

Gestion des données de recherche en sciences expérimentales

Divergences entre recherche fondamentale et appliquée

Dans le cadre d'une enquête menée à l'**Université de Strasbourg**, nous avons étudié différents projets de recherche en cours, dans diverses disciplines (des sciences humaines aux sciences et technologies). L'objectif est d'identifier quelle valeur est accordée aux données de la recherche à l'intérieur de chaque projet. Nous avons constaté que la finalité de la recherche pouvait avoir une incidence sur la manière dont sont gérées les données du projet – en particulier selon si le projet se revendique de la recherche fondamentale ou de la recherche appliquée. Nous présentons ici une sélection de **six projets de recherche** issus des **sciences expérimentales** (chimie, biologie...). Nous nous demanderons **en quoi, dans ce domaine, recherche appliquée et recherche fondamentale se différencient dans leurs pratiques de gestion des données**.

1 CADRE THÉORIQUE

Plusieurs problématiques jalonnent le chemin vers l'ouverture des données de recherche :

- 1. La définition même de « donnée de recherche » :** Pas de définition juridique, ni de consensus sur une définition unique. La donnée de recherche est difficile à circonscrire, car elle dépend du contexte dans lequel elle est utilisée. Selon Borgman (2015), une entité devient donnée scientifique, lorsqu'elle est utilisée comme preuve d'un phénomène dans un but de recherche, c'est-à-dire qu'elle est collectée, analysée et interprétée.
- 2. Le concept d'ouverture des données :** À l'heure actuelle, il n'existe pas de modèle de diffusion bien établi, comme en témoigne le caractère évasif du principe « as open as possible, as closed as necessary » dicté par la Commission européenne (2016). L'ouverture des données se heurte en effet à des enjeux éthiques, scientifiques et économiques, qui empêchent d'imposer un modèle unique applicable à l'ensemble des données.
- 3. Les stratégies de recherche actuelles :** En sciences et technologies, la publication dans des revues prestigieuses occupe une place prépondérante dans l'évaluation de la recherche. Ce qui explique pourquoi la publication a tant d'influence sur les stratégies de recherche des scientifiques et laisse si peu de place au partage des données. Publier permet de gagner en « crédibilité » (postes, subventions, confiance des pairs...). La crédibilité est ensuite réinvestie dans ce que Latour appelle un « cercle de crédibilité » (2001) : grâce à elle, le chercheur obtient de nouvelles subventions, qui seront investies dans l'achat de matériel, qui lui-même permettra de générer de nouvelles données et, par suite, de nouvelles théories ; ces nouvelles théories seront publiées dans un article ; et ainsi de suite.

2 MÉTHODOLOGIE DE L'ENQUÊTE

- 1. Critère d'échantillonnage :** une recherche d'excellence
 - ◆ Les projets de recherche ont été sélectionnés parmi les 26 unités de recherche les mieux notées par l'AERES en 2011-2012 (dernière évaluation disponible au moment de l'échantillonnage en 2017).
 - ◆ Il s'agissait de projets financés soit par l'ANR, soit par le programme H2020.
- 2. Nature de l'enquête :** des entretiens semi-directifs
 - ◆ Réalisés entre novembre 2017 et juin 2018
 - ◆ Auprès des coordinateurs scientifiques de chaque projet de recherche
- 3. Présentation des projets étudiés :**

	Discipline du projet	Nature du projet
Projet 1	Neurosciences	Recherche fondamentale
Projet 2	Neurosciences	Recherche fondamentale
Projet 3	Chimie thérapeutique	Recherche fondamentale
Projet 4	Chimie thérapeutique	Recherche appliquée
Projet 5	Biologie moléculaire	Recherche appliquée
Projet 6	Génétique	Recherche fondamentale

3 RÉSULTATS

1. La nécessité d'une recherche subventionnée

Dans les disciplines que nous avons étudiées, il ne peut y avoir de données sans financement. Dans le domaine des sciences expérimentales, la génération de données dépend en effet de matériel souvent coûteux (frais d'animalerie, séquençage d'ADN, ressources humaines chargées de réaliser les expériences...).

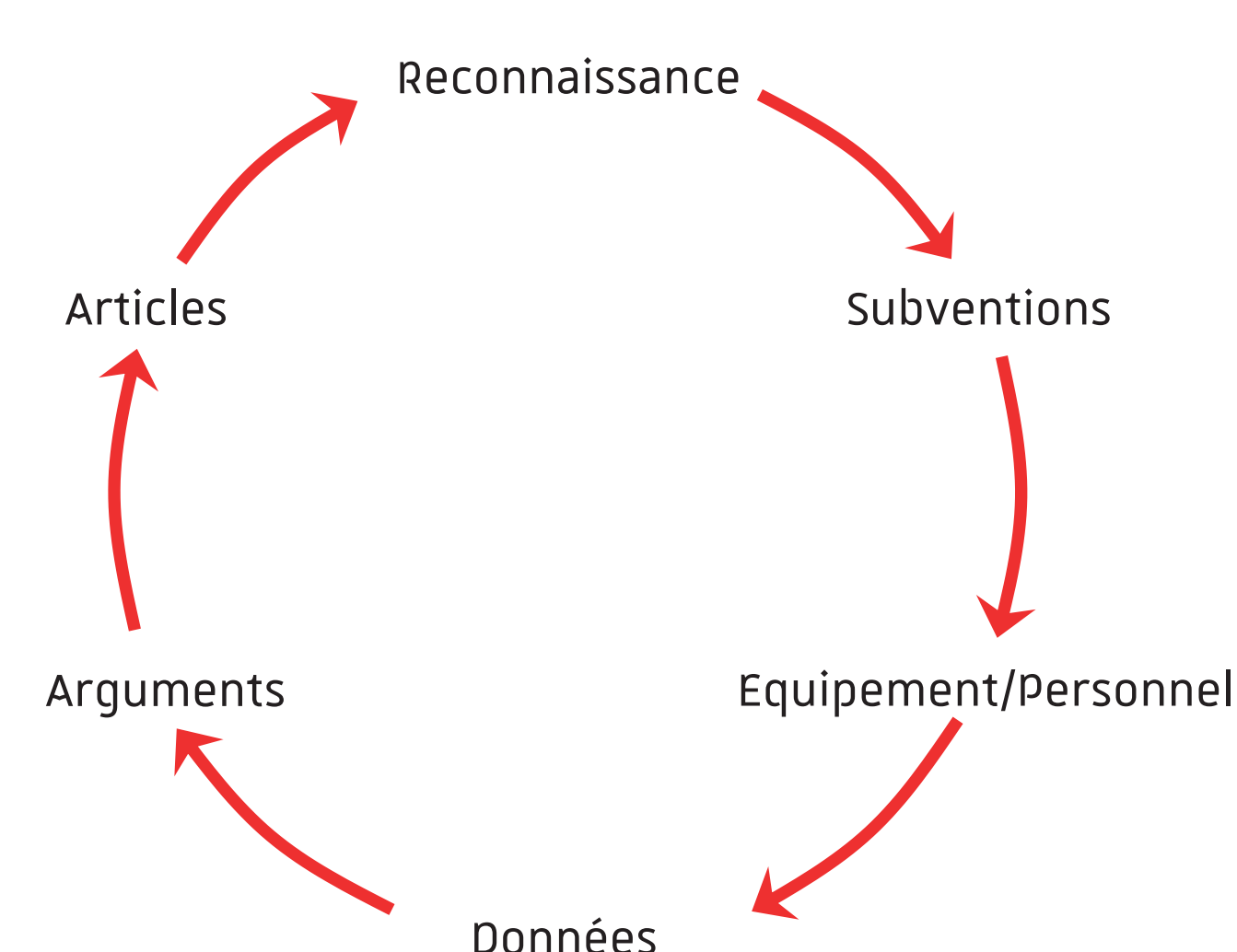
A cela s'ajoute le mode actuel de financement de la recherche, qui repose sur des subventions ponctuelles, par projet d'une durée de 3 à 6 ans.

L'obtention de financements réguliers fait donc partie des principales préoccupations des équipes de recherche que nous avons rencontrées. Or, comme nous l'avons vu, l'obtention de financements passe par le gain de « crédibilité » (voir 1.3.). C'est dans la manière d'acquiescer ce crédit que nous avons constaté des divergences entre recherche fondamentale et recherche appliquée.

2. Divergences dans la gestion des données

Etant en recherche quasi-constante de financements, les coordinateurs des projets de recherche fondamentale nous ont semblé prioriser ce qui est directement pris en considération par les agences de financement, à savoir la publication dans des revues à haut facteur d'impact. Or, comme l'expliquent Fecher et al. (2017), dans ce système donnant la priorité à la publication d'articles, les données de recherche ne possèdent qu'une faible valeur d'échange. En découle une certaine négligence quant à leur structuration et leur conservation.

Figure 1 : Cycle de crédibilité des projets de recherche fondamentale

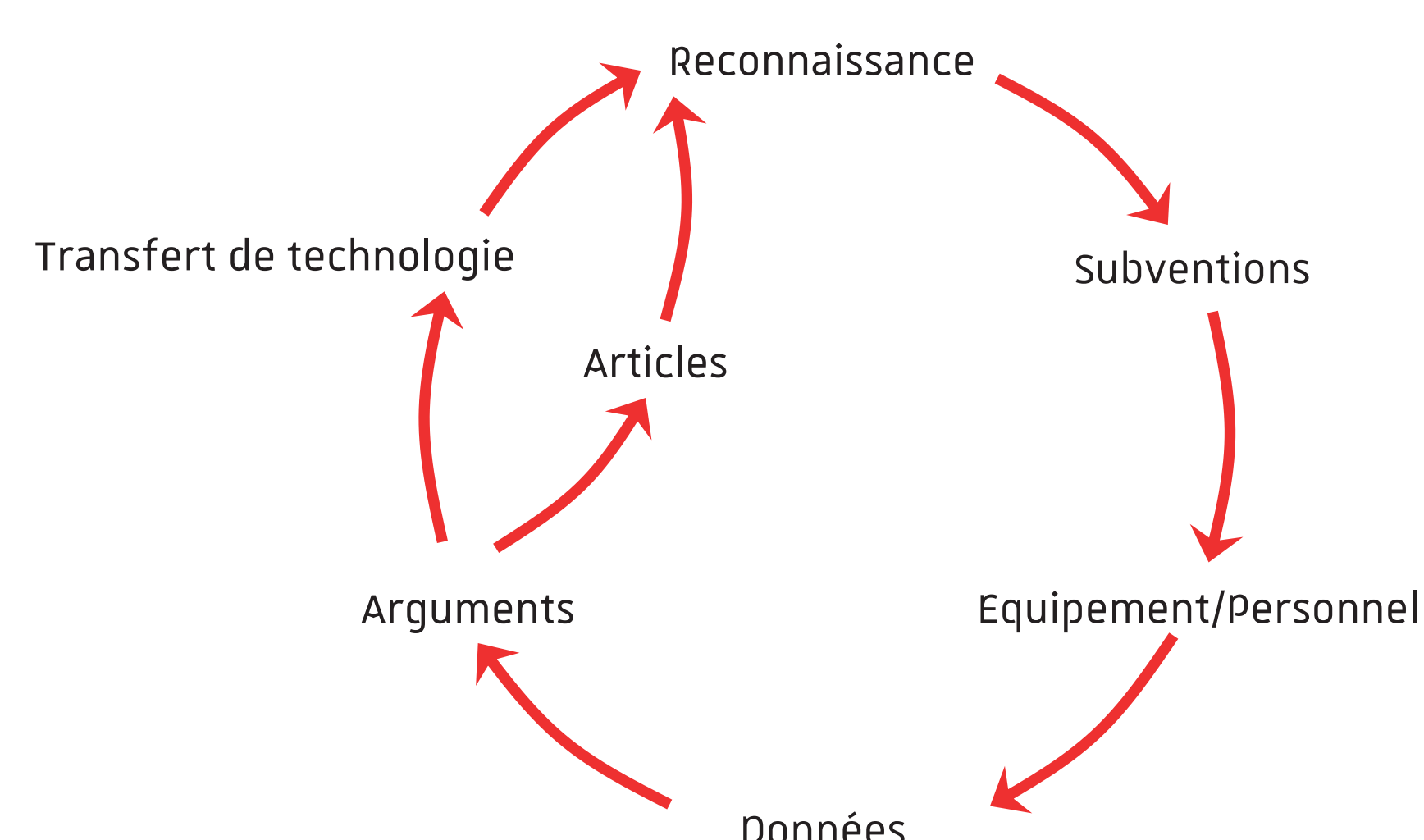


Dans la recherche appliquée, l'importance de publier dans des revues à haut facteur d'impact existe bel et bien, mais elle se double d'un souci d'assurer le transfert de la technologie qui a été créée. Car le transfert de technologie est lui aussi source de crédibilité. Il apporte des bénéfices financiers aux établissements de recherche, garantissant ainsi une confiance renouvelée dans l'équipe de chercheurs concernée.

Or cela a une incidence sur la manière dont sont gérées les données de recherche. Pour garantir la reprise d'une invention par l'industrie, il est nécessaire que le futur repreneur ait confiance dans les données. Selon les chercheurs que nous avons interrogés, cela implique que les données aient été :

- ◆ Validées (vérification de la viabilité d'un résultat avant d'entamer l'étape suivante de la recherche)
- ◆ Conservées (traçabilité des données via la tenue rigoureuse des cahiers de laboratoire + sauvegarde de l'ensemble des données générées et répliquées)
- ◆ Protégées (préservation du caractère inédit des données, en assurant leur confidentialité)

Figure 2 : Cycle de crédibilité des projets de recherche appliquée



CONCLUSION

Au vu des premiers résultats de cette enquête, il semblerait que les projets de recherche appliquée en sciences expérimentales assurent une gestion des données plus attentive et rationnelle que les projets de recherche fondamentale. Cette enquête laisse en tout cas entrevoir l'existence de modes de gestion fortement liés à l'objectif de la recherche.

Il conviendrait néanmoins de vérifier si ces conclusions se retrouvent dans d'autres projets ou d'autres disciplines scientifiques. Il conviendrait également d'appréhender ces pratiques d'un peu plus près, afin d'en comprendre le fonctionnement et les enjeux. Nous pensons en effet que c'est en s'appuyant sur l'existant et en co-construisant avec les communautés de recherche que la politique de structuration et d'ouverture des données aura le plus de chance d'être adoptée par les scientifiques.

RÉFÉRENCES

- Borgman C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge, The MIT Press.
- Commission européenne (2018). *Recommandation de la Commission du 25.4.2018 relative à l'accès aux informations scientifiques et à leur conservation*. <http://data.europa.eu/eli/reco/2018/790/oj>
- Fecher B., Friesike S., Hebing M., Linek S. (2017). « A reputation economy: how individual reward considerations trump systemic arguments for open access to data ». In *Palgrave Communications*, vol. 3. doi:10.1057/palcomms.2017.51
- Latour B. (2001). *Le métier de chercheur : regard d'un anthropologue*. Versailles, Editions Quæ. doi:10.3917/quaee.latou.2001.01