



Recommendations based on a comparative analysis of data governance policies (WP1)

Aligning strategies for research data governance policies

DOI: [https://doi.org/ 10.5281/zenodo.8436045](https://doi.org/10.5281/zenodo.8436045)

Authors: Sophie Fortuno (DGD-RS – Open Science and Research Data, CIRAD), Pascal Aventurier (Scientific and Technical Information Department/MCST, IRD), Dimitri Szabo (DipSO, INRAE), Luc Decker (Scientific and Technical Information Department/MCST, IRD), Teri Jones-Villeneuve

Contents

I. Introduction	3
II. Scope of the ANR BRIDGE project WP1	3
1. Definitions.....	3
2. Open science policy and research data policy	4
III. WP1 approach	5
1. What common framework should be adopted?	7
2. Analysis method for the RDA Data Policy IG framework	9
IV. Common framework for research data policy	12
1. Definition of research data (policy scope)	12
2. Data types that are subject to exceptions	14
3. Data embargoes.....	14
4. Supplementary materials	15
5. Data repositories.....	16
6. Data citation	18
7. Data licensing.....	19
8. Researcher/author support.....	20
9. Data availability statements	20
10. Data formats and standards (including metadata).....	21
11. Mandatory data sharing (specific types of data).....	22
12. Mandatory data sharing.....	22
13. Peer review of data	23
14. Data management plans	24
V. Other aspects to consider when creating a data policy for research institutions	25
15. Legal and regulatory aspects	25
16. Scientific integrity	26
VI. Conclusions and outlook	27

I. Introduction

Preamble

In France, public research is organized into joint research units, which means that we continually work together to produce digital content. As a result, we must coordinate policies on data sharing and open access between institutes if we are to consistently and effectively implement open science on a national and international level.

By leveraging IRD, INRAE and CIRAD's strong partnership, the ANR BRIDGE project aims to demonstrate how we can promote the application of the FAIR principles via three levers: policy, community and technical. To this end, the ANR BRIDGE project has brought together individuals from different disciplinary fields with a range of data governance, technical and scientific expertise.

To achieve this overall goal, the ANR BRIDGE project tackles the following objectives: i) at the policy level, to analyse, compare and align institutional data governance policy frameworks; ii) at the inter-community level, to develop management principles and sharing practices for researchers and data managers; and iii) at the technical level, to implement components for improved interoperability of data repositories.

In line with the main objectives of the ANR BRIDGE project, our focus is on the way the research institutions involved in the project govern their data. The aim of this project is to develop a framework that aligns with the research data policy environment in terms of national, European and international standards. This work, carried out as part of work package 1 (WP1), comes at a time when more and more scientific journal publishers and funding agencies are introducing data policies aimed at increasing the availability of research data for reuse.

II. Scope of the ANR BRIDGE project WP1

In this section, we will define several key terms with regard to WP1 of the ANR BRIDGE project.

1. Definitions

Data policy

A data policy is defined as “a set of broad, high level principles which form the guiding framework in which Data Management can operate [*authors' note: within an organization*].”

(from the OECD Glossary of Statistical Terms, <https://doi.org/10.1787/9789264055087-en>, page 123)

Data governance

Data governance falls under the umbrella of information governance (Kooper et al., 2011; Maurel, 2013), which involves determining a strategy for how information is managed and includes basic principles, rules and a management structure to ensure that information is used effectively and efficiently.

Source: Da Sylva, L., Maurel, D., Bruyère, M., Saint-Germain, M. & Gareau, G. (2019). Écosystème de la gestion de données de recherche et professionnels de l'information: présentation des enjeux, de la méthodologie et des solutions préconisées d'une enquête canadienne [The Research Data Management Ecosystem and Information Professionals: An Overview of the Issues, Methodology and Proposed Solutions of a Canadian Survey]. *Études de communication* [online], 52(June), 51–70. <https://doi.org/10.4000/edc.8615>

“Data governance refers to all the organizational structures and procedures put in place within a company [*authors’ note: or organization*] to control data collection and use. The aim is to comply with the legal obligations imposed by countries and/or the European Union.”

Source: translated from Lebigdata.fr, [Data Governance ou gouvernance des données : Qu'est-ce que c'est ?](#) [Data governance : What is it?]

“Governing” data means that (good) decisions are taken at every stage in the data life cycle (from production or reuse to final application) and that all stakeholders’ responsibilities are formally specified in this decision-making process.

Source: translated from INRAE. (2020, December 13) Datapartage - Principes pour la gouvernance des données INRAE [Datapartage – Principles for INRAE data governance]. Accessed 14 May 2022. <https://datapartage.inrae.fr/Gouvernance-des-donnees2/Principes-pour-la-gouvernance-des-donnees-INRAE>

Data governance is not data management

“Data governance is frequently confused with other closely related terms and concepts, including [data management](#).”¹

Research data management

Data governance is deals with strategy, roles, organization and rules, whereas “research data management (RDM) . . . encompasses the processes applied throughout the lifecycle of a research project to guide the collection, documentation, storage, sharing, and preservation of research data.”

Source: [Digital Research Alliance of Canada](#)

2. Open science policy and research data policy

While **institutional documents for open science policy** are starting to be published, specific **research data policy** documents have yet to be analysed in detail.

Figure 1 shows how an open science policy for research fits in with other policies that aim to make scientific output freely accessible, such as through open access, open education and free educational resources.

The ANR BRIDGE project’s work is focused specifically on examining open access and governance policies for research data, identifying the main principles for their implementation, and issuing recommendations.

¹ <https://www.talend.com/uk/resources/what-is-data-governance/>

Relation entre politique de Science ouverte et politique des Données de la recherche - Périmètre Bridge

Picture 1. The policy in relation to other national open science documents.

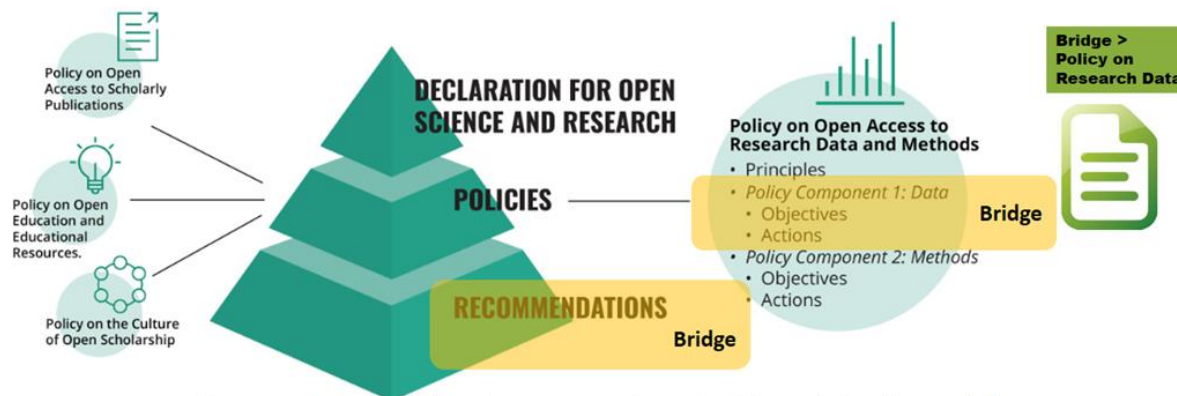


Figure 1. National Open Science and Research Coordination. 2021. "Open Research Data and Methods. National Policy and Executive Plan by the Higher Education and Research Community for 2021–2025." Responsible Research Series 7:2021. <https://edition.fi/tsv/catalog/view/64/166/575-1>

III. WP1 approach

The work carried out for WP1 of the ANR BRIDGE project initially involved:

- gathering framework and other specific documents formalizing data governance policies or components for IRD, CIRAD and INRAE. These included documents made available to research teams (guides, technical and organizational recommendations, intranet sites and informational websites), and documents used by the institutions for their internal procedures or training courses (see Table 1);
- analysing the national, European and international policy context regarding research data governance (Second French Plan for Open Science; guidelines from the Committee for Open Science, Horizon Europe, UNESCO, international funding organizations, etc.);
- monitoring and following relevant external events occurring during the project (conferences, national and international RDA webinars,² CODATA,³ etc.).

² RDA – Research Data Alliance <https://www.rd-alliance.org/>

³ CODATA – Committee on Data of the International Science Council <https://codata.org/>

Table 1. Examples of framework and specific documents that formalize policy elements or components of data governance at IRD, INRAE and CIRAD

Document	Link if available	Organization
<i>Enjeux éthiques et déontologiques de la gestion et du partage des données issues de la recherche</i> [Ethical and professional conduct issues related to sharing and managing research data]	<i>Avis numéro 8 sur les enjeux éthiques et déontologiques du partage et de la gestion des données issues de la recherche</i> , Février 2016 [Opinion no. 8 on the ethical and professional conduct issues related to sharing and managing research data, February 2016] https://www.inrae.fr/actualites/avis-partage-gestion-donnees-issues-recherche	INRAE-CIRAD-IFREMER-IRD Ethics in Common Committee
<i>Feuille de route de la science ouvert</i> [Open Science Roadmap] – June 2021	https://www.ird.fr/une-feuille-de-route-pour-une-science-ouverte-et-partagee	IRD
CIRAD is committed to Open Access to Knowledge – 2021	https://partage-connaissances.cirad.fr/en/partage-connaissances/introduction/cirad-is-committed-to/cirad-is-committed-to-open-access-to-knowledge	CIRAD
<i>Charte pour le libre accès aux publications et aux données</i> [Charter for open access to publications and data]	https://hal.inrae.fr/hal-02801732/document	INRAE
Principles of data governance for research organizations – INRAE's approach	https://hal.inrae.fr/hal-02955903	INRAE
<i>La Gouvernance des données de la recherche au CIRAD: Document institutionnel</i> [Research data governance at CIRAD: Institutional document]	Internal document – February 2019	CIRAD
<i>La Gestion des données au CIRAD : Document de politique institutionnelle</i> [Data management at CIRAD: Institutional policy document]	Internal document – February 2019	CIRAD
Organization and role of in the data life cycle	Engagement letters in accordance with the institution's internal organization (internal documents): – open science managers – functional lead and technical administrator for the Dataverse data repository, etc. Employment policy (HRP) for the field	IRD, CIRAD, INRAE
Intellectual property, using research data in real-world applications, etc.	Charters, internal documents	IRD, CIRAD, INRAE
Intranet and internet sites supporting data stewardship	https://data.ird.fr/ https://coop-ist.cirad.fr/gerer-des-donnees https://intranet-data.cirad.fr , CIRAD's Data intranet site https://datapartage.inrae.fr/	IRD, CIRAD, INRAE
Specific training programmes	Training courses	IRD, CIRAD, INRAE

We then compared the various documents we collected and assessed where they converged and diverged to identify common criteria and rules in terms of organization, stakeholders and the management of explicit data governance.

We ultimately concluded that establishing a relevant basis for the comparative analysis of data governance policies would be impossible due to the different scales of study, the disparities among the available documents, and the aims and organizational structures inherent to each institution.

1. What common framework should be adopted?

What framework should be used to support the alignment of a research data governance policy?

To answer this question, we decided to work at the research data policy level, which has the advantage of providing broad guidelines on which to base our analysis.

Several documents describe the various criteria of a research data policy in operational terms, such as i) the FAIR-enabling Data Policy Checklist 2022 report (Davidson et al., 2022⁴), ii) open science policy templates, and iii) recommendation frameworks published in the context of the European Union-funded OpenAIRE and EOSC-Pillar projects.⁵

But as noted by the Research Data Alliance's Data policy standardisation and implementation interest group⁶ (RDA Data Policy IG), as the prevalence of data policies increases, researchers and research institution support staff risk being confused by the multiple and sometimes contradictory requirements.

This is what led the RDA Data Policy IG to draw up in 2020 a checklist of 14 criteria (or "policy features"), thus establishing a research data policy framework for scientific journal publishers, with a study currently under way on research funders.

In 2021, the Research Data College of the Committee for Open Science⁷ issued recommendations to scientific journals based on this work.⁸

Contributions of the ANR BRIDGE project

The ANR BRIDGE project started with a premise: the criteria of a research data policy for scientific publishers and research funders must be easily understood by multiple stakeholders. Once established, these criteria can then form the basis of a data policy for research institutions.

This is why we chose to adopt the framework of the RDA Data Policy IG in order to converge towards common criteria.

This framework also enables us to understand the complexity of the different mandates of our three research institutions and the national and international context in which they operate.

RDA Data Policy IG framework

⁴ Davidson, J., Grootveld, M., Verburg, M., van Horik, R., O'Connor, R., Engelhardt, C., Garbuglia, F., Vieira, A., Newbold, E., Proudman, V., & Horton, L. (2022). FAIR-enabling Data Policy Checklist (1.0). Zenodo. <https://doi.org/10.5281/zenodo.6225775>

⁵ <https://www.eosc-pillar.eu/>

⁶ <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation-ig>

⁷ The Committee for Open Science is responsible for implementing national policy in France:

<https://www.ouvrirelascience.fr/home/>

⁸ <https://www.ouvrirelascience.fr/guidelines-for-journals-that-wish-to-establish-a-data-policy-related-to-their-publications/>

The RDA Data Policy IG's most recent work (November 2021) was focused on *Developing a Research Data Policy Framework for all Scientific Journals and Publishers* (Hrynaszkiewicz et al., 2020⁹).

To identify six typical data policy profiles, the RDA Data Policy IG examined the practices of scientific publishers who had published in a collective total of over 10,000 journals. This study was supplemented by an initial analysis of funding organizations' data policies, based on the global registry of grant-giving organizations maintained by Crossref.¹⁰

Using a standardized, consensus-based framework, the study identifies 14 policy features (criteria considered mandatory, recommended or not applicable) to establish a research data policy framework for scientific journals and research funding organizations (see Figure 2).

⁹ Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R. & Goudie, S. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, 19: 5, pp. 1–15. <http://doi.org/10.5334/dsj-2020-005>

¹⁰ Crossref Open Funder Registry <https://www.crossref.org/services/funder-registry/> (CC0 licence)

14 journal research data policy features arranged as 6 policy types (tiers)						
	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exceptions to policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Embargoes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Supplementary materials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data repositories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data citation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data licensing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Researcher/ author support	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data availability statements		<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data formats and standards				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (specific data types)				<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (all papers)				<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Peer review of data				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Data Management Plans (DMPs)				<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

☐ **Provide information**
The text for the policy feature will be included in the policy template but it is clear that the feature will not be enforced and checked as part of the publishing or peer review process

☒ **Provide information and action**
The text of the policy feature is included and makes clear where applicable that the feature will be checked and enforced in the publishing or peer-review process

Figure 2. Fourteen journal research data policy features arranged as six policy types <http://doi.org/10.5334/dsj-2020-005>

2. Analysis method for the RDA Data Policy IG framework

ANR BRIDGE project's work was based on an initial review of framework and specific documents (see Table 1) as well as knowledge of data management practices implemented over several years at CIRAD, INRAE and IRD.

During a first working session, in November 2021, we conducted an initial and very broad review of research data in our organizations, commenting specifically on the 14 policy features of the RDA Data Policy IG's framework. This inventory was followed by a second, more detailed review to identify the salient points.

During a second working session, in March 2022, we refined the topics and drafted a new “general” data policy profile for research institutions (see Figure 3) based on the common framework established by the RDA Data Policy IG.

Policy features for publishers and funders								Features Research institution
No.	Policy features	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06	Ex. Policy 07
1	Definition of the research data	○	○	○	○	○	○	●
2	Exceptions to policy	○	○	○	○	○	○	●
3	Embargoes	○	○	○	●	●	●	○
4	Supplementary materials	○	○	○	●	●	●	●
5	Data repositories	○	○	○	●	●	●	●
6	Data citation	○	○	○	○	●	●	●
7	Data licensing	○	○	○	○	○	○	●
8	Researcher/ author support	●	●	●	●	●	●	●
9	Data availability statements		○	●	●	●	●	○
10	Data formats and standards				○	○	●	●
11	Mandatory data sharing (specific data types)				●	○	●	○
12	Mandatory data sharing (all papers)				○	●	●	○
13	Peer review of data				○	○	●	○
14	Data Management Plans (DMPs)				○	○	○	●

Key:

○ = Information required

● = Information and action required

- = N/A

○ Provide information (recommended)
The text of the policy feature will be included in the policy template, but the feature will not be enforced and checked

● Provide information and action (mandatory)
The text of the policy feature is included in the policy template and makes clear that the feature will be checked and enforced

BRIDGE - Profile
Research institution

BRIDGE has kept all the policy features to create a data policy framework that aligns with the policy frameworks used by funding organizations and publishers.

Each organization should choose a level of requirement (mandatory, recommended, N/A) for each feature based on its stance regarding the criteria.

BRIDGE has included
+ two additional features

Legal - regulatory

Scientific integrity

Figure 3. Research institution profile

In a final phase of policy alignment work, the ANR BRIDGE project documented the criteria (or policy features) by examining the three national, European and international levels of public policy on research data.

Among the key data policy documents reviewed were:

- At the national level:
 - The *Deuxième Plan national pour la science ouverte, 2021–2024 (PNSO2)* [Second French Plan for Open Science, 2021–2024 (PNSO2)]
 - The *Plan d'action 2022 de l'Agence Nationale pour la Recherche (ANR)* [2022 Action Plan – French National Research Agency (ANR)]
 - *Vers le déploiement d'une politique commune en faveur de la science ouverte* [Towards the deployment of a common policy for open science], from the network of French funding agencies for open science
- At the European level:
 - *Practical Guide to the International Alignment of Research Data Management – Extended Edition* by Science Europe¹¹
 - The *Horizon Europe (HORIZON) Programme Guide* for 2022
- At the international level:
 - The policy work by the RDA Data Policy IG
 - And, more broadly, the *UNESCO Recommendation on Open Science* published in 2021.

These documents are listed in Table 2.

¹¹ *Science Europe* is a European association that serves as a platform for developing positions on research policy issues and creating policy messages for European institutions, researchers, national governments and the public (<http://www.scienceeurope.org/>).

Table 2. List of data policy documents

Country/ Region	Organization	Document title	Link
France	Ministry of Higher Education, Research and Innovation – MESRI	<i>Deuxième Plan national pour la science ouverte, 2021-2024 (PNSO2)</i> [Second French Plan for Open Science, 2021–2024 (PNSO2)]	https://www.ouvrirelascience.fr/second-national-plan-for-open-science/
France	French National Research Agency – ANR	<i>Plan d'action 2022</i> [2022 Action Plan]	https://anr.fr/fileadmin/document/s/2021/PA-ANR-2022-V1.1a.pdf
France	Network of French funding agencies for open science	<i>Vers le déploiement d'une politique commune en faveur de la science ouverte</i> [Towards a common policy for open science]	https://anr.fr/fileadmin/document/s/2022/Vers_le_deploiement_d_une_politique_commune_en_faveur_de_la_sciences_ouverte.pdf
Europe	Science Europe, an association representing the major research organizations and funding bodies, Brussels	Practical Guide to the International Alignment of Research Data Management – Extended Edition	https://doi.org/10.5281/zenodo.4915862
Europe	Horizon Europe, European programme to support research and innovation (2021–2027)	Horizon Europe (HORIZON) Programme Guide for 2022	https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf
International	UNESCO, United Nations Educational, Scientific and Cultural Organization	UNESCO Recommendation on Open Science, 2021	https://unesdoc.unesco.org/ark:/48223/pf0000379949
International	Data policy standardisation and implementation IG	Developing a Research Data Policy Framework for All Journals and Publishers	http://doi.org/10.5334/dsj-2020-005

IV. Common framework for research data policy

By convention, we have kept the same numbering of the policy features as in the RDA Data Policy IG's framework.

In the first section, we **define** each criterion and give an **overview** of the current situation. We also offer **recommendations** for each criterion that research institutions can adopt according to their own needs.

In the second section, we **suggest ways to supplement** this common framework, including **two additional criteria** that have been tailored to research institutions.

1. Definition of research data (policy scope)

This criterion defines the types of research data covered by the data policy. It establishes the policy scope and provides both general and discipline-specific information.

Overview

Research data can be defined in various ways.

- The *Second French Plan for Open Science* and the ANR both use the definition from the Organisation for Economic Co-operation and Development (OECD): “‘research data’ are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.”¹²
- The *Horizon Europe Programme Guide* adopts the broader terminology of “research output”, which encompasses “results generated by the action to which access can be given in the form of scientific publications, data or other engineered outcomes and processes such as software, algorithms, protocols and electronic notebooks.”
- The document *What is research data?*¹³, published by the Australian Research Data Commons, notes that “research data means data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or other research output is based . . .” and that “it may be raw, cleaned or processed. . . .”
- The RDA's Data Policy IG studied the data policy frameworks of scientific journal publishers and research funders. The study showed that these policies apply to the research data that would be needed to verify the research findings reported in the articles published in the journal. Research data includes data produced by the authors (“primary data”) and data from other sources that are analysed by the authors in their study (“secondary data”). Research data includes any recorded factual material used to produce results in digital and non-digital form. This includes tabular data, code, images, audio, video, maps, raw and/or processed data.

It should be noted here that the challenges associated with open science (open access, dissemination, legal context) mean that algorithms and source codes are now being managed in a similar way as for data.

¹² Organisation for Co-operation and Development [OECD] (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing, Paris. <https://doi.org/10.1787/9789264034020-en-fr>

¹³ https://guides.library.unisa.edu.au/ld.php?content_id=50662363 (page 1)

BRIDGE recommendations

R1. The OECD definition of research data offers a minimum basis for a shared definition.

The definition given by the OECD in its 2007 report *Principles and Guidelines for Access to Research Data from Public Funding* is widely shared:

“ ‘research data’ are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.”

Although this term should be defined in any organization’s data management policy, research institutions would be well served to identify the scope of application and specify the different types of data they produce.

The CNRS’s Institute for scientific and technical information (Institut de l’Information Scientifique et Technique – INIST) defines five types of data:

- Observational
- Experimental
- Simulation
- Derived or compiled
- Reference

Data produced as part of a research project goes through different phases during their life cycle and can therefore also be identified according to the level of processing:

- Raw data: uninterpreted and unprocessed
- Processed data: produced after raw data is calibrated or corrected
- Derived or compiled data: aggregated, calculated, reorganized or providing a specific view of the data
- Analysed data: resulting from a qualitative, statistical or mathematical analysis of the data

R2. Define research data more precisely to make it easier for research organizations to understand.

The essential aspects of research data definitions at European and international levels can be used as a basis, as referenced by the DoRANum website,¹⁴ a training platform supported by INIST-CNRS and the URFIST scientific interest group.

DoRANum has also joined the French Data Gouv Research ecosystem¹⁵, which was launched on 8 July 2022, thus making the information and resources it publishes more visible on a national level.

R3. Establish a specific policy for algorithms and source code produced by the institution. Failing that, include these components in the scope of application of the data management policy in the broadest sense of the term.

The PNSO2 promotes open access to, the application of and the reuse of algorithms and source code resulting from research.

If a research institution produces original scientific software, models or decision-support tools, it may want to establish a specific policy for algorithms and source code. This policy should set out the

¹⁴ <https://doranum.fr/plan-gestion-donnees-dmp/definitions-des-donnees-de-la-recherche/>

¹⁵ <https://recherche.data.gouv.fr/en>

institution's guidelines for data preservation and sharing policies, as well as for licences and conditions for reuse in the context of research activities, including partnerships (private/public/Global North–South), in line with its missions and final applications.

2. Data types that are subject to exceptions

This criterion defines the data that should not be made available through open access. Alternative options regarding how to obtain them could be indicated.

Overview

According to the PNSO2, “The obligation to open up public research data, required in the Digital Republic Act of 2016, should now be enacted in scientific practice with the help of appropriate infrastructures and support services. This obligation is limited by legitimate exceptions as defined by the law. For example, exceptions can be made for professional confidentiality, industrial and trade secrets, personal data and copyright-protected content. In these cases, data sharing practices should still be encouraged by defining the procedures to follow.”

Following the recommendations of the PNSO2, the members of the network of French funding agencies (ADEME, ANR, ANSES, INCa, Inserm/ANRS) are committed to following the “as open as possible and as closed as necessary” principle with regard to sharing research data, in keeping with European (Science Europe, Horizon Europe) and international efforts. This principle means that exceptions to this rule can be made when necessary.

For this criteria, the RDA Data Policy IG indicates that such a policy would not require making publicly available any “quantitative or qualitative data that could identify an individual, data which participants did not consent to be shared . . . [or] locations of endangered species.”

Alternatives to making sensitive or personal data publicly available include:

- the use of controlled access repositories;
- anonymizing data prior to public sharing;
- sharing only metadata;
- procedures for access to research data in the associated article and the management of requests for access to data from other researchers, who are the subject of the “Data Availability Statement” paragraph included in an article.

BRIDGE recommendations

R4. Apply the principle of “as open as possible and as closed as necessary” within an appropriate legal and ethical framework.

- ✓ Some data cannot be made public based on legitimate exceptions for contractual reasons or issues relating to confidentiality, regulation, public safety, ethical restrictions, intellectual property rights or commercial interests.

3. Data embargoes

A widely accepted principle of data sharing is that researchers have the right to be the first to use data produced from their research. However, what is considered a reasonable embargo period may vary depending on the discipline, data type and research.

Overview

This criterion involves determining whether embargoes are permitted for the dissemination of data and what kind of data they cover. An embargo may apply at the time of publication to protect intellectual property, such as a patent registration.

This type of information was often not mentioned in the documents we examined, as it is too specific to a particular discipline or journal.

With regard to data associated with a publication, the embargo period could be that of the publication. For example, the European Union recommends limiting the embargo period to six months for publications in science, technology and medicine and twelve months for those in the social sciences and humanities.¹⁶

For the European FAIRsFAIR project,¹⁷ the issue of authorized embargo periods must be clearly addressed in a data policy.

BRIDGE recommendation

R5. Embargoes on data sharing are permitted for reasonable periods, in accordance with best practices in the scientific discipline to which the data and project belong.

An embargo may apply to primary data in particular.

The embargo time limit must be documented.

The time limit should be established in accordance with best practices in the scientific community.

4. Supplementary materials

Some scientific journals continue to promote data sharing through supplementary materials (e.g. annexes) linked to articles.

This criterion sets out the position of a research institution's data policy with regard to this practice.

Overview

Many data policies (e.g. those of research funders) indicate that data must be stored in a trusted data repository (see §7. Data repositories). Meanwhile, many scientific journals still recommend submitting data in the form of supplementary materials (e.g. data tables in PDF format). But these practices do not allow for the FAIRification¹⁸ of data, especially in terms of making them accessible. In such cases, the data may also have the same status as a scientific article, for which the authors may have assigned their rights to the publisher.

¹⁶https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

¹⁷ [The FAIRsFAIR project https://www.fairsfair.eu/the-project](https://www.fairsfair.eu/the-project)

¹⁸ ANR BRIDGE (2023). Principaux concepts techniques [Key technical concepts]. 8 p. Zenodo. <https://doi.org/10.5281/zenodo.7474725>

BRIDGE recommendation

R6. Publishing data as part of the supplementary materials is not recommended. See also §7. Data repositories.

- ✓ Research data must be placed in data repositories that meet TRUST principles (i.e. transparency, responsibility, user focus, sustainability and technology). These repositories make data more discoverable (easier to search) and easier to reuse and cite.¹⁹

5. Data repositories

Data should be deposited in a research data repository to encourage data sharing.

When available and recognized as trustworthy, data repositories that are specific to a scientific community or discipline are preferable over generalist repositories.

Overview

The *Second French Plan for Open Science* (PNSO2) encourages research data to be structured and made public with a recommendation to distribute such data according to FAIR principles in certified data repositories. In response to the PNSO2, the French ecosystem of open science infrastructure is organized based on the European and international ecosystems. The concepts of certification and trustworthiness in data repositories are covered in the recommendations.

To deposit research data, the ANR encourages project coordinators to consult their institution's open science policy, their project set-up support unit, the DMP OPIDoR portal (data management plan resources) and discipline-specific resources such as those from the Research Data Alliance.²⁰

Science Europe recommends depositing data in **trusted repositories** where researchers can store it for sharing.

Horizon Europe recommends that when data is authorized to be shared, research projects funded by this programme deposit it in repositories with “the features of trusted repositories, and justify this accordingly in their Data Management Plans.”

Various organizations, including IRD,²¹ have published a list of criteria for choosing trusted repositories.

BRIDGE recommendations

R7. Implementing a research data policy encourages institutions to identify trusted data repositories within the national, European or international ecosystem.

Scientific communities need to consider the choice of discipline-specific repositories in relation to project goals and the type of partnership (Global North–South/private/public).

To establish and document this choice, the systematic creation of data management plans for research projects provides an opportunity for discussion and consensus among stakeholders, resulting in consideration of the different data repository options available for the project.

¹⁹ Piwowar, H., & Vision, T. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1:e175. <https://doi.org/10.7717/peerj.175>

²⁰ <https://www.rd-alliance.org>

²¹ <https://data.ird.fr/entrepots-de-donnees-2/>

R8. Communicate the criteria for selecting a trusted data repository to scientists.

Several criteria can be used to determine the level of trustworthiness of a data repository, such as the CoreTrustSeal²² certification requirements or the TRUST principles.²³

In its *Practical Guide to the International Alignment of Research Data Management*,²⁴ Science Europe sets out the minimum criteria structured around four major pillars that all trusted repositories must meet:

- 1) provision of unique and persistent identifiers;
- 2) descriptive, standardized metadata providing relevant, human- and machine-readable information;
- 3) data access and usage licences;
- 4) preservation to ensure the persistence of metadata and data over time.

Communication should focus on the requirements for CoreTrustSeal certification, which is gradually becoming the norm. This point is discussed in more detail in the following recommendation.

R9. Support the CoreTrustSeal certification process for French data repositories.

As part of the PNSO2 recommendations, data repositories should undertake the CoreTrustSeal certification process.²⁵ This certification focuses on the issues of data quality, management processes and the preservation of a repository. It attests to the reliability of the data repository services, where scientists can be confident when depositing and re-using scientific data over the short, medium and long terms.

The CoreTrustSeal repository (CTS: 2020–2022) is based on 16 evaluation requirements relating to trustworthy data repositories. The newest edition (2023–2025) has been updated to reinforce these evaluation requirements. At the time of writing, five French scientific data repositories are CTS certified and listed on the CoreTrustSeal website.²⁶

R10. Provide a repository or a list of recommended trusted data repositories.

To help scientific communities choose appropriate data repositories, a selection of trusted repositories should be drawn up based on the criteria mentioned above.

Scientists should be involved in the process in order to raise their awareness of these criteria and establish a list of repositories that suit the communities' needs.

R11. If there is no trusted, discipline- or subject-specific repository, data should be deposited in a trusted national, international or institutional multidisciplinary repository.

Several research establishments (CIRAD, INRAE, IFSTTAR, IRD, Sciences PO and Université de Lorraine, among others) have set up an institutional research data repository. This service meets the needs of many scientists who conduct their research with so-called long-tail data or for whom a suitable repository is unavailable. For CIRAD and IRD, data repositories are also a way to enhance and preserve data co-produced with partners in the Global South.

²² "Entrepôts de données de confiance : critères de conformité". Accessed 20 January 2023.

<https://www.ouvrirlascience.fr/entrepots-de-donnees-de-confiance-criteres-de-conformite>.

²³ Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M. et al. (2020). "The TRUST Principles for Digital Repositories." *Scientific Data* 7(1): 144. <https://doi.org/10.1038/s41597-020-0486-7>

²⁴ "Practical guide for international harmonization of research data management". Accessed 31 March 2022.

<https://www.ouvrirlascience.fr/practical-guide-for-international-harmonization-of-search-data-management/> and the original guide: Science Europe (2018). *Practical Guide to the International Alignment of Research Data Management*. https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf

²⁵ <https://www.coretrustseal.org/>

²⁶ <https://www.coretrustseal.org/why-certification/certified-repositories/>

As part of the PNSO2 implementation and in order to encourage active open data practices, the **Recherche Data Gouv repository**²⁷ has offered a multidisciplinary data repository since July 2022 to supplement the national and European infrastructures already used by certain disciplinary communities. The thematic reference centres²⁸ are a component of the Recherche Data Gouv ecosystem designed to support these scientific communities. Finally, a national catalogue will soon list data deposited in other trusted repositories to eventually provide a single point of visibility for all French research data.

6. Data citation

This criterion indicates whether the principle of data citation is encouraged or required by the research institution.

The data policy i) encourages or requires the provision of persistent identifiers to retrieve research data and ii) supports data citation as part of a formal reference list.

Overview

Data citation is an integral component of the FAIR principles and data citation principles.²⁹ To cite a data set, a persistent identifier (such as a DOI) and the following information are required (at minimum):

Authors (creators), Date, Title, Publisher (repository), DOI or identifier, Version.

For more on this subject, see the CoopIST fact sheet.³⁰

BRIDGE recommendation

R12. Data citation

- ✓ Cite the authors of the data set by identifying all the possible roles and contributions (design: ideas and tools, collection, curation, processing, control/verification, supervision, etc.). The authors of a data set may be different from those of a related article.³¹
- ✓ Ensure the name of the repository where the data is stored is cited.
- ✓ For metadata, it is recommended to enter the authors' affiliated organizations and entities (research units) according to the rules in force, often set out in joint research unit agreements, in order to make it easier to find them and produce lists and indicators (institutional, etc.). It is also recommended to enter the ORCID identifier for each author.

²⁷ Recherche Data Gouv, a national ecosystem for sharing and opening research data:

<https://recherche.data.gouv.fr/en>

²⁸ <https://recherche.data.gouv.fr/en/page/thematic-reference-centers-providing-expertise-for-individual-scientific-fields>

²⁹ Data Citation Synthesis Group (2014). *Joint Declaration of Data Citation Principles*. Martone, M. (ed.); San Diego, CA: FORCE11. <https://doi.org/10.25490/a97f-egykh>

³⁰ Deboin, M.C. (2021). Citer un jeu de données scientifiques, en 4 points [Citing a scientific data set: 4 steps]. Montpellier (FRA): CIRAD, 4 p. <https://doi.org/10.18167/coopist/0058>

³¹ Deboin, M.C. (2022). Reconnaître tous les contributeurs d'une publication [Recognizing all publication contributors]. Montpellier (FRA): CIRAD, 5 p. <https://doi.org/10.18167/coopist/0007>

7. Data licensing

This criterion defines the institution’s rules on the licences and copyright applicable to research data.

Overview

The type of licence (or, more generally, the conditions for data reuse), whether open or closed source, must be chosen before data production by the stakeholders as part of a consortium agreement, a contractual framework or a data management plan (DMP).

Within the same data set, and depending on the repository’s general terms and conditions of use, different licences may be assigned depending on the type of content (by file type or by groups of files).

In the case of open data, we have identified different practices from one organization to another. INRAE favours the Etalab licence by default, while IRD and CIRAD, which operate more specifically in an international context, have chosen to assign Creative Commons licences by default.³²

Journals encourage research data to be made available under open licences that allow free reuse.

Data that is not open (by agreement or due to contractual conditions) falls into the category of exceptions (§2. Data types that are subject to exceptions).

BRIDGE recommendation

R13. Data licensing

- ✓ It is advisable to use recognized licences, which offer the advantage of being legally established, machine-readable and compliant with the FAIR principles. The most widely used licences are the Etalab Open Licence (France) and the Creative Commons licences (international). Some licences may be compatible with others, such as the Etalab Open Licence and international standards (ODC-BY, CC-BY 2.0³³).
- ✓ The choice of data repository can limit the choice of licences. Check the licensing conditions before depositing the data.
- ✓ The licence attached to each data set should be discussed in advance and set out in the project’s data management plan.
- ✓ The data co-producers or data managers must agree on the type of licence required to use each data set. Note that when dealing with a jointly produced data set, rules may vary from one partner to another.

If a CC0 licence is chosen, it is important to be aware of the consequences, such as giving up all rights, and especially the right to be cited.

Since national recommendations were issued (e.g. the 2017 “*Guide d’analyse du cadre juridique en France*” [Opening research data. Analysis guide of the legal framework in France]), the implementation of the Law for a Digital Republic for research data has been covered in various publications.³⁴

³² Fily, M.F. (2015). Connaître et utiliser les licences Creative Commons. [Understanding and using Creative Commons licences]. Montpellier (FRA): CIRAD, 11 p. <https://doi.org/10.18167/xtnv-d457>

³³ <https://creativecommons.org/licenses/?lang=fr-FR>
<https://www.etalab.gouv.fr/licence-ouverte-open-licence/>

³⁴ Arènes, C., Maurel, L., & Rennes., S. (2022). Guide d’application de la Loi pour une République numérique pour les données de la recherche [Guide to implementing the Law for a Digital Republic for research data]. Committee for Open Science. <https://hal-lara.archives-ouvertes.fr/hal-03968218>

8. Researcher/author support

Research data sharing and open publication is a relatively new concept.

This criterion specifies measures adopted by the institution to support scientists on this new process: information should be made available and contact persons should be identified to provide assistance on issues related to compliance with data sharing policies, locating repositories, legal provisions, licences, intellectual property rights and routine questions.

Overview

For several years now, INRAE, CIRAD and IRD have been developing ways to support their scientists on data management and dissemination, including documents and websites, online and distance learning courses, and seminars. Many universities and organizations also offer this type of service, giving scientists from joint research units a choice between services from multiple institutions.

The national Recherche Data Gouv project,³⁵ which launched on 8 July 2022, aims to bring together the support resources for research teams across France by creating data management clusters. Its implementation is included in the *Second French Plan for Open Science* as part of the public policy on data, algorithms and source code set by the Ministry of Higher Education and Research.

“The aim of the regional **data management clusters**³⁶ is to provide a local point of contact for research teams on all matters relating to the management and dissemination of research data. To help research teams, the clusters offer general support services for managing, structuring, sharing and producing open research data. These clusters are not intended to replace the existing support services (local, national or international) in certain institutions or in certain disciplinary and subject areas. The data management clusters can draw on the disciplinary and/or subject-specific mechanisms identified as Recherche Data Gouv **thematic reference centres**.³⁷”

BRIDGE recommendation

R14. Generalist support mechanisms for data management, structuring and dissemination should be associated with or integrated into the Recherche Data Gouv data management clusters.

- ✓ Discussions on practices should be held between universities and research institutions. Generalist approaches should be aligned and shared where appropriate. By virtue of their specialization, some existing institutional support mechanisms complement the data management clusters (e.g. international partnerships, human health data).
- ✓ Keep statistics on support activities so they can be monitored over time.

9. Data availability statements

This criterion sets out the rule established by the institution regarding the provision of data availability statements (DASs).

³⁵ Recherche Data Gouv, a national ecosystem for sharing and opening research data

<https://recherche.data.gouv.fr/en/page/about-recherche-data-gouv>

³⁶ <https://recherche.data.gouv.fr/en/page/data-management-clusters-generalist-services-throughout-france>

³⁷ <https://recherche.data.gouv.fr/en/page/thematic-reference-centers-providing-expertise-for-individual-scientific-fields>

Overview

A data availability statement (DAS) is a simple and coherent way of expressing the availability of open data, or in some cases, of justifying why it may be unavailable for legal or ethical reasons.

By default, a DAS contains the data set citation and the availability conditions (contact persons, terms and conditions, licence). A DAS must be well written and contain the full citation, in a standardized format including the DOI, to make it easier to find the data.

Scientific journals are increasingly requiring a written availability statement for open data to be inserted into the published article. Some funding organizations, such as the Bill & Melinda Gates Foundation (USA), encourage this practice in their data policies.

BRIDGE recommendation

R15. Data availability statement (DAS)

- ✓ A DAS is a best practice that should accompany all data-driven publications.
- ✓ Make DAS templates available and advise scientists on how to use them.
- ✓ A DAS does not replace the need to enter metadata into the repository where the data is published. Metadata provides the same information in a FAIR-friendly format (particularly in terms of machine readability).
- ✓ A DAS could be automatically generated using metadata.
- ✓ A DAS itself is not a scientific publication. If this is the aim, researchers should publish a data paper.

10. Data formats and standards (including metadata)

This criterion indicates the institution's choice regarding the recommendation of data and metadata standards that are specific to thematic and disciplinary communities, and whether their use is encouraged or required in some or all cases.

This criterion also indicates whether certain file formats, such as open formats, are recommended or required.

Overview

Research data prepared using open formats³⁸ and community data standards are more reusable, interoperable and accessible.

In France, thematic reference centres are developing expertise in the management, structuring and dissemination of data for the scientific community. They are also participating in creating international standards on describing and disseminating field-specific data.

Generally speaking, journal publishers do not explicitly mention the FAIR principles in their data policies. Some journals encourage authors to share their research data using data and metadata formats and standards recognized by their research community by inviting them to consult the FAIRsharing.org website³⁹ for more information on established data sharing formats and standards.

³⁸ [Format ouvert ou fermé ?](#) [Open or closed format?] – DoRANum

³⁹ <https://fairsharing.org/>

BRIDGE recommendation

R16. A data policy should encourage authors to share scientific data using open formats and FAIR data and metadata standards recognized by the scientific communities.

- ✓ Monitor and track the recommendations and actions of the national thematic reference centres and international scientific communities, including those currently being developed.

The ANR BRIDGE project has identified several useful resources for scientific communities:

- Consult the FAIRsharing.org website for information on established data sharing formats and standards.
- Monitor the work carried out for the Research Data Alliance's RDA for Disciplines.⁴⁰
- To plan for and support a long-term archiving process, refer to the open format recommendations of the CINES (National Computer Center for Higher Education⁴¹), which is specialized in the long-term archiving of documents and digital data.

11. Mandatory data sharing (specific types of data)

This criterion indicates whether data sharing is mandatory for specific types of research data.

Overview

When there are established community mandates for data sharing and mechanisms by which these types of data are to be shared, the data policy may specify compliance with community standards and frameworks.

BRIDGE recommendation

R17. Mandatory data sharing (specific types of data)

- ✓ If open data sharing is a well-established and supported practice among the community of a particular subject area and discipline, this should be specified in a data policy.
- ✓ This criterion should be considered in the context of the actions of thematic reference centres with a national and disciplinary scope, supported by scientific communities that structure their systems to support the management, processing, sharing and open publication of data in their subject areas and scientific fields.

12. Mandatory data sharing

This criterion sets out the institution's rule in cases where data sharing is a condition to submit a scientific project proposal or article.

⁴⁰ [RDA for Disciplines](#)

⁴¹ <https://www.cines.fr/en/preservation/expertises/file-format/>

Overview

In the case of ANR or European projects (Horizon Europe, ERC grants, etc.), funding recipients undertake to guarantee immediate free access to scientific publications and make research data public in a way that complies with the principle of “as open as possible and as closed as necessary.” Some funding organizations, such as the Bill & Melinda Gates Foundation, require all data produced to be publicly accessible, except in exceptional cases, for which the reasons must be documented.

Journals are also increasingly requiring data to be shared and made public, both for peer reviewers and for readers, to encourage reuse.

Some scientific publishers offer their own data repositories for sharing data.

Scientific journals and funding agencies may specify the types of data that fall within this obligation.

BRIDGE recommendation

R18. Mandatory data sharing

- ✓ Check ahead of time whether the scientific journal, the project funder or a project or consortium agreement requires data to be shared.
- ✓ For data sharing, please refer to:
 - Criterion §4 – Data types that are subject to exceptions
 - Criterion §5 – Data repositories
 - Criterion §6 – Supplementary materials

13. Peer review of data

This criterion indicates whether peer review of data is provided for or required under certain conditions of data submission and dissemination.

Overview

When data are made available with research articles, they are accessible for peer review.

For journals with a strong focus on data, such as those which publish data papers, a consistent review of data for validation and reuse can improve the reproducibility and quality of published research. The instructions and assessment criteria⁴² intended for peer reviewers sometimes explicitly specify that the quality of the data set associated with an article should be reviewed.

BRIDGE recommendation

R19. The ANR BRIDGE project notes that peer review of data intended to be made publicly available could be initiated within research teams.

- ✓ The approach could be adopted as part of a review prior to the submission of data papers.

⁴² See, for example, “Guidelines for Reviewers: How to Review a Data Paper”, 23 August 2019. <https://ist.blogs.inrae.fr/afs/2019/08/23/guidelines-for-reviewers-how-to-review-a-data-paper/>.

14. Data management plans

This criterion sets out the institution's rules for drawing up data management plans (DMPs).

Overview

Creating a data management plan is an essential first step in the life cycle of a research project. It very quickly became apparent that the best way to i) limit the ethical, legal and preservation risks, ii) formalize the aspects of provenance, production, co-production in partnerships, and traceability (producer(s), contributor(s), author(s), etc.) and iii) ensure the “FAIRification⁴³” of data was through the use of data management plans.

In a project, the data management plan (DMP) is the most appropriate tool for partners to manage, monitor and control data.

- A DMP is not required by publishers.
- A DMP is required by French or European funding organizations (such as the network of French funding agencies, Horizon Europe programme, etc.).

The three partners in the ANR BRIDGE project – INRAE, CIRAD and IRD – recommend that a DMP be drawn up as part of a project and that the DMP OPIDoR platform⁴⁴ be used along with supporting instructions.

It is advisable to plan for the DMP to be updated throughout the life cycle of the research project.

It should be noted that INRAE has begun using a data management plan template for research entities (research unit, technical platform, infrastructure, etc.) so that they can document their data-related activity by having all the useful information for users in one place. The research entity can make this DMP available to users, such as by posting it on its website⁴⁵ or in a dedicated area in a data repository.⁴⁶

For specific projects, this information could then be used and included in the DMP.

A DMP does not only apply to projects where data is openly available: “A DMP can be drawn up for data sharing as well as for data with restricted, closed, total or partial access. In this case, the DMP should mention the reasons for not sharing.”⁴⁷

BRIDGE recommendations

R20. Data management plan uses

- ✓ Generalize the use of data management plans for projects and dissertations.
- ✓ The data policy should clearly state that the DMP must be updated during the project life cycle. Updated DMPs must be clearly identified by their version number.
- ✓ To create a DMP, draw from the internal or external documents indicated in your organization's recommendations and which set out the legislative, legal and ethical framework for your data.
- ✓ Offer a DMP review service whenever possible.
- ✓ If available, use the funding organization's DMP template.

⁴³ FAIRification process adopted by GO FAIR, <https://www.go-fair.org/fair-principles/fairification-process/>

⁴⁴ National portal and tool for drafting DMPs, supported by INIST–CNRS, <https://dmp.opidor.fr/>

⁴⁵ Example at URGI: <https://urgi.versailles.inrae.fr/Platform/General-Terms-of-Use/Conditions-Generales-d-Utilisation>

⁴⁶ Example at URGI: Michotey, C. (2021). “Plant Bioinformatics Facility data management plan”, <https://doi.org/10.15454/9HM5UI>, Recherche Data Gouv, V1

⁴⁷ Translated from https://doranum.fr/plan-gestion-donnees-dmp/plan-de-gestion-des-donnees-fiche-synthetique_10_13143_cg4-0k53/

- ✓ Recommend a default DMP template to use.
- ✓ Consult the supporting instructions available from DMP OPIDoR.
- ✓ Deposit the DMPs in an open archive so that they can be accessed on a permanent basis.
- ✓ Consider using “machine-actionable DMP” templates from DMP OPIDoR.

Research institutions’ data policies should encourage researchers and support staff to collectively identify relevant FAIR research data management costs. By drawing up a data management plan in advance, funding for these costs could be requested in grant applications.

R21. The policy should clearly state that the justified costs associated with FAIR research data management must be identified in a DMP prior to submitting the project.

V. Other aspects to consider when creating a data policy for research institutions

The ANR BRIDGE project also analysed the relevance of new guidelines (requirements, compliance, professional conduct, etc.) for all research institutions overseeing their research data management activities.

BRIDGE is therefore suggesting two new criteria:

1. one criterion relating to “legal and regulatory” aspects;
2. one criterion relating to “scientific integrity” in research.

15. Legal and regulatory aspects

This criterion sets out the legal and regulatory frameworks with which a research institution’s data policy must comply.

Overview

A data policy must specify any expectations associated with i) legislation on **personal data protection** in Europe and worldwide or ii) **binding legislation**.

For example:

- The General Data Protection Regulation (**GDPR**)⁴⁸ is the strictest privacy and security law in the world. Although it was drafted and adopted by the European Union (EU), it imposes obligations on organizations wherever they are if they target or collect data relating to individuals in the EU.
- Outside Europe,⁴⁹ it is important to note that different levels of protection for personal data have been introduced. Some countries have their own legislation or data protection authority, which should be consulted and respected.

⁴⁸ The General Data Protection Regulation, <https://www.cnil.fr/en/official-texts>

⁴⁹ <https://www.cnil.fr/en/data-protection-around-the-world>

- The Nagoya Protocol on Access and Benefit-Sharing (ABS)⁵⁰ “is an international agreement which aims at sharing the benefits arising from the utilization of genetic resources in a fair and equitable way.”

As part of their research activities, scientists are subject to **compulsory rules derived from legal, contractual and conventional provisions** and which are inseparable from the principles of open science.

BRIDGE recommendations

R22. The policy should clearly state that the legal and regulatory framework for data has been integrated into the institution’s research data management. It should provide links to related policies and any supporting documents. Contact persons should be identified.

16. Scientific integrity

This criterion specifies the institution’s provisions in terms of scientific integrity in research data management.

Overview

According to the 2016 report by Pierre Corvol, Honorary Professor at the Collège de France, “scientific integrity is the honest and upright conduct that must govern all research.”⁵¹

French Decree no. 2021-1572 of December 2021⁵² on compliance with the requirements of scientific integrity requires public research organizations to have a policy for conserving raw research results and for making methods, protocols, data and source codes related to research outputs available.

Research institutions:

- **must provide training, raise awareness and implement checks on scientific integrity;**
- **are encouraged to implement data management plans.**

BRIDGE recommendations

R23. The policy should clearly state that scientific integrity covers research data management. It should provide links to related science policies and supporting documents and services.

⁵⁰ The Nagoya Protocol, <https://anr.fr/en/anrs-role-in-research/commitments/the-nagoya-protocol/>

⁵¹ [Rapport Pierre Corvol](#) (2016). Bilan et propositions de mise en œuvre de la charte nationale d’intégrité scientifique [Review and proposals for implementing the national scientific integrity charter], submitted to the French Secretary of State for Higher Education and Research.

⁵² <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360>

VI. Conclusions and outlook

To begin, we did not find a directly applicable aligned framework for creating a research data policy, although framework documents have been published at the national, European and international levels.

Scientific research takes place in an increasingly complex and interdependent environment, forcing institutions, scientific research units and support services to deal with multiple variables in order to comply with:

- Numerous data policies and recommendations
- Laws and regulations governing data
- Intellectual property of results
- Scientific promotion of research work and results
- Multiple stakeholders and sometimes contradictory rules
- Digital systems and infrastructures that support research
- Scientific integrity for responsible research

The context is changing and the work carried out by the RDA Data Policy IG is helping to clarify and organize this landscape for both scientific journal publishers and funding agencies.

Our study noted that the work of the RDA Data Policy IG is an opportunity for research institutions to align themselves with a framework in order to i) better understand the complexities in which scientific research takes place, ii) develop a data policy that takes into account the rules of this ecosystem and reflects the mandates of each research institution, and iii) create a research data policy that is easier for researchers and support staff to understand.

Finally, this framework could be used within research institutions as a tool for raising awareness, exchanging ideas and consulting with the various departments that are responsible for the areas covered by research data management.

Roles and responsibilities in data governance in research institutions

This key issue focuses on the governance approach implemented by each institution.

The roles and responsibilities at this level of governance were not detailed in our document review. Drawing on the Second French Plan for Open Science, efforts could focus on examining:

- the principles (scientific, ethical, legal, best management practices, innovation) that support decision-making throughout the data life cycle as well as the stakeholders (individual, collective and institutional roles and responsibilities) laid out by INRAE in a working document;⁵³
- work already undertaken on roles and responsibilities, especially by the RDA Sharing Rewards and Credit (SHARC) IG and the FAIRsFAIR project, along with work on data stewardship carried out as part of the EOSC-Pillar.⁵⁴

Data management plans

A data management plan (DMP) plays an essential role in data management planning. We believe it is important to include the DMP among the first criteria of a research institution's data policy document.

While this is already required by French and European funding organizations, it is not a rule at the international level, which results in inconsistent management of projects outside Europe.

This is a difficulty noted by the scientists and support staff when applying this approach to projects outside the European framework, where a DMP is not required. A DMP is more of a non-binding best practice for stakeholders, which limits its usefulness and impact. More needs to be done to raise awareness about the benefits of using a DMP. CIRAD and IRD have put forward a proposal for

⁵³ INRAE DipSO. *Principes pour la gouvernance des données – Document de travail 2 INRAE version V9 – Dec 2019* [Data governance principles – Working document 2 INRAE version V9 – Dec 2019]

⁵⁴ <https://eosc-pillar.eu/data-stewardship-resources>

international training as part of the ECODOR data management clusters⁵⁵ led by the University of Montpellier.

The ANR BRIDGE project partners emphasize that their institutions have no established rule on the use or submission of the final version of a DMP. During the project working sessions, it was noted that a DMP should, at the very least, be shared with all project stakeholders. If it is to be distributed, the partners suggest that it should be deposited in a repository that complies with the FAIR principles. Open archives such as HAL or Agritrop⁵⁶ at CIRAD now accept DMP documents. The DMP should have a permanent identifier so that it can be clearly identified. This approach would support the need for monitoring and evaluation, later use and metrics for DMPs.

Adopt a common data policy framework for research institutions

The study of data policies conducted by the RDA Data Policy IG focused on scientific publishers' policies. We analysed each of the policy features given by the RDA Data Policy IG.

This framework could be developed with the higher education and research community by rearranging the order of the 16 criteria to create a suitable framework for research institutions. Finally, the data governance policy framework for a research institution needs to be assessed collectively. Indeed, it will be discussed at several RDA France conferences in 2022 and 2023.

One key point to keep in mind is the need to ensure that there are no conflicting recommendations or requirements in the data policies of the project leaders working with the joint research units. The work carried out by the three ANR BRIDGE project partners is already leading to proposals for better coordination of research data sharing, particularly within the joint research units.

* * *

⁵⁵ <https://recherche.data.gouv.fr/en/page/data-management-clusters-generalist-services-throughout-france>

⁵⁶ <https://agritrop.cirad.fr/cgi/search/advanced>