

Quelles données publier ?

Café Science Ouverte PAMIR

-

18 janvier 2024

Laurent Romary

Inria – Direction de la culture et de l'information
scientifiques



Contexte

- Séance sur le Plan de Gestion de Données : la charrue avant les bœufs ?
- Les données de la recherche, qu'est ce que c'est ?
- Droits et devoirs (surtout devoirs...)

Données de la recherche ?

- OCDE : « Les données de la recherche sont définies comme des **enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme **sources principales pour la recherche scientifique** et sont généralement reconnus par la communauté scientifique comme nécessaires pour **valider les résultats** de la recherche. »
- Une variété de formes et d'origines :
 - Corpus de référence
 - Données moissonnées
 - Données d'expérimentation (capteurs, instrumentation)
 - Simulations, données de test, modèles
 - Données issues de sources officielles (statistiques, météorologiques, hospitalières)
 - ...

[Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics](#), OCDE, 2007

N/A - Not applicable...

Risque de ne pas sentir concerné·e en pensant que nous n'avons pas de données

- La réutilisation de données est déjà de la manipulation de données

Ex. : quelle sélection, quels droits de réutilisation, quelles contraintes de sécurité y associer, quelles référence y faire pour des soucis de reproductibilité ?

- À moins de ne faire que de la théorie, on manipule forcément des jeux de données

Ex. en IA : Corpus source, paramètres d'apprentissage, modèles, sorties des modèles, données de performance, empreinte carbone

- Garder une trace de tout ce qui a contribué à l'édification d'un résultat
- Penser à celui qui passera derrière (parfois soi-même, plus tard)

Une obligation de conservation et de diffusion

Pour les personnes qui travaillent dans un organisme public :

- Tout ce que nous produisons ou recevons, documents comme données, sont des **archives publiques**
- Nous en sommes responsables mais pas propriétaires, ie nous ne pouvons pas les détruire sans l'autorisation du service producteur et de l'administration des archives de l'institution

Cadre de notre mission de service public :

- Obligation de conservation des **documents administratifs**
- [Article L.300-2 du Code des relations entre le public et l'administration](#)

Loi pour une République numérique (octobre 2016) :

- Obligation de partager nos documents et nos données, à tout public, gratuitement
- Reflet de la directive européenne [2013/37/UE](#) dite « PSI » (*Public Sector Information*)
- Le partage est donc la règle par défaut...

Quelques limites...

Le document doit être formellement achevé

- Documents préparatoires, données préliminaires etc.
- La frontière est parfois floue...
- Logiciel: « *chaque version du code source d'un même programme informatique revêt le caractère de document administratif achevé et peut être communiqué dans cet état* » (décision du tribunal administratif de Paris du 10 mars 2016)

Limites de réutilisation

- Les données doivent être rendues publiques, résulter d'un financement public d'au moins 50%, ne pas être protégées par un droit spécifique

Ne pas mettre en péril la vie privée ou le secret des affaires

- Données personnelles (cf. RGPD - Règlement Général sur la Protection des Données)
- Secret professionnel (médical, bancaire, fiscal, juridique)
- Droit d'auteur

Bilan

- Se placer dans une disposition de partage par défaut
- Penser la gestion des données dans une perspective de partage
- Mettre en œuvre l'obligation de partage à tous les niveaux (scientifique, laboratoire, infrastructures, institution)

Pour aller plus loin sur le questionnement juridique : Emilie Cotte, Fanny Sébire. Modèle de logigramme de l'Institut Pasteur relatif aux questions juridiques liées à la diffusion des données de la recherche. 2022.

<https://pasteur.hal.science/pasteur-03587216>