and/or save.

Confidentiality and privacy policy

Based on the commitment of the FAIRsFAIR project to the H2020 Open Research Data Pilot, the responses provided in this survey will be anonymised (removing all information that could directly identify an individual (personal information) or the respective institution) and made available in open access via a trusted repository. Neither you nor your institution will be identified in any publication from this study. Your participation in this study is confidential.

Personal data gathered in the course of the survey will be handled according to the [EUA privacy policy](#).

Do you agree with the Confidentiality and Privacy Policy?

- Yes

Questions

Should you have any questions or encounter technical problems filling out this survey, please contact us at research@eua.eu.

Glossary

European Open Science Cloud (EOSC): The envisaged European data commons for the scientific community for storing, sharing and re-using scientific data and results, supported by high-capacity cloud solutions with super-computing capacity. EOSC is intended to function as a federated, globally accessible environment, where researchers, innovators, companies and the general public can publish, find and reuse each other's data and tools for research, innovation and educational purposes under well-defined and trusted conditions. ([more](#))

FAIR (Findable, Accessible, Interoperable, Re-usable): "The FAIR Data Principles are a set of guiding principles in order to make data findable, accessible, interoperable and reusable". ([Wilkinson et al., 2016](#)) "These principles provide guidance for scientific data management and stewardship and are relevant to all stakeholders in the current digital ecosystem. They directly address data producers and data publishers to promote maximum use of research data." ([Liber, 2017](#)) Horizon Europe, the next European research and innovation framework programme starting in 2021 will require projects to make their research data fair. In 2018, the European Commission expert group on FAIR data published a [comprehensive report](#) on actions to implement the FAIR principles in Europe.

FAIRsFAIR: A project funded under Horizon 2020 aiming to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle for different stakeholders including universities. Particular emphasis is laid on fostering FAIR data culture in research and education and the uptake of good practices in making data FAIR. ([more](#))

Research Data Management: A "set of practices to handle information collected and created during research. [...] These practices involve, but are not limited to, data management planning, documentation, organization, storage, dissemination and preservation." ([Higman et al., 2019](#))

Part 1: General information

1 Please indicate the country of your institution *

- List of countries

[Instruction: If your country is not in the list] 1a Please enter the country

2 Please indicate the name of your institution *

3 Name and contact of the person answering the survey: *

- Name _____
- Email _____
- Position _____

4 Do you want to stay informed about the results of the survey, EUA or the FAIRsFAIR project by email? *

Your contact information will be handled by EUA ([privacy policy](#)) and the FAIRsFAIR project ([privacy policy](#)). Subscriptions to newsletters can be cancelled at any time.

- I want to receive information about the results of the survey
- I want to stay informed about EUA activities in Open Science
- I want to stay informed about the FAIRsFAIR project.
- None of the above

5 How would you describe the profile of your institution? *

- Comprehensive institution
- Specialised institution (e.g. medical science, music and arts school)
- University of applied sciences (college-type or professional education institution which does not award PhDs, or does so in only a few disciplines)
- Technical university/ University of technology
- Open university (e.g. distance learning university)

6 What is the total number of researchers (FTE), including doctoral candidates, working at your institution? *

- < 100
- 100-499
- 500-1000
- > 1 000

Part 2 - Current research data management policies, support services and competence development

7 How would you rate the awareness of the FAIR principles within your institution?

	Rather high	Neither high nor low	Rather low	Don't know
Institutional leadership	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Early-stage researchers (i.e. doctoral candidates and postdocs)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Researchers (faculty and others)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional and support staff (e.g. data stewards/managers, librarians)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Section 2a – Policies

8 Does your institution have a policy on Research Data Management? *

- Yes
- No
- Don't know

[Instruction: If Question 8 is “No” or “Don’t know”, go directly to Question 12]

9 What are the main elements of your institutional research data management policy?

Elements	Yes, included as a mandatory element	Yes, included as an optional/encouragement element	Not included in institutional policy
Requirement for research data management planning (e.g. DMPs)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions for data storage	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions on open access to data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions on FAIR data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Provisions for specific disciplinary areas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions on research integrity and ethics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specific guidelines for sensitive data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions for data protection (e.g., GDPR)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions on research data management for public-private research contracts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions on research assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions on the use of unique researcher identifiers (e.g., ORCID)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provisions on copyright and/or intellectual property	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (<i>please specify</i>)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10 Who is primarily involved in developing research data management policies in your institution? [multiple choice]

- Academic leadership/management (e.g. rector, vice-rectors, etc.)
- Faculties/research departments
- Research administration
- Individual researchers/research units
- IT department
- Library
- Legal department
- Data protection office
- Other (*please specify*) _____
- Don't know

11 At what level are research data management policies primarily implemented in your institution? [multiple choice]

- Institutional/university level
- Faculty/department level
- Disciplinary level
- Research unit level
- Other (*please specify*) _____
- Don't know

Section 2b - Support Services

12 Does your institution currently have any research data support services in place? *

- Yes
- No
- Don't know

[Instruction: If Question 12 is “No” or “Yes”, go directly to Question 18]

13 What type of support does your institution provide to researchers interested in making research data FAIR? [multiple choice]

- Training for researchers (including doctoral candidates)
- Institutional website(s) on research data management
- Planning stewardship and sharing of FAIR outputs
- Finding and reusing data from existing sources
- Using or developing FAIR research tools/services
- Preparing and documenting data/code to make outputs FAIR
- Publishing FAIR outputs on own or recommended repositories
- Recognising, citing and acknowledging contributions
- Developing open research strategy and vision
- Complying with legal and ethical requirements, FAIR principles
- Funding for implementing FAIR principles
- Finding (other) sources of training and advice on FAIR data
- Other (please specify) _____

14 Who is primarily involved in supplying research data support services in your institution? [multiple choice]

- Academic leadership/management
- Faculties/research departments (e.g. embedded data stewards and/or managers)
- Research administration (e.g. generic data stewards and/or managers)
- Individual researchers/research units
- IT service
- Library
- Legal department
- Data protection office
- Other (please specify) _____
- ☒ Don't know

15 At what level are research data support services primarily implemented in your institution? *[multiple choice]*

- Institutional/university level
- Faculty/department level
- Disciplinary level
- Research unit level
- Don't know

16 Has your institution established specific research data support roles (e.g. data stewards, research data managers)? *[multiple choice]*

- Yes, at institutional/central level
- Yes, at faculty/department level
- No
- Don't know

17 Do you have any further comments about this section?

Section 2c - RDM competences of students

This survey seeks to know how widespread the teaching of (FAIR) research data management competences and skills currently is at universities.

This section will ask if and at what level (bachelor, master, doctoral) relevant competences and skills are taught at your institution. As competences we understand the set of competences defined in the [EDISON](#) project ([CORDIS](#) entry). At the highest level, these include

- Data analytics and statistical techniques
- Data software engineering
- Data management
- Data science methods to address domain-specific research problems

18 Does your institution have an approach or a strategy to address data literacy and skills of students?

- Yes, for the entire institution
- Yes, for faculty/department(s)/scientific discipline(s)

- Yes, for both
- No
- No, but we are developing one
- Don't know

19 To your best knowledge, in which broad set of disciplines and at what level is your institution addressing one or more of the above-mentioned research data-related competences?

[Instruction: per cell in the table below, select one of the following options:]

- Never*
- Usually not*
- Sometimes*
- Usually yes*
- Always*
- Don't know*
- Not applicable]*

Domain	Bachelor	Master	Doctoral
Humanities and Social Sciences			
Life Sciences			
Natural Sciences			
Engineering Sciences			
Generic / multidisciplinary			

20 To your best knowledge, what are the more specific research data-related competences delivered by your institution at bachelor, master, and doctoral level?

[Instruction: per cell in the table below, select one of the following options:]

- Never*
- Usually not*
- Sometimes*
- Usually yes*
- Always*
- Don't know*
- Not applicable]*

Competences	Bachelor	Master	Doctoral

Using data analytics and statistical techniques to deliver insights into research problems			
Using software engineering to research, design, implement new data analytics applications and/or to support data handling during the whole data lifecycle			
Developing and implementing a data strategy , e.g., a research data management policy and Data Management Plan (DMP)			
Developing and implementing relevant data models , defining metadata using common standards and practices			
Integrating data from multiple sources			
Maintaining information on data handling, including reference to published data and corresponding data sources (data provenance)			
Ensuring data quality, accessibility, interoperability, compliance to standards, and publication (data curation)			
Understanding and following policies on data protection, privacy, IPR and ethical issues			
Applying general data science methods to domain-specific research problems ; using scientific domain knowledge to develop relevant data analytics applications;			

21 Can you provide an example of a curriculum or course that includes one or more of the data-related competences mentioned above?

22 Is your institution using specific frameworks for research data management, data science and open science when developing training activities or curricula? Please indicate all that apply to your best knowledge.

- EDISON Data Science Competence Framework (EDSF)
- EOSCpilot Skills and Capability Framework (including FAIR4S)
- Open Science Competence Assessment Matrix (OS-CAM)

- Others such as national frameworks and guidelines

-
- No frameworks, trainings or resources used
 - Don't know

23 Is your institution using resources, trainings or other services from other projects or initiatives? Please indicate all that apply to your best knowledge.

- Research Data Alliance (RDA)
- OpenAIRE
- GO-FAIR
- FOSTER
- CODATA
- Research infrastructures (e.g. ELIXIR, CESSDA)
- Others _____

24 Do you have any further comments about this section?

Part 3 - Thinking ahead: universities, research data competences and EOSC

Section 3a - Future competences

Considering the growing use of digital tools in science and other professions, an increasing number of funder mandates for research data management, as well as the opportunities that may be created by the European Open Science Cloud, it is anticipated that students and graduates across scientific domains will need more data-related competences.

25 Does your institution believe there is a need to strengthen the teaching of specific research data-related competences at the bachelor, master or doctoral level?

[Instruction: per cell in the table below, select one of the following options:]

- High need*
- Neither high nor low*
- Low*
- Don't know*

□ *Not applicable]*

Competence group	Bachelor	Master	Doctoral
Data analytics and statistical techniques			
Data software engineering			
Data management			
Applying general data science methods to domain-specific research problems			

26 Do you have any further comments about this question?

Section 3b - Universities and EOSC

27 How would you rate the awareness of the European Open Science Cloud within your institution among the following groups?

	Rather high	Neither high nor low	Rather low	Don't know
Institutional leadership	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Early-stage researchers (i.e. doctoral candidates and postdocs)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Researchers (faculty and others)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional and support staff (e.g. data stewards/managers, librarians)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

28 Where do you see potential benefits created by the future EOSC for your institution? Please tick only the three most important.

- Increasing the visibility of research data created within the institution
- Easier access to services not available within the institution (e.g. data, repositories, cloud and high-performance computing services)
- Strengthening of institutional capacity for data-driven science
- Creating opportunities for data-driven research collaboration
- Other potential benefits _____
- Don't know

29 Where do you see barriers and difficulties for your institution to engage with the future EOSC? Please tick only the four most important.

- Limited awareness of EOSC and its potential benefits
- Limited institutional capacity (e.g. skilled staff, support structures) to use EOSC services
- Concerns over long-term sustainability of EOSC and its services
- Lack of clear use cases for EOSC
- Lack of coordination among the relevant actors within the institution
- Limited training opportunities for staff
- Concerns over increased costs
- Lack of support / resistance from academic leadership
- Lack of interest / resistance from researchers
- Lack of incentivising policies or rewards from external actors (e.g. national/regional governments, research funding organisations)
- Other _____
- Don't know

Section 3c - Supporting universities

30 How can FAIRsFAIR support your institution in developing and implementing FAIR data and research data management policies?

Please tick only the three most important courses of action

- Training support staff in the area of research data management (e.g. metadata; data storage/management/curation; technical standards; FAIR principles)
- Training support staff on making e-infrastructures FAIR compliant
- Training researchers in generic research data management and the FAIR principles
- Training researchers on research data management and the FAIR principles in specific domains
- Training and support for developing, coordinating and implementing research data policies

- Sharing of good practices and peer learning across institutions
- Better availability of tools and resources to inform universities about research data management
- Other _____

31 How can FAIRSF AIR support your institution in teaching research data management competences and the FAIR principles to students and early-stage researchers?

Please tick only the three most important courses of action.

- Provision of generic competence frameworks for Data Science
- Provision of generic competence frameworks for Data Stewardship
- Provision of competence frameworks for domain-specific research data management
- Generic model courses and curricula
- Domain-specific model courses and curricula
- Sharing of good practices and peer learning across institutions
- Training of trainers
- Other _____

-END OF SURVEY-

Annex 1B - EDSF and FAIR4S in the survey questionnaire

Question 13 in section 2b) *Support Services* referred to several dimensions of the FAIR4S framework:

Question 13: What type of support does your institution provide to researchers interested in making research data FAIR? [multiple choice]	
Answer option	Equivalent FAIR4S 'key skill'⁴⁶
<input type="checkbox"/> Training for researchers (including doctoral candidates)	n/a
<input type="checkbox"/> Institutional website(s) on research data management	n/a
<input type="checkbox"/> Planning stewardship and sharing of FAIR outputs	Plan stewardship and sharing of FAIR outputs
<input type="checkbox"/> Finding and reusing data from existing sources	Reuse data from existing sources

⁴⁶ See <https://eosc-fair4s.github.io/framework.html>

<input type="checkbox"/> Using or developing FAIR research tools/services	Use or develop open research tools/services
<input type="checkbox"/> Preparing and documenting data/code to make outputs FAIR	Prepare and document for FAIR outputs
<input type="checkbox"/> Publishing FAIR outputs on own or recommended repositories	Publish FAIR outputs on recommended repositories
<input type="checkbox"/> Recognising, citing and acknowledging contributions	Recognise, cite and acknowledge contributions
<input type="checkbox"/> Developing open research strategy and vision	Develop open research strategy and vision
<input type="checkbox"/> Complying with legal and ethical requirements, FAIR principles	Apply policies to comply legal requirements, ethical & FAIR principles
<input type="checkbox"/> Funding for implementing FAIR principles	Secure funding for open science /support
<input type="checkbox"/> Finding (other) sources of training and advice on FAIR data	n/a
<input type="checkbox"/> Other (please specify)	n/a

Several questions moreover used a simplified version of several competence groups and competence definitions outlined in the EDISON Data Science Competence Framework (DS-CF), release 3, version 10.⁴⁷ This includes the broad Data Science competence and skills groups identified by DS-CF and several specific competences from the skills group DSDM: Data Management and Governance.

EDSF-CF Data Science competence and skills groups
<ul style="list-style-type: none"> ● DSDA: Data Science Analytics (including Statistical Analysis, Machine Learning, Data Mining, Business Analytics, others) ● DSENG: Data Science Engineering (including Software and Applications Engineering, Data Warehousing, Big Data Infrastructure and Tools) ● DSDM: Data Management and Governance (including data stewardship, curation, and preservation) ● DSRMP: Research Methods and Project Management for research related professions and Business Process Management for business related professions

⁴⁷ https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_CF-DS-release3-v10.pdf

- **DSDM:** Domain Knowledge and Expertise (Subject/Scientific domain related)

The decision to apply the skills and competences based on a definition from the EDISON project was taken based on the purpose of EDISON to support the development of higher education programmes in the field of data science.

Question 20 in section 2c) *RDM competences of students* used both major competence and skills groups (DSDA, DSENG, DSDK) and specific competences within the DSDM group. Definitions were simplified with the intention to be i) more fit to a general academic context and to be applicable to various disciplines beyond data science and to be ii) more easily accessible for respondents answering the questionnaire. Given the focus of FAIRsFAIR on research data management, only the specific competences defined within the DSDM group were used in more detail.

Question 20: <i>To your best knowledge, what are the more specific research data-related competences delivered by your institution at bachelor, master, and doctoral level?</i>	
Answer option	Equivalent EDSF
Using data analytics and statistical techniques to deliver insights into research problems	DSDA
Using software engineering to research, design, implement new data analytics applications and/or to support data handling during the whole data lifecycle	DSENG
Developing and implementing a data strategy , e.g., a research data management policy and Data Management Plan (DMP)	DSDM01
Developing and implementing relevant data models , defining metadata using common standards and practices	DSDM02
Integrating data from multiple sources	DSDM03
Maintaining information on data handling, including reference to published data and corresponding data sources (data provenance)	DSDM04
Ensuring data quality, accessibility, interoperability, compliance to standards, and publication (data curation)	DSDM05
Understanding and following policies on data protection, privacy, IPR and ethical issues	DSDM06
Applying general data science methods to domain-specific research problems ; using scientific domain knowledge to develop relevant data analytics applications;	DSDK

Question 25 in section 2c) *RDM competences of students* was introduced as a means to inquire concrete areas for further action of FAIRsFAIR.

Question 25: Does your institution believe there is a need to strengthen the teaching of specific research data-related competences at the bachelor, master or doctoral level?	
Answer option	Equivalent EDSF
Data analytics and statistical techniques	DSDA
Data software engineering	DSENG
Data management	DSDM
Applying general data science methods to domain-specific research problems	DSDK

Annex 2 - Focus Group in Madrid

Report available at:

<https://www.fairsfair.eu/articles-publications/outcomes-fairsfair-focus-group-universidad-carlos-iii-de-madrid>

Research Data Management and the FAIR Principles

Hosted by Universidad Carlos III de Madrid (UC3M)

30 October 2019

Sound Research Data Management (RDM) from the creation of research data to its long-term storage is considered an intrinsic aspect of Open Science. Certain RDM practices such as the preparation of Data Management Plans (DMP) and the deposition of data in repositories are becoming the new normal, intensified by funder mandates in Horizon 2020 and Horizon Europe, as well as European legislation such as the newly revisited European Directive on Public Sector Information.

Moreover, the ambitious project to create a European Open Science Cloud (EOSC) based on Findable, Accessible, Interoperable and Reusable (FAIR) data will drive the need for data stewards and managers based at universities and other research organisations, as well as skilled users and researchers producing, sharing and (re-)using data. Besides benefits for the EOSC, RDM can be considered as a main bottleneck for the adoption of Open Science – for instance as a practice supporting reproducibility and the re-use of data.

Yet, while some universities are already well advanced in professionalising their RDM services and despite a clear need to foster RDM in policy and practice, [EUA findings](#) show that only a minority of European universities have research data policies in place or employ specialised data support staff.

Within the [FAIRsFAIR](#) (Fostering FAIR Data Practices in Europe) project, EUA and UC3M are organising a focus group meeting to allow an in-depth dialogue with universities about the development and implementation of Research Data Management. The focus group will also discuss how to develop competences of academics and professional staff for the success of such policies. Main questions will include:

- What are institutional challenges, drivers and good practices in establishing and implementing research data policies in universities?
- What are challenges, drivers and good practices in educating and training students, academics and professional staff in research data management?

Participants will have an opportunity to exchange experiences and learn about practices in other participating organisations as the programme will build on small interactive groups intended to create an open exchange.

This event is mainly aimed at university leaders with an interest in designing and implementing research data policies and at various stages in this process. Other participants, e.g. representing Research Infrastructures and RDM-related projects or initiatives are also welcomed to apply.

Programme

09.00-09.15	<i>Arrival of participants</i>
09.15-10.30	Opening session <ul style="list-style-type: none">→ Welcome, participants' introduction and presentation of programme→ Introduction to FAIRsFAIR and WP7 (Lennart Stoy, EUA)→ FAIRsFAIR work on policies (Joy Davidson, DCC)→ Open Science and RDM at UC3M (Eva Mendez, UC3M)
10.30-10.45	<i>Coffee Break (15')</i>
10.45-12.15	Discussion session
12.15	<i>Lunch</i>
13.15-14.15	Discussion session
14.15-15.00	Discussion session Final wrap up
15.00	<i>End of Focus group</i>

Annex 3 - Focus Group in Amsterdam

Report available at:

<https://www.fairsfair.eu/articles-publications/fairsfair-focus-group-university-amsterdam>

Teaching data management and the FAIR data principles Hosted by the University of Amsterdam 19 November 2019

As part of the [FAIRsFAIR](#) (Fostering FAIR Data Practices in Europe) project, EUA and the University of Amsterdam are organising a focus group meeting to allow an in-depth dialogue with other universities about the development and implementation of data-related competences and the Findable, Accessible, Interoperable and Reusable (FAIR) data principles in higher education curricula.

In a world increasingly driven by data, the skills and competences to collect, manage, analyse and interpret data are becoming ever more important. Data literacy and data competences in a wide range of disciplinary curricula will become a necessity to prepare graduates for the future of work.

This is not limited to data science itself. For graduates in other fields, data competences will be important to succeed in the labour market or to conduct research. Science 2.0 and Open Science are made possible through digital technologies and rely on research data management (RDM) to ensure reproducibility and openness of research. The European Open Science Cloud (EOSC) based on [FAIR](#) data principles will be needing data stewards and managers based at universities, research organisations, public organisations and private companies, as well as skilled users and researchers producing, sharing and (re-)using the available data.

This change would require a re-thinking of how data-related competences including RDM are taught and addressed even at the level of Bachelor and Master students and early-stage researchers – in Data Science and beyond.

The focus group will address key questions, including:

- At what level and how much should students be exposed to research data management? How can the FAIR principles be addressed in education?
- What are challenges and good practices in educating and training students and early-stage researchers in research data management and the FAIR principles?

Participating in the focus group will be an opportunity to exchange experiences and learn about practices at other universities as the programme will build on interactive discussions intended to create an open exchange. Participants will have the opportunity to familiarise themselves with the ecosystem of data science initiatives and educators in Amsterdam.

Programme

09.30-10.00	<i>Arrival of participants</i>
10.00-10.30	Welcome and opening <ul style="list-style-type: none">→ Tour de table→ Presentation of programme & FAIRsFAIR (Lennart Stoy, EUA)→ EDISON & Data Science at University of Amsterdam (TBC, UvA)
10.30-11.30	First session – Input presentations <ul style="list-style-type: none">→ EOSCpilot and FAIR4S (Angus Whyte, DCC)→ Data Stewardship and Training in ELIXIR (Celia van Gelder, DTL/ELIXIR-NL)→ National Coordination of Data Steward Education in Denmark (Michael Svendsen, KB)
11.30-11.45	<i>Break</i>
11.45-13.00	Breakout discussion I
13.00-14.00	<i>Lunch</i>
14.00-15.00	Breakout discussions II
15.15-16.00	Common discussion of outcomes and ways forward

Annex 4 - Data for Figures 14a-c

Question 20: To your best knowledge, what are the more specific research data-related competences delivered by your institution at bachelor, master, and doctoral level?

a) Bachelor level

	Never	Usually not	Sometimes	Usually yes	Always	n (valid)	Don't know	Not applicable	n
Using data analytics and statistical techniques to deliver insights into research problems	2	9	15	10	4	40	11	5	56
Using software engineering to research, design, implement new data analytics applications and/or to support data handling during the whole data lifecycle	3	16	12	5	2	38	12	5	55
Developing and implementing a data strategy , e.g., a research data management policy and Data Management Plan (DMP)	10	17	5	2	3	37	11	6	54
Developing and implementing relevant data models , defining metadata using common standards and practices	6	16	9	4	0	35	13	5	53
Integrating data from multiple sources	3	12	12	8	2	37	13	5	55
Maintaining information on data handling, including reference to published data and corresponding data sources (data provenance)	6	15	5	7	5	38	12	5	55
Ensuring data quality, accessibility, interoperability, compliance to standards, and publication (data curation)	7	16	6	6	2	37	13	5	55
Understanding and following policies on data protection, privacy, IPR and ethical issues	3	8	11	10	6	38	12	5	55
Applying general data science methods to domain-specific research problems ; using scientific domain knowledge to develop relevant data analytics applications;	3	11	15	5	2	36	13	5	54

b) Master level

	Never	Usually not	Sometimes	Usually yes	Always	n (valid)	Don't know	Not applicable	n
Using data analytics and statistical techniques to deliver insights into research problems	3	6	15	13	12	49	7	0	56
Using software engineering to research, design, implement new data analytics applications and/or to support data handling during the whole data lifecycle	2	12	21	9	3	47	7	1	55
Developing and implementing a data strategy , e.g., a research data management policy and Data Management Plan (DMP)	2	12	21	9	3	47	7	1	55
Developing and implementing relevant data models , defining metadata using common standards and practices	6	20	12	9	2	49	7	0	56
Integrating data from multiple sources	5	15	14	9	1	44	10	0	54
Maintaining information on data handling, including reference to published data and corresponding data sources (data provenance)	2	11	21	8	4	46	9	0	55
Ensuring data quality, accessibility, interoperability, compliance to standards, and publication (data curation)	4	12	14	8	8	46	9	0	55
Understanding and following policies on data protection, privacy, IPR and ethical issues	2	7	13	17	8	47	8	0	55
Applying general data science methods to domain-specific research problems ; using scientific domain knowledge to develop relevant data analytics applications;	2	7	19	13	3	44	10	0	54

c) Doctoral level

	Never	Usually not	Sometimes	Usually yes	Always	n (valid)	Don't know	Not applicable	n
Using data analytics and statistical techniques to deliver insights into research problems	2	1	13	21	18	55	6	0	61
Using software engineering to research, design, implement new data analytics applications and/or to support data handling during the whole data lifecycle	3	4	25	15	5	52	6	1	59
Developing and implementing a data strategy , e.g., a research data management policy and Data Management Plan (DMP)	0	7	24	18	9	58	3	0	61
Developing and implementing relevant data models , defining metadata using common standards and practices	2	6	25	17	4	54	5	0	59
Integrating data from multiple sources	3	2	23	17	6	51	9	0	60
Maintaining information on data handling, including reference to published data and corresponding data sources (data provenance)	1	7	21	15	12	56	6	0	62
Ensuring data quality, accessibility, interoperability, compliance to standards, and publication (data curation)	3	8	21	17	6	55	6	0	61
Understanding and following policies on data protection, privacy, IPR and ethical issues	1	2	12	24	17	56	5	0	61
Applying general data science methods to domain-specific research problems ; using scientific domain knowledge to develop relevant data analytics applications;	2	4	18	18	8	50	8	0	58