

Étude sur les métiers et activités de la donnée

Connaître et reconnaître les métiers et activités autour des données de la recherche, un enjeu important.

Les données occupent une place croissante dans les activités de recherche.

Concevoir nativement des données de recherche réutilisables, y compris pour les producteurs de données eux-mêmes, est essentiel. Il est donc essentiel de développer, connaître et reconnaître les compétences en matière de gestion, préservation, circulation et ouverture des données pour l'ensemble des acteurs impliqués dans la recherche : chercheurs, enseignants-chercheurs, personnel d'appui à la recherche et les administratifs, quel que soit leur niveau de carrière.

Dans cette optique, une étude sur les métiers et activités de la donnée a été menée, pilotée par Isabelle Blanc au MESR, pour mieux comprendre les enjeux et les compétences nécessaires pour ces métiers en pleine évolution.

Un groupe de travail composé de représentants de différentes communautés métiers impliquées dans la gestion des données de la recherche a été constitué pour l'étude, travaillant en lien avec le cabinet Dataactivist, coopérative spécialiste de l'ouverture des données (constitution du groupe de travail : <https://www.ouvrirlascience.fr/definir-et-reconnaitre-les-activites-de-gestion-et-diffusion-des-donnees/?menu=4>).

UN CYCLE DE VIE REVISITÉ À L'AUNE DES ENSEIGNEMENTS ET CADRE DE RÉFÉRENCE DE LA SYNTHÈSE BIBLIOGRAPHIQUE

Une synthèse bibliographique a été réalisée, reposant sur les métiers clés couramment associés aux données de la recherche tels que par exemple le data steward, le data librarian, ou encore le data curator, avec une priorité donnée aux revues de littératures et publications post-2018 et à la prise en compte des référentiels existants.

Il ressort notamment de cette synthèse qu'une distinction doit être faite entre métiers, activités et compétences, avec un besoin de formulations non ambiguës pour décrire les activités.

Ces constats ont conduit à une étude plus fine menée sur la liste des activités constituant le cycle de vie des données de la recherche, avec une attention particulière portée dans le placement sur le cycle de vie de :

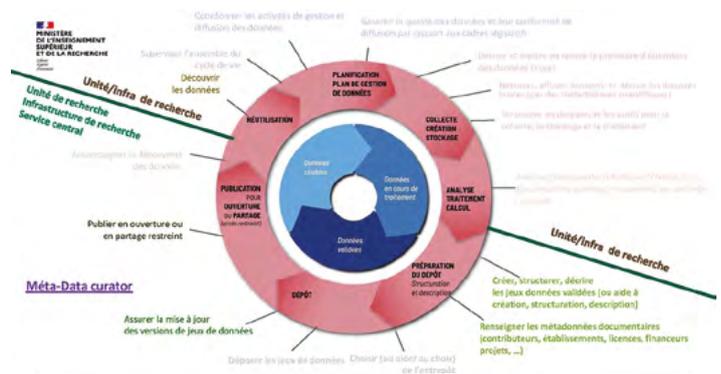
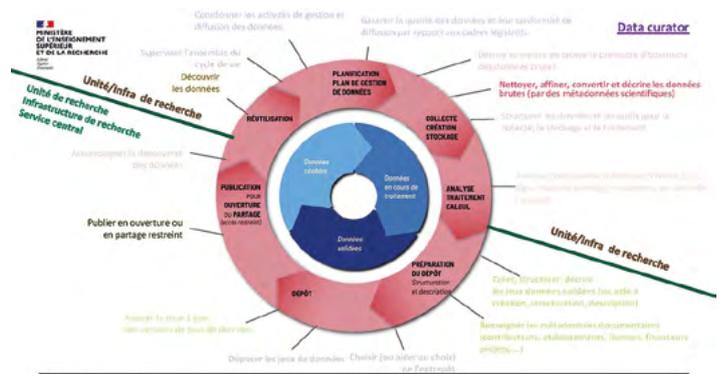
- La nature des données (en cours de traitement, validées, constituées en jeux de données, citables),
- L'endroit d'exercice des activités (laboratoires et plateformes de recherche, structures centrales, ...).

Au terme de ce travail, un cycle de vie de la donnée de la recherche a été proposé en pointant notamment les lieux de réalisation des activités, avec par exemple une distinction

entre les différents types de métadonnées reprenant les définitions de référence.

À partir de ce travail sur le cycle de vie, les métiers peuvent alors être « placés », en spécifiant quelles activités sont associées, et où elles se déroulent.

Par exemple, un « data curator » agit au plus près de la donnée au sein des unités de recherche et infrastructures, alors que le « metadata curator » peut agir également dans les services centraux, notamment sur les métadonnées documentaires.



COMPLÉMENTARITÉ DES ACTIVITÉS ET COMPÉTENCES, PARCOURS ET BESOINS DE FORMATION – RETOUR DES ENQUÊTES MENÉES

Deux enquêtes sont venues compléter l'étude bibliographique, l'une menée par questionnaire et la deuxième par entretiens.

Constituée par le groupe de travail dans l'objectif de bien identifier les pratiques, l'enquête a été diffusée très largement. Destinée à toute personne travaillant en lien avec les données de la recherche, elle a permis de recueillir 1565 réponses, dont 55 % se disant « Personnel d'appui/soutien à la recherche » et 45 % « Chercheur ou enseignant-chercheur » (doctorants et post-doctorants inclus).

Le domaine le plus représenté dans les personnels d'appui est celui de l'informatique et du calcul scientifique.

La très grande majorité des répondants (82 %) est issue d'un laboratoire, d'une équipe de recherche ou d'une infrastructure de recherche, marquant l'ancrage des activités au plus près des activités dans les structures de recherche. Par ailleurs, les répondants ont majoritairement fait appel à de l'autoformation, ce qui renforce la nécessité de réfléchir au renforcement des parcours de formation sur ces sujets. L'enquête par questionnaire a été complétée par trente entretiens menés dans l'objectif d'approfondir la compréhension des activités et des parcours, de positionner les activités dans l'organisation, d'étudier les tâches et compétences liées aux données, et de travailler sur les parcours professionnels, y compris les parcours de formation et d'auto-formation.

Les entretiens ont été menés auprès de personnes représentatives des différentes catégories de personnels, des structures dans lesquelles les activités sont exercées et des thématiques. Ces personnes ont notamment été identifiées grâce à l'enquête par questionnaire dans laquelle elles s'étaient portées volontaires.

Quelques verbatims permettent d'illustrer certains aspects importants ressortant des entretiens, l'analyse étant tou-

jours en cours. Par exemple, des difficultés au quotidien sont exprimées pour « convaincre les collègues de diffuser les données », pour « coordonner » et « motiver ». De plus, de très grands besoins de formation ont été exprimés : « // y a un besoin de formation pour un besoin d'efficacité » dit ainsi une personne interrogée.

Mais le sujet des données de la recherche réserve aussi son lot de satisfactions, avec des avancées qui sont reconnues : « Le discours [sur les données de la recherche] a changé et c'est bénéfique, on peut s'appuyer sur le ministère », et des retours positifs « comme on a vraiment beaucoup de demandes sur les données de la recherche, on a l'impression d'être vraiment utile ».

VERS DES FICHES ET VIDÉOS PERMETTANT UNE APPROPRIATION DES RÉSULTATS DE L'ÉTUDE

À partir des travaux réalisés, des livrables seront produits selon trois formats :

- Des fiches synthétiques des étapes du cycle de vie, des activités et des intervenants à chaque étape, avec leurs profils et compétences propres,
- Des fiches métiers permettant une meilleure connaissance et reconnaissance des activités et des parcours de formation,
- Des vidéos permettant de présenter les différents acteurs et de mieux comprendre les activités de chaque intervenant dans les étapes du cycle de vie.

ANNE LAURENT

Vice-Présidente déléguée à la science ouverte et aux données de la Recherche - Université de Montpellier
anne.laurent@umontpellier.fr



●●●● POINTS DE REPÈRE

DÉFINITION UTILISÉE : DONNÉES DE LA RECHERCHE OCDE 2007

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 des résultats de recherche

DÉFINITION MÉTADONNÉES

National Information Standards Organization (NISO), 2004

« Une métadonnée est une information structurée qui décrit, explique, localise une ressource informationnelle ou la rend facile à trouver, utiliser ou gérer »

La NISO distingue 3 principaux types de métadonnées : descriptives, structurelles et administratives.

Borgman, C.L., 2020

Métadonnées scientifiques : description protocole obtention, traitement, matériel, logiciels...

Métadonnées documentaires : administratives, d'utilisation, de préservation, d'accès