CESAER

The strong and united voice of universities of science and technology in Europe

ADVANCING RESEARCH DATA MANAGEMENT IN UNIVERSITIES OF SCIENCE AND TECHNOLOGY

WHITE PAPER DATED 13TH FEBRUARY 2020

Our Task Force Open Science has been instrumental in writing this white paper. We thank the following writers in particular (in alphabetical order):

- Mattias Björnmalm (CESAER)
- Federica Cappelluti (*Politecnico di Torino*)
- Alastair Dunning (Delft University of Technology)
- Dana Gheorghe (University POLITEHNICA of Bucharest)
- Malgorzata Zofia Goraczek (Technical University of Vienna)
- Daniela Hausen (RWTH Aachen University)
- Sibylle Hermann (University of Stuttgart)
- Angelina Kraft (*Leibniz Universität Hannover*)
- Paula Martinez Lavanchy (Delft University of Technology)
- Tudor Prisecaru (University POLITEHNICA of Bucharest)
- Barbara Sánchez (Technical University of Vienna)
- Robert Strötgen (*Technische Universität Braunschweig*)

To contact us, please reach out via the details provided on cesaer.org/contact/

LEUVEN, 13th February 2020

Please reference this paper using http://doi.org/10.5281/zenodo.3665372

LETTER FROM THE PAST VICE PRESIDENT

As past Vice President for Research, I am delighted to commend this report on Advancing Research Data Management (RDM) in universities of science and technology. With this, I wish to encourage all our Member institutions to reflect on their current position as regards to research data and the management thereof, and envisage how further investment in this area can strengthen research practices at their own institution.

As the report demonstrates, having a trusted technical and cultural infrastructure for supporting RDM is essential. Data-driven research underpins much of the scientific enterprise today, and its importance will only grow. Successful research institutions will have a confident, clear and coherent approach to how their research groups handle and manage their data (and related code). Only by developing the right expertise and accompanying tools will institutions be able to provide the support for the entire data life cycle: from data capture, cleaning, analysis, storage through to publishing and dissemination.

It is also apparent that developing coherent services for research data cannot be done in isolation. Working together is essential. I encourage all staff involved in delivering RDM services to maintain their working connections within CESAER, and to continue sharing strategies on policy implementation, creating shared training modules and identifying opportunities for shared infrastructure.

Furthermore, I hope to see our growing network deepen connections with the European Open Science Cloud (EOSC). It is evident from this white paper that there is still work to be done to give research data management the disciplinary ‘flavour’ needed to optimally respond to the specific ‘tastes’ of the diverse subject areas and methodologies that exist across universities of science and technology. Continued partnership with EOSC will be important here, to bring together subject communities and technical expertise, and help further develop disciplinary approaches to research data management.

Finally, I highlight that nearly 90% of respondents believed industrial collaboration to be ‘relevant’ or ‘very relevant’ for the researchers at their institutions, but that there are unique challenges when providing support to those researchers that collaborate extensively with industry. This is a fruitful and necessary area to explore further, and I am excited for continued work within CESAER (including strengthening the collaboration between Task Force Open Science and Task Force Innovation) and with our external partners towards turning these challenges into opportunities.

Karel Luyben
Past Vice President for Research of CESAER
Rector Magnificus Emeritus of Delft University of Technology

EXECUTIVE SUMMARY

Solid Research Data Management (RDM) is key for robust and reusable science. Science is increasingly being digitalised, and the data volumes produced by scientists continue to increase across all disciplines. The implementation of the FAIR data principles (Findable, Accessible, Interoperable, Re-usable) is crucial to achieving reusable and robust science. For researchers and universities, RDM is an important prerequisite and enabler for the fulfilment of FAIR data principles and as a consequence, robust and reusable science. This white paper highlights and analyses the efforts of the universities of science and technology united within CESAER in providing sustainable RDM services to their students, researchers and staff.

In chapter two, we describe the design of an online survey distributed to our member institutions, which addressed the following key aspects of RDM: (i) the establishment of a RDM policy; (ii) the provision of suitable RDM infrastructure and tools; and (iii) the establishment of RDM support services and trainings tailored to the requirements of science and technology disciplines.

In chapter three, we explore the results regarding the status of institutional RDM policies. There are significant variations in the structure of the documents, with most of the policies addressing essential RDM issues such as data preservation, reproducibility, data sharing and data management plans. The policies are primarily being implemented through the provision of centralised RDM support, data infrastructure and training, with few efforts (so far) to implement them at a faculty or departmental level. Specific guidelines that facilitate researchers from different disciplines to comply with the policies are often lacking.

In chapter four, we focus on the infrastructure which has been developed in order to support RDM. So far, institutions have focused on providing data infrastructure such as repositories, secure data storage and RDM tools such as Data Management Plan (DMP) creators, High-Performance Computing (HPC) systems and version control systems. Two main challenges are explored: (i) the lack of metadata standards that allow the description, functional preservation and ultimately re-use of the data stored; and (ii) the large and increasing size of the data produced within technical disciplines. These factors stress the need to establish large storage capacities together with services tailored for specific disciplines for data handling, sharing and publishing. Furthermore, only a few universities provide tools tailored to discipline-specific workflows and requirements.

In chapter five, we introduce the scope of RDM support services and training offered to students, PhD candidates, researchers and professors at universities of science and technology. We find that the support available for researchers engaged in industrial collaborations concerning RDM practices such as data sharing, publication and archiving is limited, which is a major challenge. The importance of providing training tailored to science and technology disciplines is also highlighted.

In addition to conclusive statements, the final chapter six contains concrete recommendations for institutions to advance RDM at universities of science and technology.

CONTENTS

1. INTRODUCTION

The growing importance of research data reinforces its high position on the agenda. Since the publication of the FAIR data principles in 2016 and the adoption of an open data policy within Horizon 2020, neither universities, research institutes nor individual scientists can afford to ignore the topic of Research Data Management (RDM).

With this paper, the universities of science and technology united within CESAER provide insight into the status of RDM services offered for researchers across Europe and identify the main challenges encountered when providing RDM support in universities of science and technology.

In contrast to some other disciplines where RDM is often driven by professional associations (e.g. DARIAH in the humanities and ELIXIR in life sciences), the approach to RDM in the engineering sciences is often bottom-up. As a result, common interfaces and protocols for managing, accessing and reusing research data from industry and academia are largely missing. Almost no broadly accepted standards exist for curating research data and the corresponding metadata.

The role of an engineer typically involves the analysis of technical-economic-social systems and using the knowledge gained through this analysis to tackle specific challenges. A major challenge in establishing standardised RDM procedures for engineering disciplines is the wide variety of engineering problems that often result in highly specialised solutions. For example, experimental measurement devices as well as simulation software, are typically custom-built and coded. At the same time, engineering research can involve the use of devices with proprietary interfaces, typically lacking common standards as well. Therefore, negotiations with vendors of such devices are necessary at a national or even an international level to agree on standards or, ideally, open Application Programming Interfaces (APIs) in order to be able to share and aggregate relevant data.

Much data is also closely connected with software and software development. In some disciplines, the software is often the main research output. While this may be more obvious in some research settings such as computational engineering and simulation science, it is also true for diverse types of experimental setups which generate large amounts of raw data (e.g. medical physics and biomedical engineering). Examples include images, videos, audio and device read-outs in various data formats.

A significant feature of science and technology disciplines is their closeness to industrial research and development as a focus on problem-solving is often of direct relevance to commercial interests. However, while academic research is often collaborative and strives to publish and share findings broadly, for commercial actors, confidentiality may bring competitive advantages (although this view is being challenged, as exemplified by the open innovation movement). This friction that sometimes exists between academic openness and industrial confidentiality brings the European Commission guiding principle to mind: data should be "as open as possible, as closed as necessary". Because of these varied influences, good management of research data is not yet a mainstream part of much engineering and related research: in many science and technology disciplines, the implementation of the FAIR principles is still in its infancy.

The backdrop of the results and discussion presented in this white paper is the work of the Task Force Open Science (TFOS), and its subgroup Research Data Management focusing on FAIR and Secure Data. The TFOS advances the understanding and implementation of open science within and outside our network, with a particular focus on Open Access (OA) to scientific publications and RDM. The TFOS collaborates closely with other international groups such as the Research Data Alliance Interest Group 'Research Data Management for Engineering' (IG RDM4Eng). Several of the authors of this white paper are also members of IG RDM4Eng, which aims to help change the culture of handling data within the engineering sector, creating awareness and bridging (sub-)communities and existing initiatives. Moreover, the IG aims to provide a platform for developing consensus on RDM best practices for engineering and to provide compelling arguments for engineers in academia and industry to introduce or to improve RDM practices in their workflows.

To increase our understanding of the common and shared issues that universities of science and technology face while embedding FAIR principles and RDM workflows, the TFOS decided to focus on challenges encountered by RDM support teams at our member institutions. These teams currently develop and provide research data management services that address all phases of the data lifecycle and are often the first point of contact regarding RDM issues for scientists and students. All CESAER member institutions were invited to participate in an online survey in April/May 2019. The TFOS asked for one completed survey per institution with typically several staff members per institution collaborating to complete the response from an institution. The specific target group of this survey were staff members responsible for RDM support and services at the institutions who typically collaborated with other relevant stakeholders within their institution to collect broad input (e.g. IT-services, the research funding office and similar).

This paper aims to show and compare concepts of RDM activities taking place at our member institutions which are all universities of science and technology, and the challenges faced when dealing with demands from students, PhD candidates, researchers and professors while providing RDM support services, training and infrastructure.

2. RDM SURVEY

The survey contained a total of 37 questions organised in four main sections: (i) policy and organisation; (ii) infrastructure and tools; (iii) support services; and (iv) training.

Of the 53 members of CESAER (all invited), 21 institutions (40%; please note that percentages are rounded and therefore do not always add up to exactly 100%) participated in the online survey (N = 21), representing 13 countries (Figure 1).

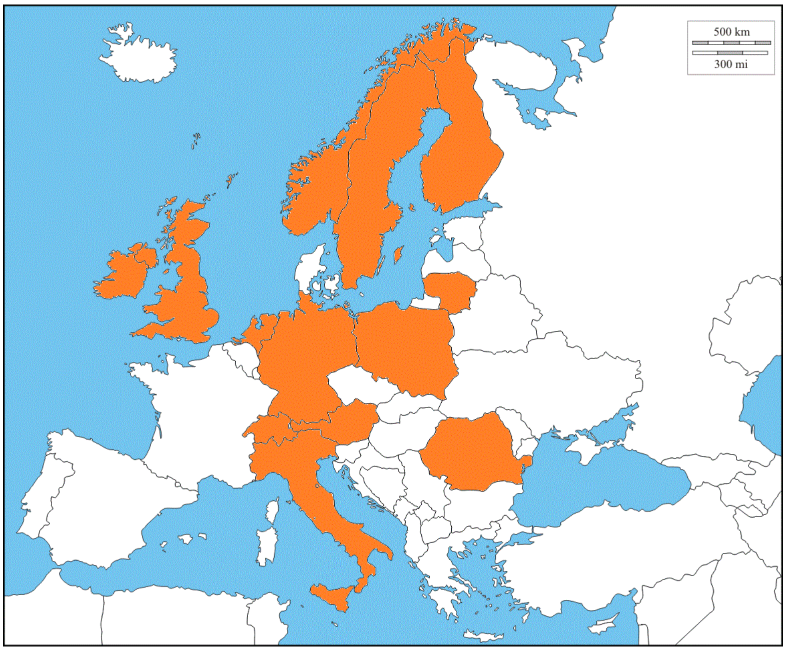

FIGURE 1. IN ORANGE, THE COUNTRIES OF MEMBER INSTITUTIONS PARTICIPATING IN THE ONLINE SURVEY 'ADVANCING RDM AT CESAER INSTITUTIONS'. 21 INSTITUTIONS PARTICIPATED IN THE SURVEY OUT OF 53 MEMBERS (ALL INVITED). THE RESPONDENTS CAME FROM 13 DIFFERENT COUNTRIES.

Individuals and teams, consisting of RDM supervisors, librarians, IT professionals, data curators, open science advisors and metadata specialists, contributed to the responses from each institution. It should be noted that in some cases, the respondents also held more than one role at their institution.

In the following chapters, the main aspects of RDM and associated services within universities of science and technology are analysed. Issues addressed include research data policies and guidelines, RDM infrastructures such as data storage, data archives and other associated tools, as well as training and support services.

The survey questions and the anonymised responses are openly available via http://doi.org/10.5281/zenodo.3616517.

3. POLICY AND ORGANISATION

In 2013, the League of European Research Universities (LERU) published a 'Roadmap for Research Data'. LERU's stated aim was to guide universities in establishing and implementing best practices that contribute to making research data open. Among its recommendations, institutions were called to create high-level steering groups for RDM, and associated roadmaps and policies. Inspired by these recommendations, the first section of our survey posed questions to monitor the status of institutional RDM policies within European universities of science and technology. This included:

- Investigating the position of those policies considering the similarity of disciplines at our Member institutions;
- Exploring how RDM is organised at different institutions;
- Identifying the main challenges encountered when implementing the policies.

3.1. STAKEHOLDERS AND POLICIES

According to the survey results, the implementation of RDM processes is typically overseen by more than one office and stakeholder. Offices and stakeholders involved in this process at the surveyed institutions typically include a combination of the library (at 86% of the respondents), the IT department (at 57% of the respondents), the research (funding) office (at 52% of the respondents) and the legal office (at 14% of the respondents). It may be obvious that libraries have a prominent role, but it is also interesting to note that the management of research data often draws on the coordinated efforts by several fields of expertise. Surprisingly, only one institution had its business and development office involved in overseeing RDM implementation. Considering the intensive focus on innovation and knowledge transfer within universities of science and technology, and the current large focus on societal impact in Europe (e.g. the agenda of the new European Commission), this highlights an area for improvement to encourage such entities to have a more significant role in order to help clarify topics such as ownership and Intellectual Property Rights (IPR) of research data while working with non-academic partners (e.g. for innovation and entrepreneurship).

More than half of the institutions (67%) participating in the survey have already established RDM policies. Only five institutions (24%) do not yet have a policy, from which a couple are currently in the process of preparing a policy. The years of publication of the policies span from 2007 to 2019, but most of them were finalised between 2016 and 2019. The policies published before 2016 (two institutions) are, or have been, under revision since the first date of publication.

The analysis of the policies allowed us to shed more light on institutional concerns about RDM. Fourteen of the respondents shared copies of the policies. As a benchmark, we used the RDM model policy developed by the EU-funded LEARN project. Our overall finding was that, although there were substantial variations in the structure of the documents (e.g. titles, structure and the extent of details of each document was heterogenous), most of the policies address the essential issues for RDM. Six of the documents were not named as 'policies' but 'guidelines', four of the documents contained the word 'open' in the title, and for three documents, RDM was a section of a broader policy (e.g. 'research integrity' or 'open research').

Additionally, each university developed its format of the policy document, and only a few sections are common for all of the surveyed documents: definitions/examples of research data (in 12 policies); definition/explanation of RDM and data stewardship (8); and division of responsibilities (16).

When asked whose responsibilities are established in the policy, the respondents indicated mainly researchers (71%) and institutions (62%), and in less frequent cases, the faculties (33%). Nonetheless, in many cases, there was a lack of detail over specific responsibilities (something indicated as useful by the LEARN template). So, while researchers were often mentioned (but without clear guidance of what they needed to do to comply with the policy), very few policies mention the role of project managers, heads of research units and PhD candidates. We strongly recommend that a successful RDM policy specifies who is responsible for each activity.

The terms used in the RDM policies were also investigated. Participants were asked which terms are explicitly mentioned in their RDM policy (Fig. 2), drawing from a list of 16 RDM topics. A free-text box was also provided for other topics not included in the list. The most common topics were: data preservation (62%), reproducibility (57%), data sharing (57%) and data management plans (57%). Open science and General Data Protection Regulation (GDPR) did not appear frequently (both 24%). Terms such as FAIR data, research integrity, open data, IPR, data ownership and data licenses are explicitly mentioned in just under half of the policies (frequency between 38% and 48% of policies).

Anything to do with software and code was notable by its absence from most of the policies, whether that be software management (in 14% of policies), licenses (10%), the publication (20%) or the preservation (20%) of software and code. In seven cases, software or source code was explicitly mentioned within the definition of research data. In four of the other policies, the presence of software can be inferred. There are ongoing discussions on software within research data management communities over different procedures and tools to manage it compared to digital data or physical materials. Software and code are important research outputs within science and technology disciplines, as they underpin much data processing and analysis. We recommend the inclusion of concrete software and code policies and guidelines for RDM, addressing its proper documentation, publication, licensing and preservation. There are still uncertainties regarding software source code, and in our experience, some researchers see the publication of source code resulting from their research as a potential breach of IPR.

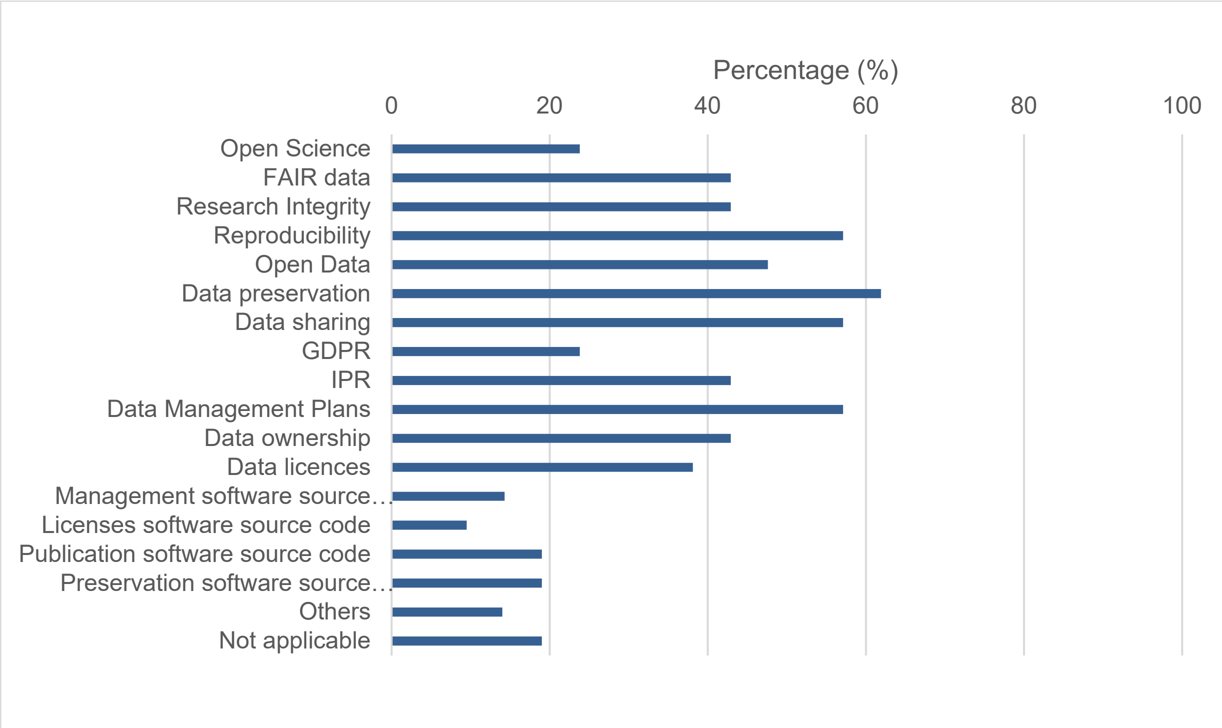

FIGURE 2. RESPONSES TO THE QUESTION: 'DOES THE RDM POLICY EXPLICITLY MENTION ANY OF THESE TERMS?' THE GRAPH SHOWS THE PERCENTAGE (%) OF INSTITUTIONS THAT EXPLICITLY MENTION IN THEIR RDM POLICIES THE LISTED TERMS.

Due to the importance of collaboration with industry for universities of science and technology, topics such as data ownership and IPR are extremely relevant. However, there were very few explicit recommendations (or links to other guidelines) embedded in the institutional policies surveyed. Five of the policies give an indication that data ownership (including IPR) belongs to the institution where the data is produced, including within externally funded projects. However, there were few explicit guidelines on how to relate data to IPR (e.g. patents and licenses) or how data resulting from partnership projects with industry should be stored, accessed and used. We, therefore, recommend that the development of guidelines for the implementation of RDM, following FAIR principles, while working with industry, should be high on the agenda.

3.2. POLICY IMPLEMENTATION

Publishing a policy is one thing, the extent to which researchers comply with it is another. It is during implementation that many challenges arise. We included three questions to investigate how our Members are dealing with policy implementation in this area.

When asked how the institution is implementing their RDM policy (where it exists) most of the respondents indicated that they provide centrally organised support (81%), training (81%) and infrastructure (76%) (Figure 3).

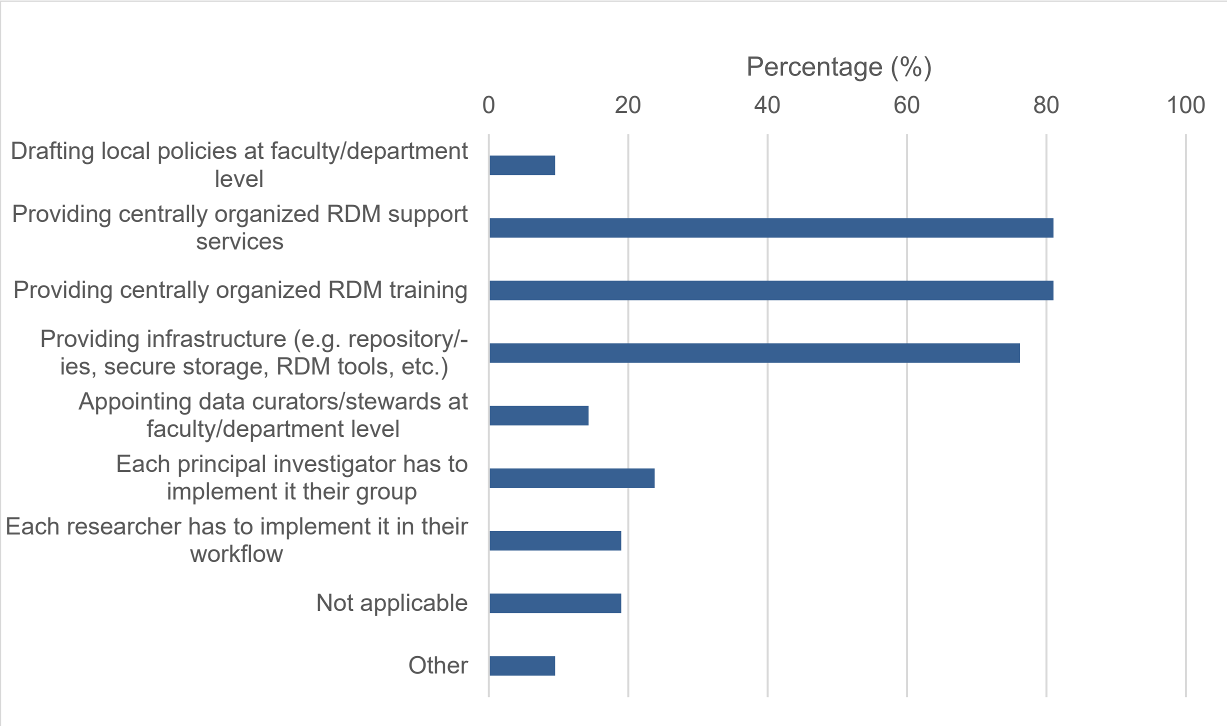

FIGURE 3. RESPONSES TO THE QUESTION: 'HOW IS YOUR INSTITUTION IMPLEMENTING THE RDM POLICY?' THE GRAPH SHOWS THE PERCENTAGE (%) OF INSTITUTIONS IMPLEMENTING THEIR POLICIES THROUGH THE LISTED ALTERNATIVES.

Only a few of the institutions were implementing the policy at a more 'local level', such as faculties or departments through the drafting of local policies (10%) or appointing data curators/stewards (14%) at those levels. That implementation and guidance is often missing at this level is a clear indication that there is still work to do in providing more discipline-specific support and guidance.

We also asked if there were initiatives to implement the policy targeting researchers directly, for example, through the provision of incentives (Figure 4). Almost half (48%) of the respondents said that their institutions do not provide formal incentives for researchers to comply with the policy. From the few positive responses, there are two types of incentives offered: (i) financial support to hire data stewards/data managers in research projects (14%), and (ii) for implementing RDM best practices in research groups (10%). Except for one university, good practice in RDM is not formally rewarded in the recruitment or promotion of academic staff.

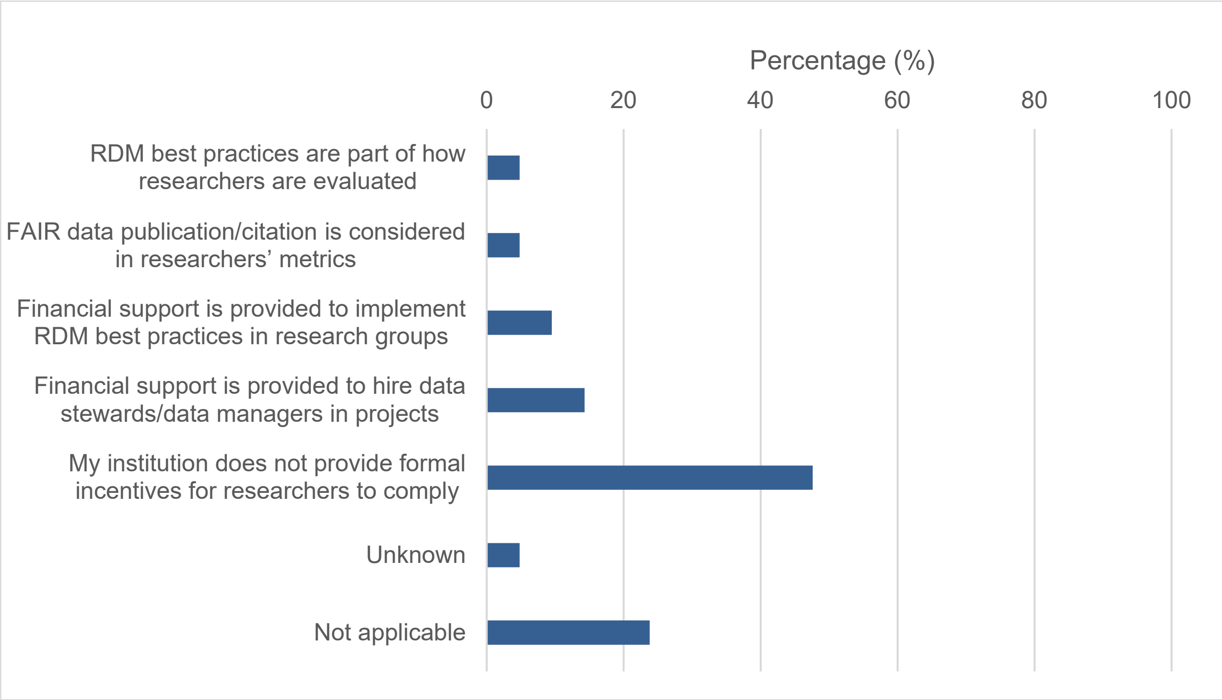

FIGURE 4. RESPONSES TO THE QUESTION: 'WHICH OF THE FOLLOWING INCENTIVES DOES YOUR INSTITUTION PROVIDE FOR RESEARCHERS TO COMPLY WITH THE RDM POLICY?' THE GRAPH SHOWS THE PERCENTAGE (%) OF INSTITUTIONS PROVIDING THE INCENTIVES LISTED ON THE X-AXIS.

Finally, participants were asked to indicate what the main challenges encountered during that process are. They could choose from seven defined issues, or select the option 'others' where additional challenges could be entered in free-text format. 'Lack of incentives for researchers to comply with the policy' was selected at the highest frequency (57%). A clear recommendation is therefore that incentives for researchers to comply with RDM policies are put in place.

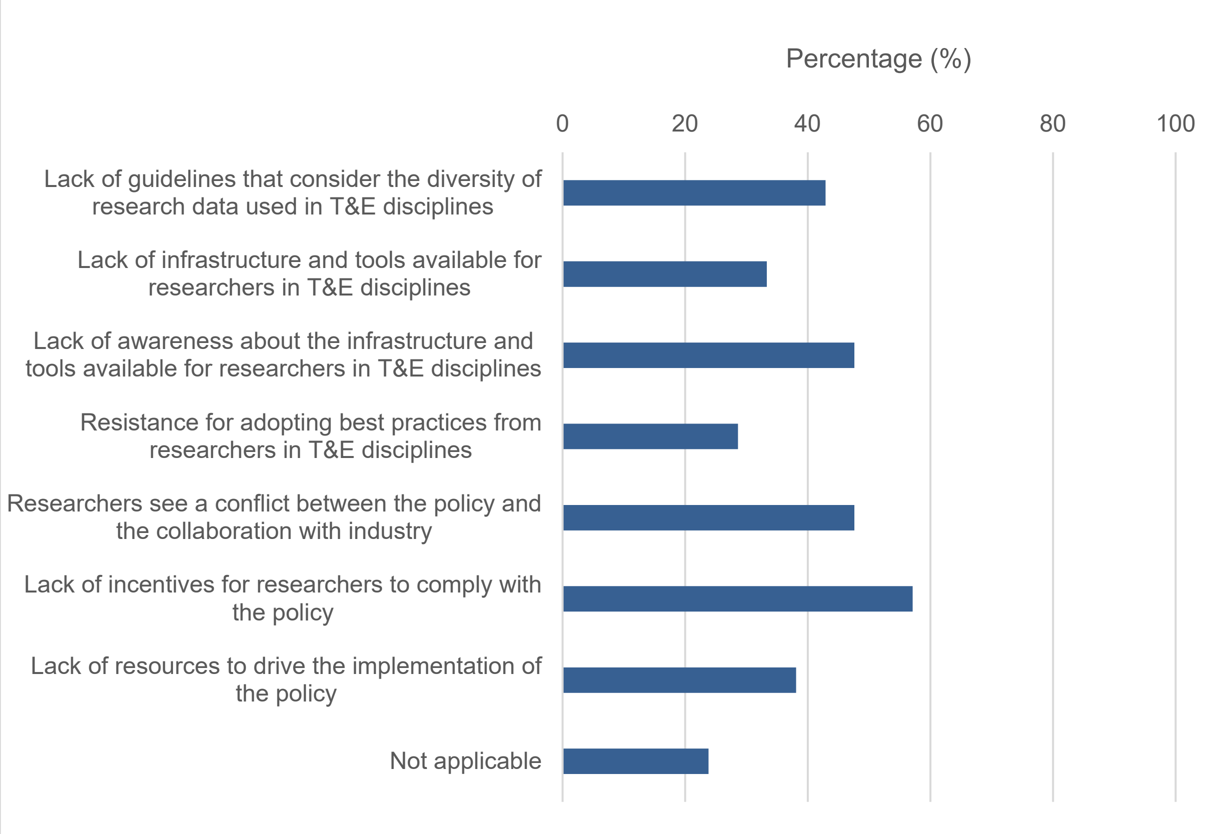

FIGURE 5. RESPONSES TO QUESTION: 'AS A TECHNICAL UNIVERSITY, WHAT ARE THE CHALLENGES FOR THE IMPLEMENTATION OF THE RDM POLICY?' THE GRAPH SHOWS THE PERCENTAGE (%) OF INSTITUTIONS THAT INDICATED THE LISTED ALTERNATIVES AS A CHALLENGE. THE ABBREVIATION 'T&E' MEANS 'TECHNICAL & ENGINEERING'.

Respondents also stated that researchers see a conflict between the RDM policy and collaboration with industry (48%). Once again, this brings to the fore the need for clarification regarding IPR considerations and the data and results obtained in collaborative projects.

Two other challenges identified were data infrastructure and tools. The development and provision of infrastructure remains a large focus at many institutional, national and even European levels (e.g. the EOSC). Despite this committed investment of resources and the related advancement in data infrastructure, lack of awareness among researchers about the availability of such infrastructure and tools is a clear challenge (as identified by 48% of respondents). On the other hand, some of the respondents (38%) identified insufficient infrastructure and tools available for researchers working in technical/engineering disciplines as a challenge.

Lack of resources was also identified as a challenge (38%). However, only two of the surveyed policy documents specified costs concerning storage, access and re-use of research data. Both documents identified researchers as responsible for their research data and urged them to foresee expenditure related to basic data management, such as storage and publishing, in their resource management of research projects (including in external grant applications). Both of these documents also mention that universities should support the preparation and curation of data sets as well as provide special support for researchers who have 'big data' storage requirements. Other institutions may have dealt with the costs of RDM in other documents, but given the need for researchers to consider the financial implications of managing and archiving data, it is telling that there are few mentions of this in current RDM policies.

4. RESEARCH DATA INFRASTRUCTURE

A reliable research data infrastructure is a central component of any RDM service. For satisfactory data management, different types of data storage, access and backup are required. Besides, the provision of tools for planning, analysing and documenting research data is fundamental when implementing best practices within research workflows. According to the results of our survey (Figure 3), institutions have focused on providing data infrastructure (e.g. repositories, secure storage and RDM tools) as a way of implementing RDM policies. In order to gain a deeper understanding about the data infrastructure already available and to identify possible shortcomings and gaps that need to be discussed and addressed, the survey contained eleven questions about institutional research data storage, research data repositories and research data management tools.

4.1. INSTITUTIONAL RESEARCH DATA STORAGE

The institutions were asked if they currently provide research data storage for their researchers. In this question, research data storage refers to temporary/project-related storage and not long-term storage thus making a distinction with data repositories/archives. The results indicate that 67% of the institutions provide data storage for researchers, while 29% do not offer such infrastructure and 5% did not know if it was offered.

The main challenges in hosting research data produced in technical/engineering disciplines was also a topic. When analysing the answers to this free-text question, it was possible to identify several recurring themes (Figure 6).

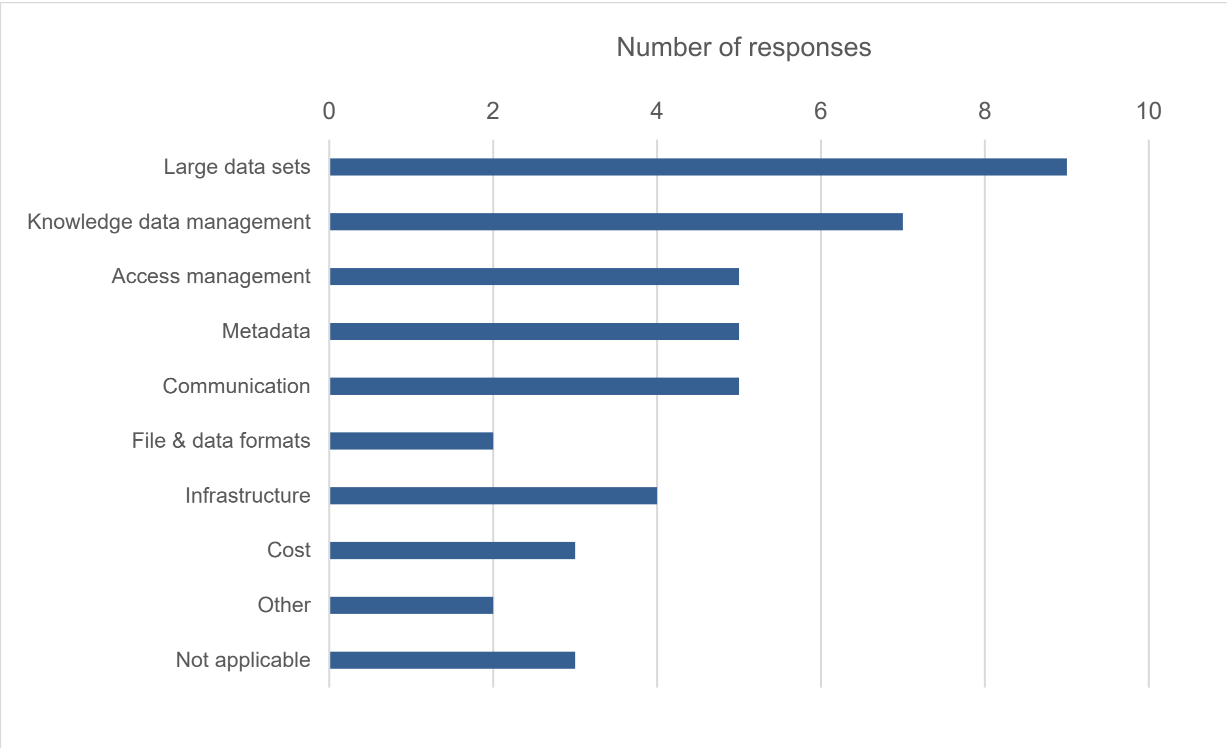

FIGURE 6. RESPONSES TO THE QUESTION: 'WHAT ARE THE MAIN CHALLENGES IN HOSTING RESEARCH DATA PRODUCED IN TECHNICAL/ENGINEERING DISCIPLINES ON YOUR INSTITUTIONAL STORAGE?'. THE RESPONSES WERE PROVIDED IN FREE TEXT. THE RESPONSES WERE GROUPED IN RELEVANT TOPICS (X-AXIS) AND THE BARS REPRESENT THE FREQUENCY THAT THE INSTITUTIONS MENTIONED THIS TOPIC.

For most respondents (9), large data sets were seen as the main challenge in institutional storage solutions. This includes problems with restricted storage capacities or large and non-transferable datasets. For seven of the respondents, the knowledge about data management and handling is seen as a main challenge. The respondents refer to aspects like knowledge about data life cycle management, data quality and data documentation. Five respondents mentioned challenges related to access management, metadata standards and communication (like strategies or guidelines) for data hosting.

The respondents were also asked to indicate if they found solutions for the challenges described in the previous question. Most of the institutions are actively working on solving these challenges, and some recurring themes from the answers are summarised here:

- Large data sets. To solve challenges regarding large data sets, respondents referred to their development of (big) data platforms or separate data environments beyond institutional storage.
- Metadata. To solve metadata challenges, some universities use (meta) data management systems. They either collect existing metadata from publications or work on the development of metadata recommending tools and automated metadata extraction.
- Knowledge and communication for access management and data management. To overcome missing knowledge in the whole life-cycle of data management, some universities are in the process of setting-up data management systems, internal knowledge-based websites and platforms (e.g. WIKI, Electronic Laboratory Notebooks (ELNs), Moodle) for documentation of standards and discipline-specific workflows, as well as increasing awareness through personalised workshops and training. At some universities, researchers also tend to create their own solutions or buy third party solutions.

4.2. HOSTING RESEARCH DATA IN REPOSITORIES

Repositories enable sharing, storing and publishing of digital resources. A distinction can be made between institutional and subject-specific repositories. Institutional repositories are offered by the institution, normally data from various disciplines can be deposited and made available according to defined conditions. The advantage is that the data is stored at the institution's location. Data description, however, is sometimes perceived as a challenge because standards within the disciplines vary a lot. Discipline-specific repositories can help to develop standards and are often better recognised in the discipline.

Currently, around half of the responding institutions (52%) are providing a data repository or archive for publishing and preservation of research data produced at their institution, although two institutions provide such infrastructure only for certain types of data or disciplines and small data sets.

When the participants were asked to identify the main challenges they face when hosting research data produced in science and technology disciplines, again the size of data sets and the lack of metadata standards were identified (Figure 7).

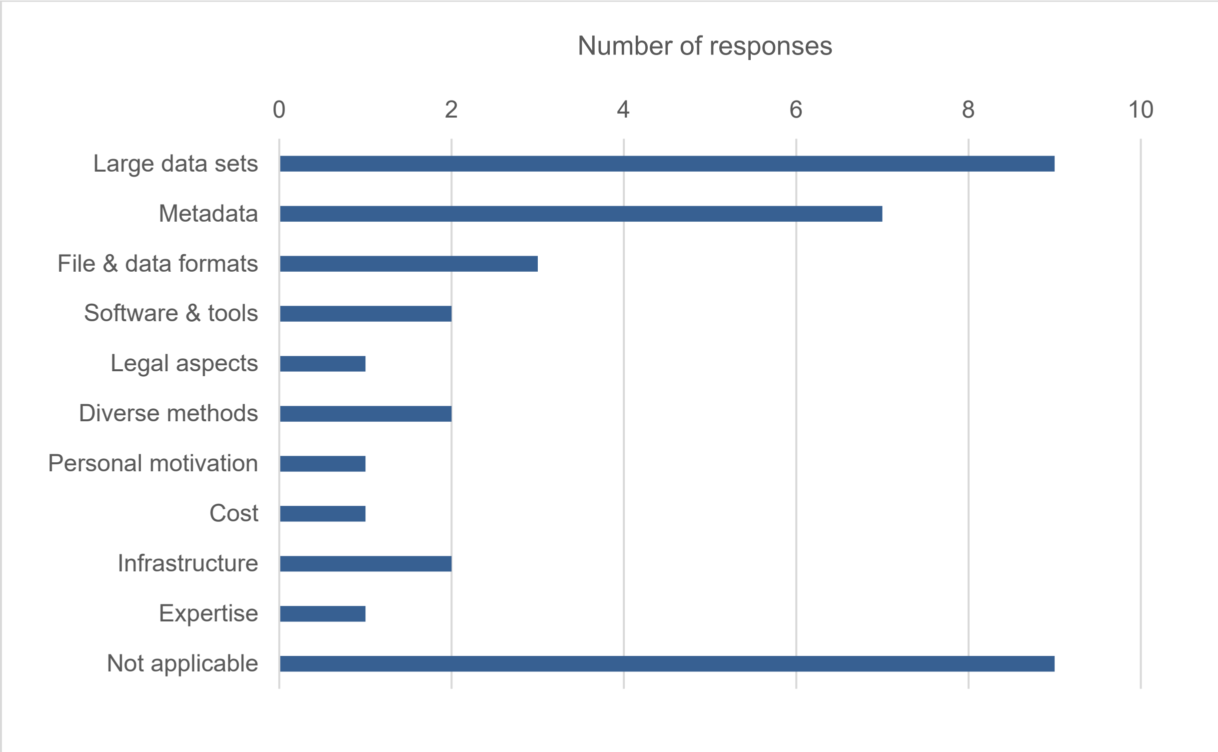

FIGURE 7. RESPONSES TO THE QUESTION: 'WHAT ARE THE MAIN CHALLENGES FOR HOSTING RESEARCH DATA PRODUCED IN TECHNICAL/ENGINEERING DISCIPLINES IN YOUR RESEARCH DATA REPOSITORY/ARCHIVE?'. THE RESPONSES WERE PROVIDED AS A FREE TEXT. THE RESPONSES WERE GROUPED IN RELEVANT TOPICS (X-AXIS), AND THE BARS REPRESENT THE NUMBER OF RESPONSES MENTIONING EACH.

Apart from this, the lack of knowledge about different file and data formats was perceived as challenging, as well as missing knowledge about software and tools for data curation.

To get a better understanding of the landscape of data repositories and archives oriented to science and technology disciplines, the participants were asked if they recommend specific solutions to their researchers. Nearly half (48%) of the respondents say they do recommend specific data repositories and archives, while 43% do not, and 10% do not know if specific recommendations are given to researchers.

Respondents were also invited to indicate the recommended platforms (as a free text answer). Although the question asked for recommendations for specifically engineering and technical disciplines, several respondents indicated that they recommend Zenodo (5 responses), which is a general repository. Next, the global Register of Research Data Repositories re3data was mentioned (4 responses) as a source to find appropriate data repositories and archives for specific disciplines. Further answers were diverse. Individual respondents also referred to general repositories like Figshare and Dryad, national platforms like MIDAS and UK data archive, the institutional archive 4TU.ResearchData or referred to the European Research Council guidelines as a source of recommendations. Only two discipline-specific data repositories were explicitly mentioned as solutions recommended for researchers: Nomad (a data repository for materials data) and the Cambridge Crystallographic Data Centre. These responses show that there is no single place or solution for science and technology disciplines to deposit data, but a mixture of repositories that fulfil the requirements for the publication specific data types.

When depositing data in a repository, a minimum set of metadata is needed in order to comply with the FAIR data principles. For example, to provide a persistent Digital Object Identifier (DOI), six mandatory fields of the DataCite metadata standard must be filled in: identifier, creator, title, publisher, publication year and resource type. However, it is strongly recommended to provide more metadata that describes a dataset clearly and concisely while aligning to the expectations of the research community for which it was created. However, not every discipline has the necessary subject-specific metadata schema in place, although good practice examples exist (for examples see the Metadata Directory and FAIRsharing) further work within research communities is needed to create metadata standards and promote their uptake.

4.3. RDM TOOLS

Providing tools for researchers in order to plan, document and analyse their data using best practices are as relevant as data storage, publication and preservation infrastructure to ensure rigorous and reproducible research. The participants in the survey were asked to indicate which RDM tools are already provided by their institution and, whenever possible, to name them explicitly in a free text field. More than half of the institutions refer to data management planning tools (e.g. DMPonline, dmptool.org, RDMO and some other national instances), HPC systems and version control systems (e.g. SVN, Git) (Figure 8). As for documentation tools, 33% of the institutions offer Jupyter Notebooks intended for sharing executable computer code enriched with text elements (e.g. paragraph, equations, links, etc.) and 27% offer ELNs (e.g. RSpace, eLabjournal, SLIMS, and some in-house developed solutions) for documentation of experimental procedures and data.

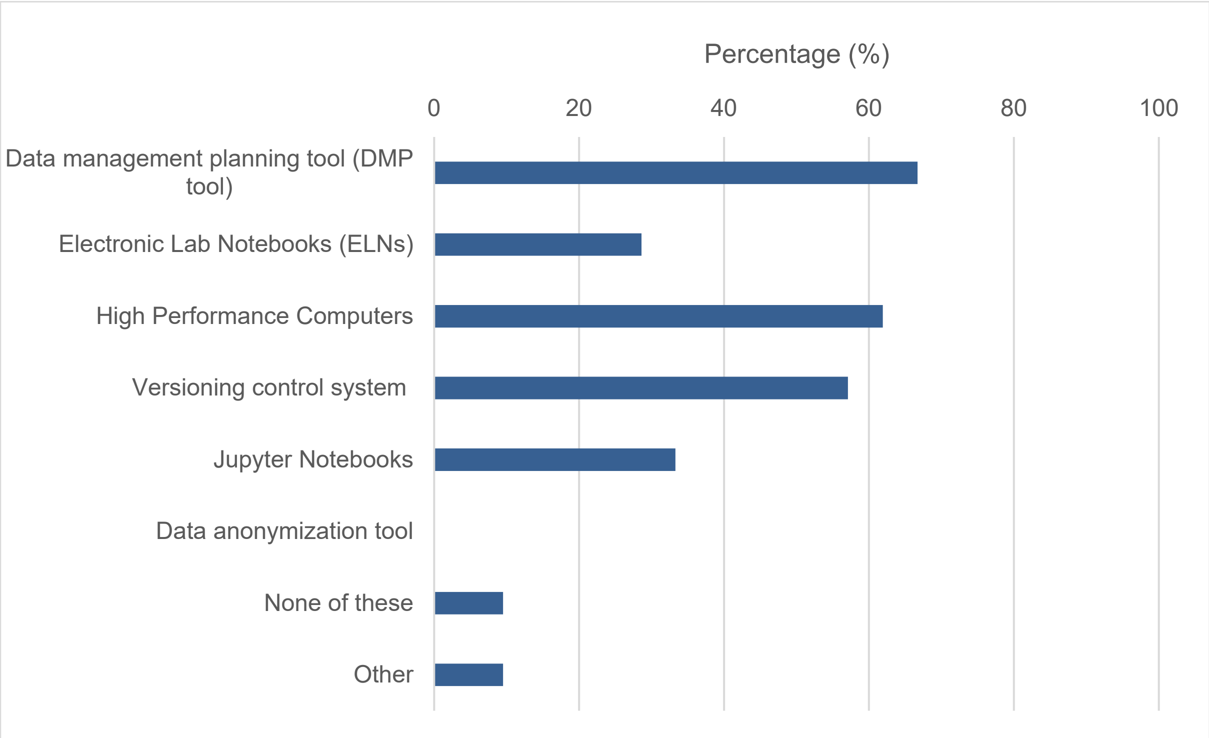

FIGURE 8. RESPONSES TO QUESTION: 'WHICH RDM TOOLS ARE PROVIDED BY YOUR INSTITUTION?'. THE GRAPH SHOWS THE PERCENTAGE (%) OF INSTITUTIONS THAT PROVIDE THE CORRESPONDING RDM TOOLS LISTED IN THE X-AXIS.

These results reveal that institutions are making a big effort to provide tools to facilitate the workflows of researchers while complying with funder requirements (e.g. DMP tools) and following best practices (e.g. data documentation and versioning of source code). However, when asked if specialised RDM tools and infrastructure are provided to researchers working in technical and engineering disciplines in their institutions, only 24% of the respondents answered in the affirmative, while most of the institutions (57%) are not (yet) providing specialised RDM tools, or lack the knowledge regarding specific tools offered by their institution. A free text field allowed the respondents to indicate the name of those specific tools targeting technical/engineering disciplines. The following tools were mentioned as examples by the respondents: a metadata harvester (in development) and a metadata manager, plus a specific data portal (for netCDF data) and a lab-data management tool.

5. SUPPORT SERVICES AND TRAINING

Together with establishing RDM policies and providing the relevant data infrastructure, universities recognise the need to support researchers in implementing those policies and making good use of the infrastructure provided. According to the results of the survey presented above, over 80% of the institutions are implementing their RDM policies through the provision of centralised RDM support and training. The third and fourth sections of the survey consisted of a total of 17 questions to gather knowledge about the type of services and training available at our member institutions. In particular, we explored the type of support that is specifically offered for science and technology disciplines, and what challenges can be identified.

5.1. RDM SERVICES AND CHALLENGES

The survey results indicate that over 80% of the respondents have a dedicated team to provide specific RDM services, a good indication that universities of science and technology acknowledge the relevance of providing RDM support. Additionally, two other respondents indicated that such a team is being set up and another one that although there is an RDM support team, the staff members are not only dedicated to this function. Only one institution declared that they do not have a centralised RDM support team. The composition of those teams highlights the crucial role of librarians (for 86% of respondents) and IT-managers (67%) in providing such support, which correlates with the offices that mainly oversee the implementation of RDM at the different institutions (Chapter 3).

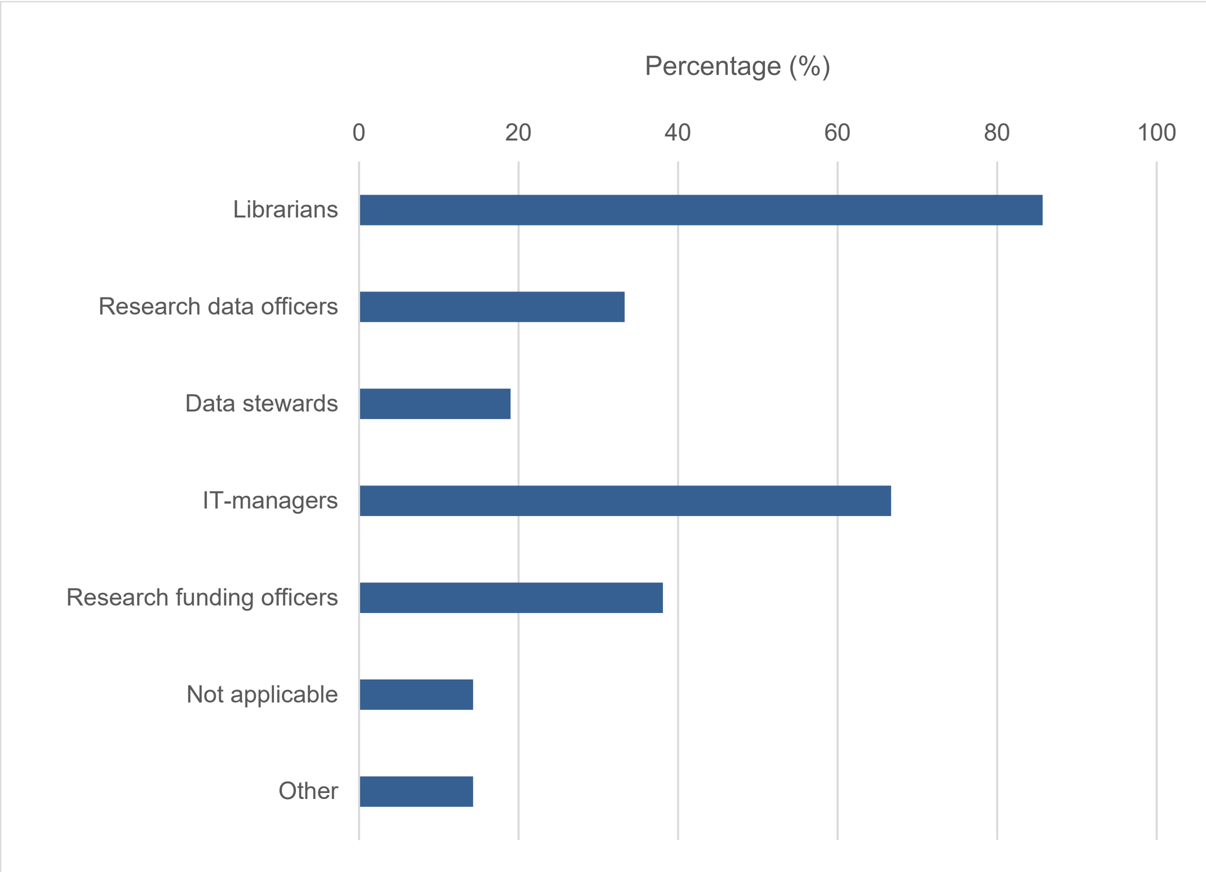

FIGURE 9. RESPONSES TO THE QUESTION: 'THE RDM TEAM INCLUDES'. THE GRAPH SHOWS THE PERCENTAGE OF INSTITUTIONS IN WHICH THE RDM TEAMS CONTAIN STAFF WITH THE DIFFERENT EXPERTISE LISTED.

The participants in the survey were also asked if they consider that there are skills missing in their RDM support team that would facilitate the task of providing support to technical and engineering disciplines. 14% of the respondents answered that they have all the needed skills, 33% do not know if they are missing skills within the team and 52% pointed out that some skills are missing. For example, six respondents mentioned that there is a lack of discipline-specific technical knowledge (including data types, research methodology and workflows) and training, which are tasks that are typically assigned to data stewards. This response correlates with the few institutions (19% of respondents) that have data stewards within their RDM teams (Figure 9). Technical skills (e.g. data curation, data integrity, software) and knowledge of legal aspects concerning data (e.g. copyright, privacy and data integrity) were also mentioned.

Besides looking at the structure of the RDM support services, we aimed at investigating the most relevant topics and challenges that these teams encounter within their day-to-day interaction with researchers working in technical and engineering disciplines.

The participants were asked to list the three most difficult RDM-related topics to address when communicating with researchers working in technical and engineering disciplines. Participants could reply using free text and the responses were diverse. Nevertheless, the following groups of topics were identified (in decreasing order of frequency):

- Documentation and metadata. When providing a more complete answer, the participants mentioned that challenges are encountered due to the lack of standards both for documentation and metadata. The lack of structured documentation, or simply enough documentation in order to publish data, can be a challenging request for researchers.
- Data Management Plans (DMP). Several respondents acknowledged the challenge of getting researchers to create and implement DMPs. One identified reason was lack of incentives, including not being recognised or rewarded in evaluations and results and development cycles.
- Data organisation and management. This covered several areas from creating awareness about the importance of RDM to implementing best practices on research workflows and discussing topics going around the whole data life cycle.
- Legal issues. Discussions around data ownership, IPR, copyright, data licenses and personal data are challenging when communicating with researchers.
- Technical issues. Including big data, the complexity of data sets, and selection criteria for determining the needed storage period of data.
- Software. In particular source code licensing, sharing and documentation were mentioned.

Other topics mentioned at least twice were industrial collaboration, incentives and collaboration within (large) projects. Regarding industrial collaboration, nearly 90% of respondents believed it to be ‘relevant’ or ‘very relevant’ for the researchers at their institutions, but there are several challenges when providing support to those researchers that collaborate with industry (Figure 10).

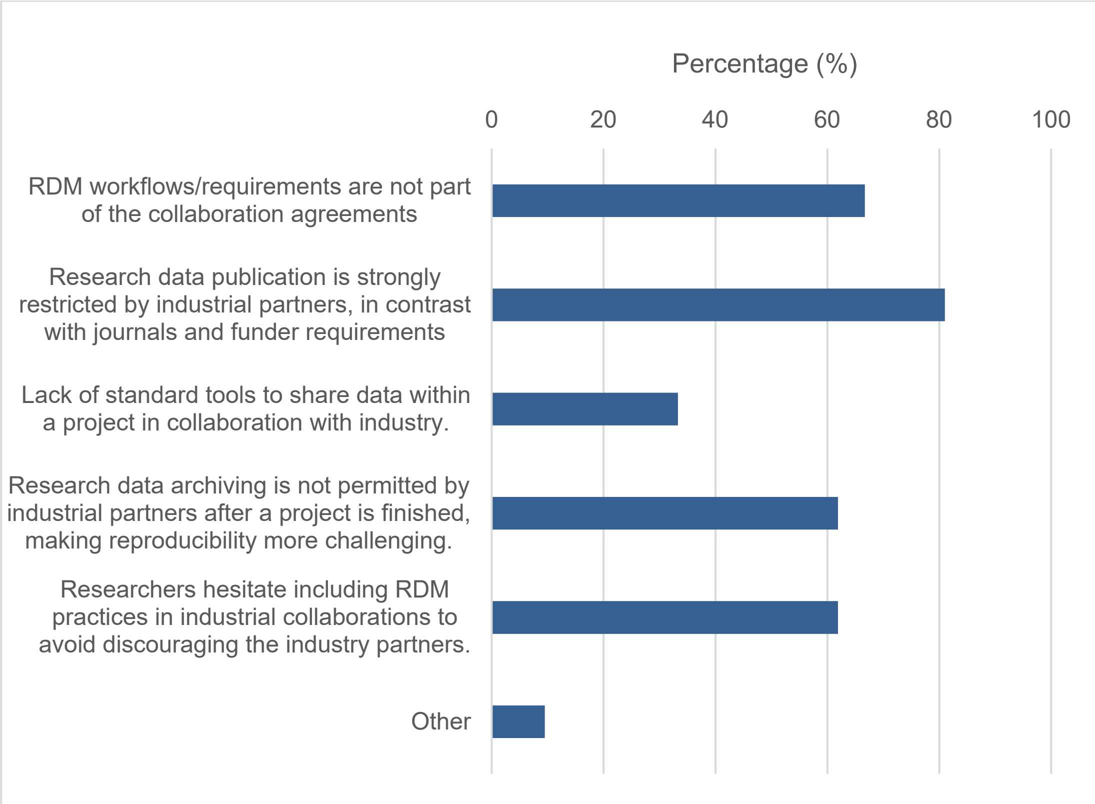

FIGURE 10. RESPONSES TO THE QUESTION: 'WHAT ARE THE MAJOR CHALLENGES IN PROVIDING RDM SUPPORT TO THOSE RESEARCHERS THAT COLLABORATE WITH INDUSTRY?'. THE GRAPH INDICATES THE PERCENTAGE OF INSTITUTIONS THAT IDENTIFIED THE LISTED ALTERNATIVES AS A CHALLENGE WHEN PROVIDING SUPPORT TO RESEARCHERS COLLABORATING WITH INDUSTRY.

81% of the respondents answered that the sharing of research data when working with industrial partners was a major issue as these partners often request restrictions on data sharing. This is in stark contrast to many funder and journal requirements which encourage or require the sharing of research data. Over 60% of the institutions stated that archiving research data after the end of a project was not permitted when collaborating with industry, which is a major barrier.

This challenge is exacerbated by the fact that for over 60% of the respondents, RDM workflows and requirements were not reported as being standard parts of industry collaboration agreements. A potential reason is that this may be seen as discouraging to industry partners. The evidence gathered here shows that there is a clear need for a dialogue between industry partners and the institutions (for example, with the input from business development unit or innovation offices at universities) on how to incorporate good RDM practices in project collaborations.

Participants were also asked to identify specific RDM tools, policies and guidelines which would be needed to provide RDM support tailored to technical and engineering disciplines. In 52% of the cases (11 institutions), participants answered listing several needs that could be classified into three main categories:

- Infrastructures and policies. The needs in this category were mostly generic in their scope and include repository, policy, cloud services, and automated metadata generation.
- Support and training. The needs were expressed in terms of data stewardship, *ad hoc* training and DMP preparation. In particular, three institutions highlighted the need for tailoring DMP to the research workflow and nature of data, e.g. measurement- or simulation-based.

- Specific tools for industrial projects. Four answers emphasised the needs of *ad hoc* guidelines for RDM, including rights and access to the data, for collaborative projects with industrial partners.

In conclusion, the issue of industrial cooperation and discipline-specific training is one of the most important aspects when it comes to subject-specific RDM support. As such, we were interested in identifying those engineering and technical disciplines that are perceived as more advanced or more hesitant in implementing good RDM practices. Answers were so diverse that it is difficult to draw any general conclusion, but some trends could be observed. Some subfields within chemistry and physics (e.g. high-energy physics), and some data-intensive disciplines, such as those related to environmental sciences, were perceived as more advanced with regards to RDM and data sharing. Subfields in some conventional engineering and technical disciplines, such as civil and mechanical engineering and architecture sub-disciplines, were seen as more hesitant to RDM and data sharing. Overall, the analysis makes it evident that there is no consensus on the RDM status in single disciplines and that this is often more dependent on the specific research community at a particular institution than on the discipline in itself.

5.2. RDM TRAINING

Participants were asked which topics were taught regularly as part of the training provided at their institutions. Results showed (Figure 11) that training is provided in most of the institutions and covers different RDM topics. For example, introduction to RDM (at 81% of respondents), data management plans (76%), data documentation (52%) and data preservation and publishing (52%). Only in a few institutions are software management (33%) and software sustainability (19%) routinely taught.

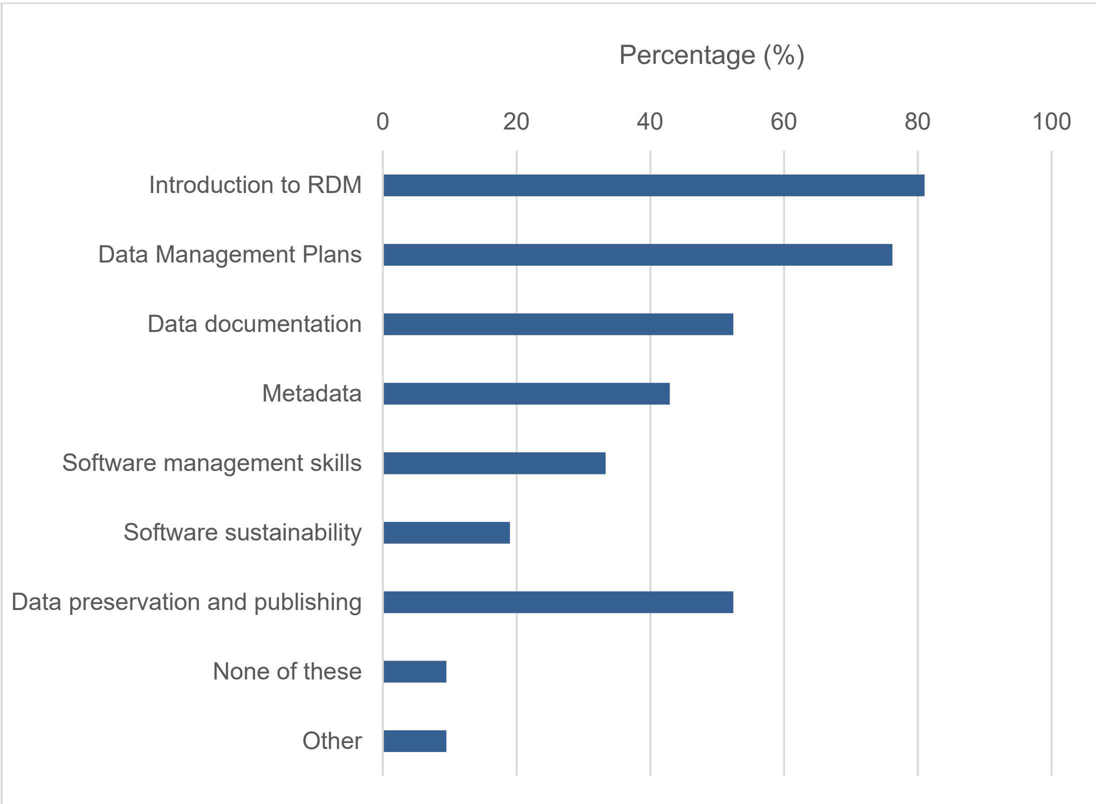

FIGURE 11. THE GRAPH SHOWS THE PERCENTAGE (%) OF INSTITUTIONS REGULARLY OFFERING RDM TRAINING ON THE LISTED TOPICS.

When asked if any training tailored to technical and engineering disciplines was given regularly, only 43% responded positively. Within this, 43% of the institutions offer training covering RDM practices including FAIR data, DMP, GDPR and ethical considerations when working with data. Three institutions mentioned they provide regular training on topics related to software management (e.g. software carpentries, good coding practices, Git and R workshops). On a more specific level, training courses on quality management and intellectual property and capitalisation of knowledge were mentioned by only one institution.

The survey results also showed that the participants in these tailored training sessions are mostly PhD candidates and postdoctoral researchers (Figure 12). Few institutions were able to engage principal investigators (only 29% of respondents) and engineers (24%). The participants were not asked directly if this was perceived as a challenge, but our experience shows that the engagement of principal investigators and engineers should be considered as relevant as training students in an institution in order to ensure successful implementation of the policy.

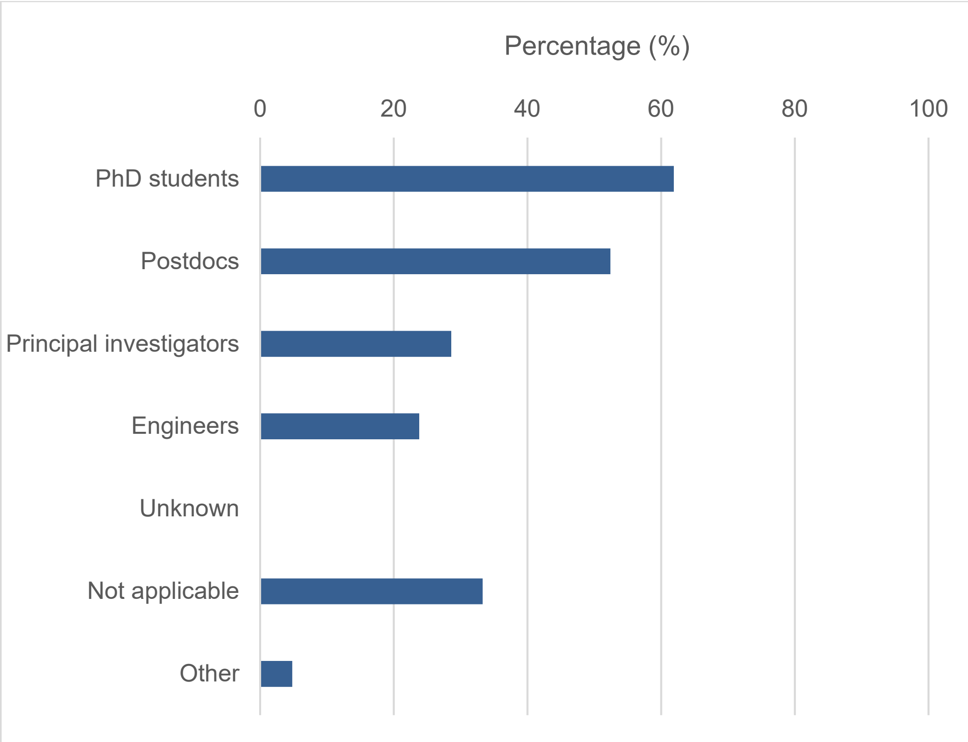

FIGURE. 12. RESPONSES TO THE QUESTION: 'WHO ATTENDS THESE TECHNICAL/ENGINEERING TAILORED TRAINING SESSIONS?' THE GRAPH SHOWS THE ATTENDING DISTRIBUTION (%) TO THE TECHNICAL/ENGINEERING TAILORED TRAINING SESSIONS.

A higher number of institutions provide training tailored to technical and engineering disciplines on request (57% of respondents). Between 2018 and 2019, courses were offered on 21 different RDM topics among the respondents. Interestingly, most topics are only mentioned once or twice when looking at the list of courses and the topics are very differently distributed across the institutions. Training on general RDM topics including data organisation, data documentation and metadata, FAIR data and DMPs are typically provided on request and tailored to the audience that requests them.

The two major challenges in providing tailored training were: (i) lack of trainers with discipline-specific knowledge (stated by 62% of respondents) and (ii) lack of interest from researchers (52%). As a third major challenge, the lack of discipline-specific material for RDM practices (24%) was mentioned.

6. CONCLUSIONS

This white paper presents insights from our survey ‘Advancing RDM practices at CESAER institutions’ conducted in 2019 and our recommendations based on those insights combined with our expertise in the area of RDM at universities of science and technology. The objective of the survey was to determine the status of RDM services provided at universities of science and technology. The results help identify the challenges that staff members of RDM support teams are confronted with when interacting with researchers from science and technology-focused disciplines, and good practices that have already been deployed in this area.

The results of the survey show that, in general, our Members are making important progress in establishing robust policies, infrastructure and support services around RDM. It is also clear that provision of good RDM support involves several offices and departments from across the university and strong support from central administration.

The results of the survey further illustrate that data produced and collected within science and technology disciplines pose several specific challenges regarding data infrastructure. Additionally, the multidisciplinary methods and the extensive collaborations (including with non-academic and commercial actors) common in science and technology disciplines provide barriers that require specific considerations for the full implementation of RDM policies and practices.

6.1. CHALLENGES

This paper has explored the following specific challenges:

- Current RDM policies ask researchers to be aware of IPR and data ownership issues. However, most policies do not provide specific instructions (or links to relevant guidelines) about who should own the data or who (should) hold the rights to them. This topic is particularly relevant for science and technology disciplines due to their close collaborations with commercial partners and the industrial sector.
- Software, code and related data practices are rarely mentioned in RDM policies. This poses a challenge for science and technology disciplines as software often has a critical role as both a tool and a research output.
- RDM policies are often implemented through the provision of centralised RDM support, data infrastructure and training. However, policies and support at a faculty level are generally lacking. Creating faculty and departmental guidelines from university-wide efforts would allow more specific and detailed guidelines and support which would accelerate the full implementation of RDM practices.
- There are far too few incentives for researchers that reward and incentivise implementation of RDM practices into everyday workflow.
- Lack of metadata standards is a major barrier for implementation of best RDM practices.
- The large size of the data produced and collected within science and technology disciplines puts strict requirements on the infrastructure needed for storing and sharing it.
- Generic RDM tools are often provided as part of central services at the surveyed institutions. However, there are only a couple of universities providing tools specific to science and technology disciplines.
- There is a general lack of trainers in RDM practices with the needed discipline-specific background and knowledge.

6.2. RECOMMENDATIONS

Using the landscaping provided in this white paper, the challenges identified in the section above, and guided by the expertise represented in TFOS and the diverse good practices deployed at our Members, we provide the following recommendations:

Develop faculty or departmental-level sub-policies and specific guidelines that are complementary to university-wide RDM policies, specifically targeting:

- Software: institutions should develop a clear workflow that guides researchers on aspects such as software licensing, publication and preservation.
- Discipline-specific workflows: institutions should provide information tailored to science and technology disciplines, e.g. data infrastructures available for the different types of data produced, different tools for documentation, implications of producing data following the FAIR principles, and when and how to publish their research data. In essence, help researchers make better sense of high-level (university-wide) requirements.

Work towards the full implementation of RDM practices in industry collaborations, including:

- Correct common misunderstandings and ensure that all partners understand the guiding FAIR principles. For example, even heavily protected and private data can follow the FAIR principles, e.g. by specifying the conditions under which data are accessible.
- Clarify the ownership of data produced at the university and within the different research collaborations, and specify the terms of use of those data by the different relevant stakeholders.
- Ensure that data and RDM has a dedicated section in collaboration agreements with industry.
- Establish a dialogue with industry partners to embed good RDM practices within collaborative projects that ensure reproducible research practices, including following the FAIR principles in the production and processing of data sets.

Provide incentives that reward researchers for implementing good practices in RDM, including:

- Departmental evaluations should include progress towards implementing good practices in RDM (e.g. availability of training, mentoring programmes and provision of support).
- Recruitment, career assessment and promotion of academic staff should consider and reward experience and usage of good RDM practices. This could include asking researchers to provide examples (during the application or assessment procedure) of their research outputs which follow the FAIR principles.

Prioritise embedding of data stewards and data managers in research teams and projects, and provide resources and information about external funding opportunities to achieve this.

Work towards the optimisation of data storage and preservation infrastructure and workflows, including:

- Implement storage facilities, sharing platforms and data repositories that can handle larger data sets, e.g. in the terabyte range.
- Use solutions with open APIs to facilitate the integration of relevant tools and software and to safeguard long-term function.
- Establish data storage guidelines that help researchers to determine which data must be stored and for how long, in order to optimise the resources available for data storage and preservation.
- Encourage RDM support services to work together with their research communities and international initiatives (e.g. RDA4Eng) in order to establish and promote uptake of metadata standards.

Engage with the research community at the institution towards building a community of data champions from different disciplines who can share their knowledge and promote best practices in RDM.

Prepare and implement training that equips students and researchers working in universities of science and technology with the knowledge and skills needed, including:

- Build a modular approach of courses to cover the different levels of knowledge needed in the different stages of a researcher's career and that considers the different types and workflows found in universities of science and technology.
- Create and make available teaching materials that facilitate continuous and targeted (discipline-specific) learning for researchers to integrate RDM practices into their everyday workflows.
- Invest resources in training RDM support staff on the relevant tools and skills needed across science and technology disciplines.
- Work together with PhD programmes to include RDM within the curriculum of students.

As the value of data for both research institutions and commercial actors is continuously growing, good RDM practices is an area of central and growing importance for universities of science and technology. The universities united within CESAER have extensive expertise in RDM, and are actively pursuing several topics in this area. For example, the topics of RDM practices during university-industry collaborations and the implementation of recruitment and promotion procedures for academic staff which rewards good practices in open science remain high on the agenda.

In the spirit of collaboration, and with the knowledge that community efforts will help take us all further, we hereby extend an open invitation for interested stakeholders who are interested in engaging in this area to reach out to us via the contact details provided on cesaer.org/contact/.

ANNEXE: ABBREVIATIONS

ABBREVIATION	MEANING
API	Application programming interface
DMP	Data Management Plan
DOI	Digital Object Identifier
DARIAH	Digital Research Infrastructure for the Arts and Humanities
EC	European Commission
ELN	Electronic Lab Notebook
ELIXIR	European life-science Infrastructure for biological Information
EOSC	European Open Science Cloud
ERC	European Research Council
EU	European Union
FAIR	Findable, Accessible, Interoperable, Re-usable
GDPR	General Data Protection Regulation
HPC	High-Performance Computing
IG	Interest Group
IPR	Intellectual Property Right
LEARN	LEaders Activating Research Networks
LERU	League of European Research Universities
Moodle	Modular Object-Oriented Dynamic Learning Environment
netCDF	Network Common Data Form
RDM	Research Data Management
RDMinEng	Research Data Management in Engineering
RDMO	Research Data Management Organiser
TFOS	Task Force Open Science