

Ouverture des données : entre injonctions et intérêts

<https://view.genial.ly/639aec4edcdd210013063a93>

29 septembre 2023

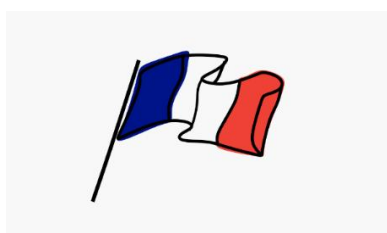
Table des matières

1. Répondre aux exigences nationales	1
2. Répondre aux exigences légitimes des revues scientifiques	2
3. Éviter la captation des données de la recherche par les éditeurs	3
4. Augmenter la visibilité de ses recherches	5
5. Favoriser les collaborations	6
6. Permettre la reproductibilité	6
7. Éviter la perte de données	8
8. Réduire le coût de la mauvaise gestion des données	9
9. En résumé	10

Tournez les pages pour découvrir pourquoi vous devriez partager vos données de recherche, plus particulièrement en les déposant dans un entrepôt de données...

1. Répondre aux exigences nationales

Diffuser les données de recherche financées par des fonds publics devient une obligation en France. Bien sûr, il existe des exceptions légales à cette obligation de diffusion.



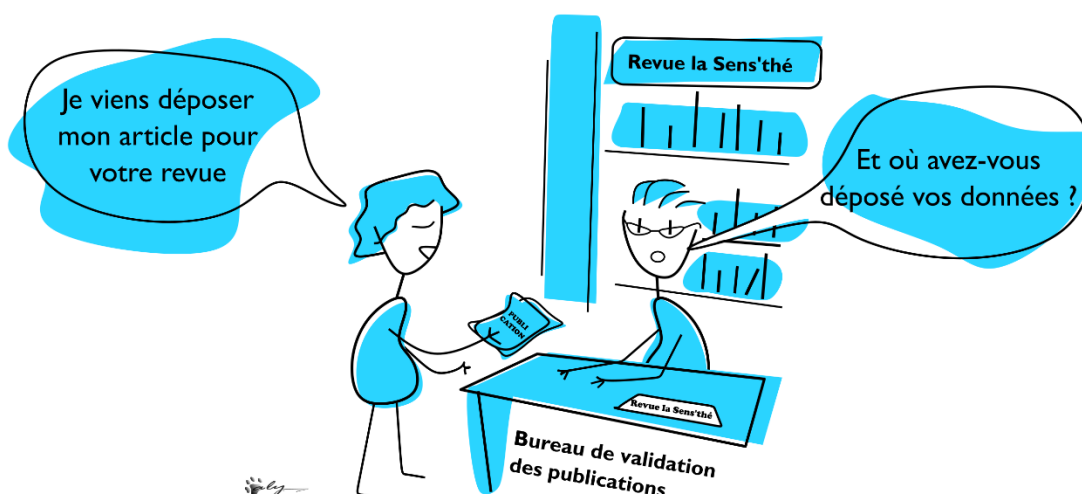
En savoir plus :

Depuis la [loi pour une République numérique](#) de 2016, les données de recherche "achevées" sont assimilées à des données administratives et font donc l'objet d'un principe d'ouverture "par défaut", c'est-à-dire qu'elles sont censées être publiées et rendues accessibles sur internet. Il existe toutefois des exceptions légales au partage et à la diffusion des données (données personnelles, sensibles, protégées par le droit de propriété intellectuelle...).

Dès 2018, le gouvernement invite dans son [premier Plan national pour la science ouverte](#) les chercheurs « à déposer les données dans des entrepôts de données certifiés, dont la gouvernance et les règles de propriété intellectuelle seront conformes aux bonnes pratiques ». En 2021, l'une des mesures du [Deuxième Plan national pour la science ouverte](#) est la mise en œuvre de l'obligation de diffusion des données de recherche financées au moins pour moitié par des fonds publics.

2. Répondre aux exigences légitimes des revues scientifiques

De plus en plus d'éditeurs de revues scientifiques exigent de pouvoir accéder aux données de recherche liées aux résultats dans les articles qui leur sont soumis pour publication.



En savoir plus :

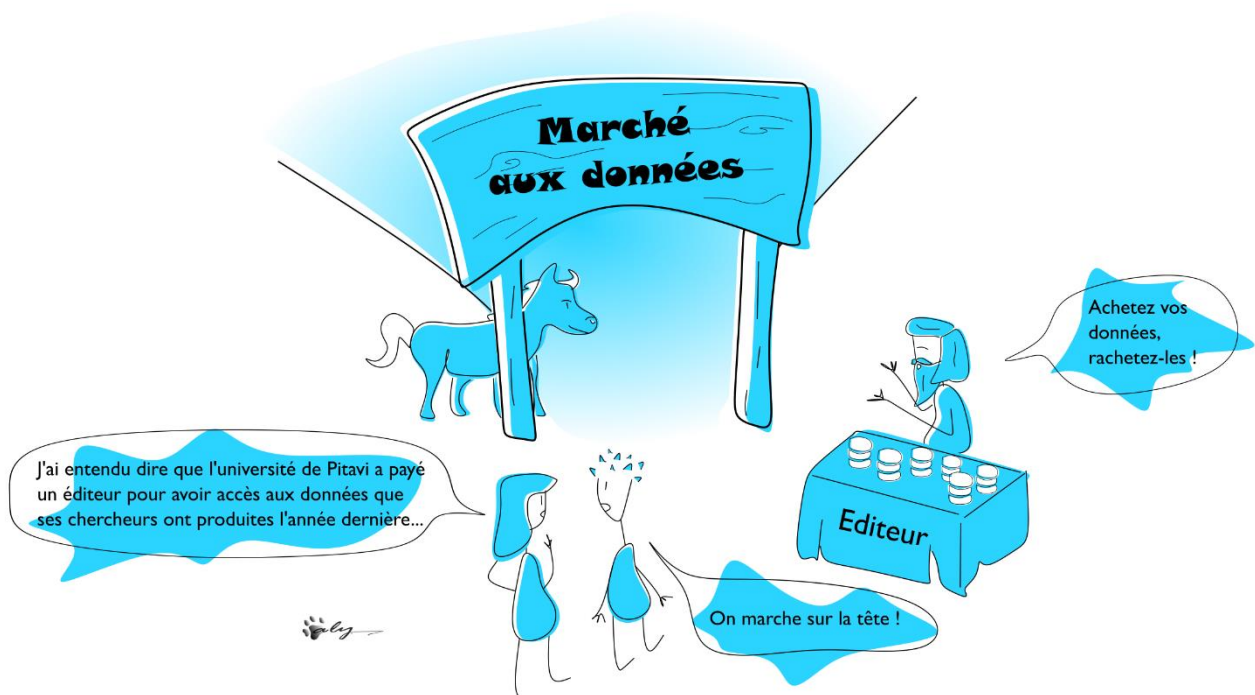
Si cette demande peut s'avérer légitime (facilitation du travail de vérification par les pairs, transparence de la recherche, fiabilité des résultats...), il faut néanmoins rester vigilant...

« Il est recommandé de ne pas confier les données à partager aux éditeurs des revues qui proposent de les publier sous forme de « supplementary data » ou de « supplementary materials » associés à l'article. Une telle publication se fait encore souvent dans un format et un environnement qui ne permettent pas de documenter correctement les données et rendent difficile leur réutilisation. Elle peut aussi s'accompagner d'une demande de transfert exclusif de droits contraire à la loi française et à l'esprit de la science ouverte. »

Cecconi B, Chatelier JY, Kuntziger B, et al. Partager les données liées aux publications scientifiques : Guide pour les chercheurs. Ministère de l'Enseignement supérieur et de la Recherche. [En ligne]. Consulté le 15 février 2023. https://www.ouvrirlascience.fr/wp-content/uploads/2022/06/Guide-Partager_les_donnees_pour-impression-A5.pdf

3. Éviter la captation des données de la recherche par les éditeurs

Pour éviter que les données ne subissent le même phénomène de privatisation que les publications, il est encore temps d'agir !



En savoir plus :

« [Certains éditeurs] ont commencé à inciter ces dernières années les chercheurs à publier non seulement le texte de leurs publications, mais aussi des matériaux complémentaires (complementary material), constitués en général par les jeux de données brutes ayant permis d'élaborer les résultats faisant l'objet de la publication. Grâce à leur position dominante, ces acteurs étaient donc en mesure d'emmagasiner peu à peu de grands volumes de données de recherche, soumises aux mêmes clauses d'exclusivité que les articles. Dès lors, le même phénomène de privatisation qui a frappé les écrits scientifiques était en train de se reproduire pour les données de recherche, avec à terme le risque pour les universités d'avoir à racheter aux éditeurs l'accès aux données que la communauté des chercheurs avaient produites. »

Calimaq. Quel statut pour les données de la recherche après la loi numérique ?. – SILex –.Blog. [En ligne]. Consulté le 01 février 2023. <https://scinfolex.com/2016/11/03/quel-statut-pour-les-donnees-de-la-recherche-apres-la-loi-numerique/>

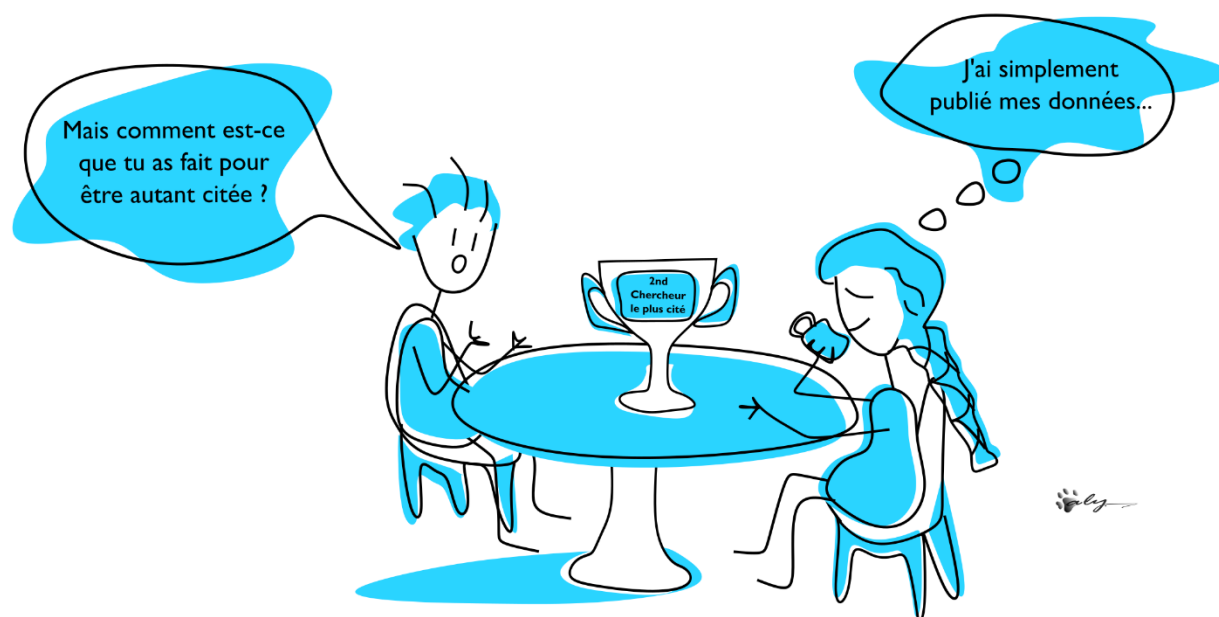
Si vous anticipez une telle demande et que vous déposez au préalable vos données dans un entrepôt de données, alors vous pourrez faire un lien entre les jeux de données déposés dans l'entrepôt et votre article. Votre éditeur pourra ainsi accéder aux données qui l'intéressent et vos jeux de données seront correctement documentés, et donc plus facilement réutilisables.

Bon à savoir :

« [...] les éditeurs ne peuvent exiger d'exclusivité et donc empêcher que les données leur ayant été fournies soient diffusées par ailleurs par les chercheurs. Les éditeurs ou les revues peuvent néanmoins exiger le dépôt des données liées aux publications dans leur entrepôt ou sur leur plateforme. Mais dans ce cas, les éditeurs ne peuvent empêcher la libre réutilisation des données par le biais de conditions d'utilisation. C'est le sens de l'article 533-4-IV qui dispose que "*les dispositions [de l'article 533-4] sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite*". Cette affirmation concerne donc également les contrats signés avec des éditeurs étrangers. »

4. Augmenter la visibilité de ses recherches

Partager ses données de recherche, c'est augmenter ses chances d'être cité !



En savoir plus :

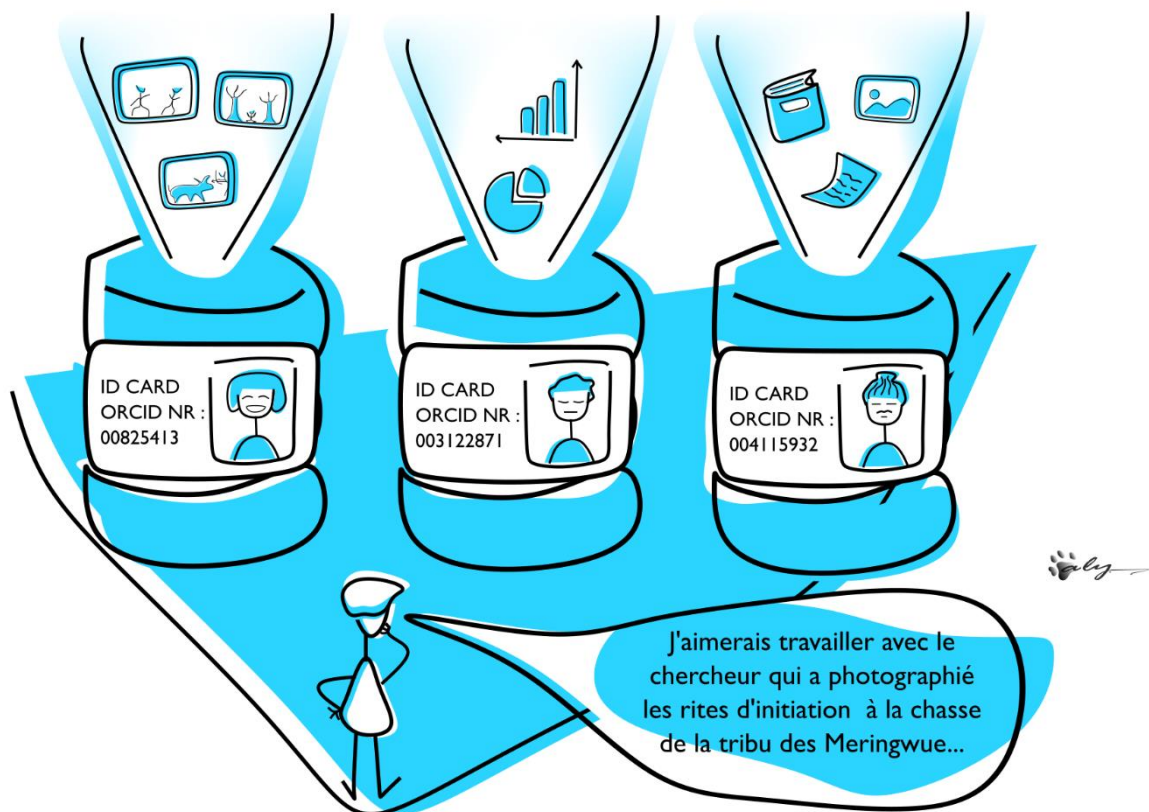
« Contrairement à une crainte ancrée relative au possible “pillage” de leurs données, ceux qui partagent les données associées à leurs articles ont un bénéfice significatif en termes de visibilité, mesurée à travers les citations. Une demi-douzaine d'études montre en effet une croissance forte des citations lorsqu'on associe des données à un papier. »

Les données de la recherche : un sujet stratégique. Collège Données de la recherche du Comité pour la science ouverte. [En ligne]. Consulté le 15 février 2023. <https://www.ouvrirlascience.fr/les-donnees-de-la-recherche-un-sujet-strategique/>

Pour en savoir plus, consultez [cette étude sur l'avantage pour la citation de lier ses publications à ses données de recherche.](#)

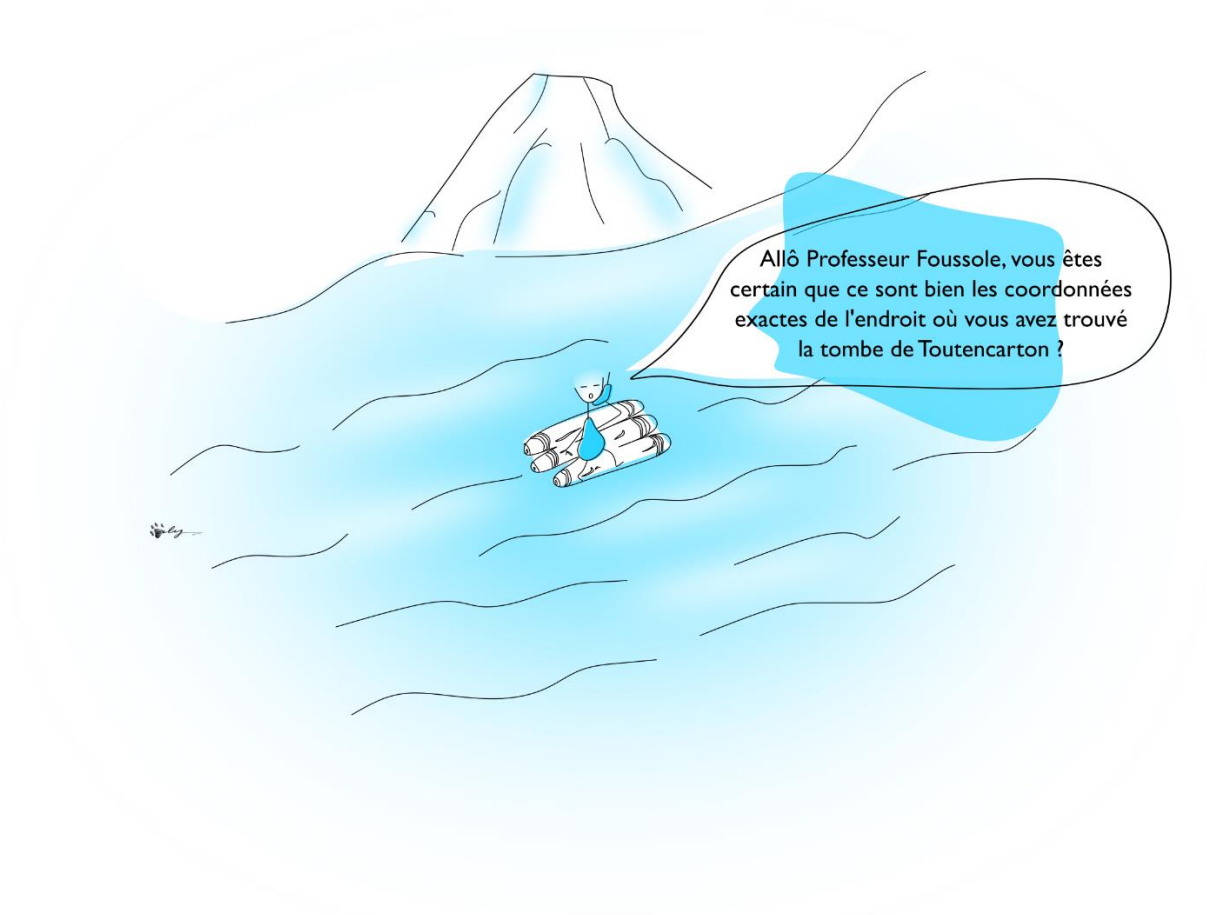
5. Favoriser les collaborations

Des données déposées dans un entrepôt de données sont toujours associées à leur(s) auteur(s), ouvrant ainsi la voie à de potentielles collaborations.



6. Permettre la reproductibilité

Le partage des données joue un rôle non négligeable dans la reproductibilité de la recherche. L'accès aux données est souvent nécessaire afin de pouvoir reproduire une expérience, ou simplement afin de pouvoir comparer les résultats obtenus.



En savoir plus :

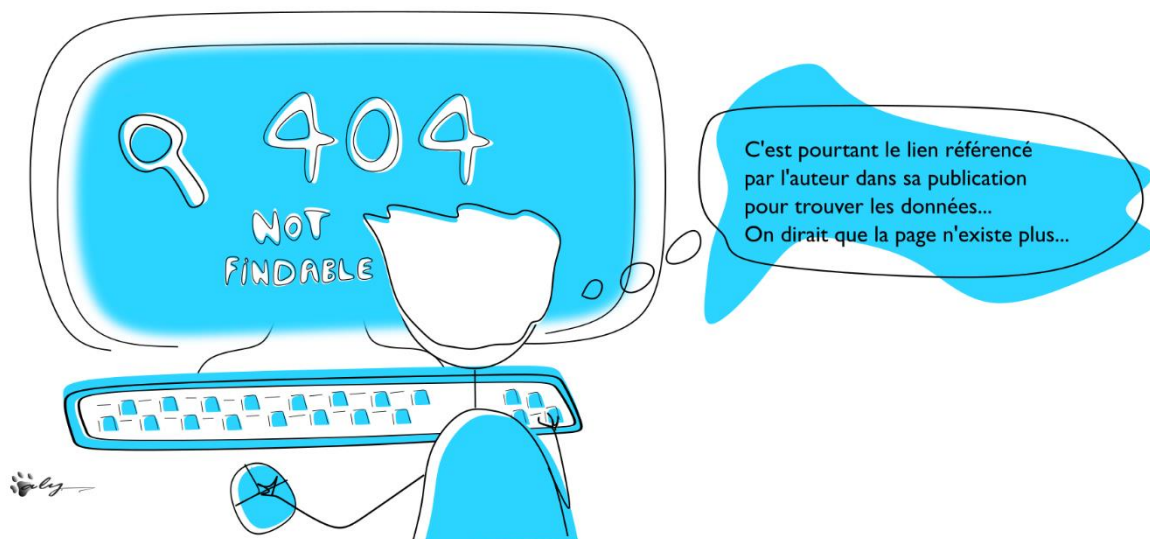
« La mise à disposition des données de la recherche est un point clé pour leur réutilisabilité et pour faciliter la reproductibilité de la recherche par vos pairs. [...] Il convient aussi d'élaborer une documentation extensive levant les ambiguïtés d'interprétation [...], et enfin de choisir un mode de partage le plus ouvert et pérenne possible. »

Desquilbet L, Granger S, Hejblum B, et al. Vers une recherche reproductible : Faire évoluer ses pratiques. Unité régionale de formation à l'information scientifique et technique de Bordeaux. [En ligne]. Consulté le 15 février

2023. https://www.ouvrirlascience.fr/wp-content/uploads/2019/06/Vers-une-recherche-reproductible_20190529_booksprinrr.pdf

7. Éviter la perte de données

Déposer ses données dans un entrepôt de données certifié offre de meilleures garanties pour rendre les données Accessibles et Faciles à trouver à court et moyen terme.



En savoir plus :

D'après [une étude publiée en 2014](#), avec les années, la probabilité de retrouver les jeux de données associés aux publications diminue (-17% chaque année) et les chances que les adresses mail de contact des auteurs soient encore valables diminuent également (-7% chaque année). Ce sont des centaines de milliers de données qui se perdent, qui ne peuvent plus être consultées, des auteurs qui ne sont plus associés à leur travail ou qui ne sont plus joignables...

Un entrepôt de données doit permettre l'attribution d'un [identifiant pérenne](#) objet (DOI, Ark, Handle...) aux jeux de données. Un entrepôt de données permet également de renseigner des PID contributeur. Les identifiants pérennes (ou PID) garantissent un lien stable à la ressource en ligne.

Remarque : gardez à l'esprit qu'un entrepôt de données ne permet pas forcément de préserver les données à long terme. Ceci est le rôle de l'archivage pérenne. La

plupart du temps, les entrepôts de données préservent les données à moyen terme (5-10 ans). Cependant, certains entrepôts proposent un service d'archivage pérenne. C'est aux chercheurs, à l'institut, au laboratoire d'évaluer la pertinence d'une préservation à long terme d'un jeu de données. Dans tous les cas, si les données ne sont pas préservées sur le long terme, les métadonnées, elles, restent toujours accessibles.

8. Réduire le coût de la mauvaise gestion des données

Un rapport de la Commission européenne publié en 2019 estime à 10 milliards d'euros par an le coût engendré par la non-application des principes FAIR dans la gestion des données de la recherche.



En savoir plus :

Le partage et l'ouverture des données de recherche permettrait par exemple d'éviter d'investir de l'argent dans la production de données qui auraient déjà été produites par ailleurs...

[Cliquer ici pour accéder au rapport de la Commission européenne.](#)

9. En résumé

Partager ses données via un entrepôt de données permet de :

- Répondre aux **exigences nationales**
- Répondre aux **exigences des revues scientifiques**
- **Éviter la captation des données** par les éditeurs
- **Augmenter la visibilité** de ses recherches
- Favoriser les **collaborations**
- Favoriser la **reproductibilité**
- **Éviter la perte** de données
- **Réduire le coût** de la non-application des principes FAIR à la gestion des données