

Collège des Données de la recherche



Mandature 2022-2025

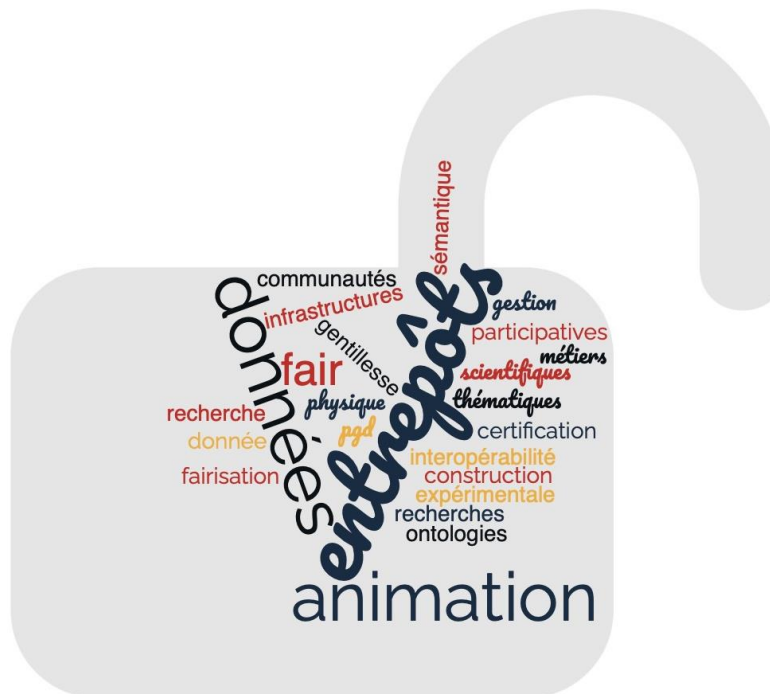
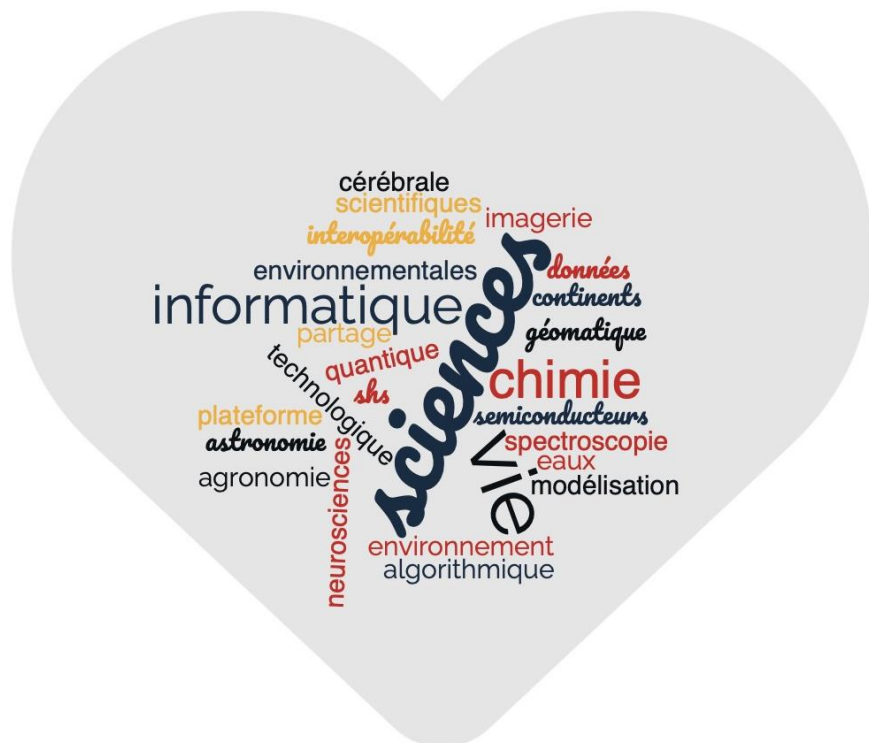
Dr Véronique Stoll & Dr Frédéric de Lamotte

22 janvier 2026

Le Collège des Données de la recherche



- 3^e mandature 2022-2025
- 20 membres



Pilotes



Véronique Stoll

Directrice de la bibliothèque de
l'Observatoire de Paris



**Frederic de
Lamotte**

Chercheur INRAE

Membres

- Cécile ARENES (Sorbonne Université)
- Romain DAVID (ERINHA)
- Stéphane DEBARD (IRD)
- Jean-François DUFAYARD (CIRAD)
- Françoise GENOVA (CNRS, Observatoire astronomique de Strasbourg)
- Christine HADROSSEK (CNRS DDOR)
- Céline HERNANDEZ (I2BC)
- Marie-Emilia HERBET (Université Jean Moulin Lyon 3)
- Hélène JOUGUET (Huma-Num)
- Thomas LEBARBE (Université Grenoble Alpes)
- Emilie LERIGOLEUR (CNRS, UMR Géode Toulouse)
- Gaëlle LEROUX (CNRS, Centre de recherche en neuroscience de Lyon)
- Jérôme MATHIEU (Sorbonne Université)
- Kenneth MAUSSANG (Université de Montpellier)
- Gilles OHANESSIAN (CNRS)
- Marie STAHL (Ecole française d'Athènes)
- Carlo-Maria ZWOLF (Observatoire de Paris)



Grands Principes

- Travail orienté sur les remontées des besoins des équipes de recherche
- Pilotage par les besoins scientifiques
- Travail en sous-groupes
- Sollicitations ponctuelles d'un 2e cercle et experts extérieurs

Objectifs et missions clés



Coordination des activités

Coordonner les activités des divers groupes et suivre l'avancement des projets.



Application des principes FAIR

Encourager l'application des principes FAIR dans la production de données de recherche.

Soutien à la politique de données

Soutenir l'adoption d'une politique de données tout au long du cycle de vie.

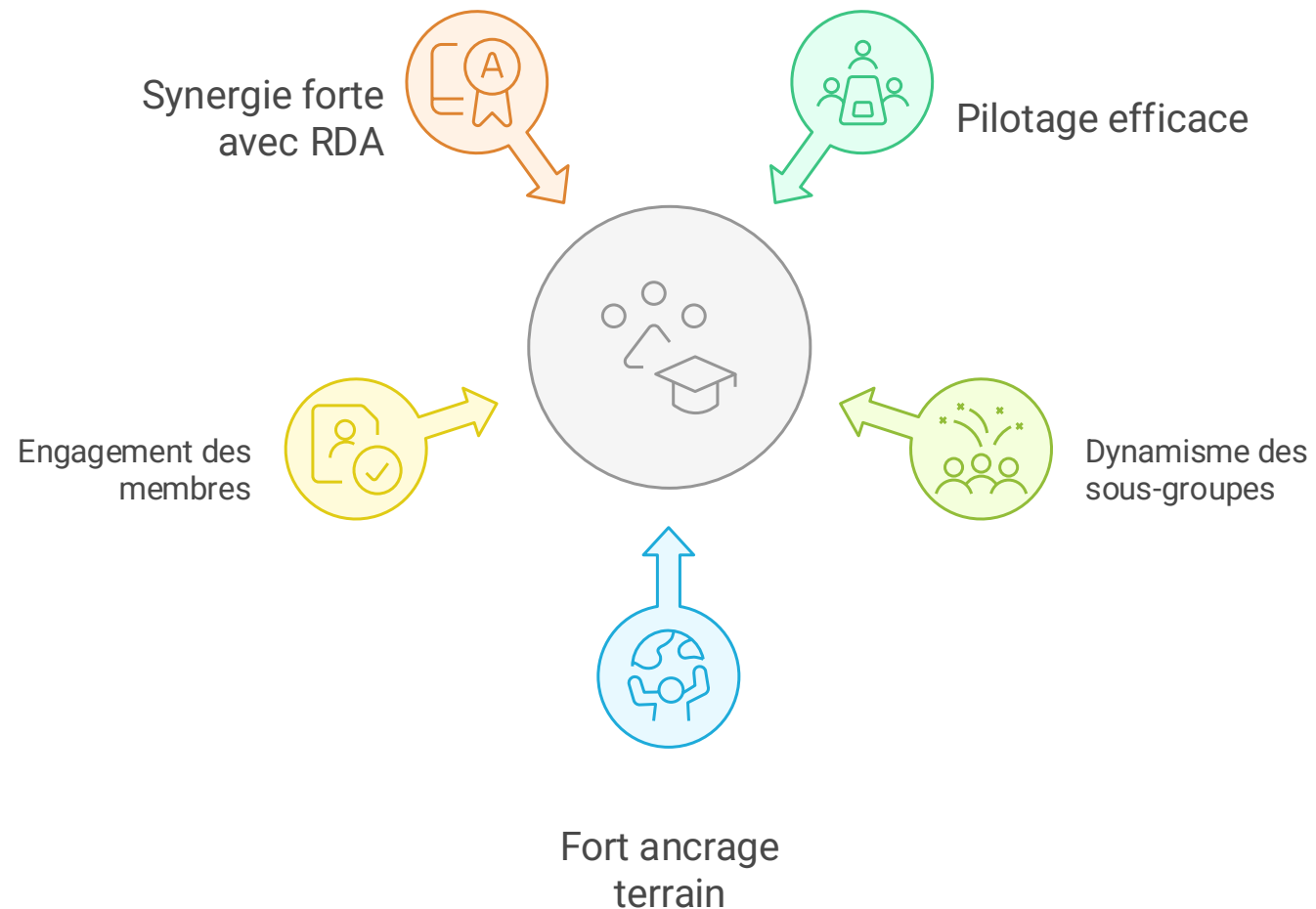


Promotion de la réutilisation des données

Promouvoir et développer des pratiques qui favorisent la réutilisation des données de recherche.

L'action des membres est guidée par les besoins des communautés scientifiques.

Points forts





Principaux travaux de la mandature

- Proposer une méthodologie pour sélectionner des entrepôts de confiance
- Etablir et maintenir la liste des entrepôts de confiance
- Elaborer des recommandations sur les PIDs
- Gérer les données de recherche dans les projets de science participative
- Soutenir la certification CTS pour les entrepôts français
- Faciliter l'appropriation des plans de gestion des données
- Faciliter l'appropriation des principes FAIR
- Evaluer la qualité d'une donnée et d'une métadonnée
- Contribution au Passeport SO et livret sur les données

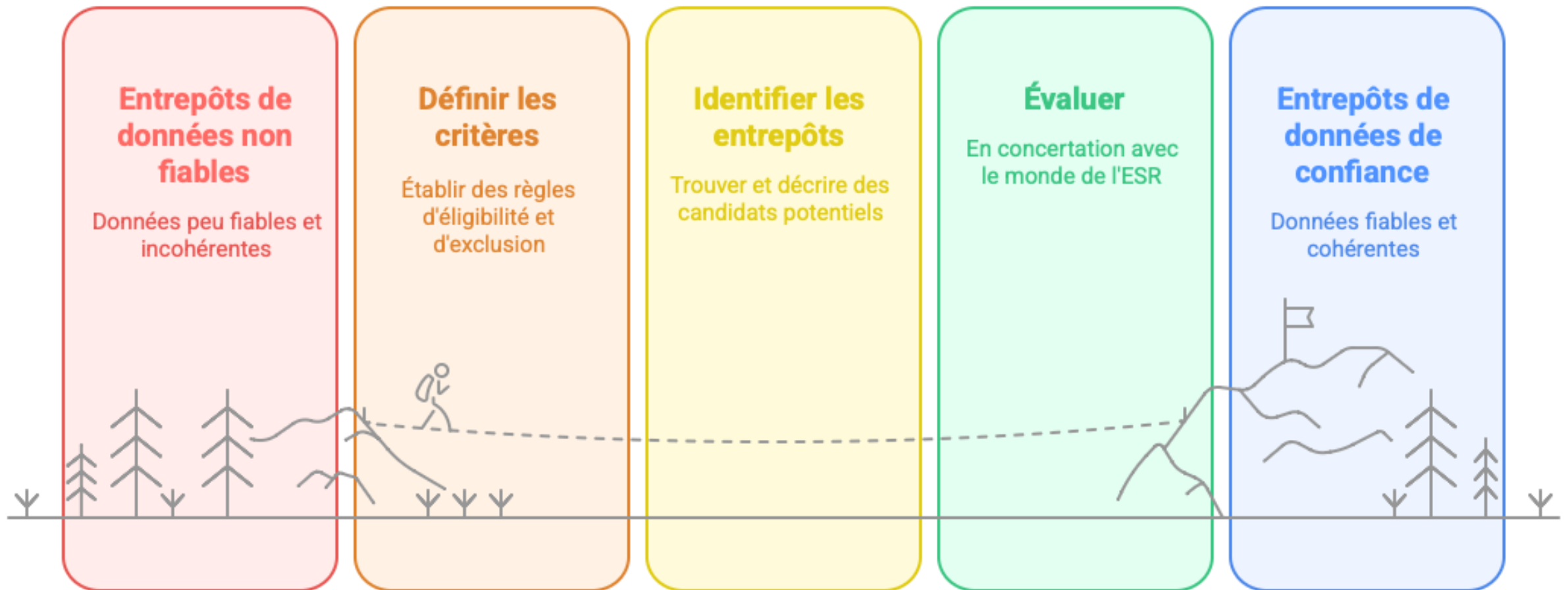
GT "Entrepôts thématiques de confiance"



- Des incitations à l'ouverture des données *sans infrastructures clairement identifiées dans la plupart des disciplines* :

« Le choix de l'entrepôt adéquat est délégué au **chercheur**. »

Rapport 2022 du Conseil des sciences et technologie du gouvernement américain



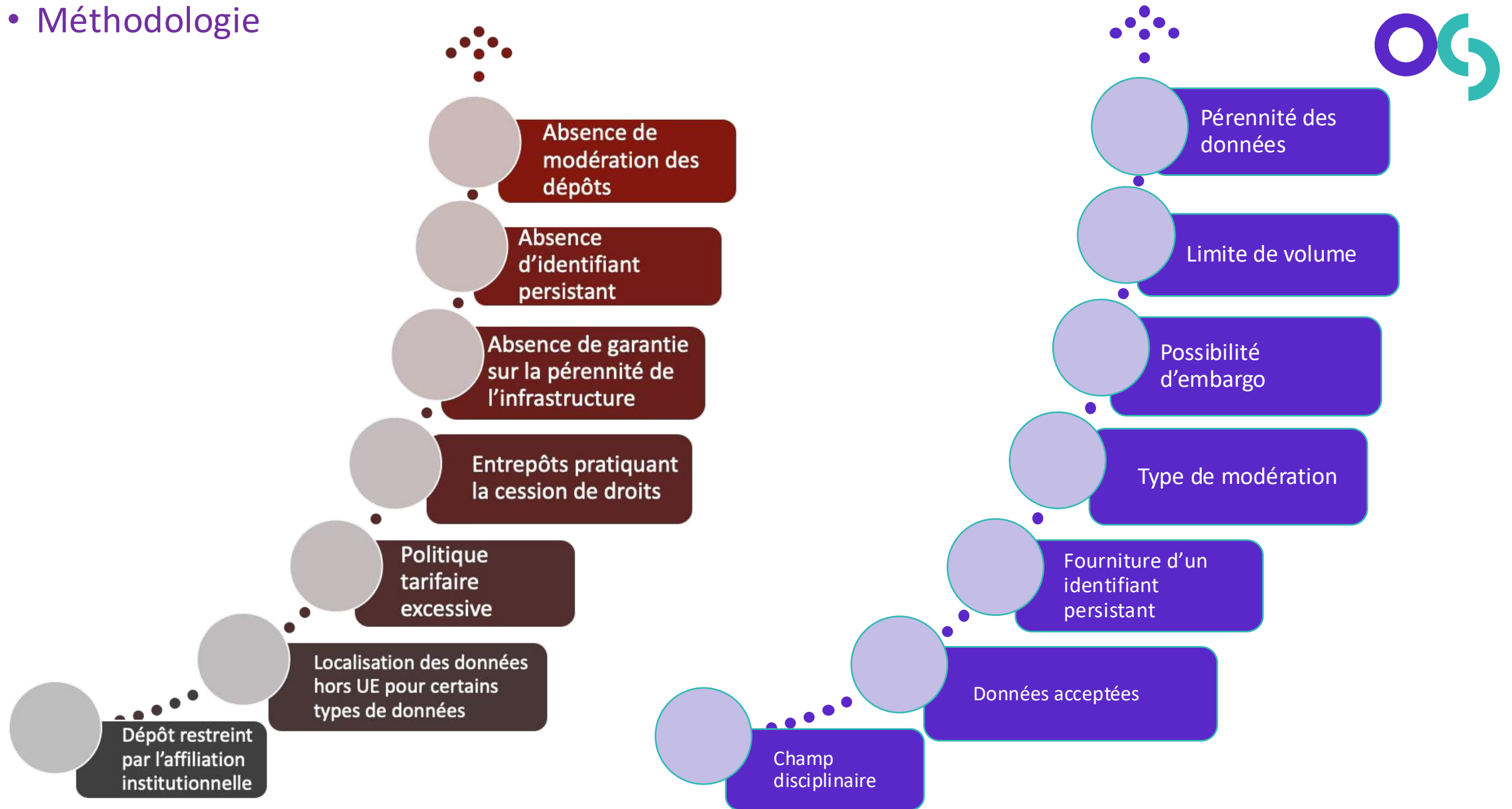


- Deux axes de travail
- Établir la liste de critères de confiance permettant d'apprécier la qualité d'un entrepôt thématique
- Proposer une note méthodologique et une liste d'entrepôts thématiques de confiance



- Les sources
- Le groupe de travail dédié du Collège des données de la recherche, composé de membres du Collège et d'experts extérieurs ;
- La littérature scientifique et la littérature grise (articles décrivant le fonctionnement des entrepôts) ;
- Les travaux de la RDA
- Les annuaires d'entrepôts (CatOpidor, Re3data, Fairsharing, Opendoar) ou liste d'entrepôts certifiés
- Les plateformes disciplinaires dédiées à la gestion des données de recherche (comme le consortium allemand NFDI, Dataacc.org, Elixir) proposant un premier recensement d'entrepôts thématiques
- Des retours d'expérience de la communauté scientifique.

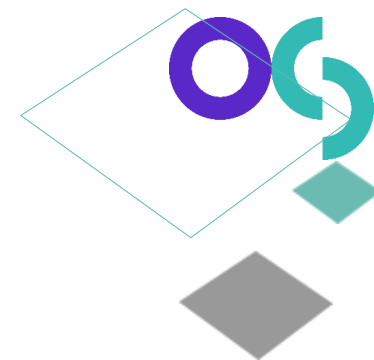
- Méthodologie



• Analyse , le Making Of



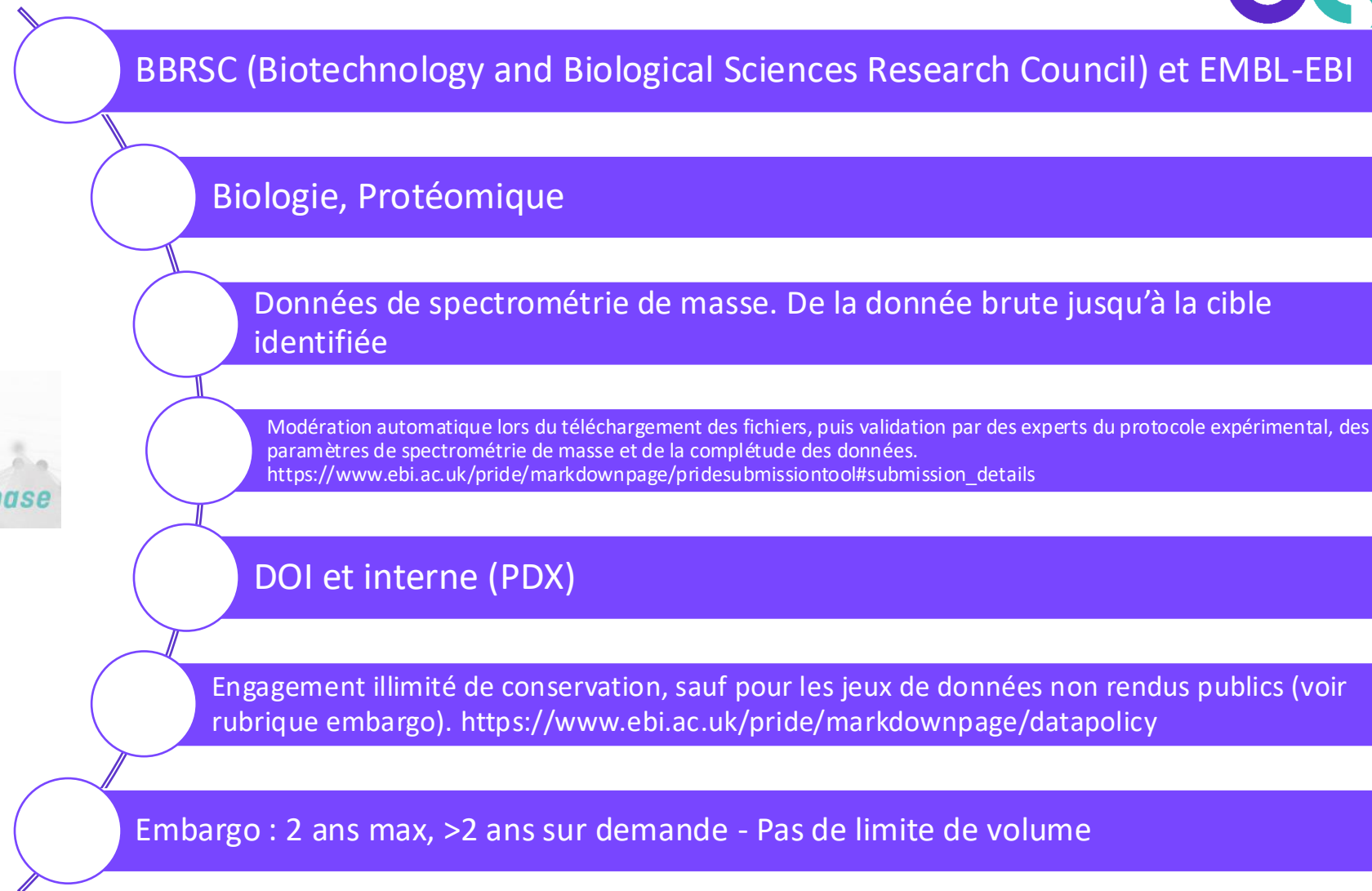
Nom	URL	Institution porteuse	Modération	Identifiant pérenne fourni	Pérennité de l'infrastructure/Engagement sur la durée de préservation des données	Discipline	Champs disciplinaires détaillés	Données acceptées= mots-clés qui parler à la	Embargo	Limite de volume
ArkeOpen	https://arkeopen.org/	Université de Strasbourg, Huma-Num	ArkeoGIS contrôle les formats d'import et la conformité des métadonnées descriptives via le protocole mis en place :	DOI	Hébergement assuré par Huma-Num	Archéologie	Archéologie, Histoire, Environnement	Données archéologiques géoréférencées	Oui	Pas de limite, toutefois l'équipe déconseille les jeux de plus de 10 000 lignes.
Cambridge Structural Database (CSD)	https://www.ccdc.cam.ac.uk/	CCDC et FIZ Karlsruhe	Automatisée (checkCIF) et humaine	DOI	Existe depuis les années 70	Chimie	Cristallographie	Structures cristallines (molécules organiques et	Oui	50 Mo par fichier, 100 Mo par versement
Centre de données astronomiques de Strasbourg (CDS)	https://cds.u-strasbg.fr/	Université de Strasbourg et CNRS	Les données déposées doivent être associées à une publication à comité de lecture ayant été acceptée.	Identifiants spécifiques à la discipline: numéro pour les catalogues, identificateur de relevé du ciel pour les collections d'images. Depuis quelques années, le CDS tente d'associer en plus un DOI à chacune de ces	Existe depuis 1972. Référence internationale dans son domaine. Certifié CoreTrustSeal en 2019, renouvellement en cours.	Astronomie	Astronomie	Catalogues (ou tables) astronomiques ; relevés images du ciel	Pas d'embargo en dehors de l'embargo standard d'un an pour les données produites dans les observatoires	Non
Chemotion	https://www.chemotion-repository.net/welcome	NFD4Chem (consortium allemand)	Une équipe de 10 reviewers assure la modération "manuelle" des données. Il y a aussi une modération automatique pour certaines données spectrales. Plus d'informations ici : https://chemotion.net/docs/repot/workflowreview	DOI	L'entrepôt est intégré à l'initiative Science Data Center du Baden-Württemberg (https://mwk.baden-wuerttemberg.de/de/service/press-und-pressemittteilung/pid/vier-science-data-centers-in-baden-wuerttemberg/) et sera prochainement considéré comme une infrastructure nationale de recherche en Allemagne. L'accès aux données est garanti pour au moins 10 ans. Un plan de stockage de long terme et d'archivage sera développé dans ce cadre sous l'égide du Centre de calcul de Steinbuch. https://chemotion.net/docs/repoff	Chimie	Chimie organique	Molécules, réactions chimiques, données analytiques (spectres) protocoles et conditions d'expérimentation. Les résultats négatifs sont acceptés (réactions ayant échoué)	Oui	Pas de limite sur la taille de fichiers, éventuellement 50 Mo à l'avenir. Pas de limite sur le volume total par utilisateur, sauf cas de mésusage détecté.
Centre de Données Socio-Politiques (CDSP)	https://cdsp.sciences-po.fr/	Sciences Po et CNRS	Le dépôt est très encadré et se fait avec l'aide d'ingénieurs du CDSP qui gèrent la curation. Voir : https://cdsp.sciences-po.fr/fr/deposer-des-donnees	DOI	Existe depuis 2005	Sciences humaines et sociales	Sciences politiques, sciences sociales	Résultats d'enquêtes nationales ou internationales, quantitatives ou	Possible en principe, ne s'est pas produit en pratique	Pas de limite formelle, les données quantitatives étant en général peu volumineuses
Collections de COpus Draux Numériques (CoCoON)	https://cocon.huma-num.fr/	UMR LACITO (Sorbonne Nouvelle Paris 3, INALCO, CNRS) et UMR LLL (Universités d'Orléans, Université de Tours, BnF, CNRS)	Vérification automatique et humaine. Syntaxe contrôlée par des schémas XML (modèle OLAC, sur-couche du Dublin-Core qualifié) et par shematron défini pour ajouter leurs propres restrictions et spécificités. Chaque demande de publication passe par une relecture des métadonnées par un administrateur. Vérification de la normalisation des champs, la complétude, la précision et la cohérence de la description. C'est lors de cette vérification qu'un alignement de certains champs sur des référentiels externes (personnes, lieux, langues, sujets, structures) est effectuée. Quelques vérifications humaines supplémentaires peuvent être faites pour s'assurer de la qualité des vidéos versées.	Identifiants OAI et DOI. Si la ressource fait l'objet d'un archivage, un identifiant ARK est attribué. Lorsque la ressource est moissonnée par ISIDORE, un identifiant HANDLE est attribué.	Démarche d'archivage pérenne (voir remarques) et partenariat Huma-Num, CIMES, CC-IN2P3	Sciences humaines et sociales	Linguistique, Anthropologie, Ethnomusicologie, Histoire	Enregistrements sonores, métadonnées descriptives	Oui (via l'accès restreint)	En fonction des formats : WAV (entre 2 et 4Go), FLAC pour les fichiers plus volumineux, mais pas formellement précisé
Conservatoire National des Données 3D	https://3d.humanities.science/	Archeovision, Consortium "3D for Humanities", CNRS-	Les métadonnées doivent être conformes au schéma déposé par le Consortium 3D SHS. Le contrôle de la qualité des métadonnées est assuré par utilisation obligatoire d'ALTAG3D. Échanges	DOI	Données hébergées chez Huma Num	Archéologie	Archéologie	Modélisations, restitutions 3D d'artefacts, de	Par défaut les métadonnées sont publiques et le dépôt lui-même	Pas de limite



Répartition disciplinaire des entrepôts sélectionnés

Domaine	Nombre d'entrepôts
Astronomie	2
Biologie	13
Chimie	9
Physique	2
Sciences de l'environnement	7
Sciences humaines et sociales	16 (dont 6 en archéologie et 3 en linguistique)

- Exemple de notice d'entrepôt



Remarques : les jeux de données sont régis par la licence CC0 depuis 2018.

Source complémentaire : Bhargava, Shruti, et Joachim Jankowski. « The PRIDE Database Resources in 2023 ».

<https://doi.org/10.1093/ndt/gfad176>.



Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante

Groupe de travail du collège Données
de la recherche

Avril 2025

•<https://recherche.data.gouv.fr/>

Rechercher sur la page

Recherche



Disciplines

Recherche

- Archéologie (6)
- Astronomie (2)
- Biologie (14)

Champs disciplinaires

Recherche

- Anthropologie (1)
- Archéologie (6)
- Astronomie (2)

Institutions porteuses

Recherche

- Archeovision (1)
- Archives de Données Issues de la Statistique Publique (ADISP) (1)

Embargo

Recherche

Recherche

ArkeOpen	Archéologie	+
Array Express	Biologie	+
Cambridge Structural Database (CSD)	Chimie	+
Centre de données astronomiques de Strasbourg (CDS)	Astronomie	+
Centre de Données Socio-Politiques (CDSP)	Sciences humaines et sociales	+
Chemotion	Chimie	+
COLlections de CORpus Oraux Numériques (CoCoON)	Sciences humaines et sociales	+
Conservatoire National des Données 3D (CND3D)	Archéologie	+
Crystallography Open Database (COD)	Chimie	+
Data Station Archaeology (DANS)	Archéologie	+
Didómena	Sciences humaines et sociales	+
EaSy Data	Sciences de l'environnement	+
ebrains	Biologie	+
EELS Data Base	Chimie Physique	+
Electron Microscopy Public Image Archive (EMPIAR)	Biologie	+
Endangered LAnguages Repository (ELAR)	Sciences humaines et sociales	+
European genome-phenome archive (EGA)	Biologie	+
European Nucleotide Archive (ENA)	Biologie	+
Geovistory	Sciences humaines et sociales	+
German Neuroinformatics Node (G-Node)	Biologie	+
Global Biodiversity Information Facility (GBIF)	Sciences de l'environnement	+
GnplS	Biologie	+



Disciplines

Sciences humaines et sociales

Institutions porteuses

« Langues et Civilisations à Tradition Orale » UMR LACITO 7107 du CNRS

Laboratoire Ligérien de Linguistique UMR LLL (7270 CNRS-Université)

Champs disciplinaires

Anthropologie

Ethnomusicologie Histoire

Linguistique

Données acceptées

Enregistrements sonores

Métadonnées descriptives

Identifiants pérennes

DOI ARK

Handle (via Isidore)

[Aller sur l'entrepôt](#) →

Modération assurée par l'entrepôt

La vérification est automatique et humaine. La syntaxe est contrôlée par des schémas XML (modèle ...

[Voir plus](#) →

Embargo

Oui (via l'accès restreint). Voir : <https://cocoon.humanum.fr/exist/crdo/missions.htm>

Limite de volume par fichier

En fonction des formats : WAV (entre 2 et 4Go), FLAC pour les fichiers plus volu...

[Voir plus](#) →

Limite de volume par jeu de données

Pas de limite précisée

GnplS

Biologie

Génomique

Génétique

Données acceptées

Ressources génétiques, polymorphismes, Données de génotypage, Données d'association, Cartes génétiques et QTL, Données de synténie



[Aller sur l'entrepôt](#) →

Limite de volume par fichier

Pas de limite précisée

Limite de volume par jeu de données

48 Mo

> [En savoir plus sur l'entrepôt GnplS](#)

6. Data sharing and long-term preservation

Expand all | Collapse all

6.1 How will data be shared?



• Describe the modalities of sharing

B *I* U A ▾ ☰ ▾ ☷ ▾ 🔗 🗃 ▾

• Reusability

e.g. Reuse for future research or other purposes (education, innovation, reproduction/transparency of science, etc.)

• Choose a data repository / catalog to deposit or describe the data

Select a value from the list or create a new one by clicking on + ▾ +

ENA : <https://cat.opidor.fr/index.php/ENA> (CoSO) ✎ ✕

• Describe each file and its characteristics

Add an element

• Persons in charge of data deposition and diffusion

Select a value from the list or create a new one by clicking on + ▾ +

Recommander des identifiants persistants pour la recherche



- Importance des PIDs dans la stratégie nationale
- Enjeu central pour la structuration et la valorisation de l'activité de recherche, notamment pour permettre une traçabilité des objets scientifiques
- Clarifier les préconisations françaises dans le paysage international

 Ouvrir
la science !



**Recommandations
pour l'adoption
des identifiants
persistants dans
l'enseignement
supérieur et la
recherche**

Collège des Données de la recherche

Septembre 2025

DOI : 10.52949/84

Bénéfices de l'usage des PIDs dans la recherche

Améliorer la visibilité et la traçabilité de la recherche

Les PIDs facilitent la découverte des produits de la recherche

Faciliter la réutilisation des produits de la recherche

Les PIDs facilitent la réutilisation des produits de la recherche

Favoriser l'intégrité scientifique

Les PIDs agissent comme des points de référence uniques pour les informations de recherche

Soutien des principes FAIR

Les PIDs fournissent des points d'ancrage techniques et organisationnels

Réduire les charges administratives

Les PIDs automatisent les échanges d'informations et standardisent les (méta)données

Contexte international : des bénéfices mesurés



Des stratégies d'adoption et d'usages actives dans certains pays (Australie, Canada, Royaume-Uni, République tchèque...)

Bénéfices en termes de gestion et de circulation des données sur la recherche et l'innovation

- Garantir l'identification, et la traçabilité des productions, des activités et des acteurs tout au long du cycle de recherche,
- Renforcer l'intégrité de la recherche, la confiance dans les résultats,
- Eviter de multiples saisies de mêmes informations,
- Développer des services à valeur ajoutée (CV complet),
- Cartographier la capacité de recherche et mesurer son impact,
-

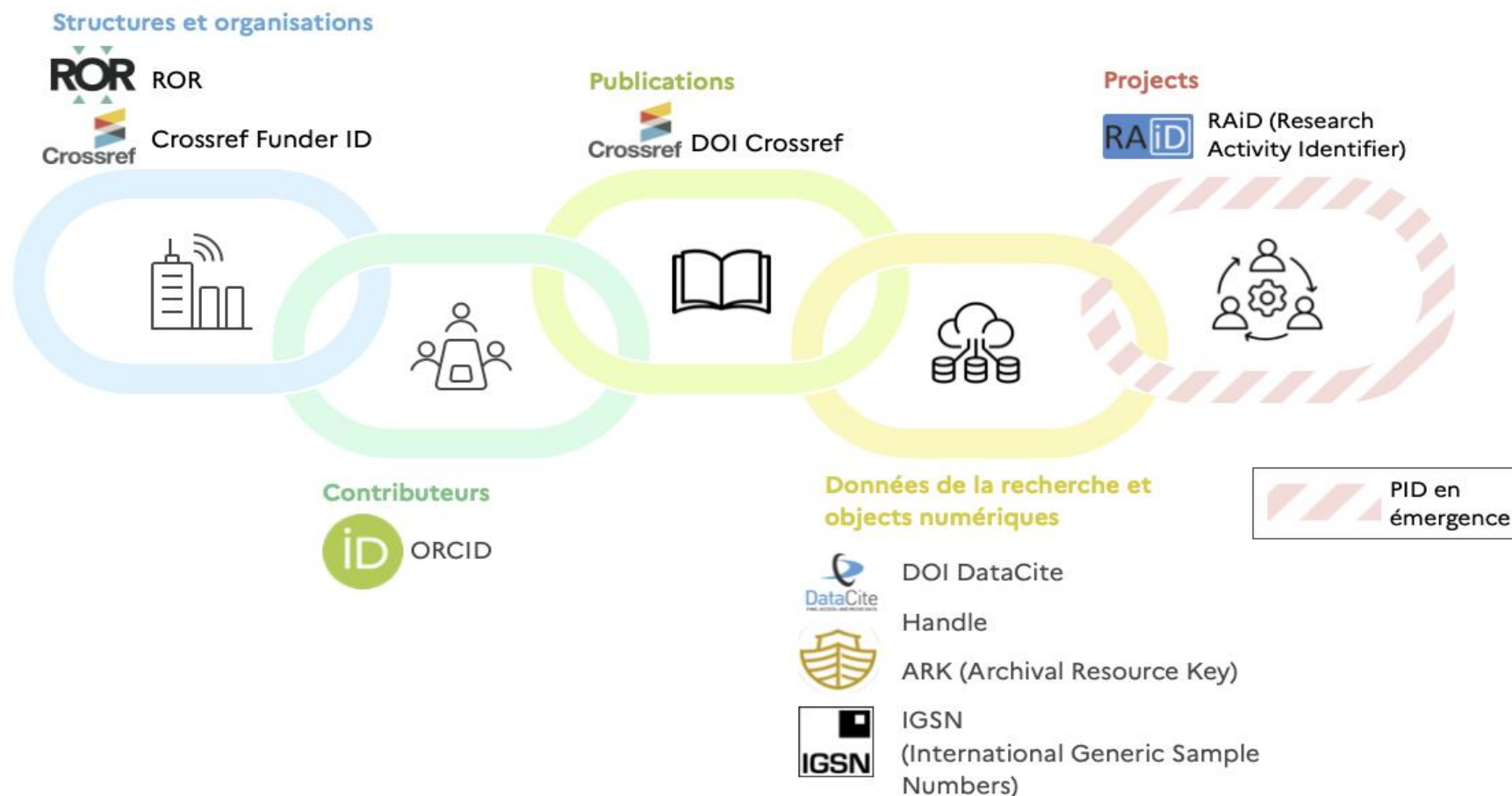


Australie : coût de la ressaisie des métadonnées relatives aux projets, publications, subventions
38 000 jours/an
24M\$
(pour 109 000 chercheurs)

<https://doi.org/10.5281/zenodo.7100578>

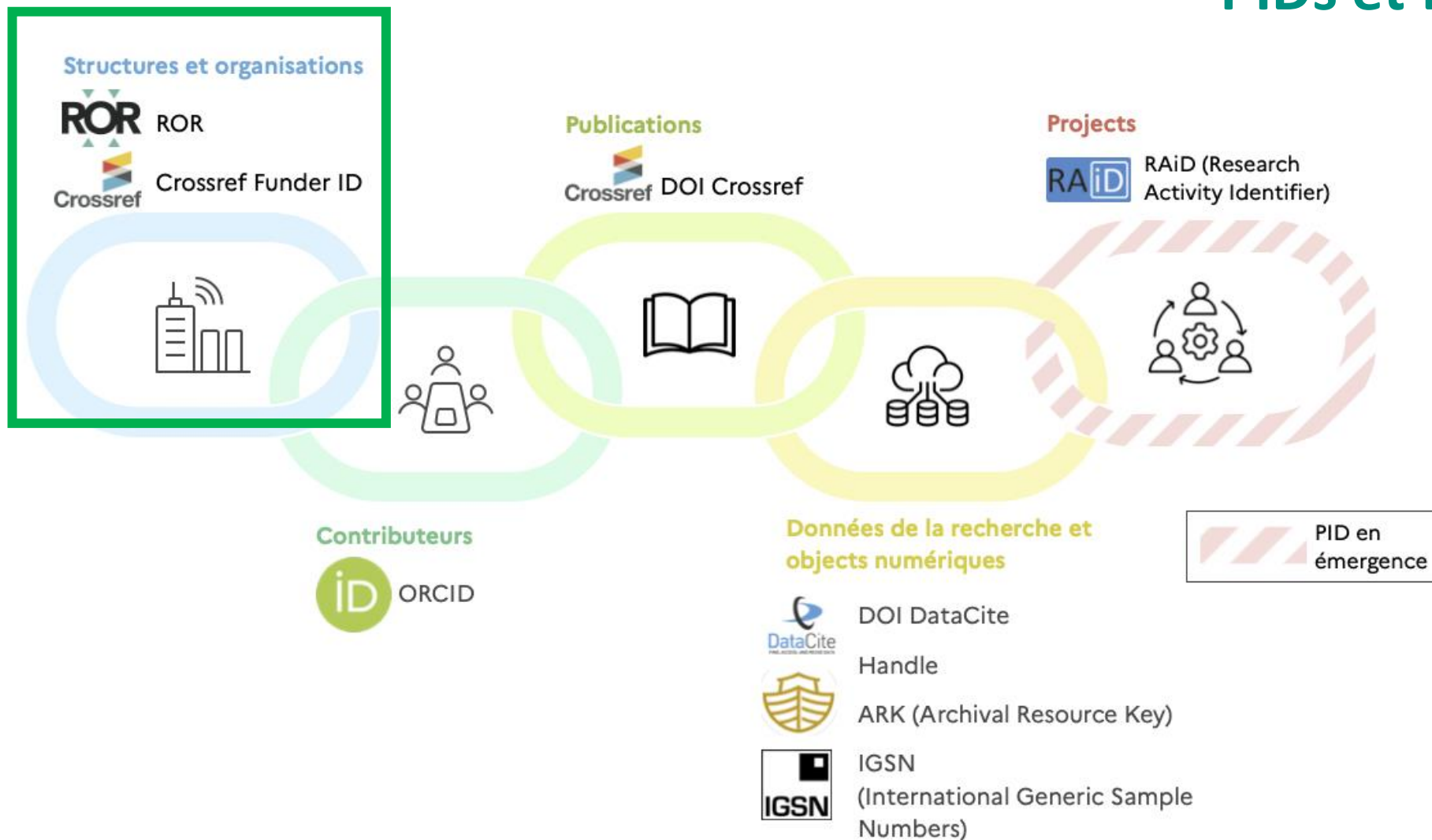


Recommander des identifiants persistants pour la recherche



DOI : 10.52949/84

PIDs et ESR



PIDs et ESR

Structures et organisations



ROR



Crossref Funder ID

Publications

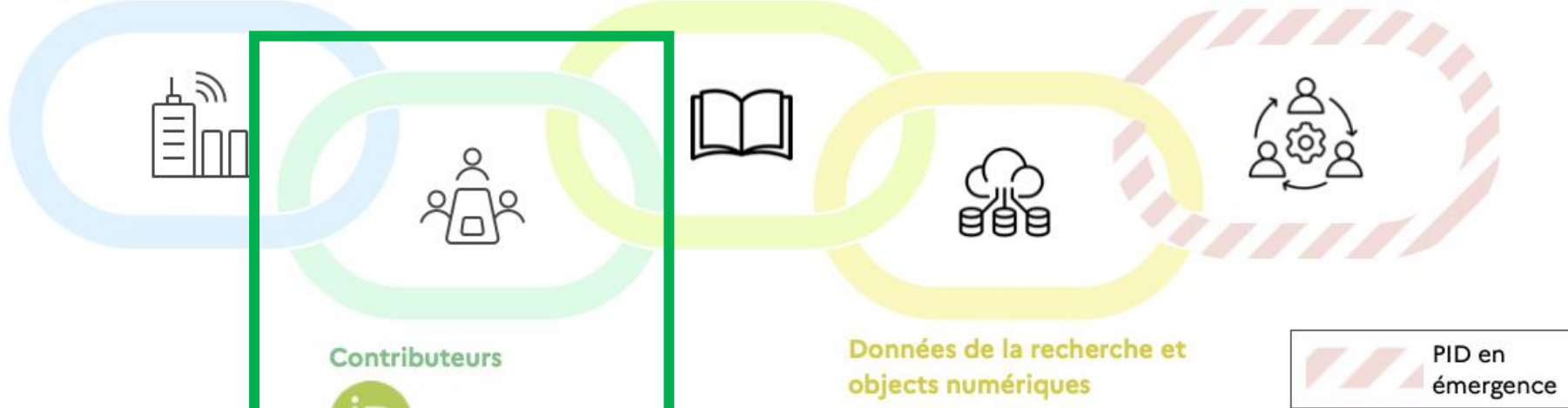


DOI Crossref

Projects



RAiD (Research Activity Identifier)



Contributeurs



Données de la recherche et objets numériques



DOI DataCite

Handle



ARK (Archival Resource Key)

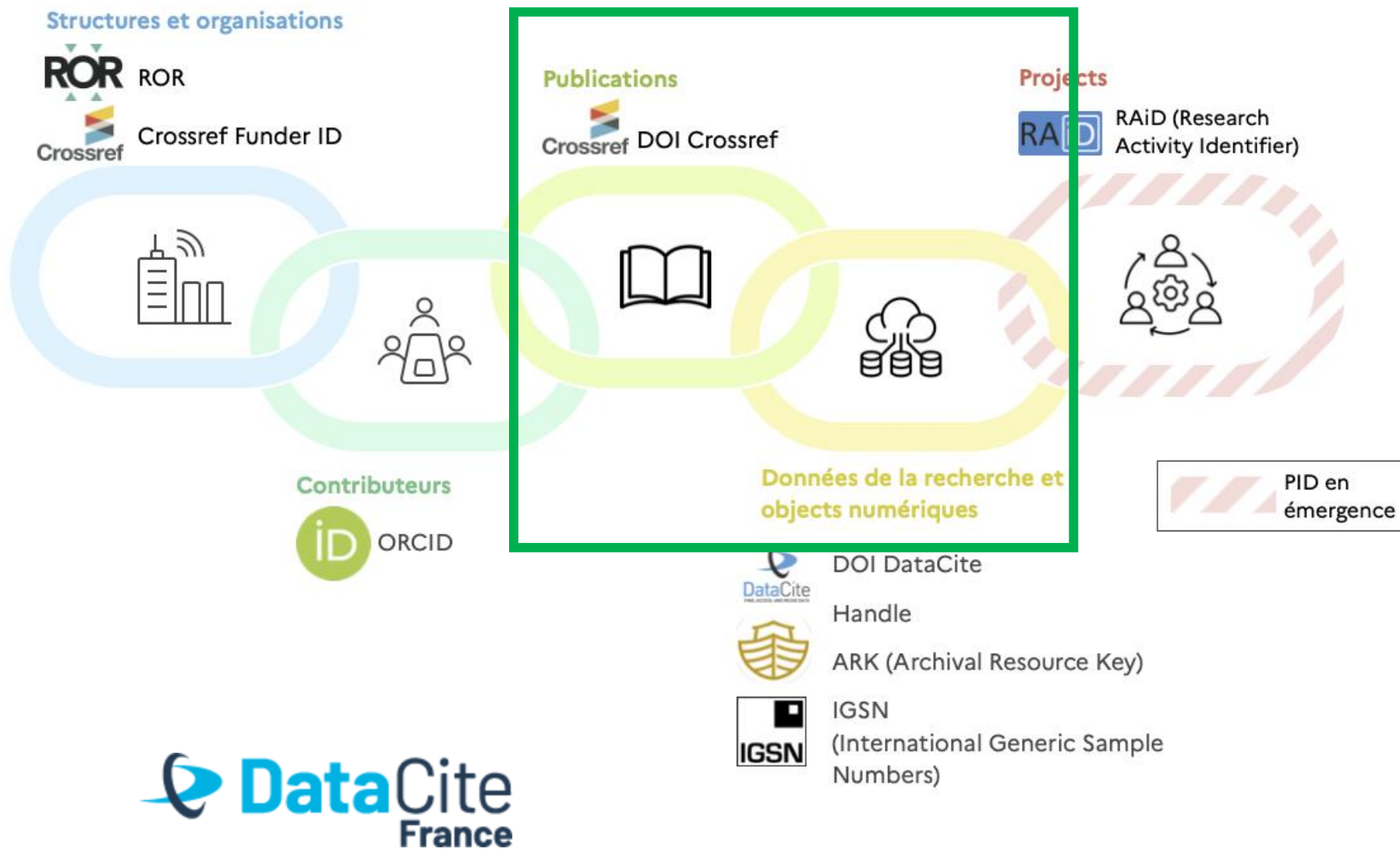


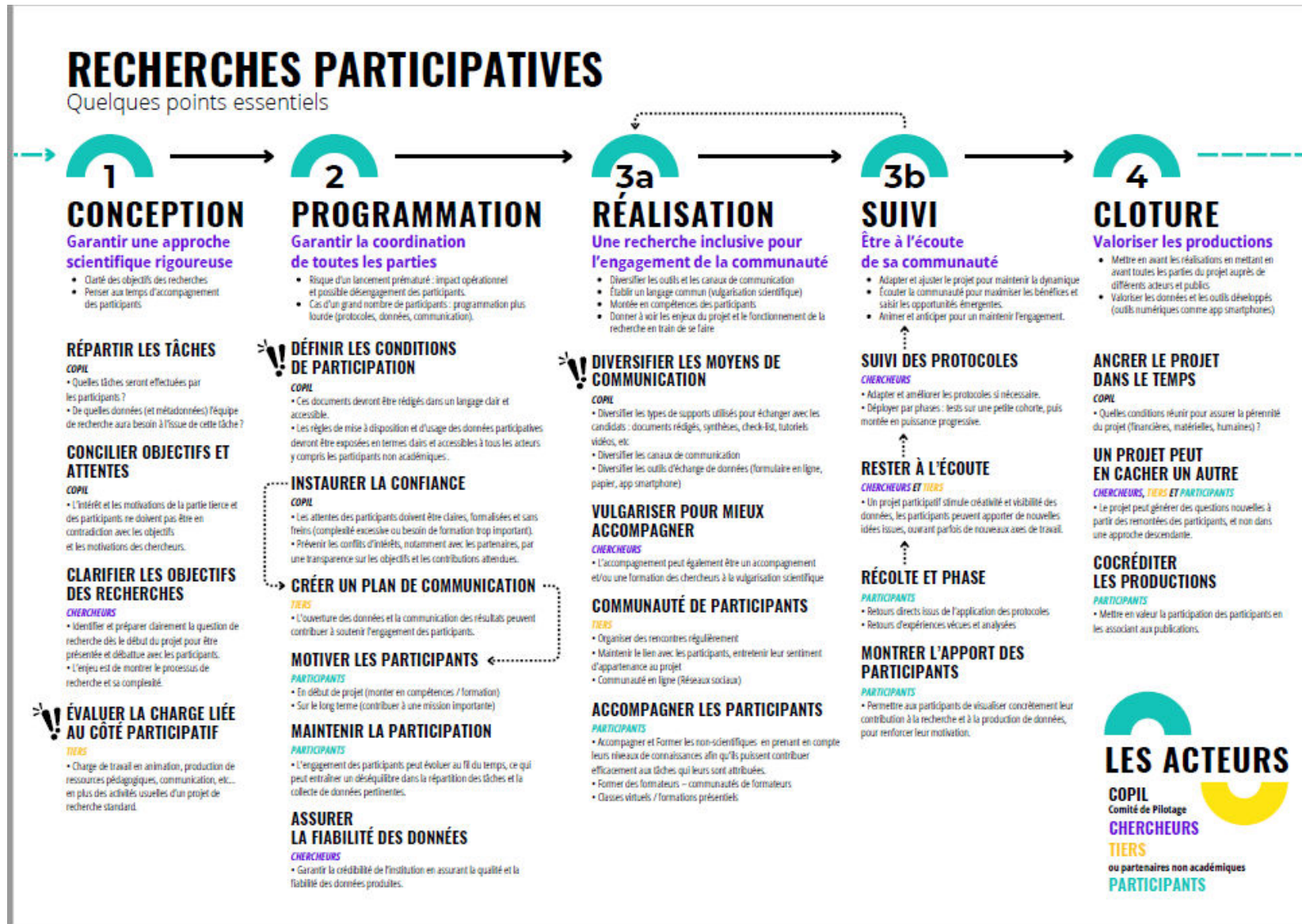
IGSN

(International Generic Sample Numbers)



PIDs et ESR



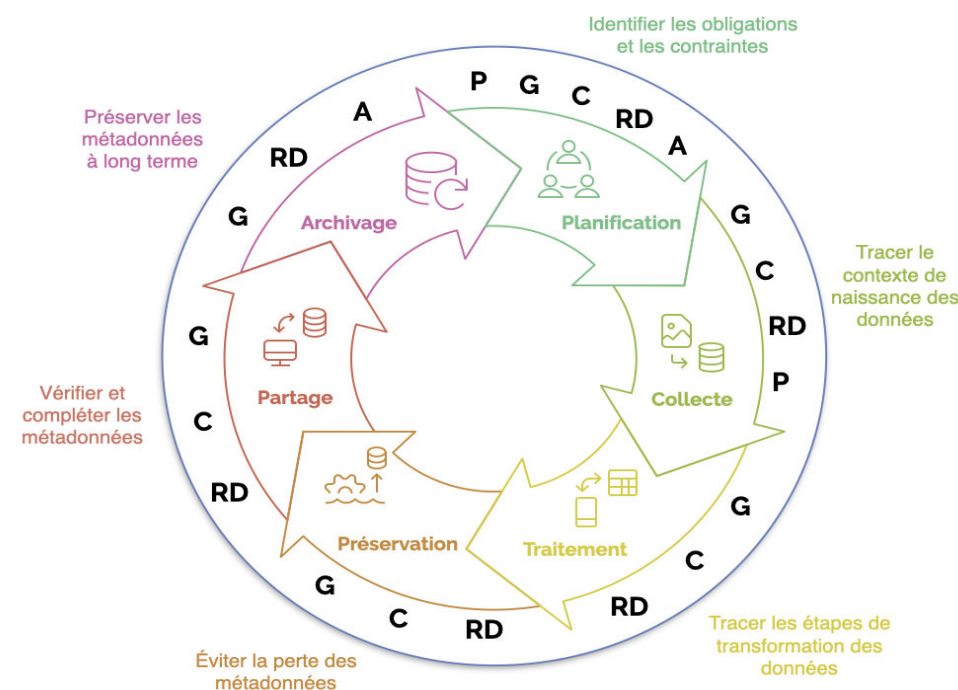


Des outils pratiques pour vérifier la qualité des données et des métadonnées



Tableau récapitulatif : étapes, actions, acteurs et outils

Étape du cycle de vie	Actions clés	Acteurs	Outils / méthodes
Planification	Rédiger le plan de gestion de données, Identifier les entrepôts/catalogues cibles	Producteurs, Gestionnaires, Curateurs	Tableurs, README, XML
Collecte/Acquisition	Documenter les jeux de données réutilisés, Transmettre les métadonnées minimales	Producteurs, Gestionnaires	Tableurs structurés, Documentation d'aide
Traitement/Analyse	Documenter les jeux de données générés, Compléter avec les métadonnées nécessaires	Responsables de l'analyse, Gestionnaires	Systèmes de gestion de versions
Préservation/Stockage	Conserver les métadonnées des jeux de données supprimés, Mettre à jour les fichiers de suivi	Gestionnaires, Curateurs	Systèmes de gestion de versions Fichiers de suivi
Partage/Diffusion	Vérifier la complétude des métadonnées, Convertir les métadonnées au format recevable	Gestionnaires, Curateurs	Outils de conversion, Entrepôts certifiés
Archivage	Vérifier l'intégrité des métadonnées, Compléter les métadonnées descriptives	Archivistes, Gestionnaires	Métadonnées archivistiques, OAIS



Les Acteurs
 P : Producteurs, G : Gestionnaires, C : Curateurs, RA : Responsables de l'analyse, D : Responsables de la diffusion, A : Archivistes

DOI : 10.52949/86

DOI : 10.52949/87

Plans de gestion des données : recommandations pour une gestion responsable et ouverte des données



- Instrument structurant des projets de recherche (gestion responsable des données, qualité, reproductibilité, partage selon les principes FAIR...)
 - Perçu comme une contrainte administrative
 - Enjeu de construire des PGD qui répondent aux exigences des financeurs, aux principes FAIR
- ET deviennent des outils opérationnels
- Ma-DMP



**Plans de gestion
des données :
recommandations
pour une gestion
responsable et
ouverte des
données**

—
Collège des Données de la recherche
—

Plans de gestion des données : d'une contrainte administrative à une opportunité scientifique



Bénéfices

- Transparence, reproductibilité
- Visibilité et valorisation des données
- Respect des exigences réglementaires
- Rationalisation du workflow
- Gain de temps
-

Devient un outil de pilotage d'un projet

- Pour les chercheurs
- Pour les équipes de recherche et les établissements
- Pour les plateformes et les équipements scientifiques
- Pour les financeurs

Plans de gestion des données : d'une contrainte administrative à une opportunité scientifique



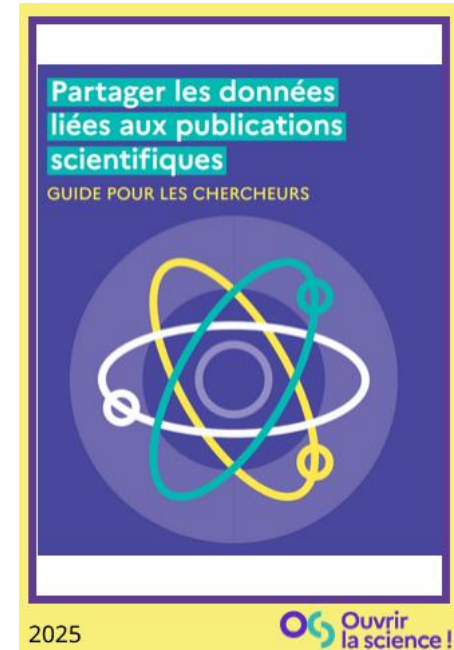
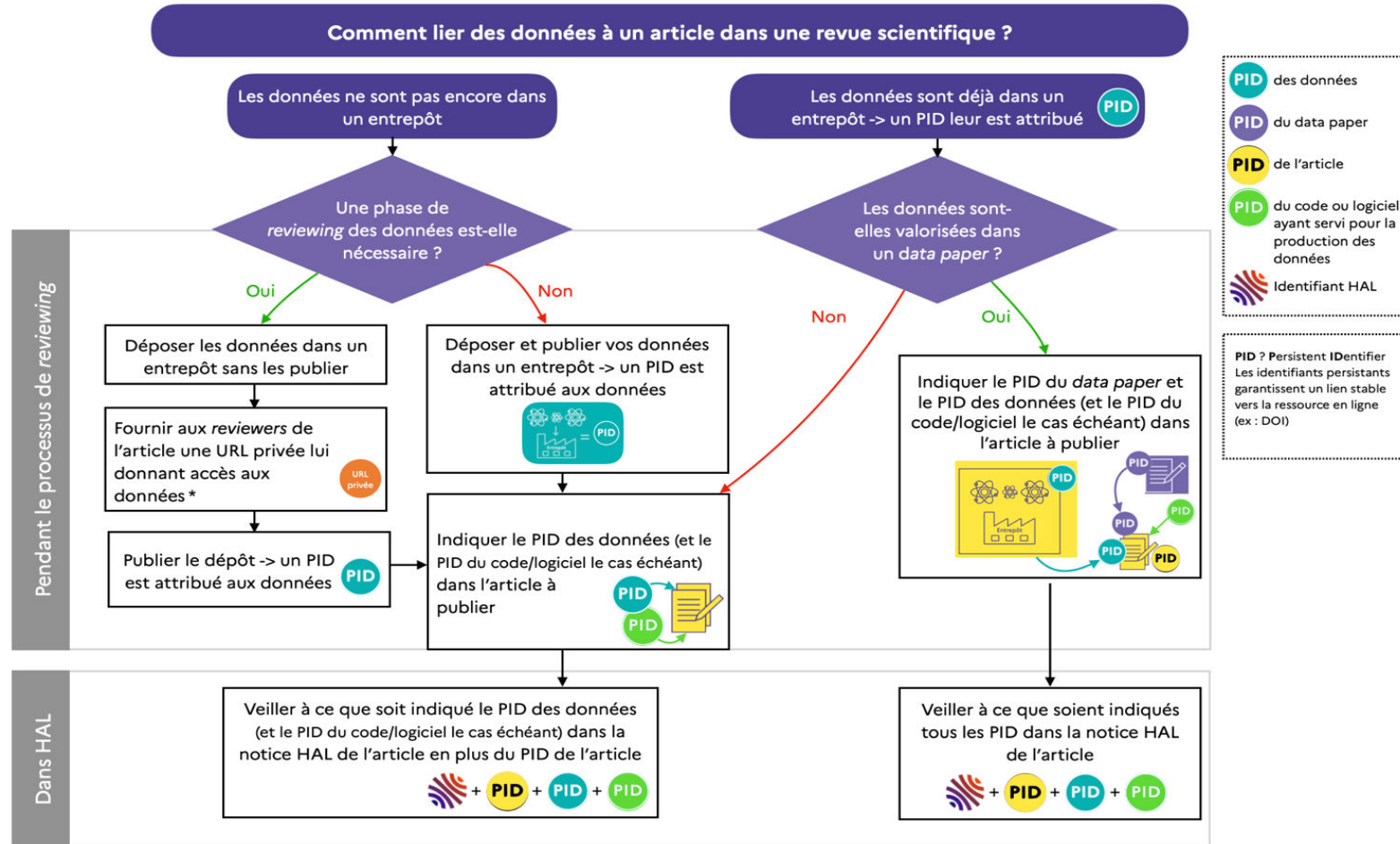
Recommandations pour l'élaboration de plans de gestion des données :

- dans le cadre de projets de recherche
 - à l'échelle d'une structure ou d'une plateforme
 - à l'échelle institutionnelle
-
- Explicitation des notions de PGD projet, structure, plateforme
 - Articulation entre les PGD et les politiques institutionnelles
 - Mise en avant forte du modèle Science Europe structuré

Mises à jour des livrables précédents



Vérification de l'intégralité des livrables de la précédente mandature
Mise à jour du guide Partager les données liées aux publications scientifiques



* Dans le cas où l'entrepôt ne fournit pas d'URL privée, vous pouvez communiquer le PID des données publiées dans l'entrepôt ou transmettre les données au reviewer par un autre moyen

20 livrables depuis 2019



Mandature 2019-2022
7 livrables

Mandature 2022-2025
13 livrables

