



# CNRS RESEARCH DATA PLAN

November 2020



## INDEX

1   Introduction	4
2   General principles	6
3   Objectives	7
4   Research data policy	8
5   Research data governance	9
6   Action plan	10

# 1 | INTRODUCTION

***This plan comes as part of the response to the requirement to accelerate the development of research towards Open Science. It therefore concerns all data resulting from research in the broadest sense as defined in the National Plan for Open Science<sup>1</sup> (PNSO).***

Not all research data is suitable for being opened or disclosed to others. There are obvious exceptions such as, for example, specific data which are confidential for personal reasons or because this would be negative for industrial competition or States' fundamental or regulatory interests. Opening data is understood according to the European research community's definition namely that data should be «as open as possible and as closed as necessary». The decision to open or protect research data needs to be taken with the CNRS departments concerned - the Partnership and Technology Transfer Department for intellectual property matters, the Data Protection Department for personal data and the Security Department for sovereignty matters. This plan and the initiatives it proposes deal with data intended to be open including raw or reprocessed data in all their formats, texts and documents, software, algorithms, protocols and workflows. The plan covers the whole data ecosystem, i.e. aspects related to both digital infrastructures and all services particularly in the context of the implementation of the FAIR principles<sup>2</sup> (Findable, Accessible, Interoperable and Reusable).

The diversity of the subjects linked to research data is obvious when we consider their 6-stage life cycle:

- creation or collection
- processing
- analysis
- preservation
- sharing
- reuse

It is important to understand all these steps to implement proper research data management throughout the whole cycle made up of several intervention levels (e.g., data generation including calculation, data storage, data FAIRisation through data processing and storage).

Storage - whether temporary or long-term is different from archiving. Storage includes data persistence, identification, indexing and optimisation of access for frequent intensive processing whereas archiving involves the preservation of data for legal or historical reasons. Storage and archiving technologies are generally different. The basic material may be the same but the rigour required in their inherent procedures and the systematism involved are different.

There are many ways of characterising research data which depend on the type of data involved, the way they are produced, the different processing stages and the different ways the data is presented to users. Types of data include numerical or symbolic data, text, images, graphs, sounds, etc. They may be the result of observations, calculations or various transformations and can be raw, curated, integrated or aggregated. Research data includes processed data and products derived from data, data included in publications or which are themselves the subject of specific «data papers».

Research data needs to be made reusable and even shareable so that other researchers who did not work on creating the original data can subsequently reuse it.

It should be noted at this point that research communities often define the terms «research data storage» and «research data repository» differently. For some, storing data on a laboratory hard disk for example does not necessarily mean the data has been processed or that associated metadata has been created to make it easier for others to reuse the data. Other communities (particularly the open science community) define a data repository as an online service which manages the overall description of data for preservation and reuse. This definition of a research data repository will be used throughout this document and does not encompass the physical space or physical infrastructure containing the data.

Data archiving and curation, or data management and sharing in broader terms, offer many advantages which will not all be set out in this plan. The availability of data linked to a scientific publication is an essential component for the understanding and validation of scientific results and also provides the basis for the reproducibility of processes that led to those results. Also, reusing existing data rather than redoing the calculations or experiments which produced them uses public investment in research more efficiently. Finally, new knowledge may emerge from combining data from very different research communities thus leading to the development of original research themes. This also requires data to be shared in the research context which implies a certain level of quality, contextualisation and even peer review.

This plan is based on studies and thought which resulted in a draft CNRS white paper in January 2018<sup>3</sup>. Detailed analysis carried out in each CNRS institute then led to the conclusion that a true «data culture» urgently needed to be promoted to provide the CNRS with a robust strategy to respond to the needs of its communities in terms of platforms for large-scale data analysis and to implement a data management, enhancement and sustainability policy.

1. PNSO, July 4<sup>th</sup>2018 : [https://cache.media.enseignementsup-recherche.gouv.fr/file/Recherche/50/1/SO\\_A4\\_2018\\_EN\\_01\\_leger\\_982501.pdf](https://cache.media.enseignementsup-recherche.gouv.fr/file/Recherche/50/1/SO_A4_2018_EN_01_leger_982501.pdf)
2. *Turning FAIR into reality, Final report and action plan from the European Commission expert group on FAIR data, European commission (2018)* : [https://ec.europa.eu/info/sites/info/files/turning\\_fair\\_into\\_reality\\_1.pdf](https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf)
3. *White paper on data at the CNRS, overview of the situation and practices, Calculation and Data Mission, January 2018* : [http://www.cocin.cnrs.fr/IMG/pdf/livre\\_blanc\\_donne\\_es\\_2018.pdf](http://www.cocin.cnrs.fr/IMG/pdf/livre_blanc_donne_es_2018.pdf)

## 2 | GENERAL PRINCIPLES

**The Digital Republic Law - or Lemaire Law<sup>1</sup> - of 2016 led to an evolution in the legal framework for research data produced by projects receiving public funding towards the principle that data should be made openly available. However, this remains relatively unknown to a large number of scientists.**

Beyond the legal obligations linked to this question, the CNRS adopted its roadmap for open science in November 2019<sup>2</sup>, following the objectives on open science defined by the European Union and the French Ministry of Higher Education, Research and Innovation (MESRI). This document already included a section on «research data» and defines the availability of research data as one of the driving forces for future advances in knowledge.

The explosion in the volume and diversity of research data, methodological developments in Big Data and the new analysis possibilities afforded by Artificial Intelligence (AI) all mean that the CNRS and its partners must adopt a policy and a strategy based on the FAIR principles and the principle of openness of data (apart from legitimate exceptions).

As one of the major European producers of research data, the CNRS needs to develop a proactive and easily understandable strategy and policy coupled with quality data services for research. An approach was proposed in the framework of the European Open Science Cloud (EOSC) initiative<sup>3</sup>. It quite naturally involves the promotion of a «French EOSC» which will act as both a contribution and the logical gateway to the broader European initiative. This approach is part of international initiatives thanks to the CNRS's involvement in very large instruments, observation systems and data infrastructures.

The Research Data Plan is primarily a response to the needs of scientists and will take varied disciplinary contexts into account. The necessary combination of the work carried out in CNRS institutes and cross-disciplinary data governance is a key strategic point for the successful implementation of the Research Data Plan.

The main elements involved are:

– the work carried out by CNRS institutes which take the diversity of approaches and communities into account and

develop national strategies to organise their research communities around data;

– the desire to provide the least advanced communities with benefits of the most advanced knowledge and experience without penalising them;

– a «trickle-down» effect from communities whose data constitute a research object to «user» communities.

The action plan is intended as a response to the requirement to change practices and mentalities while promoting the development of the right tools for the management, sharing, long-term preservation and dissemination of research data in accordance with the FAIR principles. The specific issue of the means required, particularly in terms of human resources and a training plan, is an integral part of the Research Data Plan.

1. Digital Republic Law (2016): <https://www.republique-numerique.fr/pages/in-english>
2. The CNRS Roadmap for Open Science: [https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/CNRS\\_Roadmap\\_Open\\_Science\\_18nov2019.pdf](https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/CNRS_Roadmap_Open_Science_18nov2019.pdf)
3. EOSC : <https://eosc-portal.eu/>

## 3 | OBJECTIVES

**It is essential that practices and mentalities evolve and this should be fostered as part of the general organisation of research at the individual level and above all collectively. Initiatives working towards this implemented by the CNRS and its institutes, scientific communities and research infrastructures must therefore encourage all research structures to adopt «data policies». The CNRS Research Data Plan will thus be structured around three main objectives:**

### • Disseminating the FAIR data culture

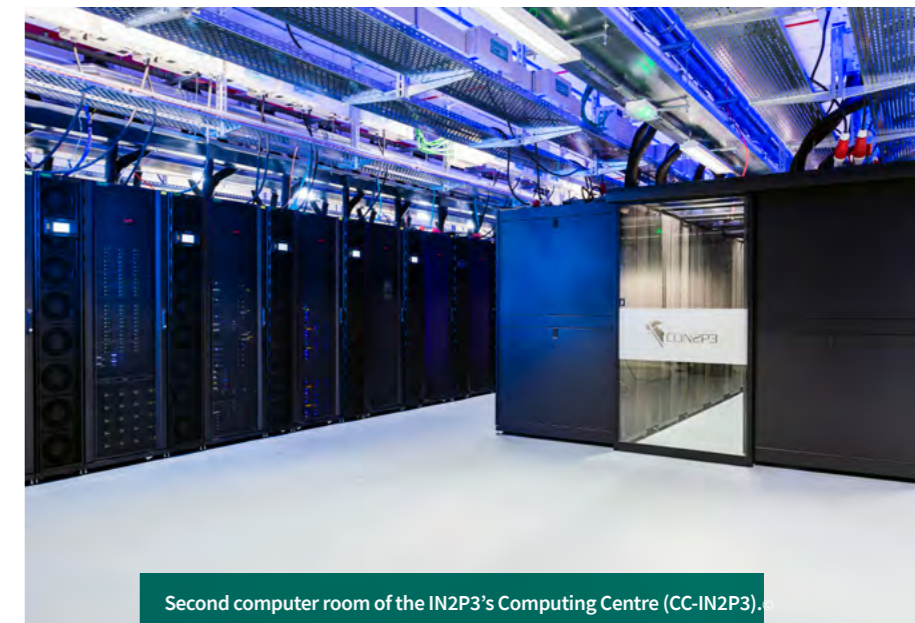
The overall objective is to adopt best practices which comply with the FAIR principles to accelerate access to and cross-referencing of data along with their use/reuse by humans and machines. This should contribute to enhancing the reliability (quality and veracity) of data and to the reproducibility of research results. The life cycle of data must be a key concern right from the initial design stage of research projects - from their production to their free availability whenever possible (open or restricted data) through data storage, curation, analysis and modelling and intermediate or even long-term archiving thereof. It is however clear that the quality and reproducibility of data also depend on other factors which are not considered here such as the quality of sources or analysis processes. To support the spread of FAIR data culture, training must be increased and enhanced to respond to a growing demand.

### • Making existing services and tools better known

These include services, «workflows», data repositories, referencing and access portals, technologies, norms and standards, etc. which make it easier for scientific communities to manage research data throughout their life cycle and also to efficiently implement the FAIR principles. The CNRS Research Data Plan must be used as a lever whenever possible because not all research communities have reached the same level of maturity in their approach to data management and in making data available to the entire community.

### • Supporting the creation of new practices, services and tools

Data archiving and curation require prior planning as to which data should be kept and how to ensure that data's integrity, quality and veracity. The CNRS could also support and lead pilot experiments on the organisation of distributed platforms for data storage, indexing and curation coupled with a set of fundamental services.



Second computer room of the IN2P3's Computing Centre (CC-IN2P3).  
Cyril FRESILLON / CC IN2P3 / CNRS Photothèque

## 4 | RESEARCH DATA POLICY

**The CNRS policy on data must be explicit and should correspond to the requirements of research itself and of scientists. Not all scientific communities are at the same level of maturity but the practice of data sharing has developed widely in certain disciplines. For example, disciplinary repositories - which are most often international - have developed over the last few decades, particularly in the field of astronomy. Other communities do not yet have a system for managing their data and have not even studied the subject in depth.**

A significant amount of data is currently irretrievably lost over a period of time when it could have been reused. Whenever possible the CNRS Research Data Plan must be based on existing data. The astronomical community has had a centre for data sharing since 1972 - the Strasbourg Astronomical Data Center (CDS)<sup>1</sup>. More recently, the Humanities and Social Sciences created a tool for managing data through the Huma-Num<sup>2</sup> Very Large Research Infrastructure which handles data from their production to their long-term archiving - if necessary - and dissemination.

**To take the FAIR principles and the need for openness of data into account, the CNRS data policy should include the following:**

- **Supporting communities to encourage them to make their data FAIR-compliant** at a very early stage of projects or when they are about to store or publish them. The issue of data management involves the storage volume required and also the way in which data are processed (cleaned, enriched, structured) and identified. This also covers the necessary services that enable data to be found, reused and shared. The evaluation of research practices will progressively incorporate criteria on data FAIRisation.
- **Recommending that data should be deposited in an open-access repository.** This is particularly important given the increasing calls for data to be deposited in repositories managed by journal publishers and so forth. Repositories are not necessarily disciplinary in scope but, when deciding whether they are reliable and relevant, we need to consider whether they serve their target community (which may be broader than a single discipline) according to user practices and expectations.
- **Compiling a directory of existing data repositories and services for researchers** with the specific aim of certifying these going forward. A repository needs to play a role in data curation and preservation and the FAIR principles are a prime objective in the context of Open Science. CoreTrustSeal core certification sets out the criteria for a

«trustworthy» repository. This makes it possible to work on improving practices based on these criteria without necessarily actually going so far as to submit a certification application.

- **The possibility of setting «reasonable» proprietary or embargo periods on data** (six months to two years for example) taking disciplinary practices into account.
- **Encouraging the re-use of relevant data**, if available, rather than the creation of new data.

The CNRS research data strategy and policy will need to dovetail with those of the MESRI and its partners in French Higher Education and Research. To work towards this, the CNRS will appoint a data administrator as recommended by the PNSO to represent it in a network being set up by the MESRI.

The CNRS will also need to ensure it takes part in national, European and broader international forums for the discussion of open science, computation and research data policies. The CNRS has made taking part in the Research Data Alliance (RDA) and the European Open Science Cloud (EOSC) an integral part of its research data policy as coordinated with the MESRI. The CNRS steers the French national node of the RDA which is funded by the MESRI. The EOSC works on computing, research data and open science issues. Internationally the CNRS is expected to become one of the first members of the EOSC association by the end of 2020.

1. CDS : <https://cdsweb.u-strasbg.fr/index-fr.gml>  
This is a revealing example: the CDS's employees are made up of 1/3 astronomers, 1/3 IT specialists and 1/3 data curators.  
2. [www.huma-num.fr](http://www.huma-num.fr)

## 5 | RESEARCH DATA GOVERNANCE

**The CNRS must acquire the necessary resources to work in a coordinated manner across the entire research continuum - from computing to publications and including massive data, the long tail of data, hardware and software infrastructures, user services, referencing and documentation. Any division of efforts in this direction would hinder the shared and transversal nature of actions taken.**

**Governance of data needs to take the three following aspects into account:**

### **The scientific aspect :**

New discoveries and even new research themes may emerge from reusing data and from the Institutes sharing best practices.

### **The technical aspect :**

This will involve promoting and implementing the FAIR management of research data, encouraging and driving professional networks and inter-network groups and defining new professional occupations.

### **The economic aspect :**

New shared resources will be required, especially human resources. The consequences in terms of costs (especially regarding longer-term savings) and the potential socio-economic impacts of FAIR data will need to be assessed.

Additionally, the idea that data should be «as open as possible and as closed as necessary» will have to be precisely defined in consultation with the CNRS's Legal Affairs Department, Innovation Office, Data Protection Department, Defence and Security Officer and Security Department.

To serve the many scientific communities involved, data governance work at the CNRS will need to be carried out in collaboration with French Higher Education and Research partners who particularly contribute to numerous data-producing infrastructures and services. This link could be managed by the data administrator who, as mentioned earlier, will represent the CNRS in an inter-institutional network currently being set up by the MESRI in accordance with the National Plan for Open Science. The CNRS will also collaborate with international projects and organisations in coordination with its Institutes who must be strongly involved in data governance.

A «sustainable business model» must be set up to support all component parts of the FAIR data ecosystem - data curation, making data available, storage and even long-term archiving. Archiving and data services represent costs in terms of financing and human resources that need to be quantified and taken into account especially after the lifetime of the projects that generated the data concerned. Federating and pooling certain data services will make it possible to optimise and reduce these costs. Defining an economic model is also a key issue for participation in the EOSC since this will closely affect the involvement of the actors concerned. Another aspect of the economic model is the potential cost<sup>1</sup> of not making data FAIR including for those working on innovation.

Having a single data governance system which is unique in the inter-institutional sense is that it provides additional resources for transversal initiatives. It therefore needs dedicated human and financial resources so that it can take actions and also requires to strengthen the structures to which these resources would be allocated (national or regional data centres, INIST<sup>2</sup>, CCSD<sup>3</sup>, etc.) This will also encourage sharing of resources and expertise as it is already the case with MICADO<sup>4</sup> for computing and the DIST<sup>5</sup> for open science. Disciplinary systems which are already managed by CNRS Institutes would not however have to change their existing way of operating.



1. A study published by the European Community attempted to evaluate such costs: Cost-benefit analysis for FAIR research data Cost of not having FAIR research data - Study DOI: 10.2777/02999 : [https://www.ouvrirlascience.fr/wp-content/uploads/2019/03/Cost-Benefit-analysis-for-FAIR-research-data\\_KI0219023ENN\\_en.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2019/03/Cost-Benefit-analysis-for-FAIR-research-data_KI0219023ENN_en.pdf)  
2. CNRS Institute for Scientific and Technical Information: <https://www.inist.fr/>  
3. Centre for Direct Scientific Communication: <https://www.ccsd.cnrs.fr/>  
4. CNRS Computing-Data Mission  
5. CNRS Scientific and Technical Information Department

A new functional Open Research Data Department (DDOR) under the authority of the CNRS Scientific Office (DGDS) will be responsible for proposing a policy and strategy for opening data at the CNRS which integrates all facets of the subject and for supporting their implementation. Overall, the DDOR defines and implements open science strategy extended to all issues linked to research data including specific themes related to digital infrastructures. The direction the DDOR will take has been defined by the 2019 Roadmap for Open Science and by this CNRS Research Data Plan. This department is the result of the merger of the DIST and MiCaDo and will provide a framework to help deal with issues related to opening up scientific publications, the management and sharing of research data and the question of massive data including storage and digital infrastructures. The CNRS Institutes will play a leading role in this new department's steering committee as it will be made up of a representative at a decision-making management

level from each Institute, namely mainly Deputy Scientific Directors (DAS).

In addition to this steering committee, a unit will be set up to deal with issues related to data closure and in particular the sharing of best practices to help differentiate between open data and data which needs to be protected. This unit will be made up of the DDOR management, the Defence Security Officer and representatives of the Innovation Office, Data Protection Department, and the Security Department. This unit's members will be invited to steering committee meetings when necessary.

## 6 | ACTION PLAN

***A plan for immediate action is outlined here. By definition this can be by no means exhaustive at this stage and will have to be jointly constructed with the Institutes. One of the first tasks of the new department's steering committee will be to review all these actions, approve them, suggest new initiatives, set priorities and manage their implementation. Some of these initiatives are linked to compliance with norms and standards (including FAIR) while others concern the Institutes' incitement policies or researchers' daily practices.***

This text deliberately does not cover certain issues that require further work. An incentive system to encourage researchers to share their data needs to be explored. Data repositories are not the only option as federation or publication-subscription infrastructures can also be used to share data. Incentives to share data must be coupled with incentives to re-use them which in turn require appropriate engineering services to facilitate this.

The action plan proposed herein is a tangible illustration of the problems arising around data management which need a transverse approach while still taking into account the specificities of disciplinary practices and the Institutes own internal policies.

### 1. Encouraging the emergence of best practices

***The FAIR principles underpin all the actions and tools presented in this section such as Cat OPIDoR<sup>1</sup> implemented by the INIST, data management plans (DMPs) or repositories which integrate approaches to certification. Laboratory notebooks are also a useful tool for achieving the levels of documentation required from a FAIR perspective.***

**Supporting scientific communities in defining how to manage their data in specific details.**

This includes defining the criteria for selecting data which need to be made FAIR-compliant. This requires taking into account the scientific value and impact for current and emerging research, the principles used by archivists and

the fact that observations may be impossible to reproduce, among other factors. Communities also require support from the Institutes associated with the transversal governance of the system in developing and maintaining their disciplinary frameworks for data sharing and FAIRisation in an international context. Supporting communities means helping them to organise and structure their work around data and data practices. This is an essential mission for the Institutes which also provide a national and international vision. The Institutes' role in this area involves dialogue with their communities and understanding their requirements and practices.

**Mapping the data, tools and services involved in data management and sharing.** It is essential to map the data produced by the research units under the authority of the CNRS. This is currently underway with a questionnaire that is being progressively distributed to unit directors by the DIST. It will enable the identification of the places where data are produced (infrastructures and/or laboratories) and stored, the quantity of data produced, data sharing and conservation policies within the units, etc. Mapping of existing structures and services for data and metadata management is also required within both the CNRS and its partners in French higher education and research. Notably, work is already ongoing on mapping the repositories and services the CNRS is responsible for. Overall, the mapping should also make it possible to identify the requirements of researchers at all the Institutes.

**Data Management Plan (DMP).** European calls for projects require a data management plan to be produced for all research projects submitted as do National Research Agency (ANR) calls for proposals henceforth<sup>2</sup>. This is a good opportunity to encourage researchers to plan the conditions for storing, curating and disseminating their data throughout their life cycles. Our aim is to support researchers in this process. The INIST is contributing to this with the DMP OPI-DoR tool<sup>3</sup> which facilitates drafting such plans, and with the organisation of the «OPIDoR tour» in all regions of France. This support needs to be based on scientific communities and local components which could be coordinated by a CNRS professional network. The generalisation and adoption of a standardised data plan for all projects (apart from the ANR and Horizon Europe) would help facilitate the implementation of the FAIR principles and good data management practices. DMPs will become easier to draft through use of all the technological possibilities enabling the extraction and reuse of data from the various information systems concerned. These are referred to as «machine-readable» DMPs insofar as they are filled in and used automatically.

**Persistent identifiers for research datasets.** It is essential to develop the usage of persistent identifiers for data, software and publications. The INIST is contributing to this by assigning Digital Object Identifiers (DOIs) to

research data for the CNRS and for French higher education and research as a whole in its role as a national Data-Cite agency. There are currently over a hundred research structures that are users at the CNRS and in higher education and research such as the National Research Institute for Agriculture, Food and Environment (INRAE), the Research Institute for Exploitation of the Sea (Ifremer), the European Synchrotron Research Facility (ESRF), the Strasbourg Astronomical Data Center (CDS), the Observatoire Midi Pyrénées, the Center for Socio-Political Data (CDS), etc. Software Heritage has recently made proposals for persistent identifiers for software<sup>4</sup>.

**Electronic laboratory notebooks.** As well as being important for matters of intellectual property or proof of scientific integrity, a laboratory notebook can act as a tool for managing data with a view to their conservation and dissemination via the notebook itself or the links it contains. The CNRS's position is that electronic laboratory notebooks should replace paper notebooks in its laboratories. It will make recommendations and set out obligations as far as the characteristics of the tool adopted are concerned in terms of conservation, traceability, durability and accessibility of the data and the tool itself. The solutions adopted by our French partners (Inserm, Ifremer, INRAE, the CNES Space Agency, etc.) and at European and international levels also need to be studied. A working group steered by the Mission for the Monitoring of and Relations with CNRS Regional Offices and Institutes (MPR) is tackling these issues at the CNRS.

**Certification of research data management systems (particularly CoreTrustSeal<sup>5</sup>).** The certification of data repositories and services is an objective set out in the National Plan for Open Science. This ensures that a data centre is «trustworthy» by examining the way it implements the whole data chain, from data ingestion to dissemination and then preservation. It can refer to networks of data centres like those of the IR Data Terra thematic data hubs<sup>6</sup>, or of the European infrastructure CLARIN<sup>7</sup>. The CNRS will be able to count on the activities in support to certification set up by the French National RDA Node<sup>8</sup>.

1. Cat OPIDoR : <https://opidor.fr/reperer/>
2. GENCI, the French Very Large-Scale Computing Centre, will soon be making a DMP compulsory for those requesting to use national computing resources.
3. DMP OPIDoR : <https://dmp.opidor.fr/>
4. Identifiers for software : <https://www.softwareheritage.org/2020/05/26/citing-software-with-style/>
5. CoreTrustSeal : [CoreTrustSeal Requirements v02.00-2020-2022](https://www.coretrustseal.org/) (doi:10.5281/zenodo.3638211)
6. Data Terra : <https://www.data-terra.org/>
7. CLARIN : [https://office.clarin.eu/vj/CE-2013-0095-B-checklist-v7\\_3\\_1.pdf](https://office.clarin.eu/vj/CE-2013-0095-B-checklist-v7_3_1.pdf)
8. RDA France : <https://www.rd-alliance.org/groups/rda-france>

## 2. Encouraging the emergence of new tools

Certain communities produce data without using a DMP (typically through laboratory experiments) and thus do «traditionally» use laboratory notebooks. These data are often poorly referenced and are generally permanently lost after the corresponding articles are published, when the doctoral student who produced them leaves the laboratory or when the storage equipment is changed. In more prosaic terms, today there is not a generic storage solution for laboratories, let alone for referencing or making data FAIR-compliant. When a laboratory asks «where could I store my data», usually it does not get an answer and has to source its own equipment independently or perhaps may at best be offered specific solutions on a case-by-case basis. How then can a process to take charge of this data be initiated and encouraged?

Electronic tools like a laboratory's «research directory» or digital notebook could link to the server which hosts data, suggest interoperable data formats and impose minimum documentation of the data in case it is disseminated later. This kind of tool needs to be simple, interoperable (and relatively cheap) and capable of being used, if necessary, as a basis for a future DMP or to facilitate a deposit in a repository at the end of the research project.

An alternative could be a CNRS institutional repository for the many data of the «long tail». These are often very diverse and of small volume but their overall volume and diversity are growing rapidly thanks to the development of imaging techniques and distributed sensors. The question of this kind of repository needs to be brought to the table in coordination with the ongoing MESRI-driven working group led by J.C. Desconnets of the Research Institute for Development (IRD) which is studying the feasibility of a national «simple» data repository. The IRD along with the National Research Institute for Agriculture, Food and Environment (INRAE), the Agricultural Research Centre for International Development (CIRAD) and other actors have already opted for the solution of an institutional repository while other institutions have made different choices. The objective on this point is to aim towards seeking national coherence and flexibility to simplify the work of depositing, referencing and making data available while respecting the full spectrum of research practices.

If this generic repository for the research data of the long tail were set up, it would need to be complementary to existing systems. These include research infrastructures, the lar-

gest of which have the capacity to develop their own tools in the national and/or international framework(s), or data infrastructures which support the FAIRisation of data and are often thematic in nature. It would also be essential to take the systems set up locally by universities or at a regional level into account. There is a clear need to closely associate the calculation, storage and processing of data especially when very large volumes are produced. National centres can play a key role as is the case of the CC-IN2P3 thematic centre<sup>1</sup> dedicated to the National Institute of Nuclear and Particle Physics (IN2P3) community. These centres could offer data and computing resources and services with the necessary capacities for increasingly heavy processing and analysis.

Certain very large-scale research facilities and infrastructures (TGIRs) - particularly in astronomy and Earth science - archive data on a long-term basis and distribute them after an embargo period. Other TGIRs do not provide storage and curation for data. Providing access to research data for scientists who do not themselves use the instruments and observation systems used to produce the data remains to be developed in most cases and particularly for inter- and/or transdisciplinary applications.

Referencing is complementary to storage strategies. It involves having a software platform of services to facilitate the visibility of and access to datasets via tools with user-friendly search functionalities. Metadata quality, accuracy and relevance are fundamental to guaranteeing good accessibility to the data.

## 3. HR - Career evaluation and management

**The CNRS must implement a coordinated response to new needs in terms of expertise, human resources and the recognition of these new transdisciplinary research data activities. It might also be wise to create a value system to encourage researchers to publish their data.**

**Make an inventory of all roles played by CNRS personnel** in the FAIRisation of data and the skills required for these roles to feed into and develop evaluation guidelines for researchers and guidelines for IT posts (Referens)<sup>2</sup>. This will lead to the identification of new profiles and professions.

1. The CNRS Institute of Nuclear and Particle Physics Computing Centre: <https://cc.in2p3.fr/>
2. Referencing list of Jobs in Research and Higher Education (REFERENS) : <https://data.enseignementsup-recherche.gouv.fr/pages/referens/>

**Target recruitment** to dovetail with the new profiles identified and thus benefit from additional human resources to support data management and sharing. The human resources required are mainly in the IT sector (data stewards) but may also include researchers (data scientists). There is a strong immediate need for engineers. This should facilitate sharing experiences and efforts..

**Accompany an in-depth change in the way scientists are evaluated.** Action will be required to ensure that investments which support FAIRisation and data sharing are considered in career evaluation. This involves tasks of collective interest like data sharing in a FAIR mode, taking part in the definition of practices and standards at disciplinary and interdisciplinary levels and in their implementation as scientific experts, working on the scientific tasks required to share data in research and data infrastructures, etc. The CNRS needs to encourage and reward the production and sharing (where possible) of FAIR data and more generally of FAIR digital objects (software) in research projects and practice, the practice of using DMPs and in finally the reuse of data which could lead to new advances in knowledge.

## 4. Training

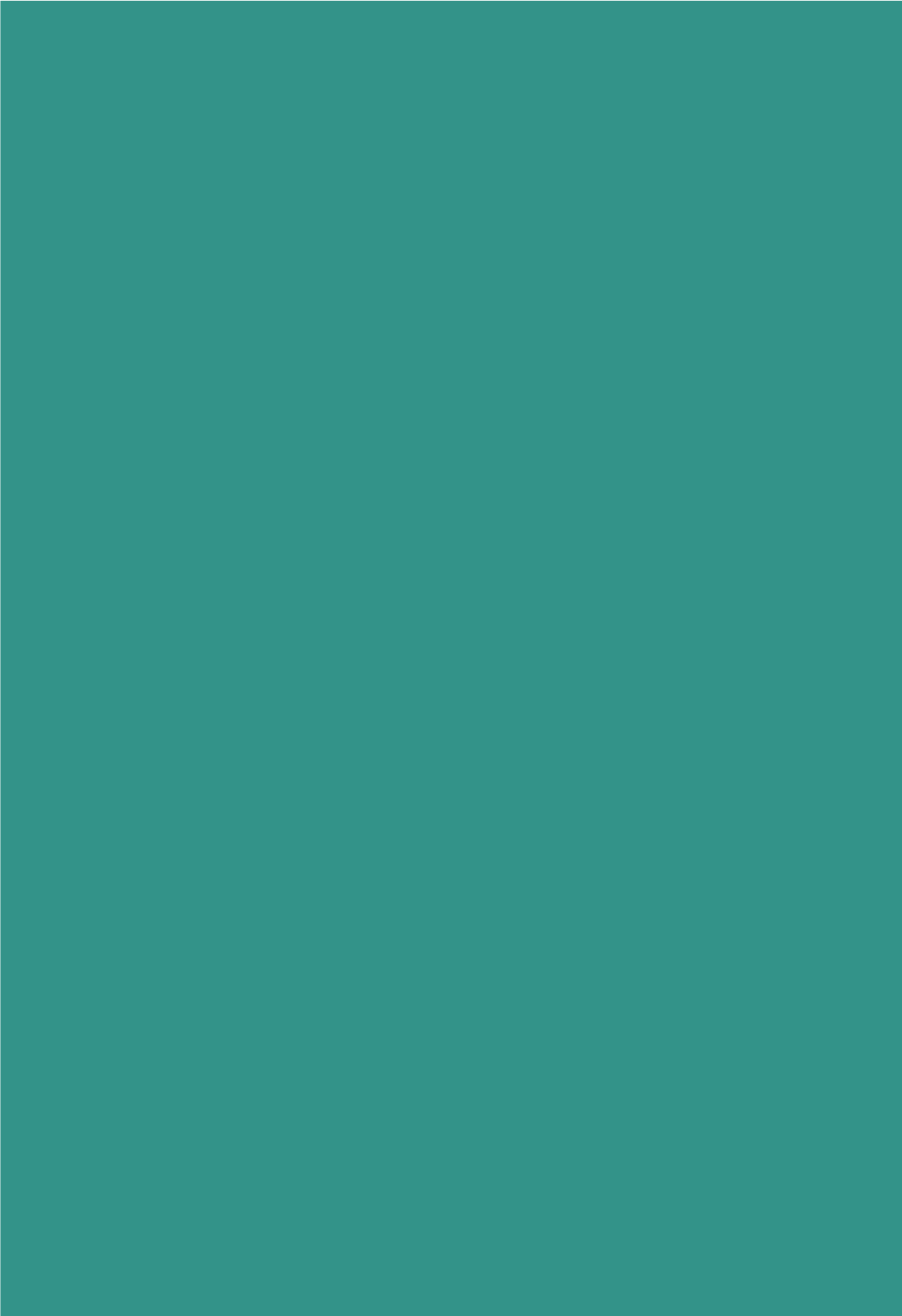
Training and communication are key elements in this action plan. The primary aim is to develop skills in FAIR data in terms of data management (FAIR principles of management, sharing, intermediate and permanent archiving) and data science (handling, processing and analysing research data). The INIST is part of this initiative providing distance learning via the DoRANum platform<sup>1</sup> and face-to-face training on data management, sharing, archiving and data management plans.

The «data culture» needs to be developed through ongoing training of all types of personnel to raise their awareness of the subject, make them familiar with standards and learn to use them and, more generally, on the various tools linked to data such as DMPs, repositories, languages and so forth.

It is also important to identify successful achievements and communicate about these to convince communities which are still not very involved to adopt data sharing. Raising awareness will be based on exchanges of best practices and lessons learnt from the implementation of the FAIR principles and the definition of the legitimate limits to data openness. One aspect of this is to collect tangible

examples and implementation guides. The same needs to be carried out for rolling out DMPs. Such exchanges can take place in the framework of disciplinary fields (Institutes) or through professional networks (MITI<sup>2</sup>) and by encouraging the personnel working on data sharing to take part in the RDA.

1. DoRANum : <https://doranum.fr/>
2. CNRS Mission for Transversal and Interdisciplinary Initiatives







© AdobeStock/ra2studio

**CNRS**

3, rue Michel-Ange  
75794 Paris Cedex 16  
+331 44 96 40 00  
[www.cnrs.fr/en](http://www.cnrs.fr/en)



Print : CNRS IFSem  
March 2021

