Science ouverte et ouverture des données

Groupe données :

Bernard Émilie -> Documentaliste

Hérault Frédéric -> Référent données opérationnel (RDO)

Moreews François -> Référent données opérationnel (RDO)

Mustière Colette -> Référente qualité

Virlet Gwendal -> Gestionnaire de données

RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

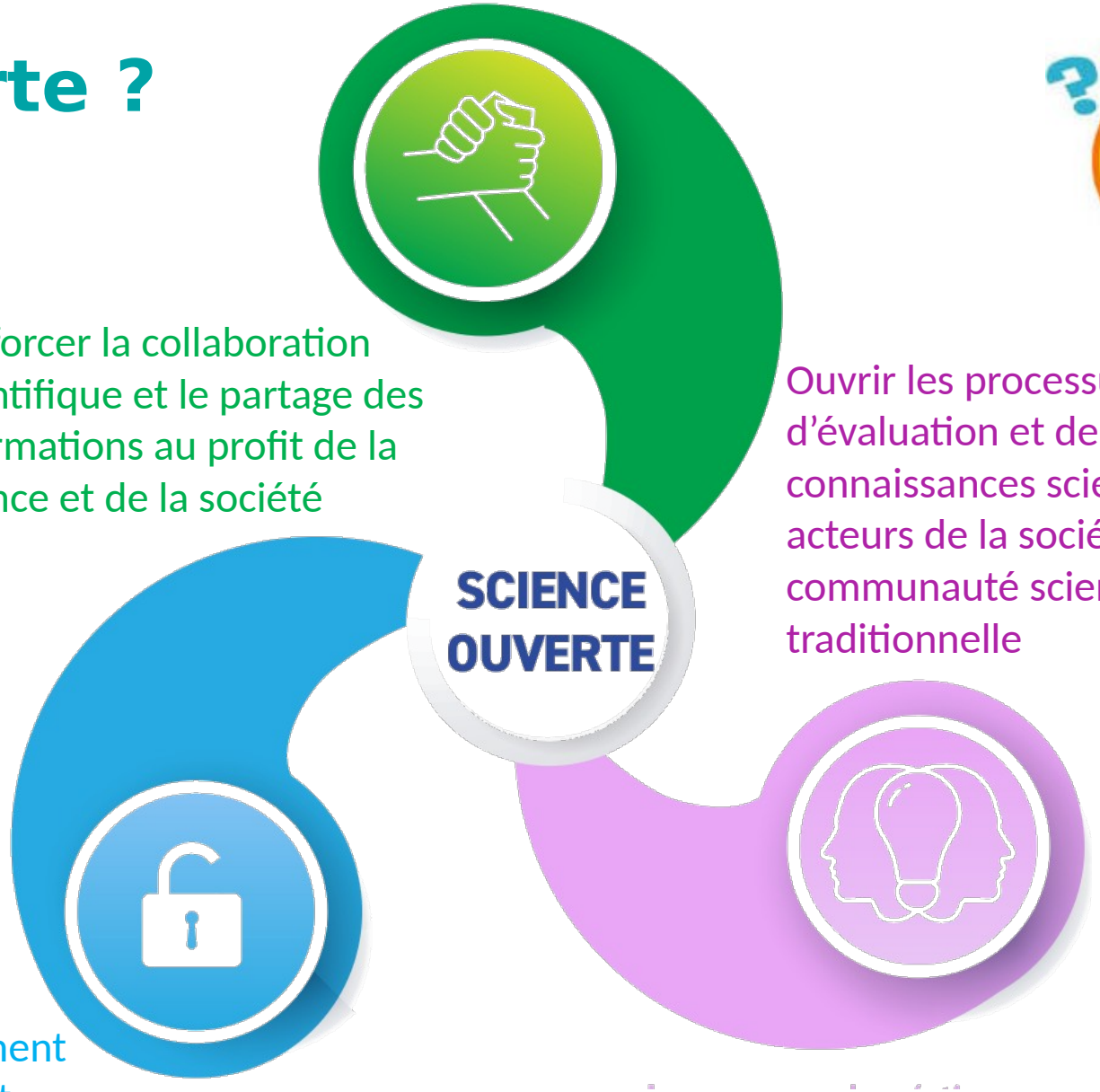
INRAE

LICENCE OUVERTE
OPEN LICENCE

La science ouverte ? Quésako ?

La science ouverte ?

Une définition de l'UNESCO.



Renforcer la collaboration scientifique et le partage des informations au profit de la science et de la société

Ouvrir les processus de création, d'évaluation et de diffusion des connaissances scientifiques aux acteurs de la société au-delà de la communauté scientifique traditionnelle

Rendre les connaissances scientifiques, librement accessibles à tous et réutilisables par tous

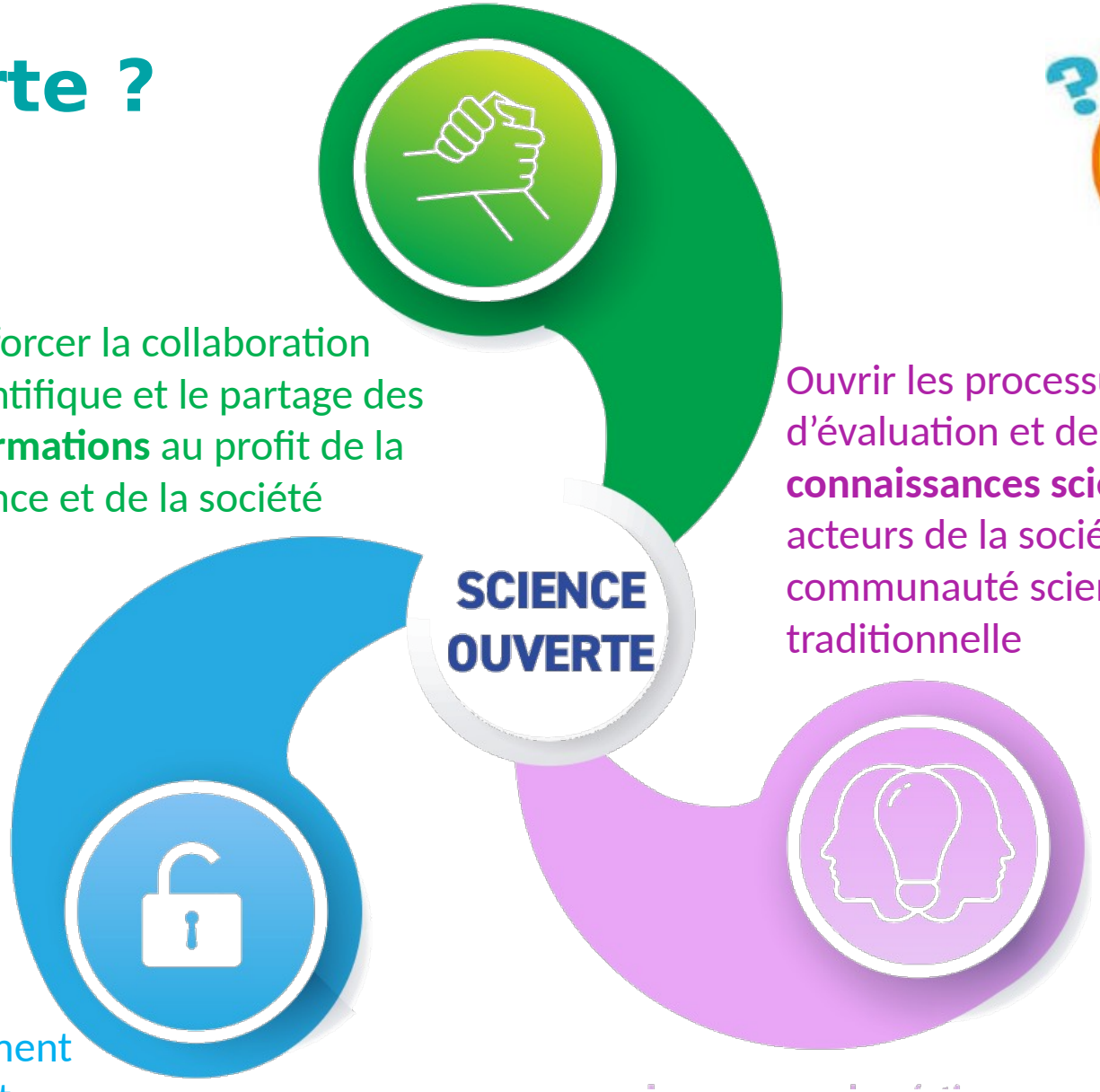
Recommandation de l'UNESCO sur une science ouverte (SC-PCB-SPP/2021/OS/UROS)
https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre consulté le 01/02/2024

La science ouverte ?

Une définition de l'UNESCO.

3 points importants

1 – Cela concerne **les connaissances scientifiques et les informations scientifiques (données)**



Renforcer la collaboration scientifique et le partage des **informations** au profit de la science et de la société

Ouvrir les processus de création, d'évaluation et de diffusion des **connaissances scientifiques** aux acteurs de la société au-delà de la communauté scientifique traditionnelle

Rendre **les connaissances scientifiques**, librement accessibles à tous et réutilisables par tous

Recommandation de l'UNESCO sur une science ouverte (SC-PCB-SPP/2021/OS/UROS)
https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre consulté le 01/02/2024

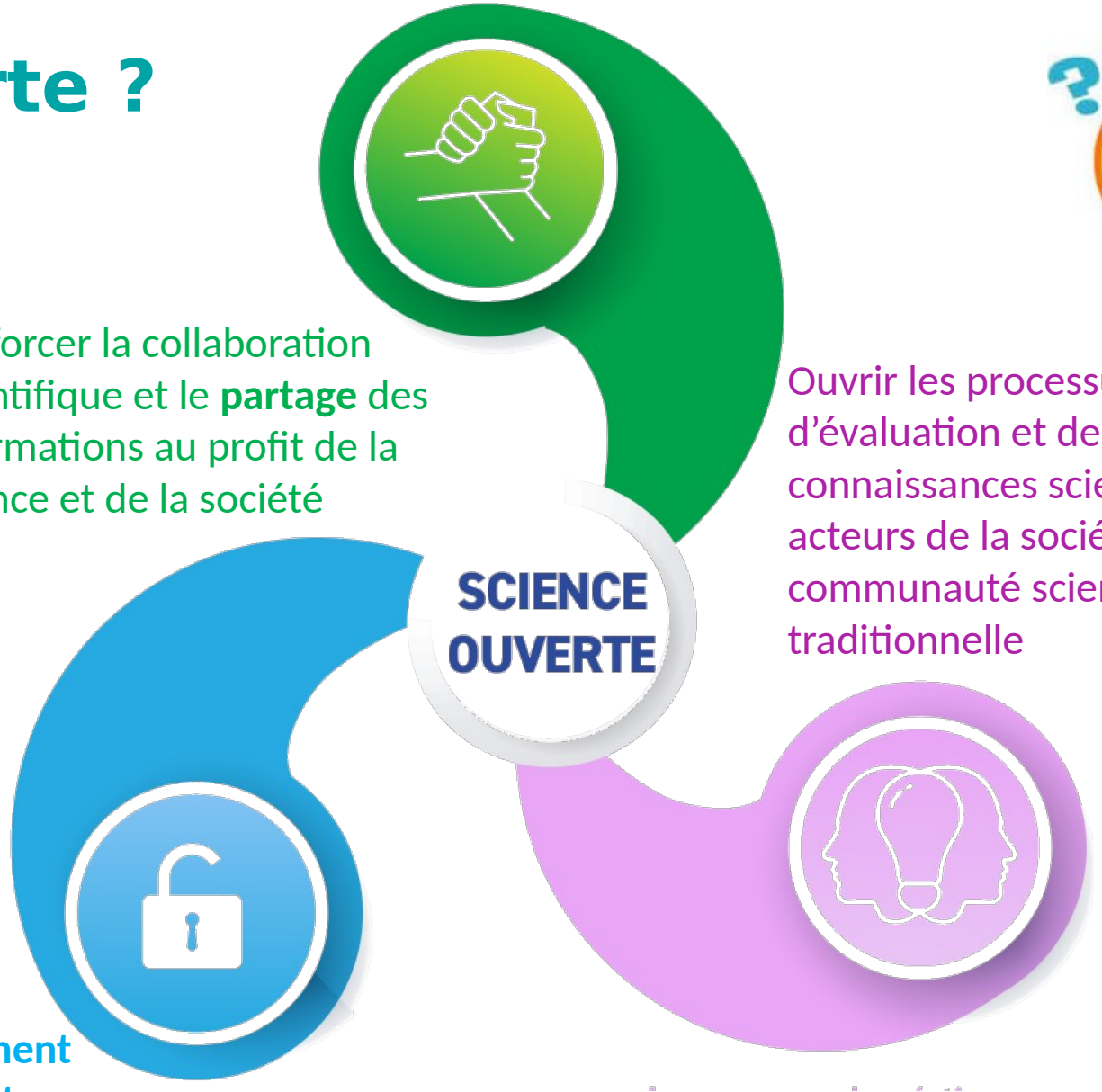
La science ouverte ?

Une définition de l'UNESCO.

3 points importants

1 – Cela concerne les connaissances scientifiques et les informations scientifiques (données)

2 – Ces connaissances doivent être **librement accessible, réutilisables, partagées, diffusées**



Renforcer la collaboration scientifique et le **partage** des informations au profit de la science et de la société

Ouvrir les processus de création, d'évaluation et de **diffusion** des connaissances scientifiques aux acteurs de la société au-delà de la communauté scientifique traditionnelle

Rendre les connaissances scientifiques, **librement accessibles** à tous **et réutilisables** par tous

Recommandation de l'UNESCO sur une science ouverte (SC-PCB-SPP/2021/OS/UROS) https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre consulté le 01/02/2024

La science ouverte ?

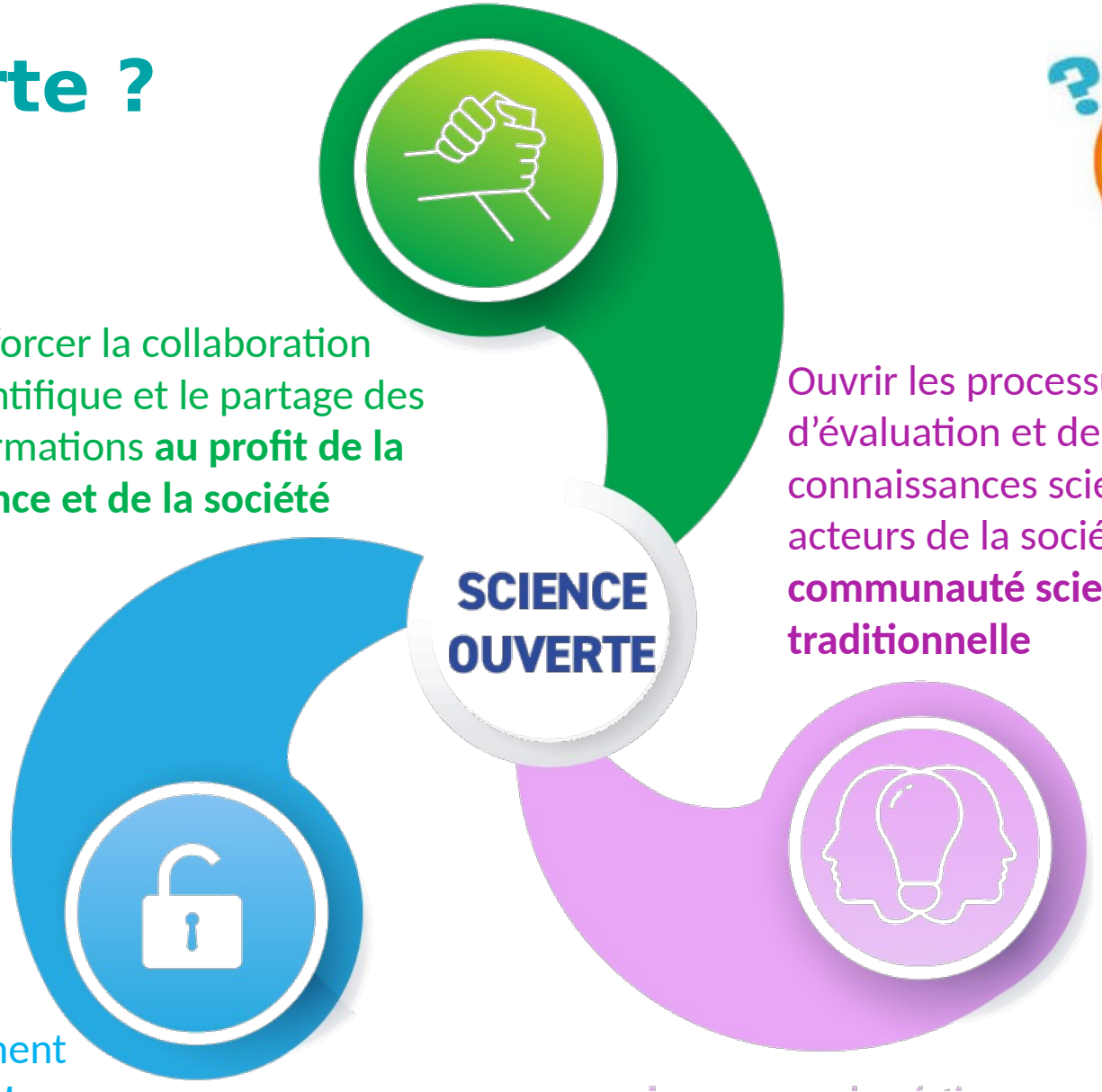
Une définition de l'UNESCO.

3 points importants :

1 – Cela concerne les connaissances scientifiques et les informations scientifiques (données)

2 – Ces connaissances doivent être librement accessible, réutilisables, partagées, diffusées

3 - **Au profit de tous ! « Au-delà de la communauté scientifique traditionnelle »**



Renforcer la collaboration scientifique et le partage des informations **au profit de la science et de la société**

Ouvrir les processus de création, d'évaluation et de diffusion des connaissances scientifiques aux acteurs de la société **au-delà de la communauté scientifique traditionnelle**

Rendre les connaissances scientifiques, librement **accessibles à tous et réutilisables par tous**

Recommandation de l'UNESCO sur une science ouverte (SC-PCB-SPP/2021/OS/UROS) https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre consulté le 01/02/2024

Quels en sont les éléments clés ?

- La science ouverte s'appuie entre autre sur :
  - Libre accès aux publications scientifiques en lignes, gratuites diffusion de la connaissance.
  - Libre accès aux données : facilite la vérification des résultats et la réutilisation des données.
  - Code source ouvert : étude et analyse du code, autorise la modification du code ou la création de codes dérivés.
  - Ressources éducatives libres : consultation, utilisation, adaptation et redistribution gratuites et sans restriction de matériels d'apprentissage, d'enseignement et de recherche.
  - Evaluation transparente et participative : identité des examinateurs, accès aux rapports d'évaluation, évaluation par la communauté au sens large (formuler des observations)…

Quel sont les objectifs de la Science Ouverte?

- Trois objectifs :
  - Accélérer les découvertes en maximisant l'accès à des données et à des logiciels fiables et solidement décrits.
  - Optimiser la réutilisation et la combinaison des données et des logiciels, qui sont des biens communs acquis grâce aux investissements publics (financement de projet ou d'infrastructure).
  - Favoriser l'appropriation du processus scientifique par tous.

L'objectif global est de favoriser l'accès et la réutilisation des produits de la recherche, et d'ouvrir le processus de recherche à d'autres acteurs de la société pour accélérer le processus de découverte.

L’ouverture des données, en quoi ça me concerne ?

Cadre légal et réglementaire

- Le libre accès et la libre réutilisation des documents et des données publiques sont la règle depuis 2016 (Loi pour une République Numérique)

  « Les administrations au sens large doivent publier gratuitement en ligne dans un standard ouvert et facilement réutilisable leurs principaux documents qu'elles produisent ou reçoivent , y compris leurs codes sources, ainsi que leurs bases de données. »

Cadre légal et réglementaire

- Le libre accès et la libre réutilisation des documents et des données publiques sont la règle depuis 2016 (Loi pour une République Numérique)

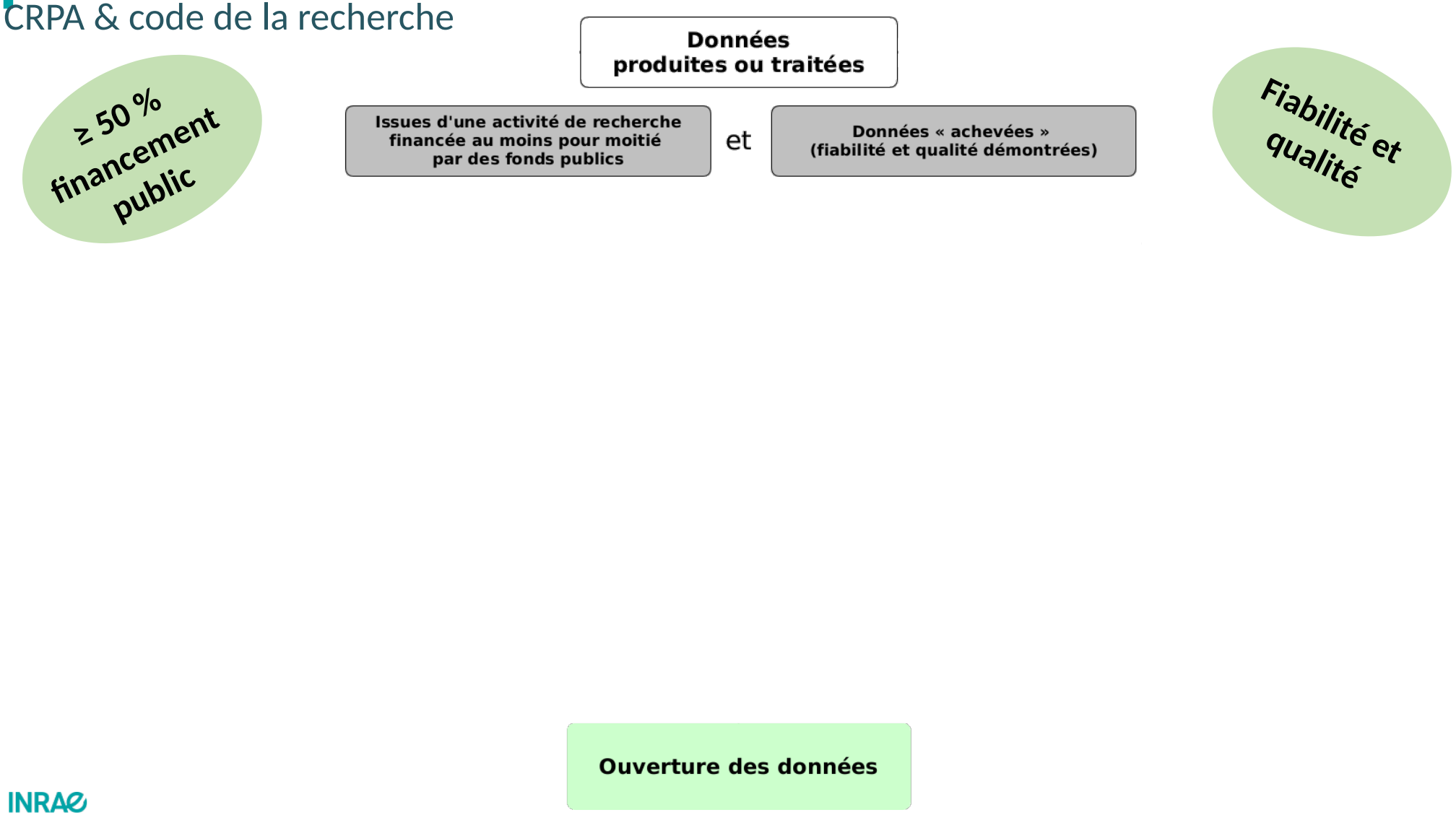
  « Les administrations au sens large doivent publier gratuitement en ligne dans un standard ouvert et facilement réutilisable leurs principaux documents qu'elles produisent ou reçoivent , y compris leurs codes sources, ainsi que leurs bases de données. »

- Les données de la recherche

  « Les enregistrements factuels (chiffres, textes, images, son, vidéo…) utilisés comme sources primaires pour la recherche et qui sont habituellement acceptés par la communauté scientifique comme étant nécessaires pour valider les résultats de la recherche »

Des données ouvertes, avec libre réutilisation par défaut !

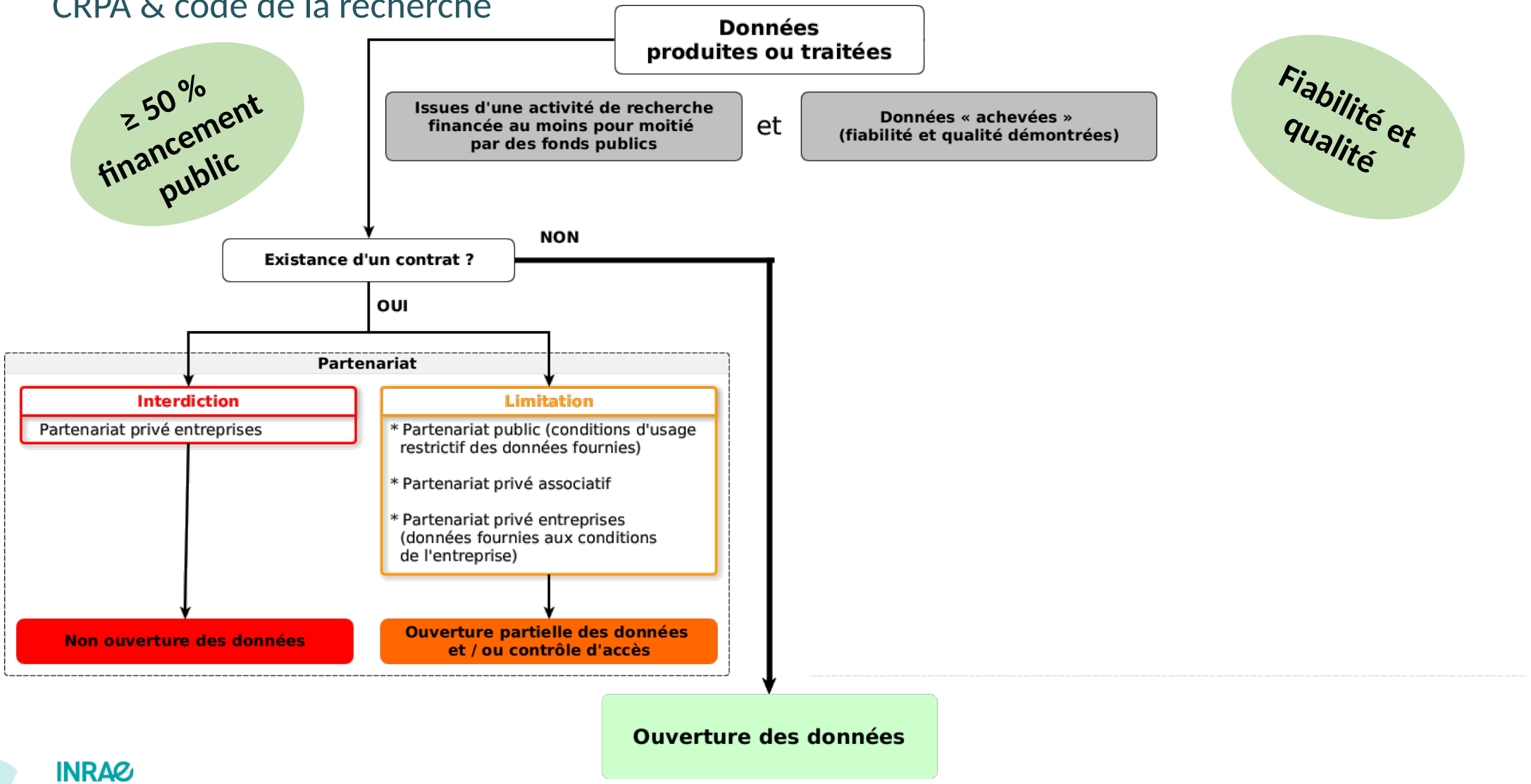
CRPA & code de la recherche



* CRPA : Code des Relation entre le Public et l'Administration

Des données ouvertes, avec libre réutilisation par défaut !

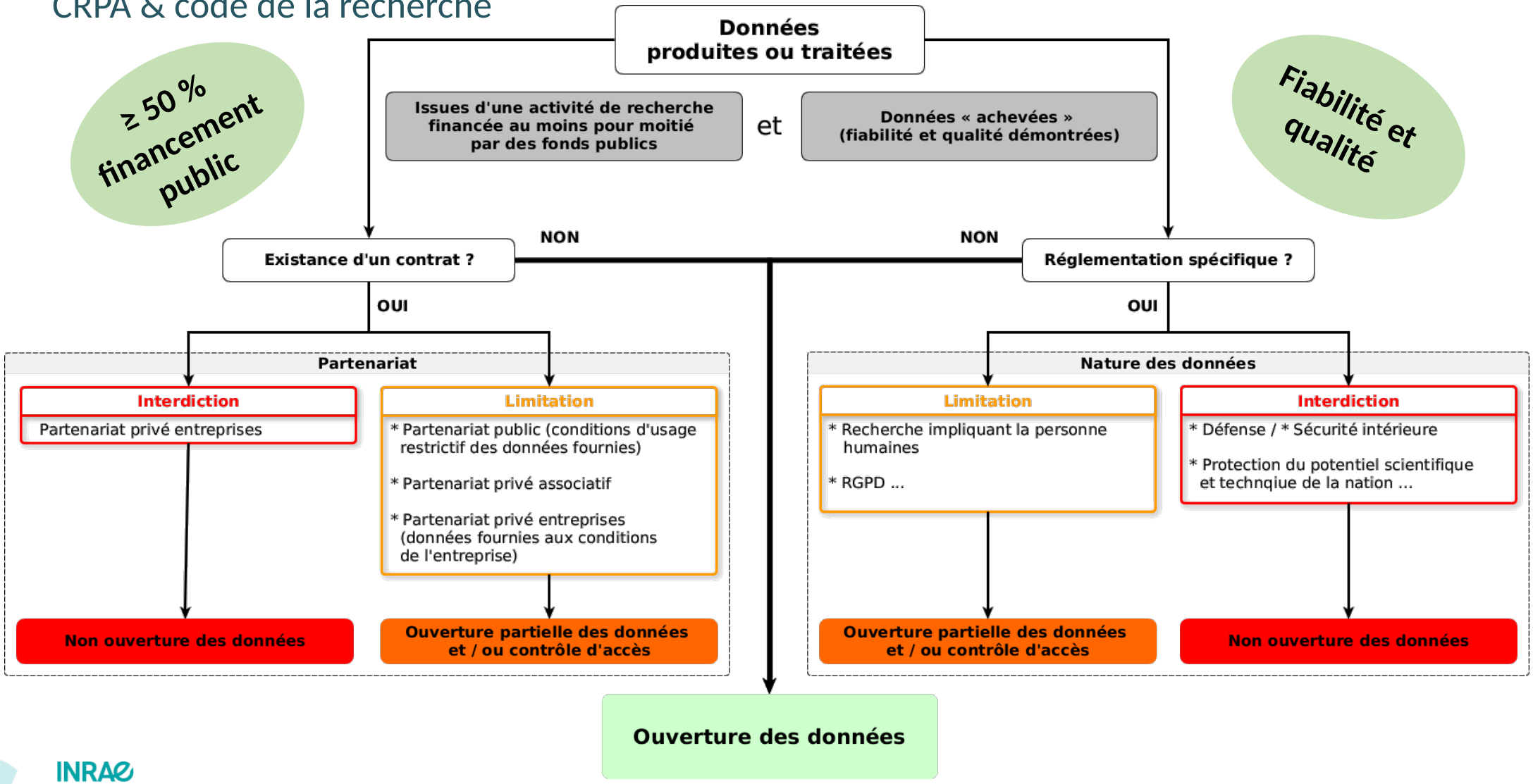
CRPA & code de la recherche



* CRPA : Code des Relation entre le Public et l'Administration

Des données ouvertes, avec libre réutilisation par défaut !

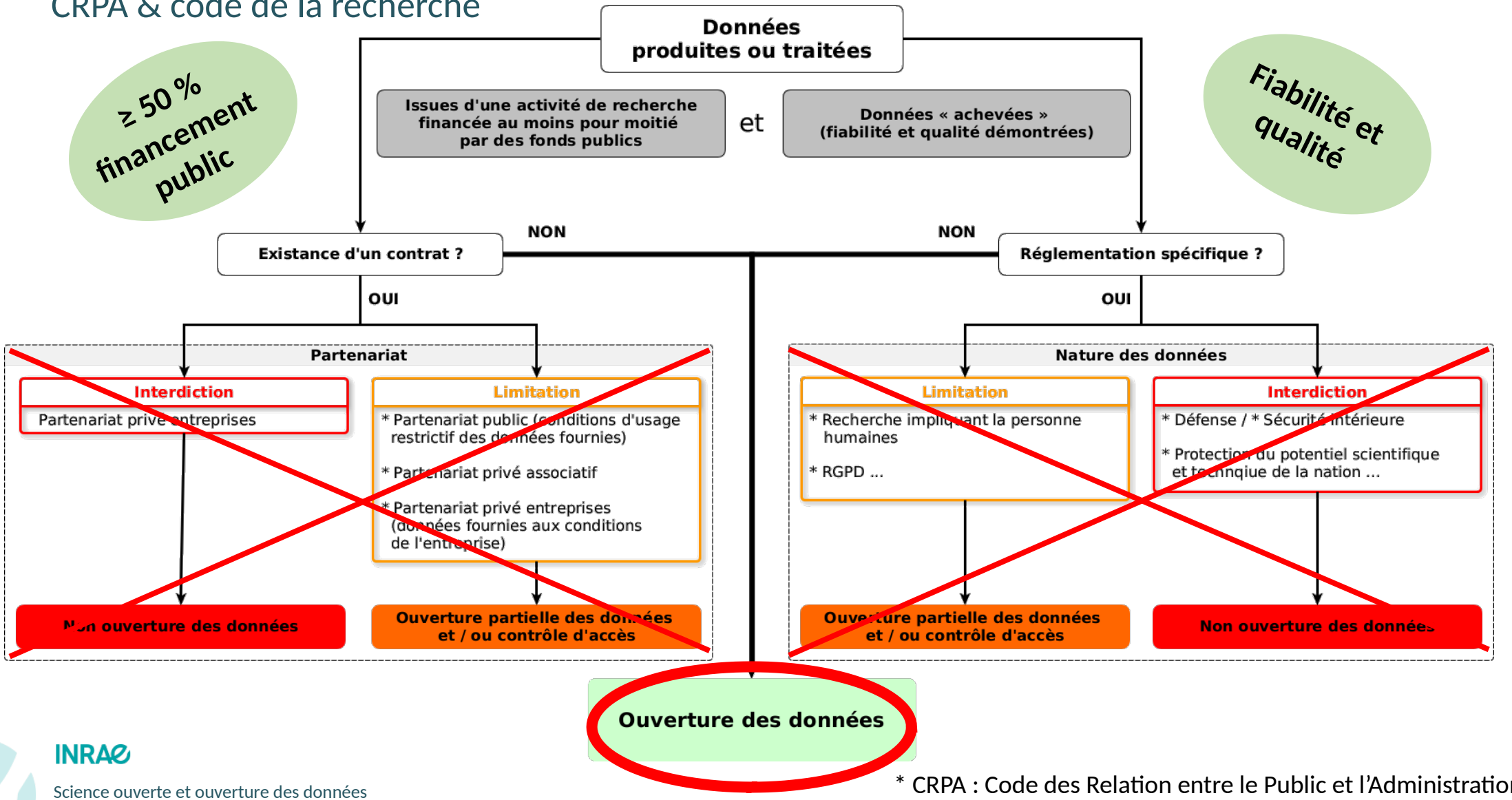
CRPA & code de la recherche



* CRPA : Code des Relation entre le Public et l'Administration

Des données ouvertes, avec libre réutilisation par défaut !

CRPA & code de la recherche



* CRPA : Code des Relation entre le Public et l'Administration

Des enjeux également économiques !

- Coût de la non ouverture des données au niveau européen :

26 milliards d'euros / an !

Des enjeux également économiques !

- Coût de la non ouverture des données au niveau européen :

  26 milliards d'euros / an !

- 10,2 milliards d'euros / an : estimation commission européenne (2018)
  - Perte de temps : accessibilité des données ou reproduction des données
  - Financement redondant d'études
  - Coût d'accès à des données qui ne sont pas accessibles gratuitement
  - Multiplication des points de stockage des données (données non accessibles)

- 16 milliards d'euros / an en retombées économiques !

European Commission, Directorate-General for Research and Innovation, Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data, Publications Office, 2018, https://data.europa.eu/doi/10.2777/02999

Et des enjeux sociétaux !

- Bien-être animal :

  Le partage de connaissances et de données favorise / accélère les 3 R (Remplacer, Raffiner, Réduire)

- Reconquérir la confiance de la société!
  - Rigueur scientifique et reproductibilité
  - Permettre à chacun de s'approprier une connaissance qu'il a indirectement contribué à financer.

Et des enjeux sociétaux !

- Bien-être animal :

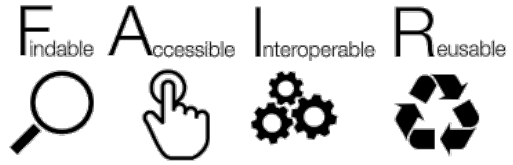
  Le partage de connaissances et de données favorise / accélère les 3 R (Remplacer, Raffiner, Réduire)

- Reconquérir la confiance de la société!
  - Rigueur scientifique et reproductibilité
  - Permettre à chacun de s'approprier une connaissance qu'il a indirectement contribué à financer.

Faire de la science ouverte c'est dire « Voici ma question , voici comment j'y ai répondu et voici les données que j'ai utilisées pour y répondre. Vous pouvez vérifier ce que j'ai fait et vous pouvez les réutiliser. »

En pratique !

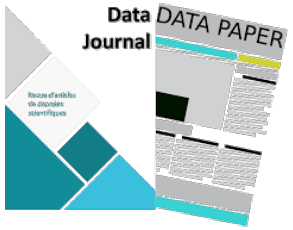
- Penser FAIR data dès la conception du projet



- Mettre les données à disposition dans un entrepôt

recherche.data.gouv.fr

- Publier un data paper



- Penser à la réutilisation de données

Facilitation du FAIR Data à PEGASE

> Produire des données FAIR

Findable Accessible Interoperable Reusable

Produire des données FAIR

- Pour favoriser

Findable

- La découverte des données humain / machine -> description et indexation des jeux de données (Métadonnées riches, Identification par DOI)

Accessible

- L'accès aux données par un stockage pérenne des données et métadonnées dans des entrepôt de données ouvert et sous des protocole standard

Interoperable

- L'interopérabilité -> Les données utilisent un langage formel, accessible et partagé et elles sont téléchargeables, utilisables, intelligibles, et combinables

Reusable

- La réutilisation -> Les données et les métadonnées correspondent aux standards des communautés et sont mises à disposition selon une licence explicite et accessible.

Vers une meilleure gestion des projets et des données biologiques

Un travail qui a débuté il y a bientôt 10 ans (2016)

- Objectifs :
  - Établir la continuité de la traçabilité entre projet / protocole / expérimentation / échantillon / analyse / résultat.
  - Faciliter l'identification des données existantes
    - Au sein d'un projet
    - Au sein de l'unité
  - Garantir une sauvegarde pérenne des données.
  - Permettre une ré-utilisation rapide des données.
  - (Faciliter la publication des données)

Un travail d'inventaire !

- Identifier les types de données générées / utilisées dans l'unité
- Identifier les métadonnées expérimentales : information environnant la données et indispensable à son interprétation
  - Identification du matériel
  - Identification des kits / solutions
  - Conditions expérimentales de la mesure…
- Distinguer les données brutes générées et les données transformées
- Identifier les formats des données

Co-construction de fichier de résultats standardisés

- Inventaire des données : ~ 420 analyses différentes
- 32 fichiers de résultats types / format excel
  - 9 genoprot
  - 2 histo
  - 1 culture cell
  - 1 hormone et cytokine
  - 17 biochimie
  - 4 technique séparative
- Une partie commune
  - méta-données : date, opérateur, mode opératoire, identification des appareils, identifications des produits …
  - Identification des échantillons
- Une partie spécifique à chaque type d'analyse
  - Méta-données spécifiques :
  - Résultats

Mise en place d’un système de sauvegarde / requêtage

- Cahier des charges
  - Basé sur les templates de résultats
  - Serveur de sauvegarde des données / pérennité des données
  - Validation automatique du type de données (numérique, date, texte)
  - Faciliter l’identification des données existantes

Maintenant que je suis convaincu comment je peux m’y prendre ?

FAIR-DS un outils informatiques dédiés