couperin.org

Consortium Unifié des Etablissements Universitaires et de Recherche pour l'Accès aux Publications Numériques

Incorporating open science into ANR projects: a practical guide

Guide compiled by the Couperin GTSO (open science working group) for Data

September 2023 (French version: February 2023)

Version 3

About the authors

Cécile Arènes (ORCID n° 0000-0002-1839-3530) is the project lead for the Research Data and Digital Humanities unit at the Sorbonne University library. She supports researchers in writing their data management plans, and teaches on open science courses for researchers and PhD students.

Laetitia Bracco (ORCID n° 0000-0002-2939-9110) heads up the ADOC Lorraine data management cluster, a data support unit for researchers at the Université de Lorraine. She leads the Couperin GTSO (open science working group) for Data and coordinates the French barometer for open science in data and software code.

Élise Lehoux (ORCID n° 0000-0003-2929-6768) is head of research support services and deputy director of the information systems and support department at the Université Paris Nanterre's SCD [*service commun de la documentation*, joint documentation service].

Alicia León y Barella (ORCID n°: 0000-0002-3983-0779) is head of the open science service at the Université de Lille's SCD. She heads up the LORD (Lille open research data) data management cluster, a unit that offers support in submitting and managing ANR [*agence nationale de la recherche*, French national research agency] and European projects in open science-related fields.

This guide is an updated version of:

Féret, Romain, Bracco, Laetitia, Cheviron, Stéphanie, Lehoux, Elise, Arènes, Cécile, & Li, Ling. (2020). Improving your ANR project thanks to Open Science (Version 2). Zenodo. https://zenodo.org/record/3769971

Proofreaders
We would like to thank **Gaëlle Gauvrit (Université de Strasbourg)** and **Cynthia Pedroja (Campus Condorcet)** for agreeing to proofread and make useful changes to this document. The authors of this guide do, however, remain fully responsible for its content.

English version
This English version was funded by the Campus Condorcet. Many thanks to the English proofreaders: **Gaëlle Gauvrit (Université de Strasbourg)**, **Ling Li (Université Lyon 2)**, and **Jennifer Morival (Université de Lille)**.

About this document

This guide was compiled by the Couperin GTSO (open science working group) for Data.

The **Couperin consortium** is a non-profit that encompasses nearly 300 members, primarily universities, research bodies, and *grandes écoles*. In addition to assessing and acquiring digital documentary resources for the benefit of its members, Couperin works to improve communication in the sciences.

The Couperin **open science working group** develops tools and resources aimed at promoting open science within the consortium's member institutions.

The **Data sub-group** develops research data resources and courses for technical and scientific information professionals as well as the scientific community.

Table of contents

1. Why and how to use open science

1.1. On getting the most out of this guide

This guide is available to researchers submitting an ANR project, or those considering it. It aims to **help project coordinators plan ahead** to comply with open science requirements and guidelines, as it can sometimes be too late to start reflecting on this once a project has been awarded funding. It can also be used to support project coordinators throughout their projects' lifecycles, providing answers to questions that may crop up along the way. Used at a project's planning stage, it aims to **improve feasibility in projects** that are submitted and awarded funding, **to facilitate collaboration** between partners in the field of open science, to help **hone a water-tight, ambitious open science strategy**, to plan ahead for risk management, and to budget (data storage and any potential APCs for example). Finally, it is a time-saving tool, in that it raises a certain number of questions as early as possible in the project's lifecycle, rather than as the project progresses.

This guide is not:

- A compendium of cookie-cutter phrases and paragraphs to be copy-and-pasted into project proposals;
- A document to be applied blindly, without considering each project's particularities;
- An ANR-approved document designed to replace or contradict the ANR's rules and policy as outlined in its official documentation;
- A detailed guide to applying the ANR's open science principles throughout the project, with the focus here being on structuring and submitting projects.

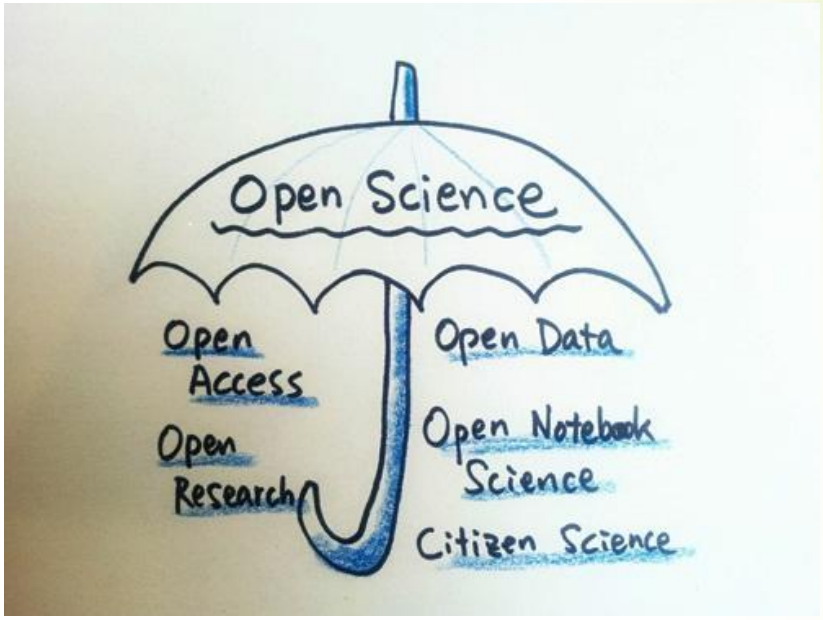
This guide draws on the documentation shared by the ANR for the **AAPG** [Generic Call for Proposals] 2023 and its authors' own experiences. It is intended as an extension of the guidelines issued by the CoSO's various colleges[1].

1.2. France's national open science policy

France's very first National Plan for Open Science was presented by the Ministry of Higher Education and Research in July 2018, and is underpinned by three pillars. It sought to pave the way for **widespread open access to publications**, **to structure and open up research data**, and finally **to contribute to a sustainable, European and international movement.** The plan took the form of a series of measures.

In 2021, a second National Plan for Open Science was issued by the Ministry of Higher Education and Research. In addition to a focus on **publications** and **data**, this plan comprised a set of new focal points on opening up **source code** stemming from research, and on **changing practices** to make open science the default approach.

[1] Data management plan – ANR recommendations. Zoé Ancion, Francis André, Sarah Cadorel, Romain Féret, Odile Hologne, Kenneth Maussang, Marine Moguen-Toursel, Véronique Stoll. June 2019. Available at https://www.ouvrirlascience.fr/plan-de-gestion-de-donnees-recommandations-a-lanr/.

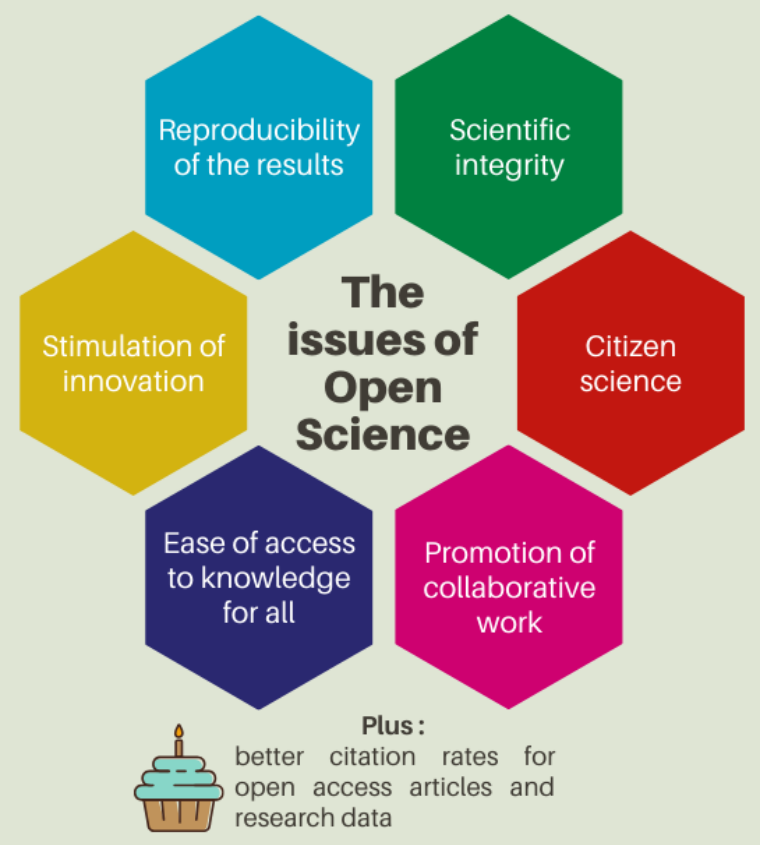


CC:BY 2.0. 지우 황. Flickr. https://www.flickr.com/photos/100477638@N03/10204741904

The notion of open science has gained traction over the past 15 years, taking the shape of a two-pronged approach: opening up the results and output of science (**publications, research data** where possible), and a new way of embarking upon research involving an opening up of **processes**[2]. This arose from a movement that opposed major publishers' monopoly by fostering open access.

Open access sets out to dissolve the financial and legal barriers to accessing publications and research data. Open science has scientific (efficiency, integrity, and reproducibility of results), social (fostering dialogue between the science community and wider society, as well as greater transparency), and economic (encouraging innovation) benefits. Finally, best practices in open science help bolster quality in research.

[2] https://www.science-ouverte.cnrs.fr/le-mouvement-pour-la-science-ouverte/



CC-BY. The benefits of open science. Wikidata.
https://commons.wikimedia.org/wiki/File:Enjeux_science_ouverte.png

Opening up access to publications

Article 30 of the French Law for a Digital Republic (enacted on 7 October 2016) enshrines researchers' right to share versions of their articles accepted for publication (postprints and AAMs) no later than six months from the date on which they are first published by the editor in science, technology and mathematics (STM), and 12 months for researchers in social sciences and humanities (SSH). This law applies irrespective of the contract signed with a journal's publisher, and whether the journal in question is French or not.

Opening up access to research data

In addition to Article 30, research data is covered by the Law for a Digital Republic and the Valter Law on the free reuse of public-sector information and modalities (enacted on 28 December 2015), which ushered in the principle of reusing public information at no cost. By default, these two laws enshrine the very concept of public open data, which includes research data. Provided they are not protected by specific rights (intellectual property, patents, personal data), research data is, by default, governed by the principles of open access and free reuse. Article 30 makes specific provisions for cases where at least half the data is publicly funded, such as partnership research.

Data made public by researchers or their institution can therefore be freely reused if not protected by a specific right[3].

[3] Cécile Arènes, Lionel Maurel, Stephanie Rennes. Guide to applying France's Law for a Digital Republic for research data. Committee for Open Science 2022. ⟨hal-03968218⟩

Opening up code and software[4]

Software is instrumental in scientific research, simultaneously serving as tool, output, and object of study. Making free licensed software source code available, with the ability to change, reuse and share it, is key to ensuring scientific results are reproducible, verifiable, and reusable. This aspect is tackled in France's second National Plan for Open Science (3rd focal point, "opening up and promoting source code produced by research").

1.3. The ANR's open science policy

In 2013, France's National Research Agency began work on an open science policy that gained speed with the very first National Plan for Open Science, followed by the second. The ANR's recommendations and approach aim to support free access to publications and contribute to opening up research data – where possible – in step with European processes. See Open science at the ANR.

Commitments to open science

In 2019, the ANR decided to bring in a series of commitments that funded projects agree to uphold with a view to promoting and fostering open science. If funding is awarded, project coordinators undertake to:

1. Share the funded project's scientific publications (full texts) on HAL

- **Where to deposit:** In an open archive, either directly in HAL itself or via an intermediary institutional archive such as LillOA or UnivOAK;
- **Types of publication:** Scientific articles;
- **Versions:** Versions accepted for publication (postprints) or publisher versions when the latter gives their explicit approval, or for open access licence publications (such as CC-BY);
- **People concerned:** All project members, including foreign authors.
  NB: all researchers attached to a French public establishment governed by French law who have authored a publication can deposit said publication in an open archive (with their co-authors' permission, where applicable). This does not, however, extend to publications written exclusively through international partnerships in the context of PRCI, cross-border projects and bilateral projects (France-Switzerland, France-Quebec, etc.), in which case the journal's policy must be checked, for example on the Sherpa/Romeo website.
- **Timeframes:** Until 2021 (inclusive): In compliance with the provisions of the Law for a Digital Republic (embargo of six months from the date of publication for STM and 12 months from publication for SSH);
  From the AAPG 2022 on: texts must be submitted to HAL immediately upon publication. In this sense, the ANR follows Plan S guidelines.

Immediate deposit with no embargo is possible via three routes:

- Publication in a **native open access journal** (gold open access with APC payment, or diamond open access with no APC payment);

[4] Sur le sujet, voir le livret « Codes et logiciels », dans la collection Passeport pour la science ouverte (https://www.ouvrirlascience.fr/wp-content/uploads/2022/10/Passeport_Codes-et-logiciels_WEB.pdf)

- Publication in a subscription-based journal in the process of shifting over to an open access format (**transitional agreement**);
- Publication in any journal, drawing on a **rights retention strategy** in which the author retains their rights to their published material, and can therefore immediately deposit their publication in an open archive. Using this strategy involves:
    - Applying a CC-BY licence to the different versions of a manuscript from the moment it is submitted;
    - Informing your publisher that you are applying a CC-BY licence to all later versions of your manuscript until the final version that is accepted for publication[5];
    - Depositing your manuscript in an open archive **from the moment it is published.**

Spotlight:

- The *Guide to applying France's Law for a Digital Republic*, published by the Committee for Open Science
- FAQs on the Law for a Digital Republic, written by Couperin
- The guide to *Implementing the rights retention strategy for scientific publications*

2. Provide a data management plan (DMP) within six months of project launch[6]. Coordinators and partners will be required to update their DMP to provide a second version at the half-way point (for projects spanning over 30 months), and a final version upon project completion.

It is not compulsory to share data from ANR projects (particularly if this poses a threat to people or legal entities), but it is **recommended**, unless the data is protected by a specific right. Much of the data generated by research projects falls under the category of public data (cf. 1.1). Research teams have, for example, received requests to access research data[7].

Whether open or not, data must, as far as possible, be produced in line with the **FAIR (findable, accessible, interoperable, reusable)** principles in order to boost visibility and facilitate reuse. In particular, the FAIR principles require that best practices be applied in terms of sharing, descriptions, and the licences applied to the data. Data support services can assist you in achieving this.

In its AAPG 2023, the ANR recommends that software developed during the research project be made available via free licensing and that the source code be archived in Software Heritage and deposited in HAL complete with the ANR project's reference (decision) code.

Finally, the ANR encourages publishing **books** and **monographs** in open access and depositing **preprints** in open archives.

[5] The manuscript submitted to the publisher must include the following line: "This research was fully or partially funded by the National Research Agency (ANR) for project ANR-nn-XXXX-nnnn. For the purposes of open access, a CC-BY licence has been applied to this document by its authors, and will be applied to all later versions up until the final manuscript accepted for publication as a result of this submission".
[6] The ANR advises using its DMP template (in French or English), or Science Europe's structured template, available in DMP OPIDoR, a platform that lets you fill in your DMP online. To view public DMPs, see https://dmp.opidor.fr/public_plans.
[7] Example of requests to access research data:
https://madada.fr/demande/donnees_relatives_a_la_concentra#outgoing-779

2. In-depth proposals: a step-by-step guide

Some aspects surrounding open science can be mentioned as early as the pre-proposal. However, the briefness of the pre-proposal format rarely allows room for developing this angle. Project coordinators who wish to make space for these aspects from the very first step of the AAPG can refer to the points that correspond to the in-depth proposal here (objectives, state of the art, partnerships and bibliography).

Open science is explicitly mentioned in the AAPG 2023 guide with respect to methodology, for example in data management, open source software development, and using persistent identifiers for all research results.

2.1. Project summary

It is important to mention the roles and responsibilities of the team members tasked with implementing the open science policy in the overview table of all individuals involved in the project. Depending on the size of the team, a single person can take on multiple roles, or they can be split among several individuals. The answers to the following questions will provide insight into the key information to include in this table.

- Sharing scientific publications: who will be responsible for sharing and archiving scientific publications and underlying data? Who will handle communication with a wider audience?

- Data management: will there be a data manager responsible for overseeing data within the project? Will each partner have a 'data lead' working alongside the data manager? Who will be responsible for checking data quality and recording (experimental and investigative protocols, production process, etc.) throughout the project? This role can encompass checking the quality of data input, checking cohesiveness, creating README files, etc.

- Code and software: who will be responsible for code readability and reproducibility? Who will oversee HAL and Software Heritage depositing?

Potential changes to the detailed proposal compared to the pre-proposal

Costs during and after the project must be accounted for in the budget. Open science activities call for working time (to be calculated by person/month) and can lead to other types of costs that must be pre-empted when drawing up the budget. This point requires particular care and attention when writing the detailed proposal.

Example: a research team is photographing archaeological artefacts kept in a museum. The coordinator has been granted the museum's permission to take these photos prior to submitting the pre-proposal, but she hadn't foreseen that fees would apply when reusing the images in publications and the project PhD students' theses. She only realises this will be the case when writing up the detailed proposal. This point will need to be specified in the changes made to the project to justify the discrepancy between the pre-proposal and detailed proposal budgets.

2.2. Context, positioning and objectives of a proposal

Part I of the detailed proposal gives space to explain the project's primary objectives, highlighting the potential reusability of the results to come out of the planned project. In terms of data research, depending on each project's particularities, this will mean indicating whether opening up the data involves particular challenges (survey data for example, which calls for complex anonymisation techniques), or adds to existing data in a meaningful way.

2.2.1. Research objectives and hypotheses (Part I. a. of the detailed proposal)

The outcomes of the research to be described in this paragraph can take different forms:

- Scientific publications (articles, chapters of books, talks from conventions, theses, etc.);
- Datasets (example here of a dataset on the impact of climate change on baobabs), linked to a data paper where applicable;
- Structured databases (example of resources made available by Ortolang on language and language processing);
- Source code for software or an algorithm (example of software code for Coq, which won first prize in the open science open-source research software awards);
- Any other output from the project intended for wider society, such as a general public conference, an event at science festivals, etc.

Example: a laboratory produces benchmark humidity statistics. Storing this data in a field-specific data repository in order to extract the benefits of this work and allow it to be reused while ensuring the data is retained can serve as one of the project's major structural open science focuses. In this case, the human, financial and technical means needed to achieve this objective must be planned for. The cost of publishing in journals that charge APCs must also be accounted and budgeted for.

Recommendations: in this section, give some examples of how the resulting data might potentially be reused.

2.2.2. Position in relation to the state of the art (Part I. b. of the detailed proposal)

Establishing a research project's state of the art involves drawing up a bibliography of existing studies, with a similar process undertaken for datasets if needed. Indeed, one of the things that open science sets out to achieve is to encourage researchers to reuse existing data. With this in mind, the project must be positioned in terms of whether or not data is available in the field of study. Where applicable, it can be a good idea to give an overview of these datasets and explain their value (benchmarks) or the problems they pose from a state of the art perspective (the data is old, new data needs to be produced) in order to potentially justify the value in embarking on a new data-gathering exercise in the context of the project. If the plan is to draw on existing data, explain how it will be reused: will you be adding to the data? Will the data be compared to the new data that emerges?

Example: an institution shares an open access prosopographical architecture database on its website. A research team working on teaching methods in the field of architecture reuses this data, supplementing it with data from an archive. The resulting database will be harvestable, meaning that its data will be automatically recoverable by other databases, which spreads it further and wider and makes it easier to find.

It is also possible to use preliminary data generated via previous research projects. This information must be mentioned in the proposal, at which point you can indicate whether this data has been shared, and how it contributes to the state of the art.

Example: a researcher has spent several years building up an unprecedented corpus of antique statuettes. Some of them have been photographed. Her proposal draws on this preliminary data to apply for funding that will enable her to finish photographing the statuettes, as well as creating an open access database that will serve as a reference in the field.

2.2.3. Methodology and risk management (Part I. c. of the detailed proposal)

Methodology

This section gives you space to outline the key stages in the project and any related scientific risks. Many open science factors can crop up in this part. For example, it may be necessary to plan for one or several tasks specifically surrounding open science within the project management work package, with associated deliverables.

Recommendations: appoint one open science lead per partner to make it easier for the person tasked with coordinating DMP drafting and publication-sharing strategy. Specify which partner will handle the coordinating.

If specific tasks surrounding making the results of research open access and usable are planned (such as publishing a data paper, uploading a database, and sharing seminar reports or posters), these can be recorded as tasks and deliverables in the corresponding work packages. If a dissemination report is created, it must explain the strategy for how the data is to be shared and how the scientific output is to be used, and include target audiences.

If the project requires access to research infrastructure (RI*), it can be useful to mention which infrastructures will be used to harvest data or access existing data. Indeed, these infrastructures may offer tools and services aimed at improving data quality and the ease with which it is shared.

Examples of RI*: Huma-Num for digital humanities, Progedo for generating and managing social sciences and humanities data.

Risk management

If, in the context of the project, you will be collecting data that requires specific precautions to be taken, you will need to specify what type of data will be gathered ahead of time (personal data, sensitive data, patient data, biological data, etc.), and how it will be processed. Data protection and data sharing are not binary opposites. It is possible, for example, to render data anonymous and restrict access to it, or restrict its use to a certain number of purposes. Reflecting on this is an upstream process that must draw on the expertise of the research unit or establishment's data protection officer, and that of other competent bodies (such as a research ethics committee). Depending on the type of data collected, an impact assessment may be requested[8]. All these steps should be tackled as early as possible in the project's lifecycle[9].

You will need to reach out to your research unit or establishment's data protection officer to have your research logged on the establishment's data processing register and implement all their recommendations over the course of your research project (informing individuals, security measures, data retention periods, etc.)[10]. These measures will also need to be outlined in the DMP's dedicated sections.

Example: beQuali, a qualitative survey data bank in the field of SSH, provides an inventory of surveys that have been reversibly or irreversibly anonymised. Generally conducted as interviews or observations, the surveys are then filed and made available to be reused for academic purposes exclusively, via controlled access.

Recommendations: Horizon Europe offers a useful guide for assessing ethical issues surrounding projects.

Risk assessments can also specify whether some scientific risks are linked to data management, such as uncertainty on whether or not permission will be granted to collect sensitive data (or on turnaround times), difficulties surrounding storing data, or sharing it back and forth between project partners.

In terms of opening up data, as early as possible you will also need to consider whether some of the data is bound by confidentiality (business, defence, health) to identify any risks and plan ahead for storage with heightened security.

Conversely, in cases where the data output can be opened up, it is vital to reach an agreement with your various partners ahead of time to determine under what conditions it will be shareable (storing datasets in one or several mutually agreed repositories, access to pseudonymised data on specific databases).

Finally, when working on projects with a private partner, the issue of data access at the end of the project and archiving it over the long term must be given careful consideration to ensure the results remain accessible to team members.

[8] https://www.cnil.fr/fr/RGPD-analyse-impact-protection-des-donnees-aipd

[9] See 'Social sciences and humanities and personal data in an open science context' SSH research guide, INSHS-CNRS, V2, 2021: https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf.

[10] https://www.cnil.fr/fr/recherche-scientifique-hors-sante

All these points must be tackled and locked down ahead of signing the consortium agreement or contracts. A lack of formal agreement on this issue may prevent output from being shared in optimal conditions should disagreements between partners arise.

Recommendations: the questions that crop up in the data management plan concern risk management and methodology. It can be useful to be aware of this from the moment a call for proposals lands to ensure that all partners agree on standardised practices. If you don't know whom to ask for help in writing up your data management plan, check SOS-PGD, a directory of services that offer support with writing data management plans in universities and research bodies, designed to help you think ahead in tackling these questions as early as the proposal stage.

To find out whether the research data generated through your project can be shared or not, one invaluable resource is the guide to legal frameworks in France for open research data. Other resources may be helpful, too:

- Guide to applying France's Law for a Digital Republic to research data (open science committee);
- Flowchart template for legal questions on sharing research data (Institut Pasteur);
- Opening up your data (MSH Lorraine);
- Research data in health (Université de Lorraine).

2.3. Organising and conducting your project

Part II of the detailed proposal is an opportunity to showcase the project consortium or team experience, by highlighting each project member's expertise and demonstrating the complementary nature of the collaboration. The ANR's open science policy requires project coordinators to identify the means and skills they already have at their disposal, as well as the ones they will need.

2.3.1. Project coordinator and consortium/team (Part II. a. of the detailed proposal)

Whether the project is individual or collaborative, the scientific coordinator will gain from highlighting their experience and past work in open science.

The balance between how data management will be centralised and decentralised should feature in the detailed proposal, with explanations of how roles and responsibilities will be shared among the partners.

The project coordinator

Scientific expertise: in their personal introduction section, the project coordinator can mention past work in open science, such as a significant number of publications deposited in open archives, datasets stored in general repositories (such as Recherche Data Gouv) and discipline-specific repositories (Protein Data Bank, Cambridge Structural Database, etc.), or contributing to major databases in their field. You can, for example:

- Give the number of open access publications you have via the Dissemin website;
- Give the number of datasets you have already shared via the DataCite search engine;
- Mention your contributions to projects that seek to share or give commercial or social value to research output (writing data papers, etc.).

Project management experience: more broadly, the coordinator can also indicate whether they have taken part in past research projects (ANR, Horizon Europe, H2020, etc.) covered by an open science

policy, or whether they have helped write a DMP in the context of a non-research project (non-funded project, thesis, etc.).

Spotlight: identify the resources available within institutions involved in the project

Just as coordinators generally turn to support units for issues concerning administrative and financial matters, it is vital they identify the support networks in place to assist them with open science aspects. See Section 3: 'Help with writing: resources and contacts'.

Individual projects

In addition to the aspects mentioned above, coordinators for individual projects will benefit from clarifying the following points:

- Their existing skills in open science, and the skill sets they intend to hone over the course of the project;
- How roles and responsibilities for data management and publication-sharing activities are allocated within their team.

Single-team projects

In addition to the aforementioned, the project coordinator can highlight open science work their team has completed (team publications deposited in open archives, datasets shared in a repository, etc.) and explain which results could be opened up as part of the project.

Collaborative projects: general coordination

In addition to the points mentioned above, coordinators for collaborative projects will need to outline interactions between each partner, particularly in terms of data management and open sharing of the publications. The detailed proposal must make clear whether partners are working on the same types of data, which involves pooling data and standardising/streamlining practices (naming rules, organisation, saving processes, etc.), or whether they are working on different data that only requires the results to be shared.

This can affect the solutions chosen for storing and sharing data, as well as the degree of standardisation needed for managing data across partners. All the questions that should be asked in this respect can be found in the data management plan. Although this isn't required when submitting the proposal, being aware of and reflecting on these issues ahead of time can be of benefit.

Collaborative projects: partner by partner

In addition to an overarching strategy, it is important that each partner provide a certain amount of information about themselves. Depending on the project at hand, each partner might, for example:

- List the main types of data they will be generating;
- Specify the team's areas of expertise in terms of open science;
- Give a breakdown of roles and responsibilities within their team;
- Describe the support their institution is providing: infrastructures, support from specific units (data management cluster, library, research and innovation service, DPO, etc.).

Recommendations: the detailed proposal should be concise. You cannot go into detail for all points listed above. Identify your project's specific characteristics to home in on the key points to be mentioned.

2.3.2. The means implemented and requested in order to meet objectives (Part II. b. of the detailed proposal)

Most open science-related expenses can be at least partially covered by the ANR, **provided they were budgeted ahead of time, upon submission.** This mean it is crucial you anticipate these costs.

The means needed to roll out an open science strategy depend on the targets the project members set themselves and the complexity level of the data involved (data sensitivity, volume, etc.). The means requested are dependent on the means that are implemented, which is why we decided to structure this section into types based on the ANR's expense categories. With regard to open access charges, known as APCs, projects are not required to draw on these to comply with the ANR's policy (the diamond model or the rights retention strategy do not incur any costs for authors). If, however, open access expenses will be incurred, they must be anticipated and budgeted for (see below).

In terms of finances, the ANR's financial guidelines apply. The eligible expenses fact sheet is a useful way of pinpointing what is and isn't eligible.

The HornEast project's data management plan, for example, offers a solid breakdown of the means implemented in order to roll out their overarching open science strategy.

Staff costs (category a[11])
More often than not, these expenses encompass the majority of all open science expenses, although they are generally included in researchers' working hours in the form of everyday research activities: sourcing data, organising and documenting data, etc.

Implementing an open science strategy does, however, call for additional work to varying degrees, depending on the level to which project members' research practices are structured and on practices in the field. This additional working time will be dedicated to tasks such as depositing publications in open archives, anonymising data, improving data documentation, systematically standardising how files are named and organised, saving files in different formats, depositing datasets in dedicated repositories, and archiving software.

These best practices are key to sharing data that can be reused by other researchers in line with the FAIR principles. To find out more on applying these best practices, read through the information available on the DoRANum platform. Quality assurance tasks can be split among the project's stakeholders or mostly entrusted to a data manager. Either way, the working time needed to manage the data must be estimated as accurately as possible, in line with the objectives set by the consortium.

Tools are available to make it easier to calculate the time needed to curate the data:

- The costing data management tool from the UK Data Service;
- The University of Utrecht's breakdown of the costs of data management across every stage of the data lifecycle.

[11] Category here refers to the ANR's expenses categories as stipulated in its financial guidelines.

Spotlight: sharing interview data that contains personal data

To comply with GDPR requirements, the project's personal data cannot be shared and must undergo an anonymisation or pseudonymisation process in order to be shared at a later stage. This is the case with data collected by interview, for example.

To make it easier to reuse and anonymise this data, you may need to transcribe and then anonymise it manually or using specialist software. The University of Manchester has uploaded a dozen toolkits guiding researchers through processing qualitative data, in particular when transcribing interview data. When calculating the time needed, it is vital that a certain number of factors be taken into account, such as the number and duration of the interviews, the number of people being interviewed at once, the quality of recording, and the transcriber's experience. Four hours are generally needed to transcribe one hour of interview time to a satisfactory standard.

Tools and material costs (category b)
In this category, open science-related costs might encompass the following needs:

- Storing and sharing data: whether the solution is provided by an IT department or external service provider. This can cover purchases of physical storage devices (NAS devices, external hard drives) or access to cloud space, as well as services for transferring high volumes of data securely. Some of these costs can be categorised as service costs (category d);
- Digitising documents if this requires buying or making heavy use of a scanner.

The EPFL's library has put together a Cost Calculator that lets you estimate costs incurred by storing and archiving data. Its calculations are merely guides and must be looked at in the context of the services on offer from your establishment. The Université de Lorraine, for example, gives its researchers free access to its high-volume data storage service, PETA.

Building and land costs (category c)
This category of expenses does not have an impact on open science aspects.

External services (and intellectual property rights) costs (category d)
In this category of expenses, project coordinators can include outsourced tasks, such as:

- Cloud access;
- Anonymisation;
- Digital scanning;
- Transcription.

They can also list costs related to purchases and maintenance for the software used to render data shareable and reusable, such as software for transcribing data, anonymising data (see our suggested tools), or converting files.

This category of expenses can also include costs related to open access sharing for articles, datasets, and more.

Spotlight: open access publishing costs (APC, BPC, BCPC, DPC)

APCs (Article Processing Charges) are fees charged to publish an article in open access.

To find out whether an open access journal applies APCs and how much these charges are, check the DOAJ (Directory of Open Access Journals) and run a search. The *Ecology and Society* journal for example applies APCs of at least $975 per article.

The QOAM (Quality Open Access Marker) database lets you compare the journal's quality (peer-reviewing process, editorialization of content, publishing times) and APC fees, and check the amounts of APCs paid by authors in practice. The *Ecology and Society* journal, for example, scores 4.4/5.

In its Publications FAQs, the ANR states that: APCs are eligible expenses when an article is published in either of the following two cases:
- Published in a native open access journal;
- Published in a transformative journal or one with a transformative agreement in place.

Publishing costs for hybrid journals (subscription-based journals with no transformative agreement in place) are therefore not eligible.
To check a publisher's open science policy and see whether it is compatible with the ANR's policy, use the Journal Checker Tool.

Other forms of scientific output are covered by open access expenses too, such as:
- Books (BPC: Book Processing Charges) and book chapters (BCPC: Book Chapters Processing Charges). The publisher Berghahn Books, for example, charges authors $15,000 to publish their book in open access, and $2,000 for a chapter.
- Datasets (DPC: Dataset Processing Charges). Data repository Dryad for example applies publishing charges in some cases. Note that it is free to deposit data in the Recherche Data Gouv repository.

As with all ANR project costs, publishing charges are only eligible if they are paid prior to the end of the project, once the service has been provided.

Recommendations: in terms of open science and as in the rest of the project, it is essential that the means requested align with the project's objectives and the consortium's existing means and resources.

Overheads (category e)
This category of expenses does not have an impact on open science aspects, with the exception of training and courses in this area that incur travel costs for the teachers or instructors used.

2.4. Project impact and knock-on effects (Part III. of the detailed proposal)

Part III of the detailed proposal examines the strategy to be used for sharing and spreading awareness of the project. As some aspects are common to all projects, project coordinators can attempt to stand out by choosing to focus on specific aspects depending on the funding scheme in question.

2.4.1. For all funding schemes

In the context of open science, project coordinators will benefit from building their sharing strategy around three underlying principles: sharing publications in open access, sharing data and code, and the connection between open science and practical applications of research.

Publications
This section is an opportunity to explore the sharing strategy that will be used for scientific articles in order to immediately share publications in open archives. Use this space to explain the balance you intend to strike between publishing approaches in order to secure immediate open access (diamond model, gold open access journals or a rights retention strategy). It is vital that the strategy chosen by the consortium's various partners be coherent.

It is also important to specify how other scientific output will be shared, such as preprints, books or book chapters, proceedings, slideshows presented in conferences, posters, and theses. Researchers should share these productions as widely as possible: they can be deposited in to open archives with mention made of ANR funding via the project code (a dedicated field can be used for this in HAL).

Research data:
Sharing FAIR data now features in ANR guidelines (cf. AAPG 2023) and is a good way of both boosting a project's visibility in the field and fostering partnerships. It is also a way of ensuring that the project output go on to be reused once the project ends.

Where possible, data sharing and disseminating should be highlighted. There are different ways project data can be shared, some linked to publications and others not:

- Sharing in field-specific and general repositories, which generates a DOI for the dataset and creates a reciprocal link between publications and datasets;
- Creating a specific database (specifying how it aligns with FAIR principles);
- Publishing a data paper.

Example: researchers publish an article in the *Dalton Transactions* journal. The corresponding dataset is deposited in DOREL and harvested in Recherche Data Gouv, and a link between the article and dataset is established via the DOI.

For some data, there is no value in sharing it, or it may not be shareable, in line with the "as open as possible and as closed as necessary" principle. It can be a good idea to specify which data cannot be shared and why, and what you intend on doing with it (deletion, confidential access).

Code and software
The 2023 generic call for proposals specifies that "in line with the Second National Plan for Open Science, the ANR recommends that software developed during the project be made available via free licensing and that the source code be archived in Software Heritage and deposited in HAL complete with the ANR project's reference code". A tutorial is available to this end.

From open science to research transfer
Research and funding bodies are increasingly focusing on putting science to work for the benefit of society. This notion encompasses realities that vary greatly by project and field.

The takeaway point here is that there is an open science and research transfer continuum, stretching from sharing scientific output aimed at research communities and professionals in the field (manufacturers, healthcare professionals, etc.) to the general public and intermediary audiences with a mediation role, such as secondary school teachers and science journalists.

Showcasing the results of your research on open access 'general public' platforms such as *The Conversation* involves both open science and research transfer, with citizen science being a key priority.

In terms of extracting commercial value from research output through patents, for example, take care to ensure that publishing data does not jeopardise securing protection for an invention. If in doubt, it is better to wait until filing the initial patent application to establish the invention's prior art, and then open up the data later, once the embargo period is over. You can discuss this topic with TTCs, ISSM, and tech transfer teams.

2.4.2. Particular approaches for specific funding schemes

PRCI

For PRCI, it is a good idea to demonstrate how the strategies chosen by partners in different countries complement one another, both in terms of sharing publications and data. Using European platforms and aggregators such as OpenAIRE can make it easier to showcase shared output.

It is important to plan ahead for this part, as each national agency can have their own specific open science requirements your partners will need to comply with. The Swiss National Fund (SNF), for example, requires a DMP upon submission[12].

PRCE

It is important to take open science aspects into account when drawing up the consortium agreement that binds the project partners, as this can impact on how ownership is shared and developing end products such as software. Striking the right balance between protection, research transfer (patents, software, databases, models, etc.) and sharing them can call for you to dig deep in developing a mixed strategy.

Example: a researcher collaborates with a start-up to develop software for a PRCE research project. The basic features will be free and run under a GPL license, while add-ons developed by the start-up will be paid. This is an example of how research output can be commercialised.

PRC, JCJC and PRME

Submissions here are similar to the sharing plans frequently required in European calls for proposals. An open science strategy must be mentioned as a key component of the approach to sharing, specifying how research results will be shared with various academic stakeholders (lecturers and researchers, students, etc.) and across other audiences (citizen science, press, decision-makers, etc.). Opening the results up to a wider public and non-academic audiences can be more time-consuming than traditional forms of academic communication, and as a result this needs to be planned ahead of time to ensure the project is cohesive.

[12] http://www.snf.ch/fr/leFNS/points-de-vue-politique-de-recherche/open_research_data/Pages/default.aspx

2.5. Bibliography and CV (Part IV. of the detailed proposal)

The bibliography, Part IV of the detailed proposal and the CV are appendices, but they too can have an open science focus.

Bibliography

- The ANR recommends providing a **link to any existing open access versions of articles**. In addition to making it easier for science experts and SEC members to access the articles, it shows an awareness of the principles of open science. The Unpaywall browser extension is a very quick and useful way of seeing whether open access versions of references in the bibliography are available;
- The ANR asks that **journal impact factor not be included in the bibliography**, as recommended by the *San Francisco Declaration on Research Assessment* (DORA). Generally speaking, and as far as possible, it is better to avoid using this type of metrics in the detailed proposal. However, some SECs are still sensitive to this, and you can bring these metrics up in your presentation of the team or consortium members.

CV

- Where possible, it is best that the **five main publications listed by the scientific leads be available in open access.** If they haven't been published in an open access journal, this is an opportunity to upload them to an open archive. You can then mark them as open access or link to an open access version. It is worth noting that HAL lets you automatically generate a CV based on publications available in open archives;
- If the **data** underpinning these publications has been shared, this should also be flagged with a link or the DOI. Don't underestimate how valuable the publications' data can be;
- You can also mention open source code and software in a Github archive or Software Heritage;
- The **research transfer** section is an opportunity to showcase action undertaken to make use of scientific output, such as sharing a database, writing a data paper or archiving and sharing open source software (in Software Heritage, for example).

ORCID ID

- An ORCID ID is a 16-digit identifier **that is unique to each individual researcher,** even when they share a name with another or publish under various names (for example following a mid-career name change). It can be linked up to other researcher identifiers such as the IdHAL, ScopusID and ResearcherID;
- You are not asked to include this number on your CV or the detailed proposal, but it should be given for the project's leads when completing the ANR's online form. **All scientific leads and the coordinator must have an ORCID ID and complete profile**, with a list of their publications as a bare minimum. University libraries and technical and scientific information departments generally offer support in creating and filling in ORCID profiles.

3. Help with writing: resources and contacts

Because each establishment has its own way of working, the following list may not be exhaustive despite our best efforts to collate all lead contacts and resources. The terms used and department names can also vary from one institution to the next.

Not all individuals are open to contact at the project planning stage: this will depend on each project's characteristic.

Planning a research project:

- Research department, departments dedicated to research project prep and planning.

Open access publications, APC budget:

- SCDs [joint documentation services] and libraries;
- Laboratory research engineers.

Data management, technical storage-related issues, choosing repositories:

- Data management clusters;
- SCDs [joint documentation services] and libraries;
- Laboratory engineers tasked with scientific data processing;
- IT services;
- Archivists;
- Field-related queries: Recherche Data Gouv thematic reference centres.

Data to be protected (personal data, national security secrets, trade secrets and business confidentiality, health data):

- DPO;
- FSD;
- ISSM;
- TTC;
- Ethics committee;
- Personal safety committee.

Glossary

AAP: *appel à projets* [Call for Proposals].
AAPG: *appel à projets générique* [Generic Call for Proposals]. This funding scheme accounts for around three-quarters of ANR-funded projects.
ANR: France's national research agency.
Consortium agreement: a contract that binds the project's partners and establishes the terms of their collaboration.
CoOS: Committee for Open Science. Drives and supports national open science policy.
Data paper: an article outlining one or several datasets, and particularly their reusability potential. A data paper can be published in specialist journals (data journals) or traditional science journals.
Diamond open access: articles published in open access, with no charges paid by either the reader or the author. Funding is provided ahead of time by an institution, a research-funding body, a non-profit, etc.
DMP: Data Management Plan. Document containing questions and giving a summary description of a project's research data and how it is managed throughout the project, in order to lay the groundwork for sharing, reusing and securing it.
DOAJ: Directory of Open Access Journals. An index of open access journals.
DOI: Digital Object Identifier. A unique identifier assigned to a dataset or publication.
DORA: The San Francisco Declaration on Research Assessment, which is calling for a change in the criteria used to assess researchers and is challenging the use of impact factors.
DPO: Data Protection Officer.
Eligible/ineligible costs: used to describe expenses that the ANR either does or does not cover. Eligible costs must align with the rules of the ANR's financial guidelines. Research body partners are generally funded at marginal cost. In particular, this excludes pay for permanent staff.
FAIR: guiding principles for opening up data, which aims to make data Findable, Accessible, Interoperable and Reusable.
FSD: the *Fonctionnaire sécurité défense*, a security/defence official within universities in France, tasked with coordinating the establishment's activities in terms of protecting scientific and technical potential and national security and defence secrets.
GDPR: General Data Protection Regulation. This regulation lays out the framework for collecting and processing personal data.
Gold open access: articles published in open access, with APC (Article Processing Charges) paid by the author.
Green open access: articles deposited and shared in an open archive, by an author (self-deposit) or a third party.
H2020: Horizon 2020, the EU's research and innovation funding programme (2014-2020).
HE: Horizon Europe, the EU's research and innovation funding programme (2021-2027).
Hybrid journal: a subscription-based journal where some articles can be made open access provided open access publishing charges are paid. Most publisher journals like Elsevier and Springer are hybrid journals.
ISSM: Information Systems Security Manager.
JCJC: the *Jeunes Chercheuses Jeunes Chercheurs* scheme [young researchers scheme]. An individual funding scheme for early-career researchers (researchers who defended their thesis less than 10 years ago).
Journal Checker Tool: a plan S-aligned compliance-checking tool.
Metadata: the information needed to describe data, generally structured in alignment with a standard.
Open access journal: journal where all articles are immediately available to the public.

Open access publishing charges: APC (Article Processing Charges), BPC (Book Processing Charges), BCPC (Book Chapter Processing Charges).
Open access: describes scientific publications made available permanently and free of charge online. There are different types of open access: green open access, gold open access, and diamond open access.
Open archive: a platform where documents that are deposited are open access. Open archives can be national (like HAL), institutional, or field-specific (like arXiv). Researchers can add a note describing their publications (title, authors, summary, etc.) and then link a file to this note depending on the rights.
Open data: data that are free to access and unrestricted.
OpenAIRE: a consortium that aims to support European scientists' research by creating and managing open access infrastructure. OpenAIRE aggregates research output (publications, datasets, software) and matches it to projects funded by European research funding bodies (H2020, ANR, etc.).
ORCID: Open Researcher and Contributor ID. A unique international identifier for researchers.
PC: Personal safety committee. The PC's opinion must be sought for research involving people. Its opinion is binding.
Plan S: an initiative sparked by a group of research funding bodies that aims to ensure that "all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo".
Postprint or author-accepted manuscript: version containing the revisions to come out of peer-reviewing by a committee. Unformatted or partially formatted file.
PRC: *Projet de recherche collaborative* [collaborative research project]. Funding scheme for multi-partner projects that do not involve non-French partners or companies.
PRCE: *Projet de recherche collaborative - Entreprises* [collaborative research project - Companies]. Funding scheme for multi-partner projects that involve one or several companies.
PRCI: *Projet de recherche collaborative - International* [collaborative research project - International]. Funding scheme for multi-partner projects that involve one or several non-French partners.
Preprint or submitted manuscript: version submitted to a journal by the authors prior to peer-reviewing.
PRME: *Projet de Recherche Mono-Équipe* [single-team research project].
Project coordinator: the person responsible for overseeing the entire project (not to be confused with the Science Lead).
QOAM: Quality Open Access Marker. A platform for assessing integrity and quality of publishing in scientific and scholarly journals that publish open access articles.
REC: Research Ethics Committee. An ethics committee issues advisory opinions to help researchers take ethical principles surrounding their research into account.
Recherche Data Gouv: an ecosystem working to support open access research data, including a data repository, a catalogue and support services (data management clusters, thematic reference centers, resource centres).
Repositories: platforms where research datasets are deposited, described, and stored. Repositories can be general or field-specific.
Research data: all scientific information produced or collected as part of a research project. Research data include photos, measurements, sounds, and more. It is evidence-based material used to validate research output and must be supported by documentation such as experimental protocols, descriptions of working methods and metadata. These data can be shared in general or field-specific repositories.
RI*: Research Infrastructures.
Science Lead: person responsible for overseeing activities conducted by the project's partners (not to be confused with the project coordinator).

SEC: Scientific Evaluation Committee. There are 48 SECs, one for every area of research (35 areas of field-specific research and 13 cross-disciplinary areas). The committees draw up the list of ANR-funded projects, drawing on reports written by scientific experts tasked with assessing the projects. These experts are external to the SECs.
SNF: Swiss National Fund. Switzerland's main scientific research funding body.
Transformative agreements: agreements between publishers and institutions (libraries, consortiums, etc.) aimed at shifting from a subscription-based financial model over to a model that enables open access to publications while ensuring fair pay for publishers.
TTC: tech transfer company tasked with commercialising research output.
Version of record, final version: article complete with final formatting and layout. Version published by the publisher.
ZRR: *zone à régime restrictif* [restricted zone], regulated access in the context of protecting France's scientific and technical potential. Generally applies to laboratories, or certain areas within a laboratory.