



HAL
open science

Gestion et ouverture des données de la recherche

Eva Legras, Gaëlle Jaouen, Romuald Lorthioir, David Gasparotto

► To cite this version:

Eva Legras, Gaëlle Jaouen, Romuald Lorthioir, David Gasparotto. Gestion et ouverture des données de la recherche. Printemps de la Donnée 2024, INRAE; Université Haute-Alsace; Université de Strasbourg; INSA; PNDB; AgroParisTech; Université de Lille; Sorbonne Université; Data Terra, Mar 2024, Paris, France. hal-04682029

HAL Id: hal-04682029

<https://hal.inrae.fr/hal-04682029v1>

Submitted on 30 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

27 mars 2024

Formation

Gestion et Ouverture des Données de la Recherche

Eva Legras - Chargée de Mission Science ouverte et Données de la Recherche

Gaëlle Jaouen - Chargée de Mission Données de la Recherche

Romuald Lorthioir – Directeur Services Numériques

David Gasparotto - Directeur DDPC

Préambule : La science ouverte, une nouvelle approche de la recherche



La science ouverte désigne l'ensemble des principes et des actions qui facilitent l'ouverture et la libre dissémination, à destination de l'ensemble de la société, des productions de la recherche scientifique, ainsi que la transparence et l'ouverture des processus de la recherche.

[Politique science ouverte d'AgroParisTech, 2020.](#)

La science ouverte, un principe soutenu par des politiques de plus en plus fermes

> Article L112-1

Modifié par LOI n°2013-660 du 22 juillet 2013 - art. 16

La recherche publique a pour objectifs :

a) Le développement et le progrès de la recherche dans tous les domaines de la connaissance ;

b) La valorisation des résultats de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie ;

c) Le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès ;

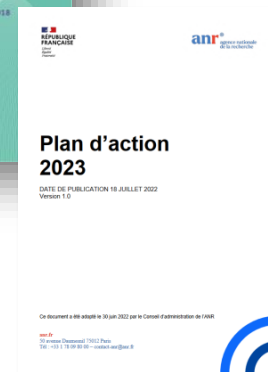
c bis) Le développement d'une capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique, et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;

d) La formation à la recherche et par la recherche ;

e) L'organisation de l'accès libre aux données scientifiques.



La science ouverte, un principe soutenu par des politiques de plus en plus fermes



- Politiques nationales
- Politiques des agences de financement publique
- Politiques institutionnelles

Données de la recherche (DR) et ouverture de la recherche

Enjeux autour des données :

- Accélération de la production et de la manipulation de DR (essor technologique)
- Définition d'un cadre légal et politique d'accessibilité et de réutilisabilité des DR
- Volonté de maîtriser les conditions de circulation et d'utilisation des DR (sécurisées, fiables, traçables...)

Volonté de la Recherche (et de notre établissement) de :

- Formaliser des règles communes de gestion et d'ouverture
- Accompagner les agents vers de bonnes pratiques et la connaissance des droits et devoirs
- Soutenir la validation des résultats de recherche
- Préserver le patrimoine de données et les DR à venir (coûteuses, uniques/non-reproductibles)

Lancement de plusieurs politiques DR intra-organismes : [AgroParisTech](#) (1^{er} Janvier 2021), [INRAE](#), [IRD](#), [CNRS](#), [CIRAD](#)...

Définition des données de la recherche pour AgroParisTech

Les données de la recherche sont l'ensemble des enregistrements factuels (informations numériques, textuelles, visuelles ou sonores...) collectés, observés ou créés dans le cadre d'une activité de recherche.

Ces données sont nécessaires à la construction de la recherche, à l'établissement et à la validation des résultats de recherche.

Lorsque les données n'ont pas encore été traitées ou contextualisées, on parle de données brutes.

Entre données brutes et résultats finaux, les transformations successives des données peuvent servir de base à d'autres travaux de recherche et peuvent donc, de ce fait, être également considérées comme données de la recherche.

Quand s'y intéresse-t-on?

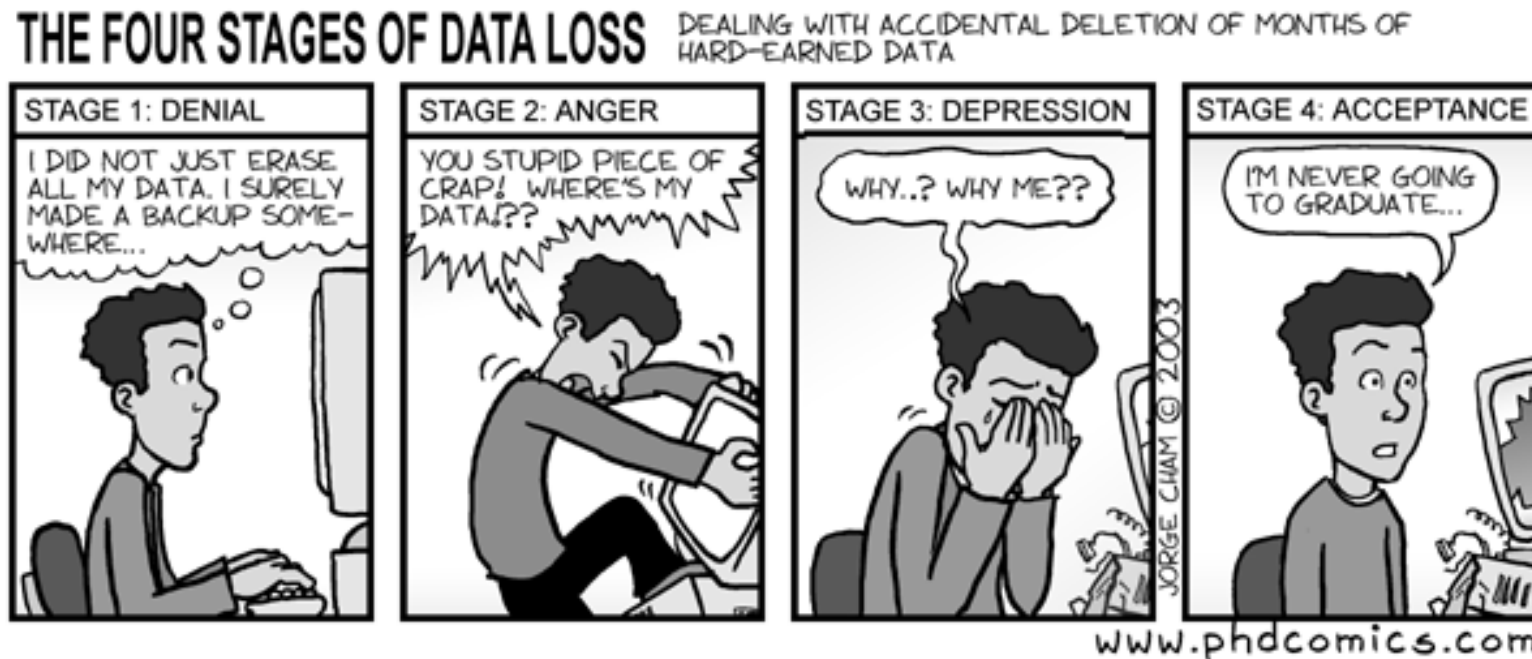
- Gestion d'un projet
- Gestion d'un laboratoire, d'une équipe, d'une plateforme, d'une unité de recherche
- Gestion du patrimoine scientifique d'un établissement

Le cycle de vie des données



Adapté d'après [University of Sidney](#) et [URFIST de Paris](#).

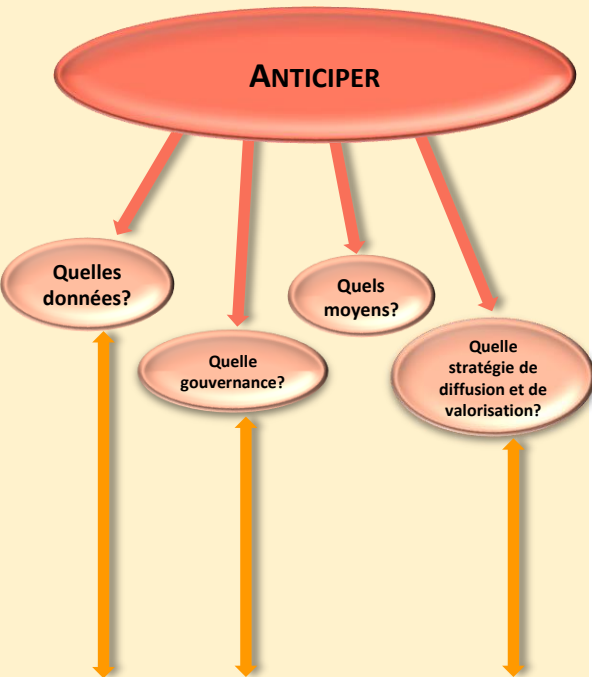
Les enjeux: de l'importance de l'anticipation ...



Pour éviter ça, reprenons point par point!

Principes clés et logigramme (exemple : cas d'un projet de recherche)

Montage du Projet



Droits ? Devoirs ? Cadre légal, réglementaire et éthique

Cadre légal de protection et d'ouverture des données de la recherche

Les questions juridiques et réglementaires, un point important à anticiper !

Définir la titularité des données : qui est responsable de quoi sur les données du projet ?

⇒ Nature des données ? Producteurs ? Partenaires ? Convention ?

Identifier les obligations à remplir depuis l'étape de collecte des données jusqu'à l'ouverture des résultats :

Qui peut voir et réutiliser les données du projet, à quel moment ?

⇒ secret ; confidentialité

⇒ communicabilité ; ouverture (à anticiper)

Sous quelles conditions particulières les données peuvent être collectées/utilisées ?

⇒ Enjeux éthiques

⇒ Enjeux de sécurité

⇒ **Enjeux réglementaires**

Cadre légal de protection et d'ouverture des données de la recherche

Ensemble de lois et règlements (français/européens) pour encourager l'ouverture de la science :

- Protéger les droits des producteurs de données (paternité, traçabilité) et des sujets d'études (confidentialité, sensibilité, anonymat)
- Cadrer les droits des utilisateurs
- Stimuler les collaborations

L'ouverture des données publiques numériques achevées est maintenant une norme et non une option !

- Données de la recherche publique = document administratif
- Les décisions reviennent aux établissements, pas au chercheur/ingénieur

Quelques cas particuliers (qu'il faut justifier), notamment : données sensibles, secret industriel, propriété intellectuelle

« Aussi ouvert que possible, aussi fermé que nécessaire »

Attention : Anticipation et entente entre partenaires sont fondamentales, surveiller les conditions des contrats et conventions

[Fiche pratique : Cadre juridique données de la recherche](#)

Cadre légal de protection et d'ouverture des données de la recherche

Données confidentielles

Données revêtant un caractère sensible élevé. Communication interdite par principe.

Sécurité :

De l'Etat, publique, de l'établissement.

Secret :

- Secret professionnel : médical, bancaire et fiscal, secret de l'instruction
- Secret industriel et commercial : procédés, informations économiques et financières, stratégies commerciales ou industrielles
- Données issues d'un contrat de prestation de service pour un acteur privé
- Connaissances propres d'un tiers communiquées pour collaboration ou prestation + Dossier Technique Secret

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions

Confidentialité de principe ; communication possible sous certaines conditions.

Sécurité :

Potentiel scientifique et technique de la nation : unités protégées, zone à régime restrictif

Secret :

Secret statistique

Propriété intellectuelle :

Textes, sons, images.

Cas particuliers : contenus sous licences libres, *text and data mining* (TDM) à des fins de recherche.

Données provenant d'un tiers :

Connaissances issues d'une collaboration avec un tiers privé dans le cadre d'un contrat de recherche

Données personnelles :

Données de santé, etc. : application du règlement général pour la protection des données (RGPD)

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur la recherche sous contrat

Avant l'ouverture des données/résultats => se poser les bonnes questions = s'assurer ne pas faire défaut à certaines de nos obligations vis-à-vis de nos partenaires.

Q1 – Dans quel cadre ont été obtenus les données/résultats ?

- **Recherche Interne** : Recherche sur fonds propre, sans contractualisation ni utilisation de données obtenues par l'intermédiaire d'un accord de confidentialité.
=> Les règles « classiques » d'ouverture s'appliquent
- **Recherche sous contrat** : Prestation, Collaboration, Consortium, ...
=> Dans ce cas la, des règles spéciales peuvent s'appliquer -> **Q2**

Q2 – Quelles sont les modalités prévues dans le contrat ?

Notamment : obligation de confidentialité pour les connaissances propres des partenaires (informations non tombées dans le domaine public)

De manière usuelle :

- **Prestation** : Les Résultats appartiennent au Partenaire, il n'est pas possible de publier sauf accord express
- **Collaboration/Consortium** : Les Résultats sont la copropriété des Parties, pendant la durée de L'Etude et une certaine durée à l'issue de celui-ci, chaque Partie devra soumettre ses projets de publication pour validation. Si les Résultats peuvent être protégés par un titre de propriété intellectuelle, un délai de 18 mois pourra être accordé.

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

Les enjeux :

Donner aux citoyens de l'union européenne des garanties de **contrôle** et de **visibilité** sur leurs **données personnelles** quelle que soit la localisation du stockage et des traitements :

- Quelles données sont collectées ?
- De quelle manière ?
- Où sont-elles ?
- Comment sont-elles mises à jour ?
- Comment les récupérer sur simple demande ?

Le périmètre :

Tous les acteurs économiques et sociaux collectant les données personnelles de citoyens européens : les entreprises, les associations, les organismes publics et leurs sous-traitants.

Deux notions importantes :

- La responsabilité de tous les acteurs d'un traitement de données nominatives
- Le contrôle a posteriori

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

Pour pouvoir être mis en œuvre, tout traitement de données doit se fonder sur l'une des « bases légales » prévues par le RGPD = la licéité du traitement (article 6 RGPD)

Le traitement n'est licite que si, et dans la mesure où, au moins une des conditions suivantes est remplie :

- a) **consentement**
- b) **contrat**
- c) **obligation légale**
- d) **intérêts vitaux de la personne**
- e) **mission d'intérêt public**
- f) **intérêts légitimes**

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

Donnée personnelle = toute information se rapportant à une personne physique identifiée ou identifiable

Ces personnes doivent en conserver la maîtrise!

L'identification peut-être :

- Directe : nom et prénom
- Indirecte : numéro de téléphone, plaque d'immatriculation, identifiant tel que le numéro de sécurité sociale, adresse postale ou courriel, mais aussi la voix ou l'image

L'identification peut-être faite :

- à partir d'une seule donnée : nom
- à partir du croisement d'un ensemble de données : une femme vivant à telle adresse, née tel jour et membre dans telle association par exemple

Par contre, des coordonnées d'entreprises (la « Compagnie A » avec son adresse postale, le numéro de téléphone de son standard et un courriel de contact générique « compagnie1@email.fr ») ne sont pas, en principe, des données personnelles.

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

Les 5 grands principes des règles de protection des données personnelles :

- **Le principe de finalité** : le responsable d'un fichier ne peut enregistrer et utiliser des informations sur des personnes physiques que dans un but bien précis, légal et légitime ;
- **Le principe de proportionnalité et de pertinence** : les informations enregistrées doivent être pertinentes et strictement nécessaires au regard de la finalité du fichier ;
- **Le principe d'une durée de conservation limitée** : il n'est pas possible de conserver des informations sur des personnes physiques dans un fichier pour une durée indéfinie. Une durée de conservation précise doit être fixée, en fonction du type d'information enregistrée et de la finalité du fichier ;
- **Le principe de sécurité et de confidentialité** : le responsable du fichier doit garantir la sécurité et la confidentialité des informations qu'il détient. Il doit en particulier veiller à ce que seules les personnes autorisées aient accès à ces informations ;
- **Les droits des personnes** : droit à l'information, droit d'accès, droit de rectification, droit à l'effacement, droit limitation du traitement, droit à la portabilité des données, droit d'opposition

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

Dérogations prévues pour la recherche scientifique :

- **Le principe de finalité** : Admission **d'une certaine indétermination des finalités initiales** des traitements à des fins de recherche: un responsable de traitement peut changer de finalités du moment que les nouvelles finalités restent « compatibles » avec la finalité initiale de la collecte de données.
- **Le principe de proportionnalité et de pertinence** : **Absence de dérogation** au principe de minimisation.
- **Le principe d'une durée de conservation limitée** : Les données peuvent être conservées au-delà de la durée qui a été nécessaire pour atteindre la finalité de recherche (par exemple, au-delà de la durée d'un projet de recherche déterminé) du moment qu'elles sont ensuite conservées uniquement pour être utilisées à des fins de recherche.
- **Le principe de sécurité et de confidentialité** : Pour éviter toute possibilité d'accès par des personnes non autorisées, il faut appliquer des procédures d'**anonymisation ou de pseudonymisation** qui permettent de garantir la protection des données personnelles dans le cadre d'un traitement de recherche scientifique.
- **Les droits des personnes** : Le RGPD et le législateur ont prévu qu'il peut être dérogé à plusieurs de ces droits sous certaines conditions si ces droits risqueraient de rendre impossible, d'entraver sérieusement la réalisation des finalités du traitement, de la compromettre, et quand une telle dérogation est nécessaire pour atteindre ces finalités. L'appréciation se fait au cas par cas.

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

Dérogations prévues pour la recherche scientifique :

- **Le cas particulier des données dites « sensibles »**

Origine raciale ou ethnique ; opinions politiques ; convictions religieuses ou philosophiques ; appartenance syndicale ; données génétiques ou biométriques aux fins d'identifier une personne physique de manière unique ; la vie sexuelle ou l'orientation sexuelle ; données concernant la santé

⇒ Traitement par principe interdit

⇒ Pour la recherche : notion de « [proportionnalité] à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée »

- **Données de santé :**

« ensemble des données se rapportant à l'état de santé d'une personne concernée qui révèlent des informations sur l'état de santé physique ou mentale passé, présent ou futur de la personne concernée. »

⇒ Le texte insiste sur l'intérêt de pouvoir traiter des données de santé dans le cadre de la recherche :

« En combinant les informations issues des registres, les chercheurs peuvent acquérir de nouvelles connaissances d'un grand intérêt en ce qui concerne des problèmes médicaux très répandus tels que les maladies cardiovasculaires, le cancer et la dépression. »

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

ANONYMISATION = appliquer un ensemble de techniques à des données personnelles pour rendre impossible toute ré-identification de la personne, par quelque moyen que ce soit. OPERATION IRREVERSIBLE

Utilité : L'anonymisation ouvre des potentiels de réutilisation des données et permet ainsi aux acteurs d'exploiter et de partager leur « gisement » de données sans porter atteinte à la vie privée des personnes. Elle permet également de conserver des données sans limitation.

Exemple: dans un fichier contenant la date de naissance des personnes, remplacer cette information par la seule année de naissance, ou un intervalle temporel (par exemple : individus âgés entre 25 et 30 ans)

PSEUDONYMISATION = traitement de données personnelles pour ne plus pouvoir attribuer les données à une personne physique identifiée sans information supplémentaire. OPERATION REVERSIBLE (possible de retrouver l'identité d'une personne si l'on dispose d'informations supplémentaires)

Exemple: remplacer les données directement identifiantes (nom, prénoms, etc.) par des données indirectement identifiantes (alias, numéro séquentiel, etc.) : une base de données de CV où seuls les noms et prénoms d'une personne auront été remplacés par un numéro qui ne correspond qu'à elle permet d'individualiser cette personne. Cette base serait considérée comme pseudonymisée et non comme anonymisée.

Cadre légal de protection et d'ouverture des données de la recherche

Données communicables sous conditions : focus sur le RGPD

Quelques règles simples d'application du RGPD

- Identifier les données entrant dans le cadre du RGPD
- Contacter le responsable de l'application du RGPD dans les organisations = Délégué à la Protection des Données (DPD) ou Data Protection Officer (DPO) et la direction responsable des services informatiques pour :
 1. S'assurer que la collecte et le traitement sont bien conformes
 2. Suivre la procédure indiquée par le ou les DPO pour déclarer les données et traitements concernés
 3. Mettre en place les garanties de sécurité adaptées au stockage et à l'accès
- Désigner un correspondant données dans le projet qui s'assurera du respect des actions indiquées

Le RGPD est aussi une démarche vertueuse d'auto amélioration dans la gestion de ses données. Profitez-en !

AgroParisTech : dpo@agroparistech.fr

Dans le cadre des UMR : désignation d'un.e DPO référent auprès de l'une des tutelles.

Cadre légal de protection et d'ouverture des données de la recherche

Données obligatoirement communicables

Données devant obligatoirement être communiquées à la personne qui en fait la demande.

Loi CADA : <https://www.cada.fr/>

Droit des citoyens à accéder aux données produites par les administrations, sous réserve de dispositions légales contraires.

Points de contact dans les administrations : personne responsable de l'accès aux documents administratifs et des questions relatives à la réutilisation des informations publiques = PRADA.

Données relatives à des émissions de substance dans l'environnement : Convention d'Aarhus :

Les données relatives à des émissions de substances dans l'environnement doivent être communiquées à toute personne qui en fait la demande, sauf cas très précis (mise en cause de la sécurité de l'Etat, impact sur des procédures juridictionnelles, communication portant atteinte à des droits de propriété intellectuelle). ([guide d'analyse juridique](#), p. 12)

Cadre légal de protection et d'ouverture des données de la recherche

Données obligatoirement diffusées

Publication d'office, sans attendre de demande. Accès à toute personne avec une connexion internet.

Intérêt sociétal :

Bases de données et données dont la publication présente un intérêt économique, social, environnemental et sanitaire, produites dans le cadre d'une activité de service public, et en l'absence de dispositions légales contraires.

Loi CADA :

Données ayant fait l'objet d'une communication à un.e citoyen.ne dans le cadre d'une demande CADA.

Codes sources :

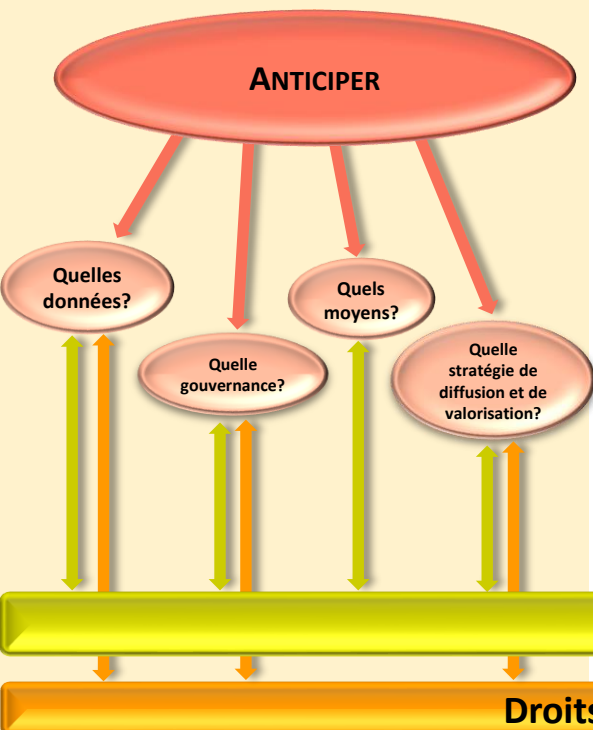
Code de tout programme produit dans le cadre d'une mission de service public.

Données géographiques :

Cf. Directive INSPIRE

Principes clés et logigramme (exemple : cas d'un projet de recherche)

Montage du Projet



Plan de Gestion des Données

Document pense-bête, synthétique, collaboratif et itératif, répertoriant tous les aspects de la gestion des DR du projet : production, documentation, organisation, stockage, partage, protection, ouverture, archivage, moyens humains et financiers

Objectifs :

- Soulever les points à discuter pour le bon déroulement du projet
- Eviter les mauvaises surprises et les oublis
- Prévoir les moyens et ressources nécessaires, anticiper les coûts
- Se mettre d'accord au sein du projet, fixer les règles

Plusieurs trames disponibles (ANR, Horizon Europe, institutionnelles, Trames AgroParisTech Projet et Entité) rassemblées sur DMP Opidor

Précautions :

- Lister les types de données et leur volume prévisionnel pour anticiper leurs particularités et les précautions à prendre (cadre juridique, infrastructures...)
- Désigner un/des responsable(s) du PGD dans le projet

[Fiche pratique : Plan de gestion de données](#)

Plan de Gestion des Données

Trame AgroParisTech pour un projet de recherche :

- 1) Informations sur le plan de gestion des données
- 2) Informations sur le projet
- 3) Information sur les données : origine ; contrôle qualité ; métadonnées
- 4) Mesures légales, contractuelles et éthiques
- 5) Fonctionnement et infrastructures : stockage ; sauvegardes ; partage et accès collaboratif aux fichiers
- 6) Ouverture des données
- 7) Archivage

Un exemple de première version de PGD pour le [projet fictif Forestcription](#)

[Modèles PGD AgroParisTech](#)

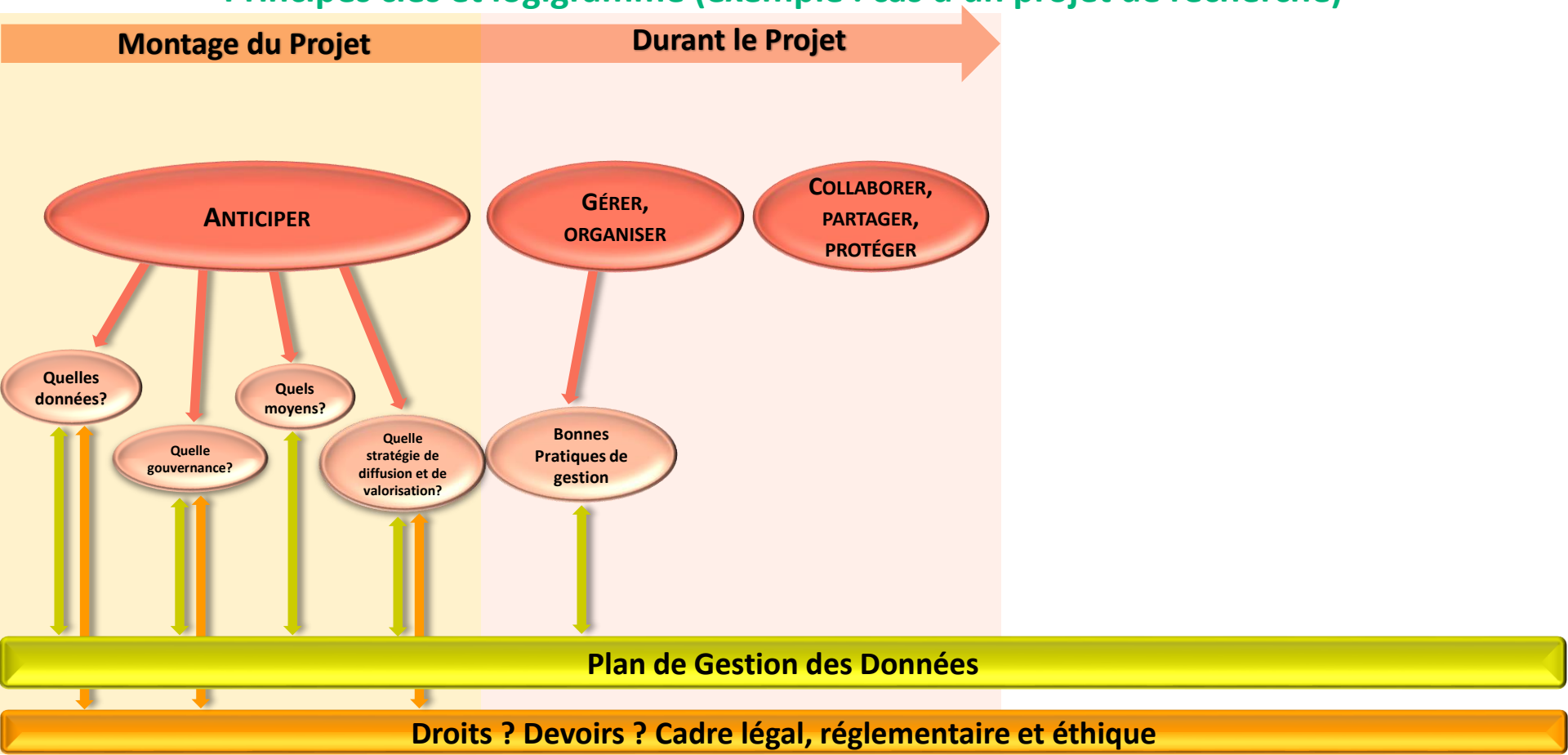
Plan de gestion de données et financeurs

- Pour de plus en plus d'agences de financement : Plan de gestion de données obligatoire.
- Europe, ANR : Première version à 6 mois après le début du projet, mises à jour en cours de projet.
- Livrable attendu, conséquences financières en cas de non complétion.



Source : [Inrae, Rédiger un PGD, 2020](#)

Principes clés et logigramme (exemple : cas d'un projet de recherche)



Bonnes Pratiques de Gestion

Ensemble de règles et réflexes pour optimiser :

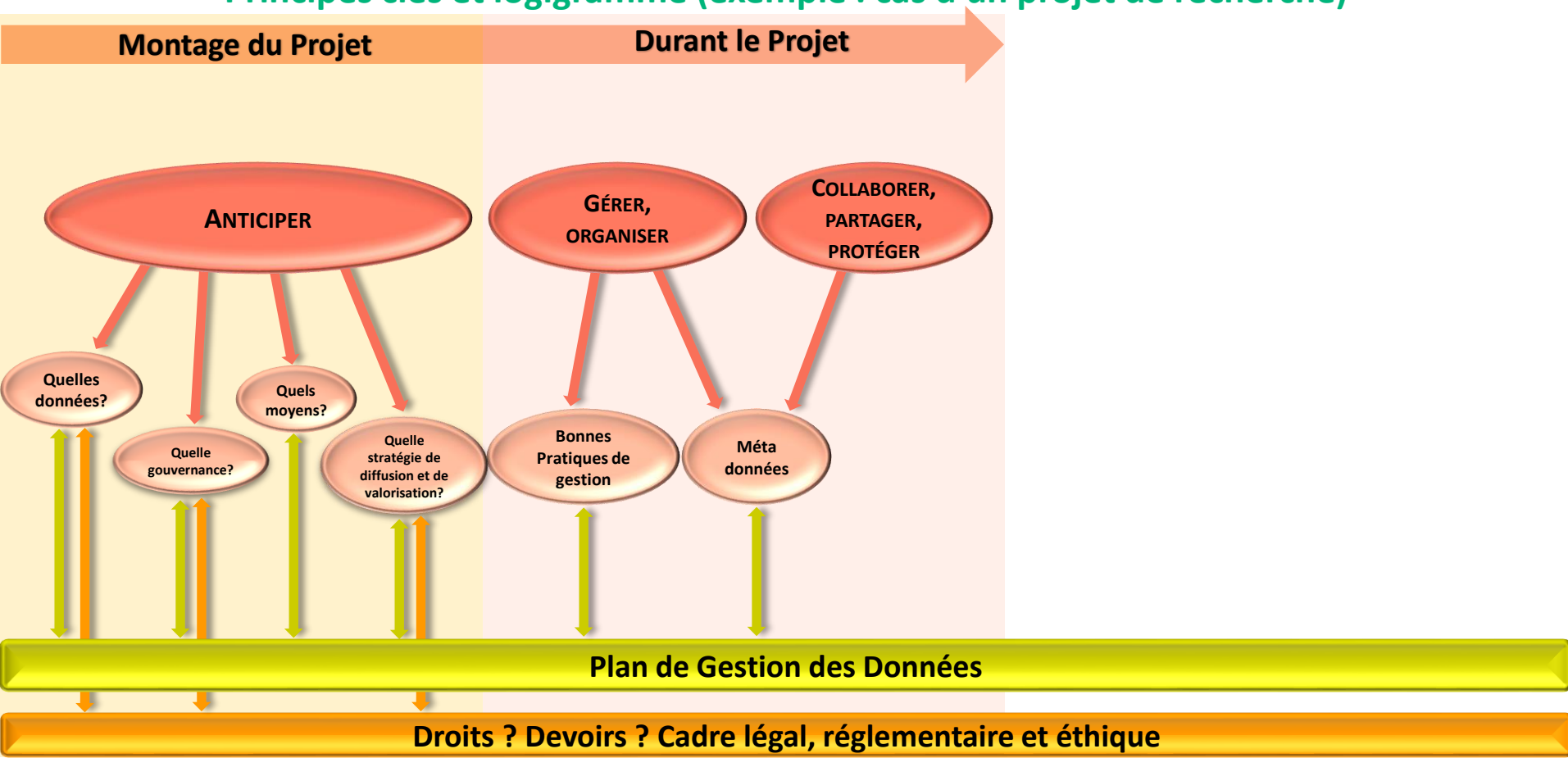
- La gestion des fichiers et des données à l'intérieur des fichiers (documentation, nommage, organisation, versionnement, homogénéité)
- La collaboration (formats ouverts, outils collaboratifs)
- La limitation des risques de confusions, d'erreurs, de pertes de données (sécurité du stockage, politique de sauvegardes)

Gagner du temps et... éviter les catastrophes !

Les précautions à prendre :

- Rester réalistes et tenir compte des moyens disponibles (humains, techniques, financiers)
- Tous les membres du projet suivent les mêmes règles pour que tout le monde se comprenne !
- Documenter autant que possible pour être transparents et accessibles aux nouveaux collaborateurs

Principes clés et logigramme (exemple : cas d'un projet de recherche)



Métadonnées

Tous les éléments de description des DR = Carte d'identité d'un jeu de données :

- contexte d'acquisition (pourquoi, comment, auteur(s)...)
- unité de mesure
- dates de collecte, de révision
- format de fichier
- Etc.

Pour assurer la compréhension et la réutilisabilité éclairée des DR, conserver les informations de genèse des DR, faciliter les collaborations

Les précautions à prendre :

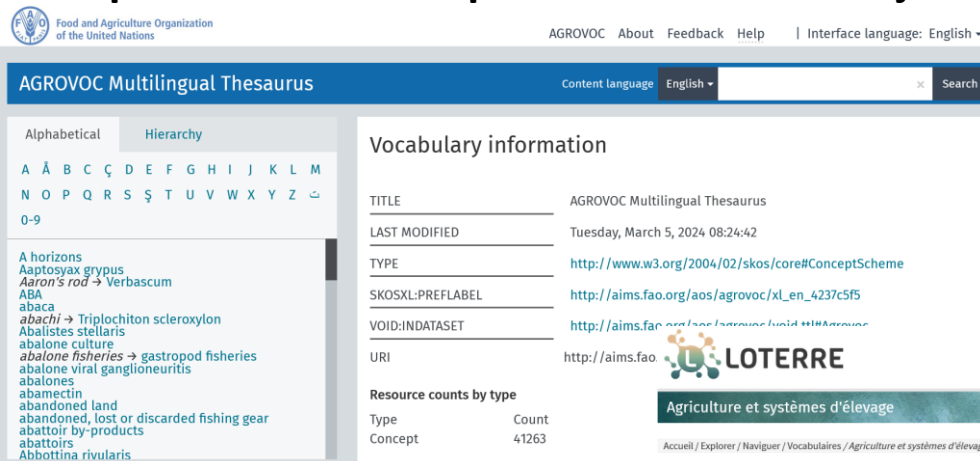
Choisir un standard de métadonnées (Dublin Core, Darwin core, EML, DataCite...) adapté au projet et à la communauté scientifique, avoir des métadonnées les plus riches possibles

Ressources : [RDAlliance](#) ; [DCC](#) ; [FairSharing](#)

Exemple de Métadonnées d'un dépôt de données en écologie, format Dublin Core:

```
<?xml version="1.0" encoding="UTF-8" ?>
<metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns="http://dublincore.org/documents/dcmi-terms/">
  <dcterms:title>Paracou forest permanent plots details</dcterms:title>
  <dcterms:identifiant>https://doi.org/10.18167/DVN1/8G8AHY</dcterms:identifiant>
  <dcterms:creator>Jaouen, Gaëlle</dcterms:creator>
  <dcterms:publisher>CIRAD Dataverse</dcterms:publisher>
  <dcterms:issued>2022-10-20</dcterms:issued>
  <dcterms:modified>2022-10-20T17:38:20Z</dcterms:modified>
  <dcterms:description><a href="https://paracou.cirad.fr/">Paracou Research Station</a>. Plots size, location and type. <a href="http://www.ecofog.gf/spip.php?article364">Guyafor
  network</a></dcterms:description>
  <dcterms:subject>Earth and Environmental Sciences</dcterms:subject>
  <dcterms:subject>Forest Ecology</dcterms:subject>
  <dcterms:subject>Forest Management</dcterms:subject>
  <dcterms:subject>Tropical Rain Forest</dcterms:subject>
  <dcterms:subject>Permanent Plots</dcterms:subject>
  <dcterms:subject>French Guiana</dcterms:subject>
  <dcterms:language>English</dcterms:language>
  <dcterms:isReferencedBy>Gourlet-Fleury Sylvie, Guehl Jean-Marc, Laroussinie Olivier. 2004. Ecology and management of a neotropical rainforest : lessons drawn from Paracou, a long-term
  experimental research site in French Guiana. Paris : Elsevier, 336 p., isbn, 2-84299-455-8, http://www.gip-ecofor.org/ecology-and-management-of-neotropical-rainforest-elsevier-ouvrage-
  telechargeable/</dcterms:isReferencedBy>
  <dcterms:contributor>Jaouen, Gaëlle</dcterms:contributor>
  <dcterms:contributor>Derroire, Géraldine</dcterms:contributor>
  <dcterms:dateSubmitted>2021-08-30</dcterms:dateSubmitted>
  <dcterms:temporal>1984</dcterms:temporal>
  <dcterms:temporal>1984</dcterms:temporal>
  <dcterms:relation><a href="https://doi.org/10.18167/DVN1/B8FHHA">Guyafor Data Dictionary</a></dcterms:relation>
  <dcterms:type>Experimental data</dcterms:type>
  <dcterms:type>Observational data</dcterms:type>
  <dcterms:type>Temporal series</dcterms:type>
  <dcterms:spatial>French Guiana</dcterms:spatial>
  <dcterms:license>NONE</dcterms:license>
  <dcterms:rights><a rel="license" href="http://creativecommons.org/licenses/by-nc-sa/4.0/"></a><br />Ce(tte) œuvre est mise à disposition selon les termes de la <a rel="license"
  href="http://creativecommons.org/licenses/by-nc-sa/4.0/">Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International</a>.
  </dcterms:rights>
</metadata>
```

Exemples de thésaurus pour l'indexation des jeux de données:



Food and Agriculture Organization of the United Nations

AGROVOC About Feedback Help | Interface language: English

AGROVOC Multilingual Thesaurus

Content language: English

Alphabetical Hierarchy

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z

A horizons
Aaptosyax grypus
Aaron's rod → Verbascum
ABA
abāca
abachi → Triplochiton scleroxylon
Abalistes stellaris
abalone culture
abalone fisheries → gastropod fisheries
abalone viral ganglioneuritis
abalones
abamectin
abandoned land
abandoned, lost or discarded fishing gear
abattoir by-products
abattoirs
Abbottina rivularis

Vocabulary information

TITLE: AGROVOC Multilingual Thesaurus

LAST MODIFIED: Tuesday, March 5, 2024 08:24:42

TYPE: <http://www.w3.org/2004/02/skos/core#ConceptScheme>

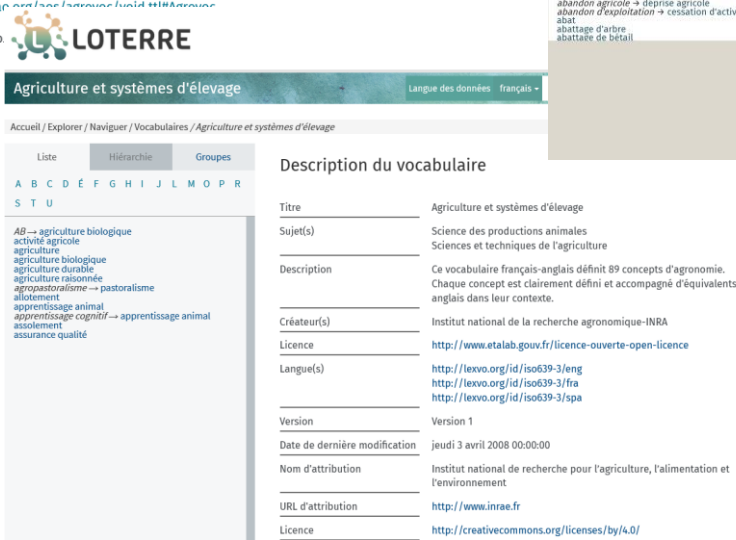
SKOSXL-PREFLABEL: http://aims.fao.org/aos/agrovoc/xl_en_4237c5f5

VOID:INDATASET: <http://aims.fao.org/aos/agrovoc/void/#Agrovoc>

URI: <http://aims.fao.org/aos/agrovoc/>

Resource counts by type

Type	Count
Concept	41263



Accueil / Explorer / Naviguer / Vocabulaires / Agriculture et systèmes d'élevage

Langue des données: français

Liste Hiérarchie Groupes

A B C D E F G H I J L M O P R
S T U

AB → agriculture biologique
activité agricole
agriculture
agriculture biologique
agriculture durable
agriculture raisonnée
agropastoralisme → pastoralisme
allègement
apprentissage animal
assolement
assurance qualité

Agriculture et systèmes d'élevage

Langue des données: français

Accueil / Explorer / Naviguer / Vocabulaires / Agriculture et systèmes d'élevage

Liste Hiérarchie Groupes

A B C D E F G H I J L M O P R
S T U

AB → agriculture biologique
activité agricole
agriculture
agriculture biologique
agriculture durable
agriculture raisonnée
agropastoralisme → pastoralisme
allègement
apprentissage animal
assolement
assurance qualité

Description du vocabulaire

Titre: Agriculture et systèmes d'élevage

Sujet(s): Science des productions animales
Sciences et techniques de l'agriculture

Description: Ce vocabulaire français-anglais définit 89 concepts d'agronomie. Chaque concept est clairement défini et accompagné d'équivalents anglais dans leur contexte.

Créateur(s): institut national de la recherche agronomique-INRA

Licence: <http://www.etalab.gouv.fr/licence-ouverte-open-licence>

Langue(s): <http://lexvo.org/id/iso639-3/eng>
<http://lexvo.org/id/iso639-3/fra>
<http://lexvo.org/id/iso639-3/spa>

Version: Version 1

Date de dernière modification: jeudi 3 avril 2008 00:00:00

Nom d'attribution: Institut national de recherche pour l'agriculture, l'alimentation et l'environnement

URL d'attribution: <http://www.inrae.fr>

Licence: <http://creativecommons.org/licenses/by/4.0/>



Vocabulaires Ouverts @ INRAE

Thésaurus INRAE

Langue des données: français

Accueil / Vocabulaires / Thésaurus INRAE

À propos | Nous contacter | Aide | In English

A-Z Hiérarchie Groupes Nouveautés

A B C D E F G H I J K L
M N O P Q R S T U V W X Y Z
0-9

AC → analyse canonique
ADV → virus adeno-associé
abaissément de nappe → rabattement de nappe
abandon agricole → déprise agricole
abandon d'exploitation → cessation d'activité
abat
abattage d'arbre
abattage de bétail

Description du vocabulaire

TITRE: Thésaurus INRAE

DESCRIPTION: Thésaurus ouvert et partagé couvrant les domaines de recherche d'INRAE. Il sert de référentiel au sein de l'institut pour indexer et annoter des documents, pages web, descriptions d'activités, jeux de données, à des fins de recherche ou d'analyse de l'information. Contribuez en proposant des modifications, ajouts ou corrections du thésaurus en cliquant sur 'Nous contacter'

LANGUE: <http://id.loc.gov/vocabulary/iso639-1/en>
<http://id.loc.gov/vocabulary/iso639-1/fr>

VERSION: 2.0

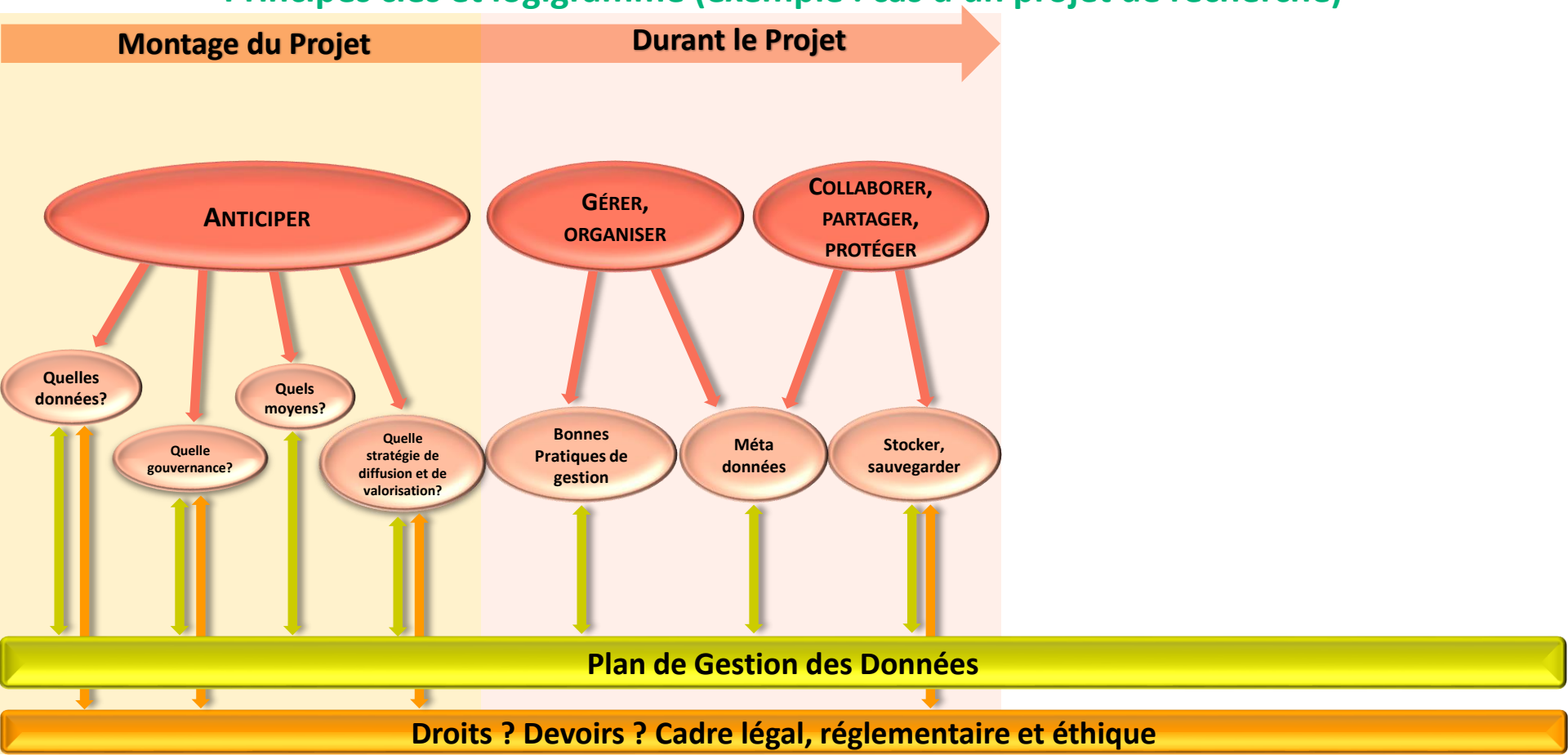
DATE DE PUBLICATION: mercredi 17 mars 2021 00:00:00

DATE DE DERNIÈRE MODIFICATION: lundi 5 février 2024 00:00:00

Une petite pause?

AgroParisTech 

Principes clés et logigramme (exemple : cas d'un projet de recherche)



Stratégie de stockage des données

Quelques définitions:

Stockage d'une donnée numérique : désigne l'ensemble des méthodes et technologies permettant de conserver des données numériques

Données vivantes : elles changent régulièrement et sont la plupart du temps des données de production (projet, expérimentation, collaboration)

Données mortes : elles ne sont (quasiment) plus modifiées (données d'archives)

Données brutes : données non travaillées et interprétées (ex: données produites par des instruments). On travaille toujours sur des copies de données brutes.

Stocker c'est bien mais pas suffisant. Il faut :

Il faut avoir **recensé** les données stockées, pouvoir les **retrouver**, pouvoir **y accéder** rapidement, et pouvoir les **lire**.
=> un enchaînement nécessaire sinon le stockage est inutile !

L'accès à une donnée numérique se fait TOUJOURS à travers un ou des logiciels compatibles avec le format des données à lire :

- de la simple commande (type edit, more,...)
- ou de logiciels (Arcview, OMERO, Excel, R,)
- jusqu'au LIMS (plateforme numérique intégrée de gestion de laboratoire).

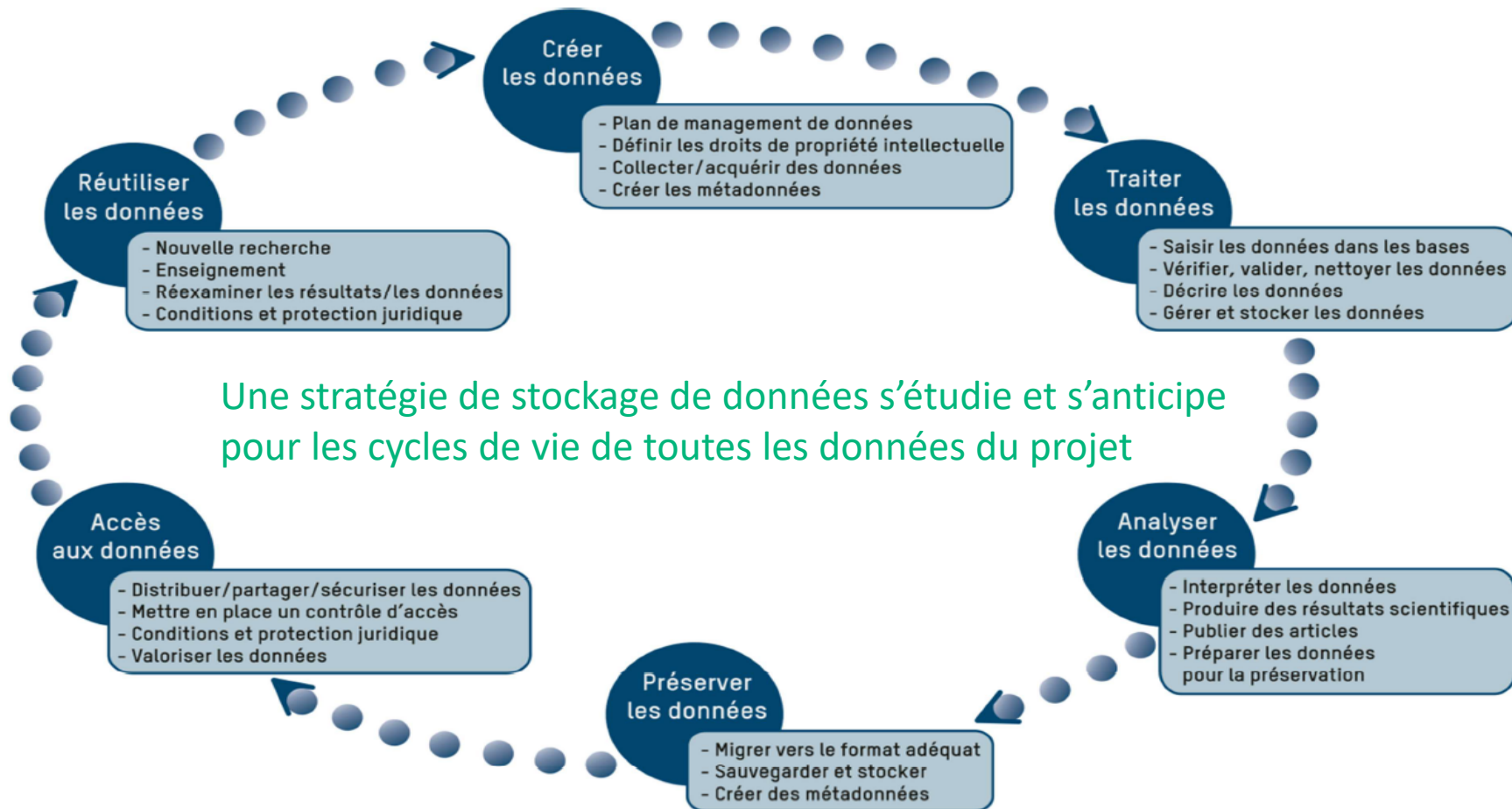
Les grands types de support de stockage de données au format numérique :

- Local : clés USB, ordinateurs, appareils de laboratoire
- Local sécurisé : disques RAID en DAS
- Les espaces de stockage réseau internes de type NAS (Novell, Samba, Windows server...)
- Les clouds privés (Seafile, Teams internes, Owncloud...)
- Les entrepôts de données
- Les clouds publics (Azure, Google, AWS...)

Le coût d'acquisition et de fonctionnement, les modalités d'accès, l'impact écologique, les problématiques de sécurité, ... sont très différents d'une solution de stockage à une autre.

Le cycle de vie des données

39



Elaborer la stratégie de stockage démarre par la question « magique » à se poser pour chaque étape du cycle de vie des données:

=> Le projet a besoin de

Quel(s) volume(s) de données

Pour stocker quel type de données

Selon quelles modalités d'accès

Pour quelle durée

Avec quel niveau de sensibilité

Selon quels modèles économiques et écologiques

Avec quelles contraintes réglementaires et contractuelles

Les réponses permettent de sélectionner des types de stockage candidats : existants ou à prévoir.

Définir cette stratégie ne veut pas dire qu'une réponse pourra y être apporter de manière simple et rapide

=> *l'anticipation est la meilleure façon de garantir la faisabilité de cette stratégie*

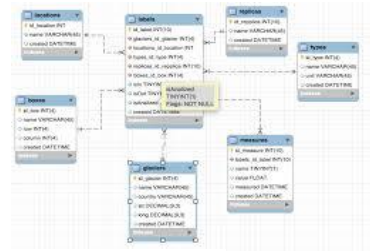
Les volumes de données:

- < 1 To: nombreuses solutions possibles
- Entre 1To et 10 To: à étudier en amont
- > 10 To: à anticiper très en amont, selon les besoins définis par la stratégie de stockage

Les types de données:

- les fichiers plats
- les bases de données
- les données non structurées :

A	B	C	D	E	F
1 ANNÉE	Président	Vice-président	Secrétaire	Treasorier	Administrateur
2 1988	Michel Sardou	Bruno Pelletier	Daniel Lavoie	Johnny Hallyday	Gilles Vigneault
3 1989	Michel Sardou	Daniel Balavoine	Bruno Pelletier	Johnny Hallyday	Gilles Vigneault
4 1990	Michel Sardou	Daniel Balavoine	Bruno Pelletier	Gilles Vigneault	Jacques Brel
5 1991	Michel Sardou	Joe Dassin	Bruno Pelletier	Gilles Vigneault	Jacques Brel
6 1992	Michel Sardou	Joe Dassin	Bruno Pelletier	Gilles Vigneault	Jacques Brel
7 1993	Luc Plamondon	Michel Sardou	Bruno Pelletier	Gilles Vigneault	Jacques Brel
8 1994	Luc Plamondon	Michel Sardou	Bruno Pelletier	Gilles Vigneault	Jacques Brel
9 1995	Joe Dassin	Isabelle Boulay	Bruno Pelletier	Gilles Vigneault	Michel Sardou
10 1996	Joe Dassin	Diane Dufréne	Roch Voisine	Gilles Vigneault	Michel Sardou
11 1997	Bruno Pelletier	Diane Dufréne	Patrick Bruel	Gilles Vigneault	Michel Sardou
12 1998	Bruno Pelletier	Diane Dufréne	Patrick Bruel	Gilles Vigneault	Michel Sardou



LinkedIn

facebook



Les modalités d'accès: comment accéder aux données et pour quel usage ?

Les accès :

- mono-utilisateur
- multi-utilisateurs sur un seul site, plusieurs sites, d'un même institut, de plusieurs instituts, indéfinis...
- accès manuel ou automatisé (via un programme)

Les usages :

- Collaboratifs: accès à plusieurs selon des droits différents (lecture, modification, création...)
- Traitements: calculs en local, calculs intensifs distribués (**attention à la performance d'accès aux données, vitesse et latence notamment**)

La durée nécessaire du stockage et **les contraintes réglementaires et contractuelles** sont souvent liées => RGPD, éléments de preuve, reproductibilité des expérimentations...

La durée dépend de la finalité du stockage et n'est pas forcément identique à la durée du projet de recherche.

Les questions de format et d'accessibilité des données impactent la durée de stockage.

La sensibilité des données s'évalue (a minima) selon deux axes: la **sûreté** et la **sécurité**

Quels sont les impacts d'une perte, altération, divulgation des données stockées ?

La sûreté vise à prendre en compte les risques de perte accidentelles des données.

La sécurité vise à prendre en compte les risques d'attaques sur les données en vue de les dérober, altérer, détourner ou nuire au bon déroulement du projet.

La stratégie de stockage doit prendre en compte ces aspects.

Les pays Européens sont de plus en plus la cible de cyberattaque, y compris dans les domaines de la recherche.

Les solutions propres à AgroParisTech:

Le cloud privé Seafile: permet le stockage et partage de bibliothèques, dossiers et fichiers sur des équipements centralisés et sauvegardés. 100 Go par utilisateurs du SI AgroParisTech. Espace de stockage dédié par département, service, UMR, projets sur demande et après étude de faisabilité.

Un projet de recherche peut contribuer à l'achat d'espace de stockage supplémentaire sur Seafile => relation gagnant - gagnant

Des solutions de stockage / archivage locales possible à travers l'achat de disques locaux (à la charge des unités mais pouvant être accompagnés par la DSI locale): disques sécurisées à connecter sur un ordinateur et permettant du stockage / archivage avec un coût financier et écologique faible.

Pas de solutions simples et disponibles pour des accès à des données structurées type bases de données ou logiciels. Les besoins doivent s'évaluer au cas par cas et donc être très anticipés.

En résumé, quelques bonnes pratiques du stockage d'un projet de recherche:

En amont:

- Identifier les cycles de vie des données du projet
- Poser la question magique pour chacune des étapes du cycle
- Identifier les solutions de stockage proposées par les différents partenaires susceptibles de convenir sur chacune des étapes du cycle
- Echanger avec les parties prenantes et établir le modèle économique du stockage (usage gratuit, investissement nécessaire, location, partenariats...)

Pendant la vie du projet:

- Évaluer l'adéquation entre les moyens de stockage alloués et les besoins
- Garder la maîtrise du stockage, de la cartographie des données (localisation, modalités d'accès) et de leur lecture (maîtrise des formats)

[Fiche pratique : Stratégie de stockage des données numériques](#)

En résumé, quelques bonnes pratiques du stockage d'un projet de recherche:

En amont:

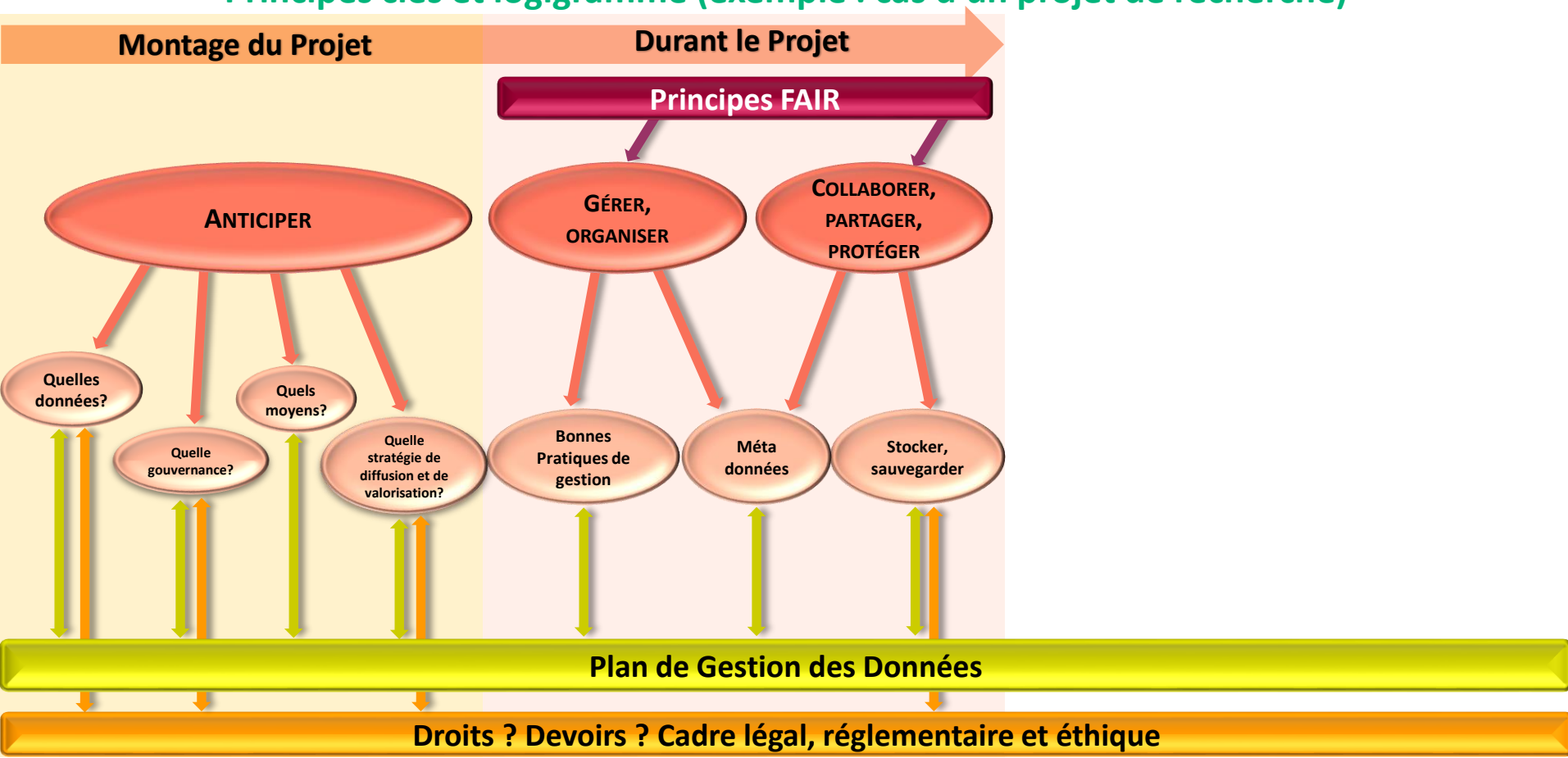
- Identifier tous les ensembles de données du projet
- Poser la question magique pour chacun
- Identifier les solutions de stockage proposées par les différents partenaires susceptibles de convenir pour chacun
- Echanger avec les parties prenantes et établir le modèle économique du stockage (usage gratuit, investissement nécessaire, location, partenariats...)

Pendant la vie du projet:

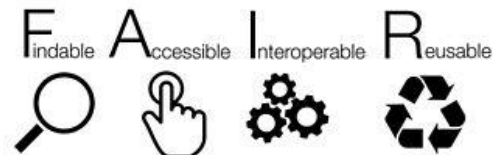
- Évaluer l'adéquation entre les moyens de stockage alloués et les besoins
- Garder la maîtrise du stockage, de la cartographie des données (localisation, modalités d'accès) et de leur lecture (choix des formats, gestion des droits)

[Fiche pratique : Stratégie de stockage des données numériques](#)

Principes clés et logigramme (exemple : cas d'un projet de recherche)



Principes FAIR



Qu'est-ce que c'est ?

Encore un pense-bête! 4 ensembles de précautions à prendre pour que les DR soient Faciles à trouver, Accessibles, Interopérables et Réutilisables

À quoi ça sert ?

Standardiser (formats libres et ouverts, vocabulaires, unités...), documenter (métadonnées), tracer (identifiant pérenne...)

 **Principes FAIR ≠ ouverture obligatoire : des données FAIR peuvent être confidentielles**

Les gros avantages ?

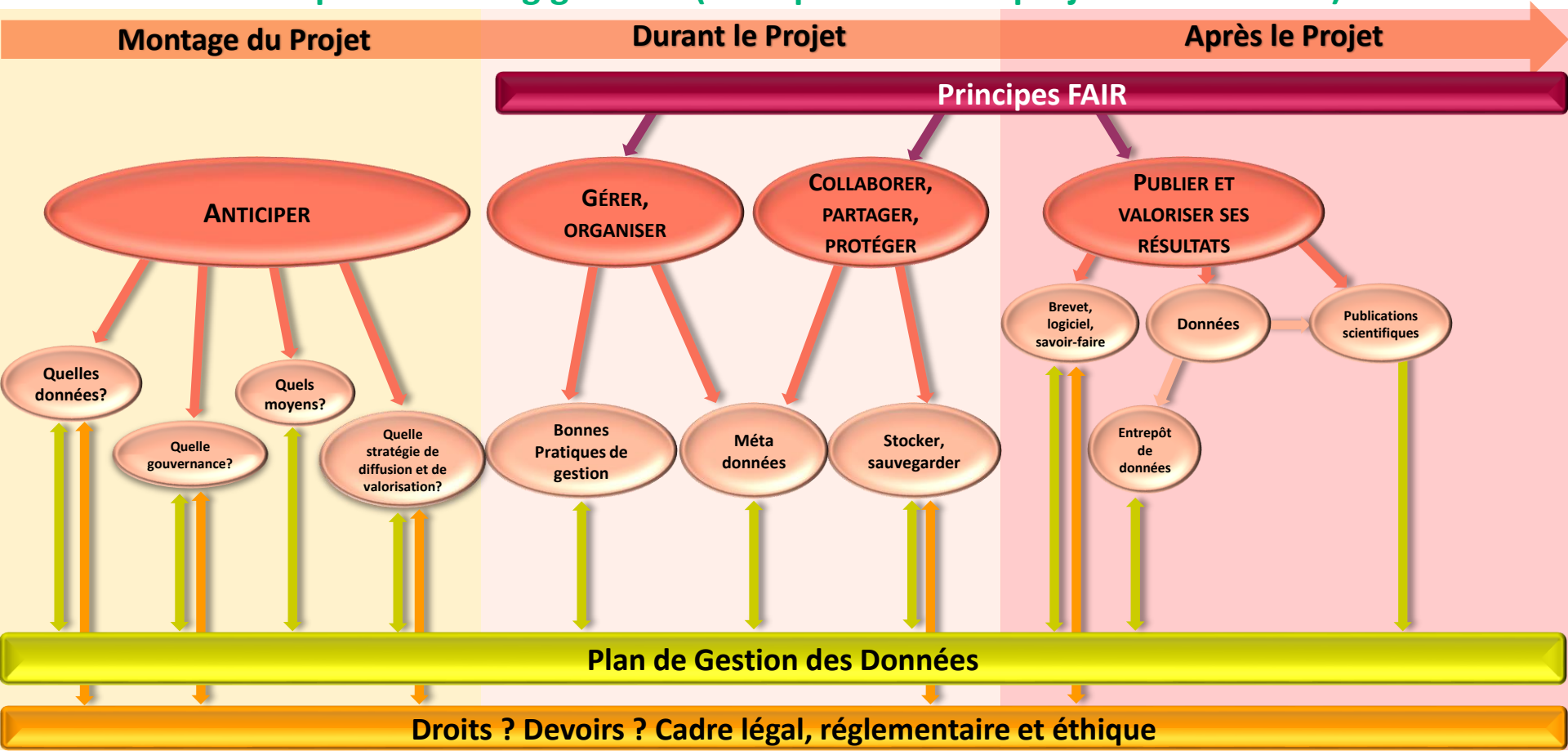
Facilite la conservation, la traçabilité, la publication, les collaborations pour l'ouverture et même avant !

Les précautions à prendre ?

Respecter les standards

[Fiche pratique : Principes FAIR](#)

Principes clés et logigramme (exemple : cas d'un projet de recherche)



Entrepôts de Données

Outils clés en main pour publier ses données = base de données en ligne + principes FAIR



Publier ≠ Partager

Publier = Ouvrir durablement, efficacement et de façon juridiquement cadrée ses DR à une communauté ciblée ou plus large

Tout est guidé et cadré : grilles de métadonnées, formatage des données, attribution d'identifiants pérennes (DOI, PURL...), outils de recherche et de découverte...

- Choisir son entrepôt : Thématique/Généraliste? Popularité? Durabilité? Certification? (ex CoreTrustSeal)
- S'assurer de son droit à ouvrir ses données
- Lier ses données à ses publications (via identifiants pérennes)

Annuaire d'entrepôts : [re3data](#), [FairSharing](#), [OAD](#), [EOSC](#), [OpenDOAR](#), [MESRI](#)...

Exemples d'entrepôts : **disciplinaires** : [GBIF](#) (biodiversité), [GenBank](#) (génétique), [Nakala](#) (SHS) ; **généralistes** : [Recherche Data Gouv](#), [Dryad](#) ; **institutionnels** : [Cirad DataVerse](#), [IRD DataSuds](#)...

[Fiche pratique : Entrepôts de données](#)

Portail Cirad du libre accès

Retrouvez l'espace des données de recherche du Cirad dans la Collection Cirad sur l'entrepôt national fédéré des données de la Recherche → [recherche.data.gouv](https://recherche.data.gouv.fr/)

Dataverse

Search ▾ User Guide Support Log In



UMR EcoFoG (AgroParisTech, CIRAD, CNRS, INRAE, Univ Antilles, Univ Guyane)

Ecologie des Forêts de Guyane

[CIRAD Dataverse >](#)[✉ Contact](#) [🔗 Share](#)

The Joint Research Unit Ecology of Guianan Forests (EcoFoG) gathers AgroParisTech, INRAE, Cirad, CNRS, the University of the French West Indies and the University of French Guiana. The purpose of the unit is to integrate approaches in ecology and material sciences to understand the relations between diversity and ecosystem functioning, considering the effects of logging, anthropic pressure and global change and to impulsive innovation to valorize forest resources, taking into account its large diversity and the particular constraints of moist tropical environment.

Search this dataverse...

Find

[Advanced Search](#) [Dataverses \(5\)](#) [Datasets \(15\)](#) [Files \(50\)](#)**Dataverse Category**[Research Group \(5\)](#)**Publication Year**[2022 \(4\)](#)[2021 \(1\)](#)

1 to 5 of 5 Results

Sort ▾

GuyaforONF (ONF)

24 oct. 2022 Guyafor



ONF permanent plots included in the Guyafor network.

**Paracou (CIRAD)**

20 oct. 2022 GuyaforCIRAD



The Paracou Research Station is Cirad's primary site for the study of the long-term effects of forest disturbance in Amazonia.

**GuyaforCIRAD (CIRAD)**

20 oct. 2022 Guyafor

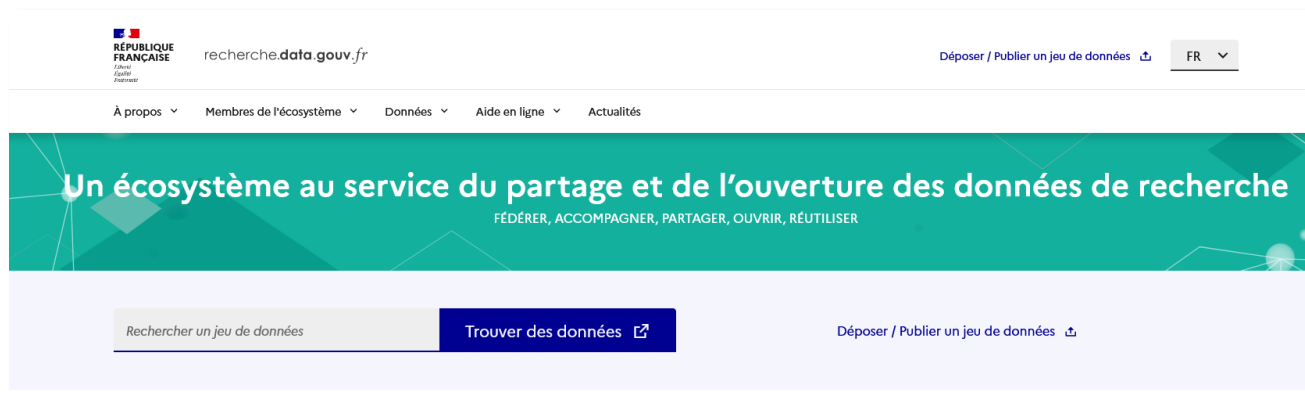


CIRAD permanent forest plots included in the Guyafor network.



Exemple d'un [entrepôt de données d'une UMR](#) sur un DataVerse institutionnel

Focus sur une solution nationale : Recherche Data Gov



- Ouvert en juillet 2022 : projet piloté par INRAE
- Solution technique : Dataverse (similaire à DataINRAE)
- Inclut des espaces institutionnels pour plusieurs ONR et universités
- Objectifs : accueillir les jeux de données qui n'ont pas d'entrepôt thématique reconnu ; recenser les jeux de données de la recherche française
- Classes virtuelles : <https://recherche.data.gouv.fr/fr/page/classes-virtuelles>
- Associé à une structuration en cours de l'accompagnement DR : ateliers de la donnée

Entrepôts de Données et financeurs



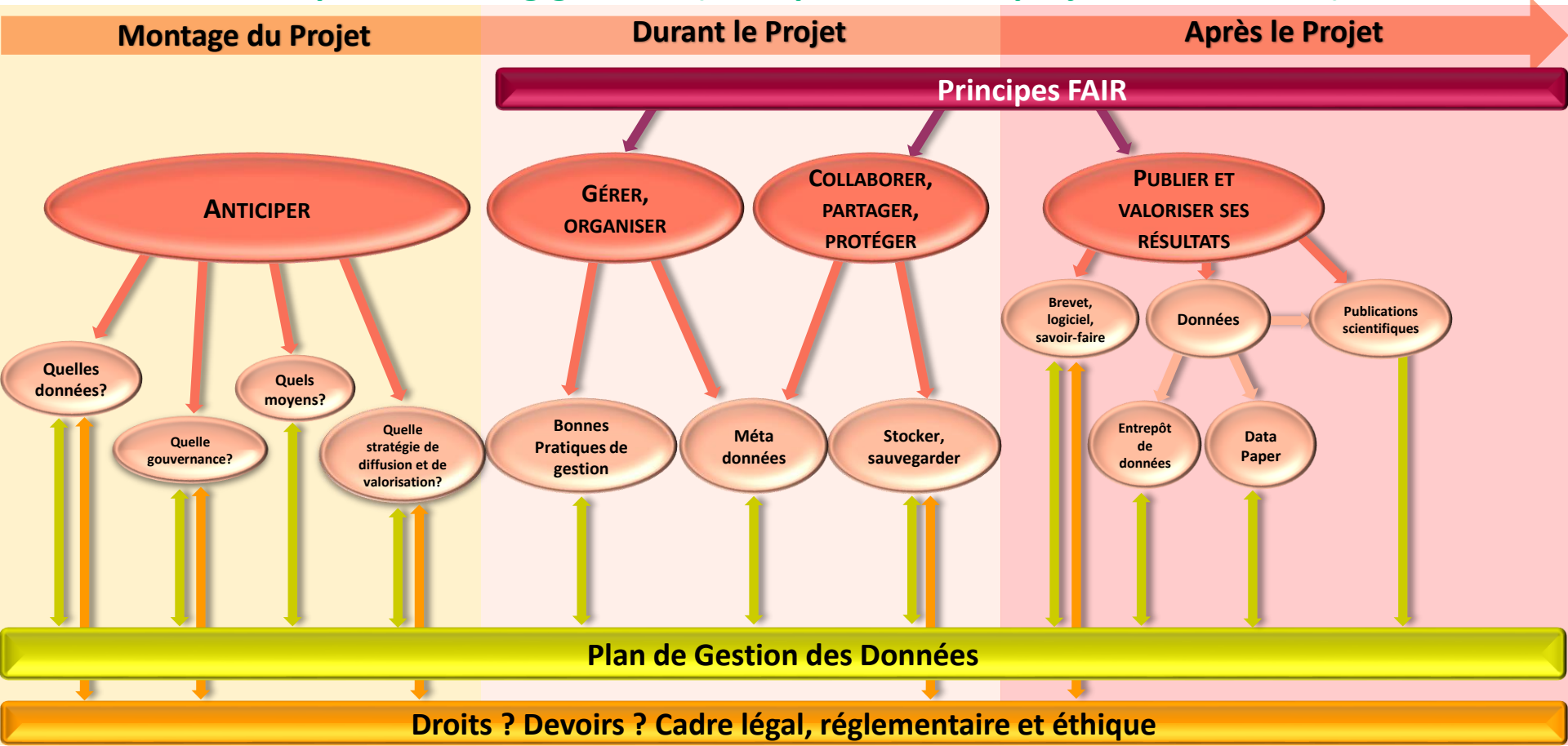
- Ouverture des données = principe par défaut ; justifier les cas de non-ouverture (exceptions autorisées, à préciser dans le PGD)
 - Licence CC-BY ou CC-0 (ou équivalent)
 - Demande de dépôt des données dans un entrepôt de confiance, en précisant dans les métadonnées les outils nécessaires à la réutilisation/validation des données (logiciels, etc.)
 - Accès numérique ou physique obligatoire pour les données nécessaires à la validation des conclusions présentées dans des publications.
 - Coûts relatifs à la gestion des données éligibles.
- Ouverture des données = recommandation, pas d'obligation à ce stade. Application du principe « aussi ouvert que possible, aussi fermé que nécessaire ».
 - Coûts relatifs à la gestion des données éligibles.

Plus d'informations :

[Europe](#)

[ANR](#)

Principes clés et logigramme (exemple : cas d'un projet de recherche)



Article de Données ou Data Paper

Article qui décrit précisément un jeu de données : genèse précise et complète, potentiel de réutilisation, accès aux données. MAIS pas de résultats scientifiques, d'analyses ou d'interprétations ; peut être publié en association avec un article scientifique classique pour cela.

- Informe la communauté de l'existence et de la valeur de ces données
- Synthétise et simplifie la communication des données
- Complète le dépôt en entrepôt, connectable à un article scientifique
- Valorise le temps passé à gérer les données

Les précautions à prendre :

- Choix du journal (thématique, audience, comité de lecture...)
- Choix de l'entrepôt (compatibilité)
- Respect des consignes du financeur

Ressources : listes de revues spécialisées publiées par [INRAE](#), le [Cirad](#), le [GBIF](#)...

Exemple d'un [DataPaper](#) et du [dépôt de données](#) correspondant (GBIF):

SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

Fungi of French Guiana gathered in a taxonomic, environmental and molecular dataset

Gaëlle Jaouen¹, Audrey Sagne², Bart Buyck³, Cony Decock⁴, Eliane Louisanna², Sophie Manzi⁵, Christopher Baraloto⁶, Mélanie Roy⁵ & Heidy Schimann²

In Amazonia, the knowledge is particularly the case in coastal localities. Here, v particularly focusing on t with a stratified and r...

Fungi of French Guiana

Published by [Unité Mixte de Recherche EcoFoG \(Ecologie des Forêts de Guyane\)](#)

Schimann H • Roy M • Jaouen G

[DATASET](#) [PROJECT](#) [METRICS](#) [ACTIVITY](#) [DOWNLOAD](#)

5,219 OCCURRENCES

33 CITATIONS

Fungi of French Guiana: a dataset to gather taxonomic, environmental and molecular data thanks to a comprehensive, stratified and reproducible sampling scheme in 126 plots positioned representative of the main forests habitats

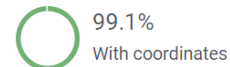
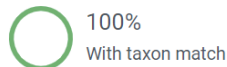
Publication date: April 18, 2019

Metadata last modified: April 18, 2019

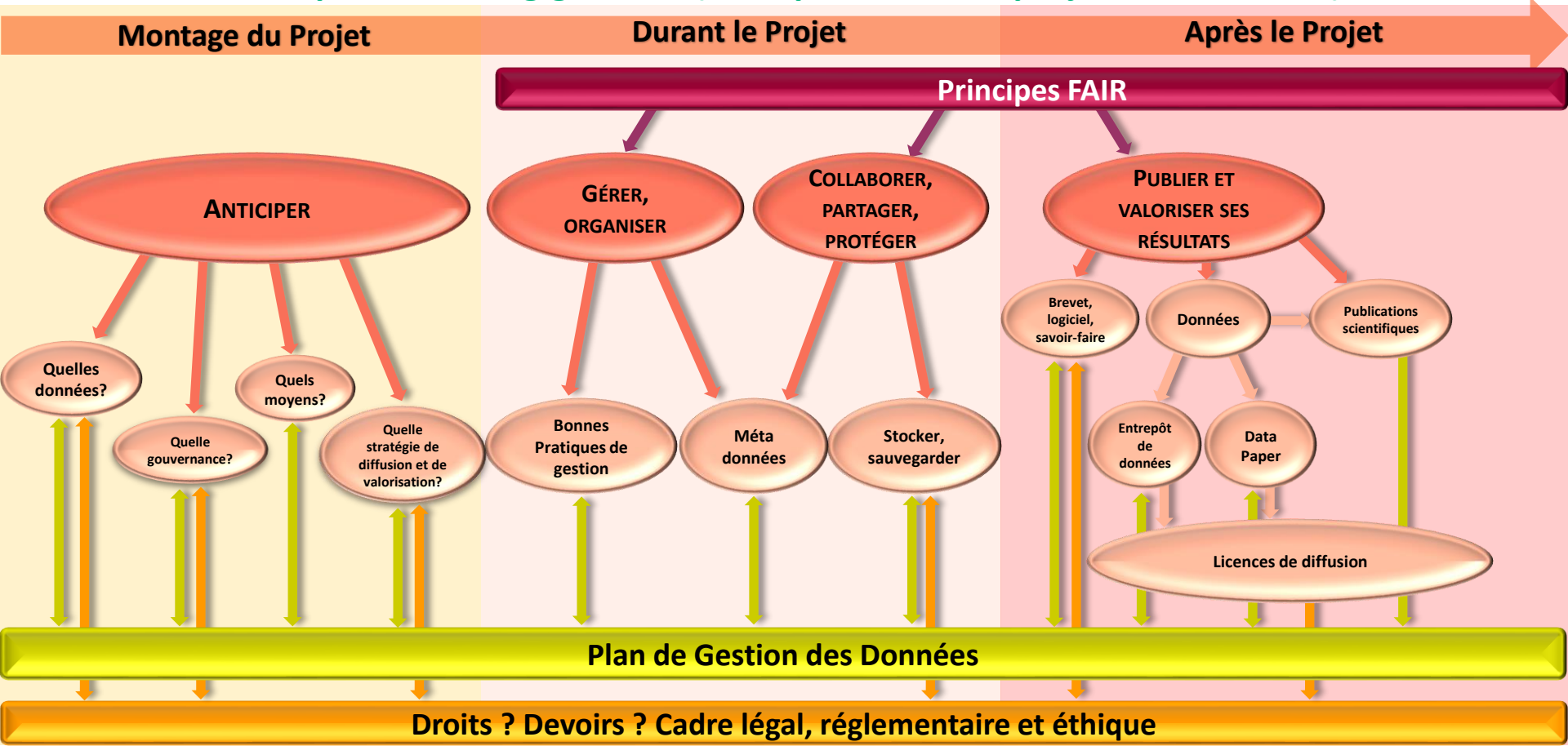
Hosted by: GBIF France

Licence: CC BY 4.0

[How to cite](#) [DOI](#) 10.15468/ymvlrp



Principes clés et logigramme (exemple : cas d'un projet de recherche)



Licences de Diffusion

Qu'est-ce que c'est ?

Outil juridique pour spécifier publiquement les formes de diffusion et de réutilisation autorisées par le détenteur des droits d'un contenu

À quoi ça sert ?

A choisir ce qu'on autorise une fois pour toutes avec les DR publiées (reproduction, réutilisation, transformation...)

Les gros avantages ?

Synthétise les conditions, gratuit, facile d'utilisation. Protège les producteurs de données.

Les précautions à prendre ?

Vérifier les particularités des DR, les accords des partenaires, les conditions demandées par les financeurs du projet et les journaux de DataPapers

Licences courantes : [Creative Commons](#) (ex: ).

Pour le droit français : <https://www.data.gouv.fr/fr/licences>

[Fiche pratique : Licences de diffusion](#)

Exemples de licences sur des dépôts de données:

Fungi of French Guiana

Published by [Unité Mixte de Recherche EcoFoG \(Ecologie des Forêts de Guyane\)](#)

Schimann H • Roy M • Jaouen G

DATASET PROJECT METRICS ACTIVITY  DOWNLOAD

5,219 OCCURRENCES

Fungi of French Guiana: a dataset to gather taxonomic, environmental and molecular data thanks to a comprehensive, stratified and reproducible sampling scheme in 126 plots positioned representative of the main forests habitats

Publication date: April 18, 2019


Metadata last modified: April 18, 2019


Hosted by: GBIF France

Licence: CC BY 4.0

 How to cite  DOI 10.15468/ymvlrp

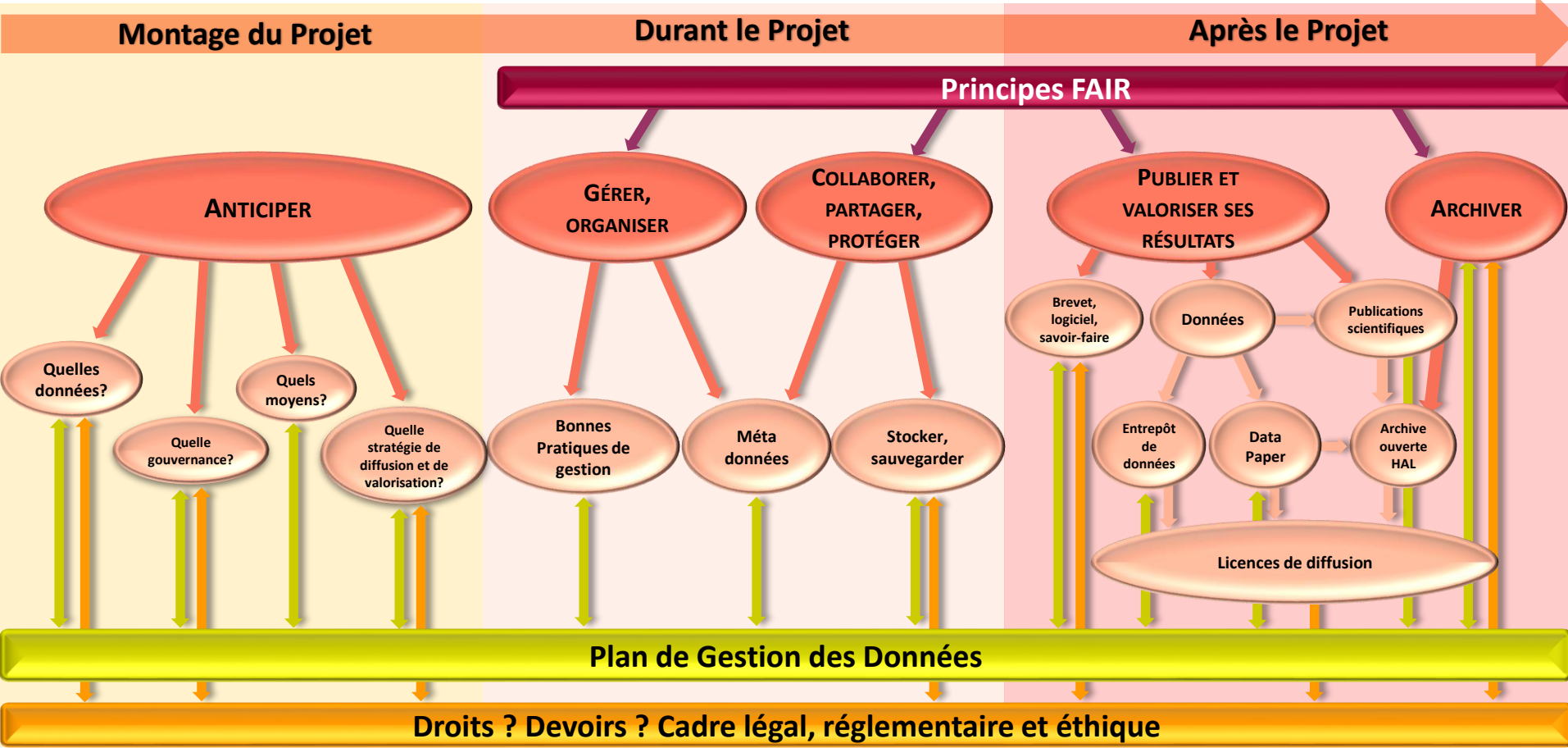
 5,219
Occurrences

 100%
With taxon match

 99.1%
With coordinates

 100%
With year

Principes clés et logigramme (exemple : cas d'un projet de recherche)



De l'ouverture à la naissance d'un nouveau projet : Rechercher des données

Pourquoi ?

Généralement, le chercheur produit ses propres données, néanmoins la recherche de données équivalentes peut vous faire gagner du temps au lieu d'en collecter de nouvelles

- > Pour répondre à d'autres questions de recherche,
- > Pour la participation à des méta-analyses,
- > Pour créer ou pour valider des modèles,
- > Pour réévaluer ou dupliquer des résultats,
- > Pour optimiser des données produites dans un cadre éthique strict (exemple : experimentation animale)
- > Pour apporter de la crédibilité et des arguments à votre travail.

Rechercher des données

Les sources peuvent être diverses ; 4 voies principales se dégagent néanmoins :

1. La littérature scientifique :

La recherche bibliographique d'articles peut amener à trouver des données dans les articles eux-mêmes ou les annexes d'un article ("supplementary materials, files" ou "data papers")

Les données peuvent aussi être publiées séparément. Dans ce cas, l'article mentionnera la référence du jeu de données publiée, généralement accessible dans un répertoire de données.

2. Entrepôt de données (*Data repositories*)

C'est une base de données qui permet de déposer, conserver et rendre accessible des jeux de données standardisés, décrits par des métadonnées riches.

L'entrepôt assure ainsi **l'accessibilité, la traçabilité, la pérennité.**

Rechercher des données

3. Portail de données d'organisations

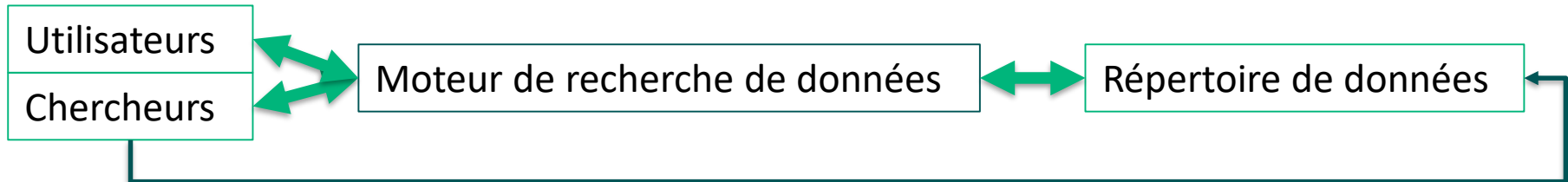
Les organismes qui récoltent régulièrement des données (ex : statistiques) peuvent offrir leurs données par leur propre portail.

Ex : Eurostat, gouvernement français : <https://data.gouv.fr>

4. Moteurs de recherche de données (*Data search engines*)

Un moteur de recherche vous permet de trouver et naviguer rapidement entre les jeux de données ou le “matériel d’accompagnement” publiés par les chercheurs.

Il en existe de nombreux.



Rechercher des données






La description des métadonnées, au service de la réutilisation des données

Pour assurer la compréhension et la réutilisabilité éclairée des DR, conserver les informations de genèse des DR, faciliter les collaborations

- > Conserver la mémoire du contexte de production
- > Permettre le travail collectif dans le cadre du projet ou au-delà
- > Faciliter le partage technique et l'interopérabilité des données

Quelques outils ...

Datacite


 Works  People  Organizations  Repositories

Search works by keyword(s) or DOI.

Search people by name, keyword(s), or ORCID iD.

Search organizations by organization name, keyword(s), or ROR ID.

Search repositories by repository name or keyword(s).

Quelques services : dispose d'APIs, fonctionne avec Crossref (demande ou recherche de DOI), possibilité d'implémenter son ORCID (comme Hal)

Re3data : découvrir des entrepôts de données (cf. datacite repositories)

Filter

- Subjects ⊕
- Content Types ⊕
- Countries ⊕
- AID systems ⊕
- API ⊕
- Certificates ⊕
- Data access ⊕
- Data access restrictions ⊕
- Database access ⊕
- Database access restrictions ⊕
- Database licenses ⊕
 - Apache License 2.0 (38)
 - BSD (12)
 - CC (149)
 - CC0 (72)
 - Copyrights (139)
 - ODC (13)
 - Public Domain (23)
 - other (146)
- Data licenses ⊕
 - Apache License 2.0 (47)
 - BSD (35)
 - CC (984)
 - CC0 (203)

Facette

Navigation


Browse by subject

Browse by content type

Browse by country

Syntaxe

Search...

 Search

[Toggle short help](#)

- * at the end of a keyword allows wildcard searches
- " quotes can be used for searching phrases
- + represents an AND search (default)
- | represents an OR search
- represents a NOT operation
- (and) implies priority
- ~N after a word specifies the desired edit distance (fuzziness)
- ~N after a phrase specifies the desired slop amount

[Recherche.data.gouv.fr](https://recherche.data.gouv.fr)

Chercher dans cette collection...



Recherche avancée

  Collections (793)  Jeux de données (3 805)  Fichiers (47 108)

1 à 10 de 4 598 résultats

↑↓ Tri ▾

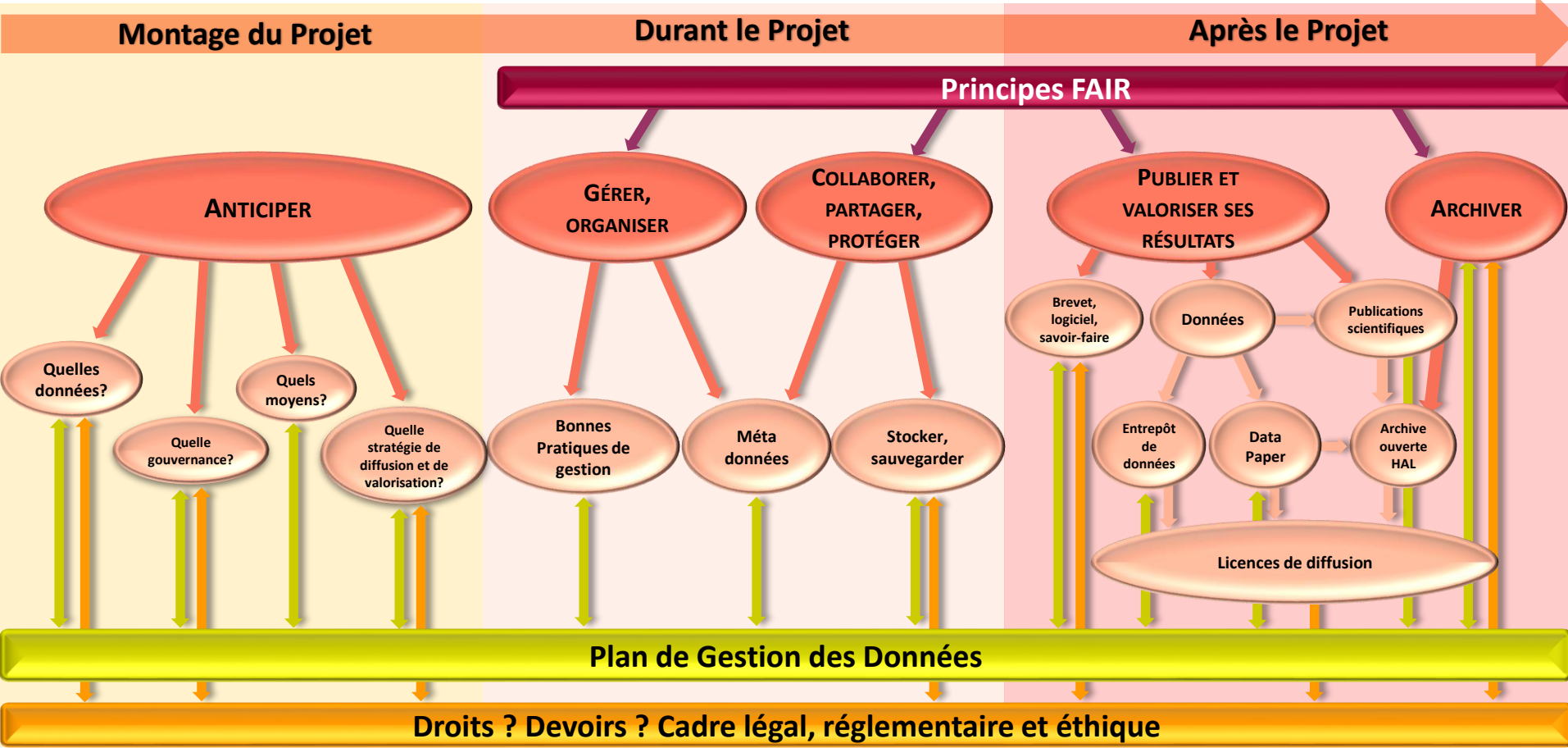
 LASIE (La Rochelle Université)
27 mars 2024 La Rochelle Université

Vers une convergence des entrepôts de données en France

Basé sur le logiciel libre « dataverse »

Entrepôt bac à sable (pour les tests) : <https://demo.recherche.data.gouv.fr/>

Principes clés et logigramme (exemple : cas d'un projet de recherche)



Politique DR à AgroParisTech

- **Formalisation de la gestion des données** aux niveaux projet et unité ; PGD vivement recommandé
- Mise en application des **principes FAIR**
- **Formalisation des procédures de stockage et d'archivage**
- **Ouverture par défaut des données produites par l'établissement**, sous réserve de l'absence d'un droit contraire ; attention particulière apportée aux **données utilisées dans les publications** (*limiter le syndrome « data upon request »*)
- **Anticipation systématique des questions liées à l'ouverture ou à la confidentialité** des données dans tous les projets de recherche
- **Choix de l'entrepôt** : privilégier les entrepôts thématiques de référence de sa discipline ; en 2^e choix, les entrepôts des partenaires au sein des UMR (Data Inrae, etc.) ; en dernière alternative, les entrepôts généralistes (Recherche Data Gouv, etc.). Critères complémentaires à prendre en compte (certification, gouvernance, etc.)

Politique DR à AgroParisTech

- **Encouragement à la rédaction de Data Papers** dès que pertinent.
- **Recours à des données ouvertes pour nourrir les supports pédagogiques**, exercices pratiques, etc. quand pertinent.
- Prise en compte de la question des DR dans les conventions de stage étudiants.
- **Demande de rédaction de PGD pour les doctorants** et d'ouverture/restitution des données produites au cours de la thèse en l'absence de barrière réglementaire.
- **Encouragement à la réutilisation de données ouvertes.**
- Anticipation et attention aux enjeux de sécurité et de calendrier de diffusion en cas de **données menant à valorisation économique.**

Moyens :

- Formation
- Accompagnement
 - ⇒ cellule données de la recherche d'AgroParisTech : donnees-recherche@agroparistech.fr
 - ⇒ En savoir plus : [cliquez ici.](#)
 - ⇒ Réseau de référents données de la recherche dans les unités
 - ⇒ En savoir plus : [cliquez ici.](#)
- Prise en compte dans les indicateurs de suivi de l'activité de recherche
- Politique d'ouverture associée aux outils de soutien à la recherche.

[Les Données de la Recherche à AgroParisTech](#)

Ressources complémentaires :

[Fiches pratiques](#) AgroParisTech :

[Cadre juridique données de la recherche](#) ; [Cadre juridique RGPD Recherche](#) ; [Plan de gestion de données \(PGD\)](#) ; [Modèles PGD AgroParisTech](#) ; [Bonnes pratiques de gestion](#) ; [Métadonnées](#) ; [Stratégies de stockage des données numériques](#) ; [Principes FAIR](#) ; [Entrepôts de données](#) ; [Data paper](#) ; [Licences de diffusion](#)

[Dorandum](#)

[Inrae DataPartage](#)

[CoopIST](#) Cirad

Des questions ?



**Pour aller plus loin, place aux travaux
pratiques demain même heure!**