

Un «connectome de données de recherche» pour sécuriser et utiliser à long terme le «patrimoine national de recherche»

Une première idée de SWITCH

Sommaire exécutif

À l'heure actuelle, des données de recherche accessibles et utilisables constituent la principale ressource permettant de garantir la force d'innovation et la capacité de recherche de la Suisse. La question de savoir comment rendre ces données de recherche largement accessibles et durablement utilisables comme éléments du **patrimoine national de recherche** a fait germer l'idée d'un **connectome de données de recherche**. Un connectome est un objet abstrait qui relie les recueils aux objectifs prédéfinis. Le présent article décrit ses particularités. Cet article cherche à travailler collectivement à la concrétisation de la vision d'une science ouverte et donc de faire progresser cette concrétisation.

Situation actuelle

Il est particulièrement important pour la Suisse, un pays pauvre en ressources naturelles, d'exploiter au mieux les potentiels de l'avenir numérique en vue de renforcer la capacité d'innovation et donc de sécuriser à long terme le positionnement de recherche et d'innovation, de maintenir l'emploi et de favoriser le bien-être de la Suisse [1, 2, 3]. Il faut pour cela des données accessibles et utilisables. Ces données représentent les matières premières les plus précieuses pour dégager de nouveaux résultats de recherche et pour faire éclore une nouvelle capacité d'innovation, en particulier dans le contexte de la mutation du secteur suisse de la formation et de la recherche qui fait face à de nouveaux défis complexes [4, 9]. Cette mutation a déjà été abordée partiellement au niveau fédéral par des initiatives nationales [5, 6, 9, 10, 11] ou programmée pour les années à venir [13]. Outre les domaines d'actions identifiés, les interactions d'infrastructures TIC de base jouent un rôle particulier vis-à-vis du stockage à long terme, d'une large accessibilité et d'une utilisation durable des données de recherche^a. L'accessibilité des données de recherche est considérée comme un aspect essentiel à l'amélioration de l'efficacité, de la transparence et de la reproductibilité de la recherche scientifique [5, 6, 14]. L'accessibilité des données de recherche n'est pas suffisante à elle seule dans le contexte de la science ouverte. De nouvelles technologies, telles que le Big Data et l'intelligence artificielle, et de nouveaux services pourraient rendre ces données de recherche durablement utilisables et donc les transformer en catalyseur permettant d'obtenir de nouveaux résultats de recherche et d'améliorer la capacité d'innovation économique.

Les données scientifiques de projets de recherche financés par les pouvoirs publics devraient par conséquent être considérées comme un précieux **patrimoine national de recherche**. Le FNS a reconnu l'importance de cette évolution pour les projets de recherche nationaux et exige que les demandes de subventions publiques comportent obligatoirement un «Data Management Plan» (DMP). Pour une utilisation correspondante, le FNS exige la publication et l'accessibilité de données de recherche dans des recueils publics qui n'ont pas de visée lucrative.

Formulation de la problématique

Les données de recherche sont actuellement déposées dans des recueils soit spécifiques à un domaine (également à l'international) soit institutionnels. L'écosystème qui en résulte se développe continuellement et de manière dynamique, si bien qu'il devient complexe et qu'il manque de clarté [8]. Cette complexité a un impact sur les coûts et sur l'efficacité de l'ensemble de l'écosystème de recherche. Les lacunes existantes doivent être comblées et les doublons, notamment dans les petits recueils locaux, doivent être éliminés par une harmonisation adaptée.

Les «collecteurs de données» commerciaux tels que Google ou les maisons d'édition telles que Reed-Elsevier discernent eux aussi la valeur de ces données de recherche et étendent leur portefeuille de produits par des collections de données de recherche. D'importantes sources d'informations financées par les pouvoirs publics tombent ainsi dans l'escarcelle d'opérateurs commerciaux et sont privatisées et commercialisées de manière indésirable, faute d'alternative. L'économie du pays s'en trouve compromise. Les hautes écoles et les établissements de recherche suisses cherchent par conséquent à stocker de manière centralisée les données de recherche de leurs propres chercheurs et à les rendre accessibles, tout du moins institutionnellement.

a «Forschungsdaten sind erhobenes, beobachtetes oder generiertes Faktenmaterial, das in der wissenschaftlichen Gemeinschaft allgemein als notwendig für die Dokumentierung und Validierung von Forschungsergebnissen angesehen wird.» [5], cf. aussi [6].

On constate que de précieuses données de recherche nationales sont actuellement stockées de manière décentralisée^b et qu'elles ne sont pas reliées au-delà des domaines ou des institutions, si bien que d'autres acteurs n'y ont pas accès et ne peuvent donc les utiliser.

Proposition de solution

Pour libérer des capacités d'innovation insoupçonnées dans des domaines connus et nouveaux mais aussi interdisciplinaires (par exemple en médecine ou dans le changement climatique) et donner plus de poids à la science en Suisse face à la concurrence internationale, il faut réviser entièrement notre manière de réfléchir. Le **patrimoine national de recherche** doit être protégé et les données de recherche précieuses doivent être largement rendues accessibles et utilisables. Ce n'est qu'ainsi que nous pourrions jeter des bases durables pour la mise en œuvre d'une vision de la science ouverte en Suisse. L'énumération suivante définit, de manière non exhaustive, les principales phases de cette perspective globale:

1. Mise en place d'un connectome des données de recherche

Les données de recherche stockées de manière décentralisée mais financées à l'échelon national doivent être identifiées selon une approche uniforme, interconnectées au travers d'un **connectome des données de recherche**, protégées, rendues accessibles à tous les destinataires potentiels en fonction du groupe cible et rendues durablement utilisables par des méthodes novatrices (voir l'illustration 1 en page 3).

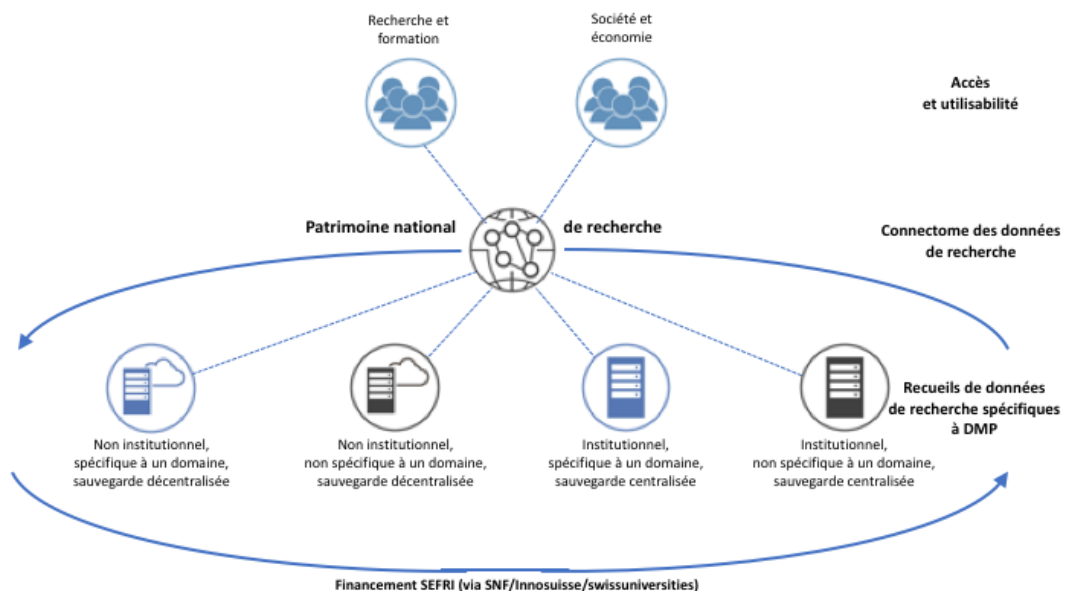


Illustration 1: Patrimoine national de recherche

^b Stockage/archivage décentralisé du fait de domaines de recherche, d'institutions, d'éditeurs, etc. individuels

2. Unir les forces nationales

Cette idée ne peut se concrétiser qu'avec une approche uniforme et concertée impliquant les «bons partenaires» (intérêts, ressources, compétences et bilans). L'interconnexion de données de recherche des recueils financées à l'échelon national est dans l'intérêt des exigences en matière de recherche du FNS et d'Innosuisse. Ces institutions doivent soutenir la démarche et s'engager avec vigueur pour poursuivre des intérêts et des efforts communs. À titre d'exemple, le FNS doit introduire des consignes correspondantes destinées aux chercheurs lorsqu'il attribue des subventions et imposer les «Data Management Plans» (DMP). D'autres acteurs tels que les académies des sciences de Suisse devraient être impliqués sur la base de leurs propres atouts (par exemple le regroupement des domaines). Ce n'est qu'à ce prix que le circuit de financement (voir l'illustration 1) des données pourra être bouclé durablement.

3. Repenser les modèles de financement actuels

L'approche pratiquée jusqu'à présent consistant à délivrer des «financements initiaux» liés à des projets^[9,12] pour des projets (de petite envergure) initiés par le bas peut être intéressante pour le lancement de premières idées. Pour assurer le fonctionnement à long terme et le développement de ces services nationaux, une base institutionnelle correspondante et un financement durable sont indispensables^[17]. Il faut en outre tenir compte du fait que les initiatives individuelles actuelles ne peuvent être transférées que de manière limitée dans un ensemble supérieur, si bien que des lacunes et des doublons continuent d'exister. Une approche initiée par le bas ne fonctionne pas pour l'idée esquissée ci-dessus d'un **connectome des données de recherche**. La complexité et les coûts de l'écosystème s'en trouveraient accrus et la puissance d'impact de la science ouverte serait largement réduite. La réalisation de cette idée doit donc être détachée des structures de financement en place.

4. Financement centralisé

La satisfaction des besoins économiques de la science ouverte ne peut par conséquent être obtenue qu'au travers d'intérêts supérieurs de l'État mis en mouvement à la vitesse et sous la pression requises. La mise en œuvre nécessite un financement centralisé du **connectome des données de recherche** associé à un mandat national de mise en œuvre qui doit être suivi stratégiquement par un consortium disposant des compétences et des qualifications nécessaires. Le SEFRI, le FNS et Innosuisse sont particulièrement sollicités ici.

Pourquoi SWITCH lance cette idée

La fondation SWITCH fait partie intégrante de la communauté des hautes écoles suisses et s'engage en faveur de la qualification des utilisateurs à l'intérieur et à l'extérieur du monde académique sur la base des compétences clés en matière de réseau, de sécurité informatique et de gestion des identités. L'idée esquissée ci-dessus d'un connectome des données de recherche correspond à notre vision consistant à délivrer des solutions TIC sûres, efficaces et globales pour le site de formation et de recherche qu'est la Suisse.

Sources

- [1] Cf. SEFRI – «Herausforderungen der Digitalisierung für Bildung und Forschung in der Schweiz» ([lien](#), accès 17.01.2019).
- [2] Cf. OFCOM – «Strategie Digitale Schweiz» ([lien](#), accès 17.01.2019).
- [3] Cf. Le Conseil fédéral – «Bundesrat verabschiedet Bericht zu Rahmenbedingungen der digitalen Wirtschaft», communiqué de presse ([lien](#), accès 17.01.2019).
- [4] Cf. SEFRI – «Digitalisierung - Aktionsplan im Bereich Bildung, Forschung und Innovation in den Jahren 2019-2020» ([lien](#), accès 17.01.2019).
- [5] Cf. FNS – «Open Research Data» ([lien](#), accès 17.01.2019).
- [6] Cf. FNS – «Open Science» ([lien](#), accès 17.01.2019).
- [7] Cf. FNS – «Data Management Plan (DMP) – Leitlinien für Forschende» ([lien](#), accès 17.01.2019).
- [8] Cf. FNS – «Call for tender for a landscape and cost analysis of data repositories» ([lien](#), accès 17.01.2019).
- [9] Cf. CRUS – «Bündelung der Kräfte in der wissenschaftlichen Information, Nationale Strategie» ([lien](#), accès 17.01.2019).
- [10] Cf. CRUS – «White Paper for a Swiss Information Provisioning and Processing Infrastructure 2020» ([lien](#), accès 17.01.2019).
- [11] Cf. Swissuniversities – «Hauptstossrichtung Publikationen Umsetzungsmassnahme EP-10, Digitalisierung – Angepasste Strategie und Umsetzungsmassnahmen» ([lien](#)(en allemand), accès 17.01.2019).
- [12] Cf. Swissuniversities – «Wissenschaftliche Information: Zugang, Verarbeitung und Speicherung» ([lien](#), accès 25.01.2019).
- [13] Cf. Swissuniversities – «Stratégie 2021-2024 en «information scientifique» (lien n. d., accès 22.01.2019).
- [14] Cf. UKRI – «Concordat on Open Research Data» ([lien](#), accès 17.01.2019).
- [15] Cf. SWITCH – «Governance» ([lien](#), accès 25.01.2019).
- [16] Cf. SWITCH – «Gemeinsam für mehr Leistung, Komfort und Sicherheit in der digitalen Welt» ([lien](#), accès 25.01.2019).
- [17] Cf. NICT – «Whitepaper - An NICT View on the Research Data Management Landscape in Switzerland » (ébauche, accès, 28.01.2019).