

SYNTHÈSE

La gestion des données du savoir pour la recherche et l'innovation au 21^{ème} siècle

Le problème

Les données de la recherche représentent la nouvelle monnaie d'échange à l'ère numérique. Des sonnets aux statistiques, en passant par les gènes et les données géographiques, la quantité de matériel créé et stocké croît de façon exponentielle. Cependant la Feuille de Route de LERU pour les Données de la Recherche¹ a identifié d'importantes inégalités entre les organismes de recherche en termes de préparation pour faire face à ces questions. L'écart est particulièrement sévère dans les domaines suivants : l'élaboration de politiques de gestion des données, la sensibilisation aux difficultés actuelles, le développement des compétences, la formation, les coûts, la création de communautés de pratique, la gouvernance, ainsi que les différences entre les disciplines, systèmes législatifs, terminologies et espaces géographiques.

La solution

La présente **Synthèse du projet LEARN** aidera les décideurs à prendre des décisions solides. En outre, le Kit de Bonnes Pratiques de LEARN est à disposition de tous les acteurs de la recherche; organisé autour de cas d'étude, ce kit permettra aux organismes de recherche de faire face au déluge de données. Le projet LEARN a également développé un outil d'auto-évaluation².

Politique des données de la recherche

Tout organisme de recherche devrait adopter une politique des données de la recherche énonçant les responsabilités dont sont investis les chercheurs lorsqu'ils reçoivent un financement. Le projet LEARN a créé un modèle de politique pour la gestion des données dans les organismes de recherche; ce modèle est accompagné de conseils pour mettre en place une telle politique. Le modèle de politique proposé peut être adapté et adopté par chaque organisme mais aussi par des consortiums régionaux, nationaux et/ou internationaux.

Les données « FAIR »

Selon les bonnes pratiques, les données de la recherche doivent être « FAIR »³ :

- **F**aciles à trouver (Findable) – **A**ccessibles (Accessible) – **I**nteropérables (Interoperable) – **R**éutilisables (Reusable)

Pour être faciles à trouver, les données doivent être correctement décrites, si possible à l'aide de taxinomies et d'ontologies. Pour être accessibles, elles doivent, dans l'idéal, être ouvertes (open data) et disponibles pour le partage et la réutilisation. Les données de la recherche ne peuvent pas toutes être ouvertes mais les bonnes pratiques indiquent que ces données doivent être «aussi ouvertes que possibles et aussi fermées que nécessaires » (“as open as possible, as closed as necessary”)⁴.

1 http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf

2 Tous ces outils sont disponibles à l'adresse suivante: <http://learn-rdm.eu> (consulté le 12/01/17).

3 Voir www.force11.org/group/fairgroup/fairprinciples (consulté le 12/01/17).

4 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf p.4



learn

www.learn-rdm.eu



Ces données doivent aussi être interopérables et lisibles par des machines utilisant un vocabulaire conforme aux principes « FAIR ». Pour être réutilisables, les métadonnées qui décrivent les données doivent respecter les standards établis par le champ de recherche dont il est question.

Plans de gestion des données de la recherche

Il est recommandé aux chercheurs de planifier la collecte, le traitement, la description et la diffusion de leurs données dès le début de leur recherche. Ecrire un Plan de gestion des données de la recherche (Data Management Plan) permet de rassembler ces éléments et d'établir un programme de gestion durant la recherche⁵.

Infrastructures

Afin de traiter leurs données les chercheurs et organismes de recherche doivent pouvoir accéder à un écosystème numérique adéquat. Ces infrastructures peuvent être gérées par chaque institution; elles peuvent aussi être fournies par des services commerciaux ou par des entrepôts de données spécialisés dans un champ de recherche particulier; ou bien elles peuvent prendre la forme de plateformes régionales, nationales ou internationales. Chaque champ de recherche et chaque pays devra trouver le système qui lui convient. De manière générale, ces infrastructures doivent proposer les services suivants :

- Le stockage de données, pour les chercheurs qui en collectent activement;
- Une plateforme de publication, où les données de la recherche et les logiciels associés peuvent être partagés et réutilisés;
- Un système d'archivage, permettant que les données soient traitées et préservées sur le long terme, en conformité avec les exigences des financeurs;
- Un système permettant de chercher parmi les données entreposées, afin que chercheurs et citoyens puissent découvrir ces données (qu'elles soient accessibles in situ ou en ligne).

La Commission Européenne encourage l'usage de l'European Open Science Cloud (EOSC)⁶. L'EOSC est une métaphore qui entend exprimer un processus cohérent et sans obstacle, ainsi que l'idée que les données de la recherche constituent un bien commun (commons). L'EOSC sera un environnement fédérateur permettant de partager et réutiliser ces données. Cet environnement sera fondé sur un ensemble d'infrastructures déjà en place ou émergentes dans les Etats Membres et aura une supervision internationale allégée; une large part de liberté sera accordée aux questions pratiques d'utilisation.

Formation

La prévalence des données de la recherche oblige tous les chercheurs, qu'ils soient débutants ou expérimentés, à s'armer de compétences et d'outils leur permettant de travailler en confiance dans cet environnement où les données abondent. Les organismes de recherche doivent prendre l'initiative en termes d'offre de formation; souvent, la tâche en revient aux bibliothèques universitaires ou de recherche.

⁵ Pour plus d'informations, voir www.inist.fr/donnees/co/module_Donnees_recherche_26.html ou www.dcc.ac.uk/resources/data-management-plans (consulté le 12/01/17)

⁶ Voir <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud> (consulté le 12/01/17)



learn

www.learn-rdm.eu



Financement

La gestion des données a un coût. Il n'y a pas de méthode unique pour estimer ce coût mais plusieurs modèles de calcul, tels que le 4C Project, sont disponibles⁷.

Les risques

Les acteurs de la recherche s'exposent aux dangers suivants si les Bonnes Pratiques en matière de gestion des données ne sont pas respectées. Les chercheurs risquent de perdre des financements s'ils ne se conforment pas aux exigences des financeurs. La négligence peut causer la perte d'importants résultats de la recherche, sous peine de rendre difficile voire impossible la validation des conclusions du projet. Enfin, les partenariats et collaborations ne peuvent pas se développer si les résultats de la recherche ne sont pas partagés.

Les avantages

La gestion intelligente des données de la recherche offre de nombreux avantages. La mise en place de Bonnes Pratiques améliore la rigueur et la transparence des résultats. Les organismes de recherche peuvent rejoindre des initiatives internationales majeures telles que l'European Open Science Cloud. Les données de la recherche peuvent en effet devenir le nouvel élément clé de la communication scientifique et, tout comme les publications, contribuer à résoudre les grands défis auxquels la Société est confrontée : la pauvreté, les maladies et le réchauffement climatique.

Conclusion

Les données de la recherche peuvent favoriser l'innovation et stimuler de nouvelles découvertes pour le plus grand bénéfice de la Société. Tous les acteurs du cycle de la recherche ont un rôle à jouer. Cette Synthèse met en valeur ce que les chercheurs et organismes de recherche doivent faire pour être à la hauteur de ces défis passionnants.



⁷ Pour plus d'informations, voir www.4cproject.eu/summary-of-cost-models (consulté le 12/01/17).