

Script vidéo « Point de vue Référent.e.s Intégrité Scientifique » du cas Data : to me or not to me ?

Quels manquements peut-on relever dans la conduite de nos personnages ?

La gestion rigoureuse et cohérente des données de la recherche constitue aujourd'hui un enjeu de taille pour la production de nouvelles connaissances scientifiques.

D'une manière générale et plus particulièrement dans le cadre de la science ouverte, il est recommandé de partager les données brutes. Cependant, ce partage de données brutes doit être réfléchi en amont dans le cadre d'une stratégie de valorisation. Cette stratégie de valorisation déterminera les modalités d'ouverture et la temporalité, notamment pour les données de longue traîne, acquises depuis plusieurs années.

La transparence conférée par un accès aux données vient renforcer les conditions d'une recherche intègre.

Alors comment faire ? Il est important de contextualiser et d'associer des métadonnées décrivant les méthodologies d'acquisition. Cela permet de définir les contraintes associées, les domaines d'application et de réutilisation de ces données brutes et leurs limitations. On peut, dans le cadre de la science ouverte, attribuer un DOI, c'est-à-dire attribuer un identifiant numérique correspondant à ces données brutes. Pour cela, on peut déposer ces données dans un entrepôt de confiance.

La méthodologie associée au traitement des données doit être décrite avec le plus de précision possible et doit permettre de pouvoir refaire les expériences et le traitement des données brutes. Cela contribue à la robustesse et la reproductibilité mais aussi cela permet la vérification des résultats publiés. Un DOI peut également être attribué aux méthodologies afin de les décrire de façon exhaustive dans un document annexe et les rendre ainsi citables, voire de les publier dans un article de façon autonome dans certains cas.

Cette transparence et une description honnête, fiable et précise de la collecte et du traitement des données permettent de garantir que les résultats sont reproductibles et robustes – Cela donne indirectement de la crédibilité à l'équipe qui partage ses données.

Dans quels cas est-il possible d'émettre des réserves sur le partage des données ? Une temporalité relative à l'ouverture est-elle possible ?

Le régime juridique applicable à la science ouverte est gouverné par le principe général, issu de la Commission européenne, selon lequel les données de la recherche doivent être aussi ouvertes que possible et pas plus fermées que nécessaire.

Dans le cas de projets soumis à des clauses de confidentialité, la diffusion des résultats et données peut être restreinte. Les modalités de diffusion doivent être préalablement décrites dans le texte du projet et ou l'accord de consortium.

Des moratoires peuvent être mis en place pour ouvrir les données. La durée de ces moratoires sera à définir par les responsables lors de la publication des résultats afin de garantir leur utilisation exclusive pendant le temps nécessaire à leur valorisation par la publication scientifique. Cela présente l'avantage d'une protection temporaire tout en garantissant une ouverture des données à venir. En attendant, il est possible de prévoir une ouverture progressive des données selon un processus compatible avec les pratiques de la communauté scientifique et qui permette un premier niveau d'usage par des tiers.

Dans des projets reposant sur ou générant des données personnelles, le partage de ces données est soumis au Règlement Général sur la Protection des données, le RGPD, auquel il faut strictement se conformer.

La nature politisée du sujet de recherche et la contestation possible de la qualité des données et leur interprétation sont-elles des arguments légitimes pour ne pas partager les données ?

Dans cette situation, les arguments avancés par Bernard ne sont effectivement pas recevables. Une politique de science ouverte appliquée tant aux données brutes qu'à leur traitement aurait peut-être pu limiter la polémique et, en tout cas, asseoir la crédibilité de l'équipe de recherche et sa reconnaissance.

De plus, ces données brutes auraient pu être utilisées de façon indépendante par d'autres équipes, dans le but, par exemple, de vérifier leur validité, de confirmer les résultats, de poser d'autres questions ou encore d'envisager d'autres recherches. Le travail de cette équipe aurait été valorisé par ces retombées et les citations des études en aval.

Pour protéger temporairement le travail de cette équipe en vue de sa valorisation par publication scientifique, un moratoire sur l'ouverture des données aurait pu également être envisagé. Il est alors important de définir à quelle échéance les données seront totalement exploitées.

La science procède par la formulation de différentes hypothèses, leur vérification et leur confrontation, plaçant le principe contradictoire au cœur de la démarche scientifique.