SCIENCE EUROPE

PRACTICAL GUIDE TO **THE INTERNATIONAL ALIGNMENT OF RESEARCH DATA MANAGEMENT**

Extended Edition
with DMP Evaluation Rubric

SCIENCE EUROPE
Shaping the future of research

January 2021

'Practical Guide to the International Alignment of Research Data Management' (Extended Edition)

Author: Science Europe

For further information please contact office@scienceeurope.org

Table of Contents

Foreword by Dr Thierry Damerval 2
Introduction 4

GUIDANCE FOR ORGANISATIONS:
CORE REQUIREMENTS FOR
DATA MANAGEMENT PLANS 7

GUIDANCE FOR ORGANISATIONS:
CRITERIA FOR THE SELECTION OF
TRUSTWORTHY REPOSITORIES 11

GUIDANCE FOR RESEARCHERS: 15
Translating the Core Requirements into a DMP template
Guiding the Selection of Trustworthy Repositories

GUIDANCE FOR REVIEWERS: 31
Evaluation Rubric for Data Management Plans

Notes and References 51
Annex: Compatibility with the FAIR Data Principles 52

Foreword by Dr Thierry Damerval

Member of the Science Europe Governing Board and President of the French National Research Agency

January 2021

The objective of Open Science is to make knowledge accessible for all. It promotes the unhindered dissemination of research outputs and aims to improve research efficiency by supporting data discovery, accessibility, interoperability, and re-use. At the French National Research Agency (ANR), we support the European and international alignment of efforts on the sharing of research data, following the principle 'as open as possible, as closed as necessary'. We encourage our researchers to consider research data management (RDM) and data sharing from the development phase of a research project and throughout its lifecycle.

The first edition of the Science Europe Practical Guide to the International Alignment of Research Data Management was published in January 2019 and was quickly taken up by numerous research funding and performing organisations across Europe, including the European Commission, who use it to define their own RDM policies and as an educational resource for researchers. Key to this success was the engagement of the Dutch Research Council (NWO), led by its President and former Science Europe Governing Board member, Professor Stan Gielen, and the commitment he made in 2017 to champion the alignment of RDM among research funding organisations in Europe.

Two years after its launch, Science Europe presents the second, extended edition of its guide. Following requests from numerous research stakeholders, it now includes an additional fourth chapter that provides guidance for those who are called to evaluate DMPs.

I am proud that ANR was among the first organisations to implement Science Europe's recommendations and that we succeeded to ensure its uptake at national level. I am confident that this extended second edition will prove to be even more useful than the first. With more and more research funding and performing organisations committed to improve data management, this resource will truly support the further alignment of RDM policies across Europe and beyond.

Dr Thierry Damerval

Introduction

The research sector is undergoing an important paradigm shift towards Open Science and aims to make research outputs available for use and re-use by other researchers. Quality-assured research data are key for good knowledge creation. Responding to the need to make the research system at large more efficient in terms of using existing knowledge, data should be made available according to the FAIR principles, meaning that data should be Findable, Accessible, Interoperable, and Re-usable. Instrumental initiatives such as the European Open Science Cloud (EOSC) promote FAIR research data and will require sound institutional practices to ensure that data are shared in optimal conditions. It is therefore essential that the public research sector plays a leading role in establishing and implementing research data management (RDM) policies and practices for FAIR data.

Research funding organisations, research organisations,[1] individual researchers, and reviewers have different needs and requirements from RDM policies and practices. This guide presents core requirements for data management plans (DMPs), criteria for the selection of trustworthy repositories, guidance for researchers on how to comply with these requirements, and a DMP evaluation rubric to support DMP review. The guidelines provide organisations and research communities with a common basis from which they can develop their own RDM policies. They should be considered minimum requirements that can be amended to accommodate institutional or disciplinary policies and practices.

The content of this guide has been developed to support researchers in ensuring that data are FAIR, where appropriate. There may be legitimate reasons (including project-specific or privacy-related ones) to delay or restrict access, which calls for a balanced approach towards openness to research data. The guide even goes beyond 'FAIRness' on other aspects such as data storage, backup during the project, and long-term preservation.[2]

This guide has been developed by experts from Science Europe's Member Organisations, in consultation with stakeholders from the broader research community to take their diverse needs into account.[3] It is intended to provide resources that are useful for all research funding organisations, research organisations, and researchers. It focuses on content-related questions and leaves flexibility for adaptation to organisational and disciplinary policies and procedures.[4]

This guide is divided into four parts:

Core Requirements for Data Management Plans: six aspects that every DMP should cover, with detailed guiding questions.

Criteria for the Selection of Trustworthy Repositories: four topics detailing criteria that every trusted repository should meet.

Guidance for Researchers: more detailed information and examples to support researchers in complying with organisational requirements.

Guidance for Reviewers: guidance to support the evaluation of DMPs by reviewers, aligned with the DMP core requirements presented in previous chapters.

HOW TO USE THIS GUIDE

Research funding organisations, research organisations, and research communities are encouraged to use the **core requirements for data management plans** as a basis to set up their own DMP templates and the **DMP evaluation rubric** to review DMPs.

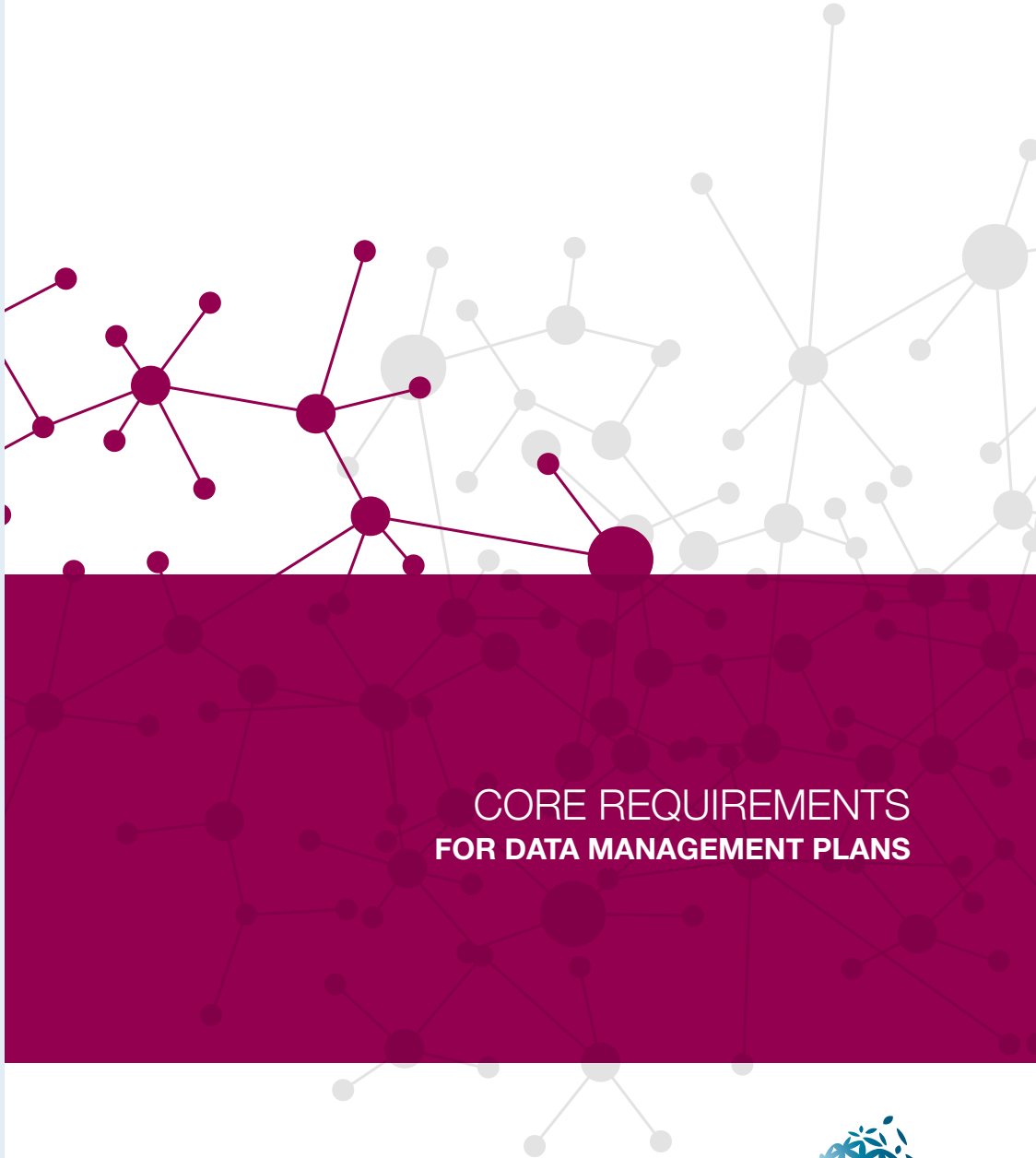
The **guidance for researchers** supports researchers in drafting and updating DMPs and managing their data throughout the research lifecycle. It also provides additional information for research organisations that aim to support their researchers in this endeavour.

For other actors in the research sector, this guide serves as a reference document on how a DMP should be structured and used.

The **criteria for the selection of trustworthy repositories** and the respective guidance will help research funding organisations, research organisations, and individual researchers identify repositories for storing and sharing data.

The core requirements for DMPs and the criteria for the selection of trustworthy repositories can be seen as two stand-alone documents and used independently. It is however recommended to take both into consideration when developing or amending institutional or discipline-specific policies in order to reach the best possible alignment among institutions.

Organisations wishing to adapt the guidance to their organisational or disciplinary policies can find templates for all parts of the guide in an adaptable format on the Science Europe website at http://scieur.org/rdm.



SCIENCE EUROPE
Shaping the future of research

Introduction to the Core Requirements for Data Management Plans

Research funding organisations and research organisations increasingly require researchers to develop data management plans. These plans support the researcher in considering all relevant aspects of data management from the very beginning of a research project. A DMP should stimulate researchers to think about optimal handling, organising, documenting, and storing of their data.

Currently, there is a lot of variation in research data management policies. Many research funding organisations, research organisations, and research communities have developed their own rules and templates. This can be confusing for researchers and is especially problematic as many researchers acquire their funding from different sources; they are increasingly confronted with different grant requirements and institutional policies. There is an urgent need to align data management policies in order to provide more clarity for researchers. DMPs should not be a bureaucratic burden for them, but a useful means of support when planning and conducting a research project.

The following list presents six topics that should be covered in DMPs, each specified with several guiding questions. These topics and questions for setting up a DMP form the core requirements that every research funding organisation should request in order for the researcher to develop a useful DMP. The order of the core requirements can be changed according to specific needs and organisational focal points. However, all six core requirements need to be addressed in a DMP.

An example template providing guidance on which aspects to further consider in a DMP can be found on Page 17 of this guide.

CORE REQUIREMENTS FOR **DATA MANAGEMENT PLANS**

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. **Data description and collection or re-use of existing data**
   a. How will new data be collected or produced and/or how will existing data be re-used?
   b. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. **Documentation and data quality**
   a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
   b. What data quality control measures will be used?

3. **Storage and backup during the research process**
   a. How will data and metadata be stored and backed up during the research process?
   b. How will data security and protection of sensitive data be taken care of during the research?

4. **Legal and ethical requirements, codes of conduct**
   a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
   b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
   c. How will possible ethical issues be taken into account, and codes of conduct followed?

5. Data sharing and long-term preservation

a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

c. What methods or software tools will be needed to access and use the data?

d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

SCIENCE EUROPE
Shaping the future of research

Introduction to the Criteria for the Selection of Trustworthy Repositories

Providing access to data is one of the pillars of sound, reproducible scientific research. More and more research funding organisations, institutions, and journals require researchers to deposit their research data in repositories. Researchers need to be able to identify trustworthy repositories where they can store their data for sharing. There is currently no generally accepted list of such repositories, whereas general registries of repositories list more than 2,000 of them. However, the maturity and trustworthiness of these repositories are difficult to assess. Many repositories have not yet sought to be certified by an acknowledged certification body. Identifying an appropriate repository can therefore be a challenging task for researchers, their organisations, and funding organisations.

In some disciplines, researchers work with discipline-specific repositories which already have certain policies and standards in place that meet the needs of the specific community. Other repositories serve a more general research public, and their policies and standards are necessarily more generic as well.

Some repositories have been certified as trustworthy repositories by one of several acknowledged certification bodies. In order to facilitate the recognition of trustworthy repositories for researchers, it is strongly recommended that repositories that have not yet been certified seek certification by such a body.

It is always recommended to refer to broadly recognised discipline-specific or certified repositories in the first place. The criteria for the selection of trustworthy repositories presented in this guide should be used in cases where no such repository can be identified.

The list of criteria presented in this guide consists of a number of minimum criteria, organised on four major topics, that all trustworthy repositories should fulfil. This list does not prioritise one criterion over another.

More detailed explanations on the criteria for the selection of trustworthy repositories can be found on Page 26 of this guide.

CRITERIA FOR THE SELECTION OF **TRUSTWORTHY REPOSITORIES**

Trustworthy repositories should meet the following minimum criteria:

1. **Provision of Persistent and Unique Identifiers (PIDs)**
   a. Allow data discovery and identification
   b. Enable searching, citing, and retrieval of data
   c. Provide support for data versioning

2. **Metadata**
   a. Enable finding of data
   b. Enable referencing to related relevant information, such as other data and publications
   c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
   d. Use metadata standards that are broadly accepted (by the scientific community)
   e. Ensure that metadata are machine-retrievable

3. **Data access and usage licences**
   a. Enable access to data under well-specified conditions
   b. Ensure data authenticity and integrity
   c. Enable retrieval of data
   d. Provide information about licensing and permissions (in ideally machine-readable form)
   e. Ensure confidentiality and respect rights of data subjects and creators

4. **Preservation**

   a. Ensure persistence of metadata and data

   b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)

SCIENCE EUROPE
Shaping the future of research

Translating the Core Requirements into a DMP template

The following example of a data management plan template is based on the core requirements for DMPs. These core requirements should be considered as a minimum standard, leaving the flexibility to formulate additional guidelines according to the needs of specific domains or to national or local legislation.

The template presented below refers to the 15 questions covering six core requirements for good data management. Additional guidance and explanations are provided to help researchers fill out such a template and to assure that all relevant aspects of research data management are covered. The below table is an example of how the core requirements can be transformed into a DMP template. It will be up to the individual organisations and disciplines to develop templates that fit their needs.

GENERAL INFORMATION	
Administrative information	• Provide information such as name of applicant, project number, funding programme, version of DMP.
1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA	
1 a **How will new data be collected or produced and/or how will existing data be re-used?**	• Explain which methodologies or software will be used if new data are collected or produced. • State any constraints on re-use of existing data if there are any. • Explain how data provenance will be documented. • Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.

1b

What data (for example the kind, formats, and volumes), will be collected or produced?

- Give details on the kind of data: for example numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.
- Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf).
- Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.
- Give preference to open and standard formats as they facilitate sharing and long-term re-use of data (several repositories provide lists of such 'preferred formats').
- Give details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns).

2 DOCUMENTATION AND DATA QUALITY

2a

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- Indicate which metadata will be provided to help others identify and discover the data.
- Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
- Use community metadata standards where these are in place.
- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
- Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.
- Consider how this information will be captured and where it will be recorded (for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks).

2b

What data quality control measures will be used?

- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.

3 STORAGE AND BACKUP DURING THE RESEARCH PROCESS

3a

How will data and metadata be stored and backed up during the research?

- Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations.
- Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

3b

How will data security and protection of sensitive data be taken care of during the research?

- Explain how the data will be recovered in the event of an incident.
- Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.
- Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed.
- Explain which institutional data protection policies are in place.

4 LEGAL AND ETHICAL REQUIREMENTS, CODES OF CONDUCT

4a

If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

- Ensure that when dealing with personal data data protection laws (for example GDPR) are complied with:
  - Gain informed consent for preservation and/or sharing of personal data.
  - Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data).
  - Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible).
  - Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data, for instance by a trusted third party).
  - Explain whether there is a managed access procedure in place for authorised users of personal data.

4b

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

- Explain who will be the owner of the data, meaning who will have the rights to control access:
  - Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and re-use licenses.
  - Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement.
- Indicate whether intellectual property rights (for example Database Directive, *sui generis* rights) are affected. If so, explain which and how will they be dealt with.
- Indicate whether there are any restrictions on the re-use of third-party data.

4c

What ethical issues and codes of conduct are there, and how will they be taken into account?

- Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning.
- Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.

5 DATA SHARING AND LONG-TERM PRESERVATION

5a

How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

- Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism).
- Outline the plan for data preservation and give information on how long the data will be retained.
- Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents.
- Indicate who will be able to use the data. If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. Explain what action will be taken to overcome or to minimise restrictions.

5b **How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?**	• Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes. • Indicate how it will be decided what data to keep. Describe the data to be preserved long-term. • Explain the foreseeable research uses (and/or users) for the data. • Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the DMP that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.
5c **What methods or software tools are needed to access and use data?**	• Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data. • Indicate whether data will be shared via a repository, requests handled directly, or whether another mechanism will be used?
5d **How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**	• Explain how the data might be re-used in other contexts. Persistent identifiers (PIDs) should be applied so that data can be reliably and efficiently located and referred to. PIDs also help to track citations and re-use. • Indicate whether a PID for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.

6 DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES

6a **Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**	• Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible. • For collaborative projects, explain the co-ordination of data management responsibilities across partners. • Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised. • Consider regular updates of the DMP.
6b **What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**	• Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in. Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges. • Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered.

Guiding the Selection of Trustworthy Repositories

The following table provides guidance for the selection of trustworthy repositories by criteria structured according to four main topics.

1 PROVISION OF PERSISTENT AND UNIQUE IDENTIFIERS (PIDS)

A trustworthy repository should:

1a Allow data discovery and identification

- ensure that PIDs are included in the corresponding metadata.

1b Enable searching, citing, and retrieval of data

- consistently assign PIDs (for example a DOI,[5] URN,[6] ARK[7]) to the data it holds, allowing the corresponding data and metadata to be found, referred to, and retrieved, even if the location where the data are stored changes.

1c Provide support for data versioning

- ensure that the version of the data stored in the repository is clearly specified and documented via a permanent audit trail in order for the provenance to be traced.

Note: Not all repositories use an accepted and universal PID system such as the ones mentioned above. Instead, they use a local identifier or administrative number that the repository itself maintains. This increases the risk that the data cannot be found anymore if they are moved to another location, or if the repository ceases to exist, reorganises, or changes its governance.

2 METADATA

The data should be accurately described with rich metadata. The metadata should document how the data were generated, under what license and how they can be re-used, and provide the context for proper interpretation by other researchers.

A trustworthy repository should:

2a Enable finding data

- ensure interoperability and re-use of data by others by providing the data and metadata in an accessible language, based on a well-established formalism. Data and metadata should be described using standard vocabularies and formats allowing computer systems to search for them, combine them in an automatic way, and distinguish the metadata from the research data file(s).

2b Enable referencing to related relevant information

- ensure that in the metadata information it is possible to declare links to other relevant or associated information by providing the PID and a description of the scientific relation. One particular kind of information is details on the associated researcher, for which permanent research IDs exist (such as ORCID,[8] ISNI,[9] or DAI[10]).

2c Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data

- ensure that metadata are archived for the long term and that metadata always remain retrievable, even if the corresponding research data are not or no longer available (for example due to privacy restriction, legal obligations, or other protective measures).
- ensure that retracted data due to poor research practices or misconduct are still findable through the metadata and preserved in order to allow examination of the research record.

2d Use metadata standards that are broadly accepted (by the scientific community)

- ensure that the metadata maintained by the repository are machine-retrievable and use standards that are broadly accepted (by the scientific community).

- ensure that community standards or best practices for data handling are followed if these exist. Note that repositories that are specialised in a particular research field may have community standards regarding the data and metadata that are uploaded.

2e Ensure that metadata are machine-retrievable

- encourage that the information contained in the metadata are structured in a way that allows machines to retrieve it, for example by providing a form with specific fields to be completed.

3 DATA ACCESS & USAGE LICENSES

A trustworthy repository should:

3a Enable access to data under well-specified conditions

- be clear about the terms under which the data can be re-used. Such (license) information is usually included in the metadata.

3b Ensure data authenticity and integrity

- ensure that the metadata contain detailed information about the provenance of data, including how they were generated, how they were processed, in which context they may be re-used, and how reliable they are.

3c Enable retrieval of data

- allow retrieval of data or at least metadata using an open and standardised protocol (not a proprietary communication protocol).

3d Provide information about licensing and permissions (in ideally machine-readable form)

- allow license information to be referred to in a structured way, so that the conditions of use are clear, preferably to humans as well as to machines. Where possible, common or broadly accepted licensing systems should be used (such as Creative Commons) which can be referred to by URL.

3e Ensure confidentiality and rights of data subjects and creators

- provide a way for authentication and authorisation of human and machine-users, allowing to set user (or group) specific access rights to account for data with confidentiality issues and other restrictions.

4 PRESERVATION

A trustworthy repository should:

4a Ensure persistence of metadata and data

- ensure the preservation and continued availability and access to the data and metadata entrusted to it by its users.

4b Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)

- manage the preservation of data and metadata in a documented way. In particular, it should have a preservation policy that details the mission and scope of the repository, governance aspects, financial sustainability, outsource partners, and retention periods (the timeframe of preservation).
- have a publicly available contingency plan and ensure preservation of data and metadata beyond the lifetime of the repository (for example by allowing easy extraction and transfer of data and metadata to another repository).

SCIENCE EUROPE
Shaping the future of research

Data Management Plan Evaluation Rubric

1. WHAT IS THE DMP EVALUATION GUIDANCE FOR?

When research (funding) organisations require researchers to develop DMPs they must be able to also follow up with evaluation and feedback for the researchers. This guidance is designed to support and guide the evaluation of DMPs, prompting the analysis of whether all required aspects have been covered. It is drafted in a generic way and deliberately written in plain language. Completely aligned with the guidance for researchers, the DMP evaluation rubric also aims to ensure the 'FAIRness' of data, even though this is not explicitly stated in all sections of the rubric.

2. WHY THIS FORMAT?

This DMP evaluation guidance takes up the requirements and guiding questions from previous chapters of this guide. It provides criteria to help the reviewer assess whether the information provided in the DMP is sufficient to ensure that the research team will manage data as expected. It is presented in the form of a rubric and lists the different criteria and performance levels that indicate to what extent the criteria are met. This rubric contains two performance levels: 'Sufficiently Addressed' and 'Insufficiently Addressed'. Insufficiently addressed refers either to a lack of information or information deemed incorrect.

3. CAN IT BE ADAPTED?

Following the same structure as the previous chapters, the rubric provides the core criteria for analysis while leaving flexibility for organisations to accommodate legislative frameworks, institutional circumstances, or disciplinary requirements. The different elements of this guide can and should be adapted accordingly. These changes should be referred to both in the guidance for researchers and for DMP reviewers. It is therefore important to stress that this rubric must be seen as guidance, not as a ready-made tool (such as a checklist). Organisations that want to develop checklists for the evaluation of DMPs can use the rubric as a framework and adapt it accordingly.

4. WHO IS IT FOR?

The guidance will be helpful for anyone who is called to evaluate a DMP. This includes research officers, reviewers, or institutional data managers. Researchers will also find it useful as an additional source of information (building on the previous section). Reviewers using the rubric are strongly encouraged not to use it only as a tick-box exercise, but instead to use it to capture their comments and ratings. Sharing these comments with the researchers will provide additional support and clarification so that they can improve their DMPs.

5. WHEN CAN IT BE USED?

DMPs are reviewed at different stages of the research project lifecycle, depending on institutional policies. The rubric is designed to work alongside this process and can be used each time a DMP is reviewed. Reviewers must keep in mind that a DMP is a living document. The level of detail provided in a DMP might vary depending on which version is being assessed; for example, the first version included in a funding application, or a later version documenting the deposition in a repository.

DMP Evaluation Rubric

DMP Question	DMP Guidance	Performance Levels	
GENERAL INFORMATION			
Guidance for Researchers		**Sufficiently Addressed The DMP…**	**Insufficiently Addressed The DMP…**
Administrative information	• Provide information such as name of applicant, project number, funding programme, version of DMP.	• Contains the minimal information required to identify the applicant and the references of the project.	• Provides no or limited information, which makes it hard to identify who is responsible for the project.
1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA			
Guidance for Researchers		**Sufficiently Addressed The DMP…**	**Insufficiently Addressed The DMP…**
1 a **How will new data be collected or produced and/or how will existing data be re-used?**	• Explain which methodologies or software will be used if new data are collected or produced. • State any constraints on re-use of existing data if there are any. • Explain how data provenance will be documented. • Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.	• Gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used. • Explains, if existing data are re-used, how these data will be accessed and any constraints on their re-use. • Explains clearly, if applicable, why new data must be collected, instead of re-using existing data.	• Provides little or no details on where the data come from and what data will be collected or re-used. • Does not, if applicable, provide sufficient rationale for generating new data.
1 b **What data (for example the kind, formats, and volumes) will be collected or produced?**	• Give details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media. • Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf).	• Clearly describes or lists what data types will be generated (for example numeric, textual, audio, or video) and their associated data formats, including, if needed, data conversion strategies. • Explains why certain formats have been chosen and indicates if they are in open and standard format. If a proprietary format is used, it explains why.	• Provides no or little details on what data types will be generated and does not provide a valid reason for this omission (for example a statement that no data will be produced or generated). • Only lists/describes the kinds of data without specifying their formats.

	• Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used. • Give preference to open and standard formats as they facilitate sharing and long-term re-use of data (several repositories provide lists of such 'preferred formats'). • Give details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns).	• Provides information about the estimated data volume. • Clearly states, if applicable, that no new data will be produced or generated by the project. NB. Information derived from previously existing data sources (namely output, processed, and analysed data) are to be considered new data under this question.	• Only lists formats, without specifying the kinds of data. • Does not provide an estimate of data volume.

2 DOCUMENTATION AND DATA QUALITY

Guidance for Researchers		Sufficiently Addressed The DMP…	Insufficiently Addressed The DMP…
2a **What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?**	• Indicate which metadata will be provided to help others identify and discover the data. • Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used. • Use community metadata standards where these are in place. • Indicate how the data will be organised during the project mentioning, for example, conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use. • Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on. • Consider how this information will be captured and where it will be recorded (for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks).	• Clearly outlines the metadata that will accompany the data, with reference to good practice in the scientific community (for example uses metadata standards where they exist). • Clearly outlines the documentation needed to enable data re-use, stating where the information will be recorded (for example a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks). • Indicates how the data will be organised during the project (for example naming conventions, version control strategy and folder structures).	• Provides little or no details on the metadata that will accompany the data. • Provides no information, or only a very vague mention of documentation, without providing any detail or explanation.

2b **What data quality control measures will be used?**	• Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.	• Clearly describes the approach taken to ensure and document quality control in the collection of data during the lifetime of the project.	• Provides no information or only a vague mention on how data quality is controlled and documented during the lifetime of the project.

3 STORAGE AND BACKUP DURING THE RESEARCH PROCESS

Guidance for Researchers		Sufficiently Addressed The DMP…	Insufficiently Addressed The DMP…
3a **How will data and metadata be stored and backed up during the research?**	• Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations. • Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.	• Clearly (even if briefly) describes: › The location where the data and backups will be stored during the research activities. › How often backups will be performed. › The use of robust, managed storage with automatic backup (for example storage provided by the home institution). or • Explains why institutional storage will not be used (and for what part of the data) and describes the (additional) locations, storage media, and procedures that will be used for storing and backing up data during the project.	• Provides no information or very vague reference to how data will be stored and backed up during the project.
3b **How will data security and protection of sensitive data be taken care of during the research?**	• Explain how the data will be recovered in the event of an incident. • Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.	• Clearly explains: › How the data will be recovered in the event of an incident. › Which institutional and/or national data protection policies are in place and provides a link to where they can be accessed. › Who will have access to the data during the research.	• Provides little or no details on how the data will be recovered in the event of an incident, which institutional data protection policies are in place, and who will have access to the data during the research.

	• Consider data protection, particularly if your data is sensitive (for example containing personal data, politically sensitive information, or trade secrets). Describe the main risks and how these will be managed. • Explain which institutional data protection policies are in place.	• Clearly describes the additional security measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe, when sensitive data are involved (for example personal data, politically sensitive information, or trade secrets).	• Provides little or no details about data protection and risk management, or the explanation is too vague, when sensitive data are involved (for example personal data, politically sensitive information, or trade secrets).
4 LEGAL AND ETHICAL REQUIREMENTS, CODES OF CONDUCT			
Guidance for Researchers		**Sufficiently Addressed The DMP…**	**Insufficiently Addressed The DMP…**
4a **If personal data are processed, how will compliance with legislation on personal data and security be ensured?**	• Ensure that when dealing with personal data, data protection laws (for example GDPR) are complied with: › Gain informed consent for preservation and/or sharing of personal data. › Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data). › Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible). › Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data, for instance by a trusted third party). › Explain whether there is a managed access procedure in place for authorised users of personal data.	• Clearly indicates if personal data will be collected/used as part of the project, and, if applicable, how compliance with applicable legislation will be ensured (for example by gaining informed consent, considering encryption, anonymisation, or pseudonymisation). • Describes the procedure to manage access to only authorised users.	• Provides little or no details to demonstrate that personal data, if any, will be managed in compliance with applicable legislation.

4b **How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?**	• Explain who will be the owner of the data, meaning who will have the rights to control access: › Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and re-use licenses. › Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement. • Indicate whether intellectual property rights (for example Database Directive, sui generis rights) are affected. If so, explain which and how will they be dealt with. • Indicate whether there are any restrictions on the re-use of third-party data.	• Clearly explains, if applicable: › Who will have the rights to control access to which part of the data. › What access conditions and re-use licenses will apply to the data. • Clearly explains, if applicable, how intellectual property rights will be managed. • Explains for multi-partner projects and multiple data owners how these matters are addressed in the consortium agreement. • Alternatively, there is a clear statement that there are no such restrictions on the data. • Indicates, if applicable, whether there are any restrictions on the re-use of third-party data.	• Does not address legal issues (or only for a subset of the data), and does not provide good reason or explanation for not doing so. • Does not address matters of rights to control access to the data in case of a multi-partner project and does not provide good reason or explanation for not doing so.
4c **What ethical issues and codes of conduct are there, and how will they be taken into account?**	• Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning. • Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.	• Provides details of what ethical issues have been considered that may affect data storage, transfer, use, sharing and/or preservation, and demonstrates that adequate measures are in place to manage ethical requirements. • Mentions, if applicable, whether ethical review is being pursued. If ethical approval has been obtained, refers to the relevant committee and documents. • Refers to relevant ethical guidelines and/or codes of conduct or alternatively provides a clear statement that explains why ethical issues have not been considered.	• Provides little or no details to demonstrate that ethical implications and codes of conduct have been considered, and does not explain why they did not need to be considered.

5 DATA SHARING AND LONG-TERM PRESERVATION			
Guidance for Researchers		**Sufficiently Addressed The DMP…**	**Insufficiently Addressed The DMP…**
5a **How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?**	• Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism). • Outline the plan for data preservation and give information on how long the data will be retained. • Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents. • Indicate who will be able to use the data. If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. Explain what action will be taken to overcome or to minimise restrictions.	• Clearly describes how the data and/or metadata will be made discoverable and shared. • Specifies when data will be shared and under which licence. • Includes the name of the repository, data catalogue, or registry where data will or could be shared. • Includes information on how long the data will be retained and gives precision on its timely release. • Clearly explains, if applicable, why data sharing is limited or not possible, and who can access the data under which conditions (for example, only members of certain communities or via a sharing agreement). • Explains, where possible, what actions will be taken to overcome or to minimise data sharing restrictions.	• Provides little or no details on how and when data will be shared, or the explanation is not adequate or technically viable.
5b **How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?**	• Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes. • Indicate how it will be decided what data to keep. Describe the data to be preserved long-term. • Explain the foreseeable research uses (and/ or users) for the data.	• Provides details of what data collected or created in the project will be preserved in the long term and clearly indicates for how long. This should be in alignment with funder, institutional, or national policies and/or legislation, or community standards.	• Provides no further information or lacks adequate explanation on what provisions would be made for data preservation.

	• Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the DMP that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.	• Provides details of which (versions of) data and accompanying documentation will be retained or destroyed, and explains the rationale (for example contractual, legal requirements, or regulatory purposes). • Provides details of how the selection is made, and what possible interest there would be for re-use (or not). • Provides details on how the data, accompanying documentation, and any other required technology such as copies of software in specific versions will be archived in the long term. • Explains how data will be managed in a sustainable way beyond the lifetime of the grant. • Provides the name of the archive or trustworthy repository – or the way to curate and preserve data – that will be used to make data available for re-use.	
5 c **What methods or software tools are needed to access and use data?**	• Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data. • Indicate whether data will be shared via a repository requests handled directly, or whether another mechanism will be used?	• Clearly indicates which specific tools or software (for example specific scripts, codes, or algorithms developed during the project, version of the software) potential users may need to access, interpret, and (re-)use the data. • Provides information, if relevant, on any protocol to access the data (for example if authentication is needed or if there is a data access request procedure).	• Provides little or no details on which software developed during the project will be necessary to access and interpret the data, how it will be made available, or why that may not be possible.

5 d **How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**	• Explain how the data might be re-used in other contexts. Persistent identifiers (PIDs) should be applied so that data can be reliably and efficiently located and referred to. PIDs also help to track citations and re-use. • Indicate whether a PID for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.	• Specifies how the data can be re-used in other contexts. • Clearly indicates if and which PIDs are provided for all datasets, individual datasets, data collections, or subsets. If PIDs will not be used, it explains why. • Clearly presents the approach, and the choice of identifiers is justified and refers to international standards.	• Makes no mention of PIDs nor provides a valid reason for not providing them. • Provides no clear information on what type of PID will be assigned to the data and whether individual datasets and/or collections or datasets will be issued with PIDs.

6 DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES

Guidance for Researchers	Sufficiently Addressed The DMP…	Insufficiently Addressed The DMP…
6 a **Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**	• Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible. • For collaborative projects, explain the co-ordination of data management responsibilities across partners • Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised. • Consider regular updates of the DMP.	• Clearly outlines the roles and responsibilities for data management/stewardship (for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing), naming responsible individual(s) where possible. • Clearly indicates who is responsible for day-to-day implementation and adjustments to the DMP. • Explains, for collaborative projects, the co-ordination of data management responsibilities across partners.	• Does not discuss responsibility for data management/stewardship activities and/or does not indicate who is responsible for day-to-day implementation and adjustments to the DMP. • Provides no description, in case of a collaborative project, on how data management responsibilities will be co-ordinated across partners.
6 b **What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**	• Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in. • Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges. • Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered.	• Provides clear estimates of the resources and costs (for example storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges) that will be dedicated to data management and ensuring that data will be FAIR and describes how these costs will be covered. Alternatively, there is a statement that no additional resources are needed.	• Provides no answer or is vague about the resources required for data management and ensuring that data will be FAIR (for example resources are not listed or costed inappropriately), and/or does not describe how the costs will be covered.

Notes & References

1 The term ‘research organisations’ refers to research performing organisations, universities, and research institutes.

2 For further information on how the FAIR Principles are translated into the core requirements and criteria, please see the Annex.

3 The concept of aligning RDM policies and practices was presented at an open event on 30 January 2018. Two consultation rounds were organised in April 2018 and August–September 2018 on the draft of the first edition of the Guide. Another consultation was held in July–September 2020 on the additional chapter containing the DMP evaluation rubric.

4 For procedural elements of implementing DMPs, see the RDA DMP Common Standards Working Group: https://www.rd-alliance.org/groups/dmp-common-standards-wg

5 Digital Object Identifier

6 Uniform Resource Name

7 Archival Resource Key

8 Open Researcher and Contributor ID

9 International Standard Name Identifier

10 Digital Author Identifier

Annex:
Compatibility with the FAIR Data Principles

THE FAIR DATA PRINCIPLES		Core Requirements for DMPs (CR)	Criteria for the Selection of Trustworthy Repositories
To be Findable			
F1	(meta)data are assigned a globally unique and eternally persistent identifier	CR 5d	Criterion 1
F2	data are described with rich metadata	CR 2a	Criterion 2
F3	metadata clearly and explicitly include the identifier of the data they describe	CR 5d	Criterion 1, Criterion 2
F4	(meta)data are registered or indexed in a searchable resource		Criterion 2
To be Accessible			
A1	(meta)data are retrievable by their identifier using a standardised communications protocol	CR 5c	Criterion 1, Criterion 2
A1.1	the protocol is open, free, and universally implementable	CR 5c	Criterion 2
A1.2	the protocol allows for an authentication and authorisation procedure, where necessary	CR 4b, CR 5a, CR 5c	Criterion 3
A2	metadata are accessible, even when the data are no longer available	CR 4c, CR 5a, CR 5d	Criterion 2c

THE FAIR DATA PRINCIPLES		Core Requirements for DMPs (CR)	Criteria for the Selection of Trustworthy Repositories
To be Interoperable			
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	CR 1b, CR 2a	Criterion 2d
I2	(meta)data use vocabularies that follow FAIR principles	CR 2a, CR 2b	Criterion 2
I3	(meta)data include qualified references to other (meta)data	CR 2a, CR 5a, CR 5c	Criterion 2b
To be Re-usable			
R1	meta(data) are richly described with a plurality of accurate and relevant attributes	CR 2a, CR 2b	Criterion 2
R1.1	(meta)data are released with a clear and accessible data usage license	CR 4b, CR 5a	Criterion 3d
R1.2	(meta)data are associated with detailed provenance	CR 1a, CR 1b, CR 2b	Criterion 1c, Criterion 2, Criterion 3b, Criterion 4a
R1.3	(meta)data meet domain-relevant community standards	CR 1b, CR 2a	Criterion 2d

Science Europe
Rue de la Science 14
1040 Brussels
Belgium

Tel +32 (0)2 226 03 00
Fax +32 (0)2 226 03 01
office@scienceeurope.org
www.scienceeurope.org

SCIENCE
EUROPE
Shaping the future of research