



Principes communs de FAIRisation et gouvernance
des données entre institutions de recherche



ANR-Bridge 2019-2022

Adoption de méthodes et produits de la RDA

Sophie Fortuno | Cirad – DG DRS
Pascal Aventurier | IRD – IST/MCST
Dimitri Szabo | INRAE - DipSO



Equipe projet

IRD : Pascal Aventurier (coordinateur), Luc Decker, François Sabot, Jérôme Perez.

Appui : Jean-Christophe Desconnets, Hugo Catherine

Cirad : Sophie Fortuno, Marie-Claude Deboin, Pascal Bonnet, Gwenael Doux,
Céline Barthélemy

INRAE : Dimitri Szabo, Esther Dzale-Yeumo, Christian Mougine, Michael Alaux,
Yannick de Oliveira

Objectifs du projet

❑ Objectif général

Fournir des recommandations pour la FAIRification des données dans un cadre de gouvernance interinstitutionnel

❑ Objectifs spécifiques

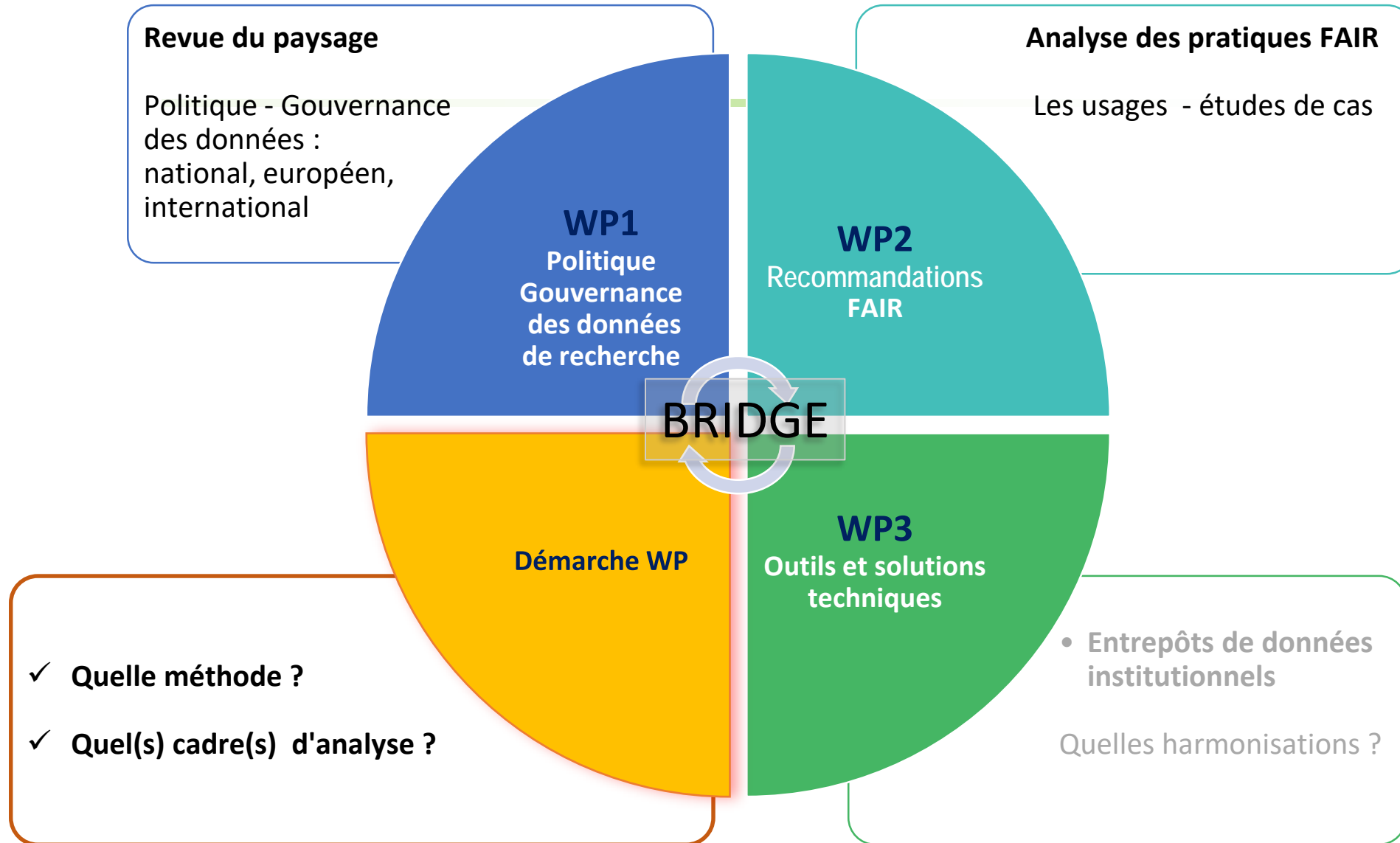
- WP1 : Analyse des politiques pour un cadre harmonisé de gouvernance des données de la recherche
- WP2 : Recommandations FAIR
- WP3 : Interopérabilité et harmonisation entre entrepôts de données

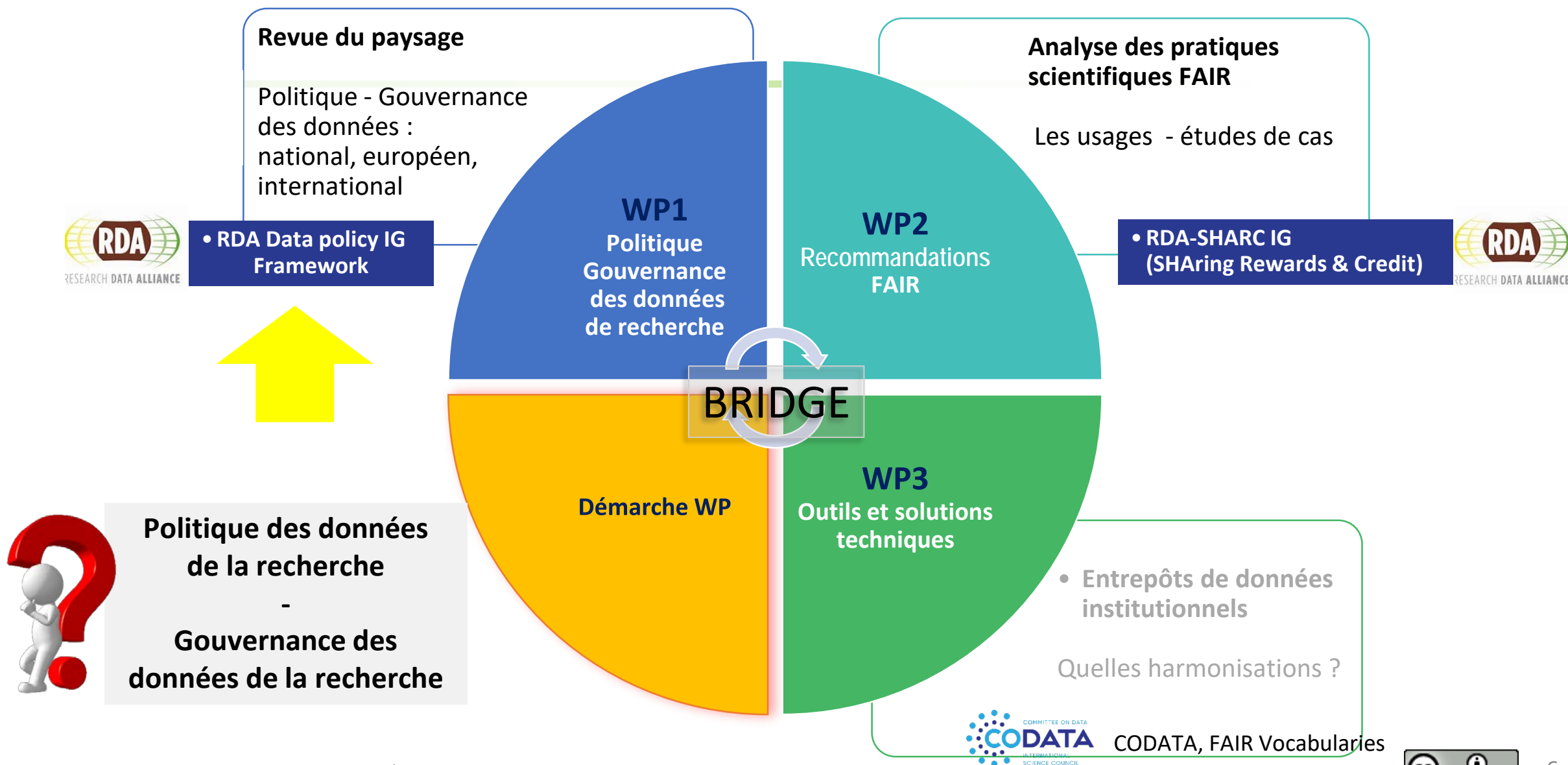


Démarche générale

- Revue, état des lieux par WP
- Veille biblio, revue d'initiatives...
 - ✓ Bibliographie, conférences RDA, CODATA, etc
 - ✓ Identifier des approches méthodologiques
 - ✓ [Outputs & Recommendations | RDA \(rd-alliance.org\)](https://rd-alliance.org)







Quelques définitions WP1

Politique des données

Un ensemble de principes généraux de haut niveau qui forment le cadre directeur dans lequel la gestion des données peut fonctionner.

(traduit de glossary OCDE <https://stats.oecd.org/glossary/detail.asp?ID=4454>)

La **gouvernance des données** correspond à l'ensemble des organisations et des procédures mises en place au sein d'une entreprise/d'un organisme afin d'encadrer la collecte de données et leur utilisation. Il s'agit autant de respecter les obligations légales imposées par les pays et/ou l'Union européenne.

[Data Governance ou gouvernance des données : Qu'est-ce que c'est ? \(lebigdata.fr\)](http://lebigdata.fr)

« **Gouverner** » les données signifie que des (bonnes) décisions sont prises à chaque étape du cycle de vie des données (de leur production ou réutilisation à leur valorisation) et que les responsabilités de chacun sont définies dans ce processus décisionnel.

« Datapartage - Principes pour la gouvernance des données INRAE ». s. d. Consulté le 14 mai 2022. <https://datapartage.inrae.fr/Gouvernance-des-donnees2/Principes-pour-la-gouvernance-des-donnees-INRAE>.

Un constat partagé

- ✓ De plus en plus d'éditeurs de revues et de bailleurs de la recherche introduisent des politiques de données visant à augmenter la disponibilité des données de recherche à des fins de réutilisation
- ✓ **Des exigences politiques multiples générant de la confusion auprès des chercheurs et les personnels de soutien**
- ✓ Plusieurs politiques sont similaires mais pas identiques (pas de règle convenue)



Data Policy Standardisation & Implementation IG

RDA Plenary 18 (virtual)

Enquête menée sur les politiques de données de recherche auprès de 10 000 revues scientifiques



[Developing a Research Data Policy Framework for All Journals and Publishers](#)

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

A partir d'un cadre normalisé faisant consensus, l'étude **RDA Data policy IG** relève **14 caractéristiques** (obligatoire, recommandé ou sans objet) **définissant un cadre de politique de données de recherche des éditeurs de revue scientifique**

➤ **6 profils types de politique dégagés**

Enquête (en cours) auprès de 100 bailleurs de la recherche

Framework - Constat Bridge

14 journal research data policy features arranged as 6 policy types (tiers)

n°	Caractéristiques= Articles	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
1	Definition of the research data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	Exceptions to policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
3	Embargoes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
4	Supplementary materials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
5	Data repositories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
6	Data citation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
7	Data licensing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	Researcher/ author support	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
9	Data availability statements		<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
10	Data formats and standards				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
11	Mandatory data sharing (specific data types)				<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
12	Mandatory data sharing (all papers)				<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
13	Peer review of data				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
14	Data Management Plans (DMPs)				<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Capacité à définir les termes d'une politique de données pour les établissements de recherche en cohérence avec les caractéristiques des politiques des éditeurs *[et des bailleurs]*

➤ Ex. Policy 07

Légende :

- = Sans objet

Fournir l'information (recommandé)
Le texte de la caractéristique sera inclus dans le modèle de politique, mais la caractéristique ne sera pas obligatoirement appliquée et vérifiée

Fournir l'information et agir (obligatoire)
Le texte de la caractéristique est inclus dans le modèle de politique et indique clairement que la caractéristique sera appliquée et vérifiée

Policy 01 – template example

Data Policy Standardisation & Implementation IG
 RDA Plenary 18 (virtual)

Appendix: Policy feature standard texts for policy template construction

These policy texts are available for reuse and creation of derivative works, under the Creative Commons Attribution license (CC BY).

Please cite the associated paper when reusing the policies:
 Hrynaszkiewicz, I, Simons, N, Hussain, A, Grant, R and Goudie, S. 2020. Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, DOI: <http://doi.org/10.5334/dsj-2020-005>

Standard text for policy 1

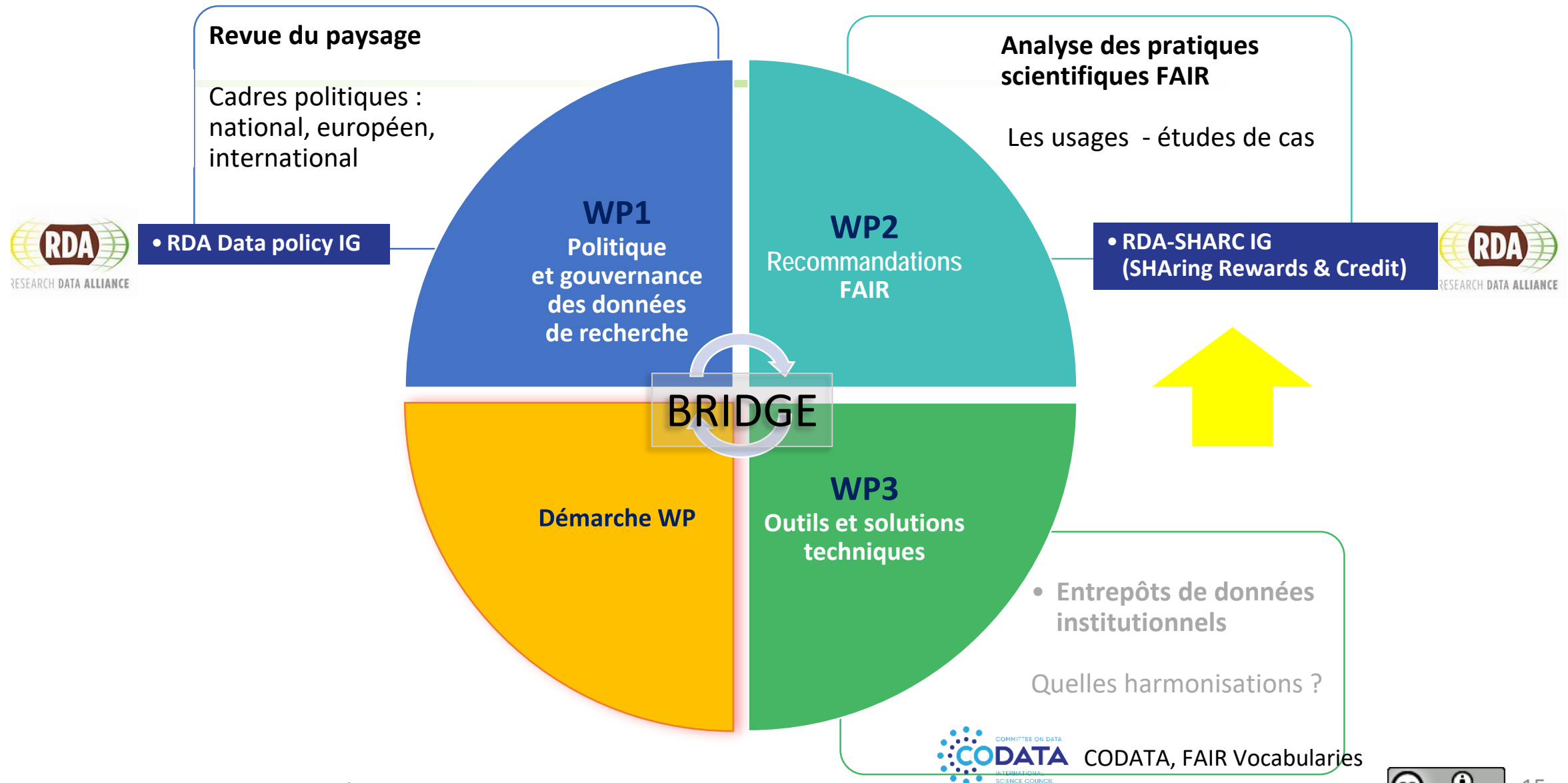
Policy summary for authors	
Authors are encouraged to make the research data that support their publications available but are not required to do so. The decision to publish will not be affected by whether or not authors share their research data.	
Required <ul style="list-style-type: none"> No action Optional <ul style="list-style-type: none"> Data citation Data sharing via repositories for all research data 	
Policy summary for Editors	
The journal will encourage authors to share data but the editor's decision to publish will not be affected by whether or not authors share their research data.	
Required <ul style="list-style-type: none"> Respond to questions about this policy. Optional <ul style="list-style-type: none"> Data citation Data sharing via repositories for all research data 	
Feature	Text

Definition of research data	This policy applies to the research data that would be required to verify the results of research reported in articles published in the journal. Research data include data produced by the authors ("primary data") and data from other sources that are analysed by authors in their study ("secondary data"). Research data includes any recorded factual material that are used to produce the results in digital and non-digital form. This includes tabular data, code, images, audio, documents, video, maps, raw and/or processed data.
Definition of exceptions	Research data that are not required to verify the results reported in articles are not covered by this policy. This policy does not require public sharing of quantitative or qualitative data that could identify a research participant unless participants have consented to data release. The policy also does not require public sharing of other sensitive data, such as the locations of endangered species. Alternatives to public sharing of sensitive or personal data include: <ul style="list-style-type: none"> Deposition of research data in controlled access repositories Anonymisation or deidentification of data before public sharing Only sharing metadata about the research data Stating the procedures for accessing your research data in your article and managing data access requests from other researchers
Embargoes	Embargoes on data sharing are permitted.
Supplementary materials	Sharing research data as supplementary information files is discouraged.
Data repositories	The preferred mechanism for sharing research data is via data repositories. Please see <recommended repository list> or https://repositoryfinder.datacite.org/ for help finding research data repositories.
Data citation	The journal encourages authors to cite any publicly available research data in their reference list. References to datasets (data citations) must include a persistent identifier (such as a DOI). Citations of datasets, when they appear in the reference list, should include the minimum information recommended by DataCite and follow journal style.
Data licensing	The journal encourages research data to be made available under open licences that permit reuse freely. The journal does not enforce particular licenses for research data, where research data are deposited in third party repositories. The publisher of the journal does not claim copyright in research data.
Researcher/ author support	Questions about complying with this policy should be sent to [email].

Contribution Bridge sur la base des 14 caractéristiques du cadre *RDA Data policy IG*

- ✓ « **Une vision de référence** » de **Politique des données de la recherche lisible**
- ✓ Bridge définit 2 nouvelles caractéristiques (articles) centrales pour un profil organisme de recherche
 - Légal - réglementaire
 - Intégrité scientifique
- ✓ Propose un « alignement documenté » des textes des principales caractéristiques basé sur une revue des documents de politiques nationale, européenne et internationales
- ✓ Chaque organisme choisit pour chaque caractéristique, un niveau d'exigence (obligatoire, recommandé, sans objet) selon sa position sur la caractéristique

https://zenodo.org/communities/briddge_anr_so/



Grille RDA-SHARC IG - Etudes de cas WP2

Périmètre

- 10 jeux de données variés, fournis par les partenaires du projet
- Projet ANR-Data4C+, UMR AMAP, UMR Génétique Quantitative et Évolution-Le Moulon, URGI
 - ✓ Favoriser les jeux de données inter-établissement

Conduite d'entretien, synthèse et analyse réalisées par **DATAACTIVIST**

- Un entretien détaillé par jeu de données avec les producteurs
- Remplissage de la grille RDA-SHARC (extensive template)
- Atelier et rapport interne de synthèse

A	B	C	D	E	F
Scientist Name / Project Name, Date	Tick the proper case using '1' (so that it will be automatically counted)	Never/NA	Mandatory	Sometimes	Always
1) FINDABLE (8 essential criteria)					
Indexed identifier? (F01_E)					
Identification					
Are each data/dataset identified by an indexed and independant identifier?					
	Unique, global, persistent identifier? (F02_E)				
Identification					
Are the data identifiers unique, global and persistent?					
	Identifier scheme? (F03_E)				
Identification					
Has any identifying schema been used for data (e.g., DOI)?					
Persistent metadata / data link? (F04_E)					
Metadata traceability					
Are the metadata linked to the dataset through a persistent identifier?					
	Persistent metadata & authority link? (F05_E)				
Metadata traceability					
Are the metadata of each dataset linked to a unique authority (responsible for the datasets at a given time)?					
	Persistent datasets & authority link? (F06_E)				
Metadata traceability					
Are all datasets linked to an authority over time (legal entity, e.g., institution, association or established body) through a unique and persistent identifier?					
Standards / dictionary for data description? (F07_E)					

Questionnaire d'évaluation FAIRness

45 questions

18 questions

[Templates for FAIRness evaluation criteria - RDA-SHARC ig V1.3 - Google Sheets](#)

June 16, 2022

Presentation Open Access

Data FAIRification in a cross-institutional governance framework : recommendations from the ANR-BRIDGE project

Aventurier, Pascal; Fortuno, Sophie; Szabo, Dimitri; Alaux, Michael; Bartelety, Céline; Bonnet, Pascal; Catherine, Hugo; Dzalé, Esther; Deboin, Marie-Claude; Decker, Luc; Desconnets, Jean-Christophe; Doux, Gwenaël; Mougine, Christian; Perez, Jérôme; Sabot, François

<https://doi.org/10.5281/zenodo.6652404>

Data FAIRification in a cross-institutional governance framework : recommendations from the ANR-BRIDGE project

Authors: Aventurier, Pascal (1); Alaux, Michael (2); Bartelety, Céline (3); Bonnet, Pascal (3); Catherine, Hugo (1); Dzalé, Esther (2); Deboin, Marie-Claude (3); Decker, Luc (1); Desconnets, Jean-Christophe (1); Doux, Gwenaël (3); Fortuno, Sophie (3); Mougine, Christian (2); Perez, Jérôme (1); Sabot, François (1); Szabo, Dimitri (2)
 (1) IRD, Institut de Recherche pour le Développement, France
 (2) INRAE, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, France
 (3) CIRAD, Centre de coopération internationale en recherche agronomique pour le développement, France

<https://bridge-science-ouverte.fr/bridgean@ird.fr>

Context
 Open science is a priority for the European Commission, fostering research data management policies and infrastructures. In France, the Second National Plan for Open Science aims to ensure that open science becomes a common and shared practice, encouraged by the entire international ecosystem of higher education, research, and innovation. In addition, the French Ministry of Research is building a national federated data repository : Recherche Data Gouv, as part of its Open Science Plan.

Objectives
 In this context, three French research organizations, each maintaining an institutional data repository based on Dataverse software, initiated the BRIDGE project funded by the French National Research Agency following a special Open Science call.
 One challenge is that French research is mainly organized into joint research units : within a same laboratory, project or publication datasets can be managed by different research organizations, necessitating choosing between several institutional and disciplinary data repositories that can cause data scattering.

Three priorities
 The goal of the Bridge project is to provide guidelines and harmonize research data policies and repository management in a reusable approach for other institutes or contexts, focusing on :

1. Analyzing and improving institutional data governance policies
2. Providing and endorsing common guidelines for data producers and managers
3. Choosing FAIR vocabularies and developing IT tools to improve FAIRness of repositories with some shared metadata schemas

Main outputs

- Define and adopt common rules for all the project partners. The DMP (Data Management Plan) must be built in the light of these rules.
- Overcome some difficulties related to the practical implementation of data FAIRification. For example : bringing data into standard; define sharing conditions, common rules for harvesting metadata repositories.
- Align with common ontologies and controlled vocabularies (data type, disciplinary fields, etc.) available via the semantic web.

The project builds on the widely supported recommendations of three Research Data Alliance (RDA) and Committee on DATA (CODATA) interest groups : [Data Policy IG](#), [Sharing Rewards and Credit \(SHARC\) IG](#) and [CODATA FAIR Vocabularies](#).

20 JUN 2022 RDA 19th Plenary Meeting, Part Of International Data Week, 20–23 June 2022, Seoul, South Korea

<https://www.rd-alliance.org/rdas-19th-plenary-poster-exhibition>



<https://bridge-science-ouverte.fr/>



https://zenodo.org/communities/bridgge_anr_so/