Préserver la souveraineté numérique grâce à la gestion et à la conservation au niveau local

Traduction en français de : Safeguarding digital sovereignty through local stewardship and preservation
https://zenodo.org/records/20718121

CCSD – June 2026

La montée en puissance de la souveraineté numérique

Au cours du dernier siècle, l'écosystème de la communication savante est passé d'un paysage composée de nombreuses petites revues indépendantes servant des communautés scientifiques discrètes à un environnement dominé par quelques grands éditeurs commerciaux, chacun gérant des catalogues de milliers de revues et continuant à acquérir des concurrents et des services associés (Larivière et al., 2015). Le résultat est un marché quasi monopolistique dans lequel cinq entreprises (basées principalement aux États-Unis et en Europe) contrôlent désormais la majorité de la production scientifique mondiale. Les communautés scientifiques ont ainsi perdu une grande partie de leur gouvernance et de leur influence sur les plateformes et les infrastructures dont elles dépendent (Fyfe et al., 2017). Pour aggraver encore la situation, les éditeurs commerciaux ont tendance à privilégier les contenus en anglais et les sujets qui intéressent les communautés de recherche occidentales, souvent au détriment des travaux pertinents au niveau local.

Pour ces raisons, entre autres, de nombreux pays réévaluent la manière dont ils hébergent et préservent leurs productions de recherche (Jansen et al., 2023). Ce changement reflète une tendance plus large dans le paysage géopolitique, où les nations investissent de plus en plus dans des infrastructures numériques souveraines, non seulement pour la recherche, mais aussi comme alternative aux grandes plateformes commerciales dont leurs populations sont devenues dépendantes.

Ces préoccupations vont bien au-delà des publications. Les jeux de données, les outils d'analyse et les plateformes sur lesquels les chercheurs s'appuient dépendent souvent d'infrastructures numériques externes, hébergées et gouvernées en dehors des régions qui les produisent et les utilisent (Ruckstuhl, 2022 ; Gu, 2023). À mesure que les données de recherche deviennent de plus en plus précieuses, notamment en tant que matière première pour les systèmes d'intelligence artificielle (Digital Science et al., 2025), l'emplacement où elles sont stockées et ceux qui gouvernent leur utilisation ont des implications qui vont bien au-delà de la communauté de la recherche.

L'urgence de ces efforts en faveur de la souveraineté devient évidente lorsque des perturbations réelles surviennent. Par exemple, lorsque l'Agence américaine pour le

développement international (USAID) a subi des réductions drastiques de financement au début de 2025, des chercheurs de plusieurs pays ont soudainement risqué de perdre l'accès à des jeux de données qu'ils avaient contribué à créer. Cet épisode a illustré une vulnérabilité structurelle : la dépendance à des plateformes étrangères peut transformer une décision politique lointaine en une crise locale immédiate, sans avertissement ni recours.

Dans le domaine de la recherche, la revendication de la souveraineté des données a été particulièrement marquée en Amérique latine et en Afrique, deux régions qui ont longtemps été soumises à des relations d'exploitation de la part de la science occidentale. Les chercheurs issus de pays plus riches se sont régulièrement rendus dans ces régions pour y mener des études, puis sont rentrés chez eux avec les résultats obtenus, ne laissant pratiquement rien derrière eux.

Une publication brésilienne récente résume plusieurs dimensions structurelles de ce problème :

> *« Lorsque les chercheurs du Sud global déposent leurs données dans des archives contrôlées par des entreprises étrangères, utilisent des plateformes d'analyse propriétaires ou s'appuient sur des infrastructures cloud situées dans d'autres pays, ils cèdent, dans une certaine mesure, leur souveraineté sur une ressource stratégique. »* (da Silva, 2025)

Des réponses à cette dynamique ont émergé dans ces deux régions. En Amérique latine, depuis plus de trois décennies, des investissements significatifs ont été réalisés dans des infrastructures scientifiques locales, aboutissant à la création de deux plateformes majeures de publication et d'hébergement (Redalyc et SciELO), ainsi que de services régionaux de découverte et d'indexation. Il convient de noter que, en plus de garantir la souveraineté locale, ces infrastructures soutiennent les publications en espagnol et en portugais, les deux principales langues de l'Amérique latine. En Afrique, des objectifs similaires ont été poursuivis, bien que de manière moins coordonnée, à travers des initiatives telles que African Journals Online et le programme LIBSENSE de WACREN, qui visent à garantir que les recherches produites sur le continent soient gérées et préservées localement (Abbott, 2025 ; Abbott et al., 2025).

Ces discussions ont désormais lieu dans d'autres régions également. La Chine, devenue le premier producteur mondial de publications scientifiques (Pisani et al., 2025), a principalement publié des articles dans des revues hébergées en dehors du pays, avec seulement un très faible pourcentage (environ 5 %) publié dans des revues détenues par des Chinois (Yang & Huang, 2025). Pour remédier à ce déséquilibre, la Chine a récemment exprimé son intention de développer son infrastructure nationale pour la publication et la gestion de ses productions de recherche (Owens, 2024 ; Yang & Huang, 2025). Au Canada, dans le but de garantir la conservation de copies en local, les principaux organismes fédéraux de financement de la recherche ont proposé une politique d'accès ouvert exigeant que tous les articles évalués par des pairs issus de recherches financées par des fonds publics soient déposés dans un dépôt institutionnel canadien dès leur publication, indépendamment de leur disponibilité sur la plateforme de l'éditeur (Conseils des trois organismes de recherche canadiens, 2025). Cette politique vise explicitement à garantir que des copies de toutes les recherches financées par le Canada soient préservées au Canada et accessibles au public canadien.

En Europe, des questions similaires se posent concernant le contrôle des infrastructures de recherche critiques. Le Groupe de réflexion du Parlement européen sur l'avenir de la science et de la technologie a récemment organisé une discussion sur la souveraineté des données dans la recherche, mettant en garde contre les risques stratégiques que la dépendance aux bases de données externes et commerciales pourrait poser pour la liberté académique, l'innovation et la continuité de la recherche (Parlement européen, 2025). Par la suite, le Comité de pilotage de l'EOSC a publié un avis recommandant à l'Europe de renforcer sa capacité à gouverner et à protéger les productions de recherche en investissant dans une infrastructure souveraine fédérée (Commission européenne 2025). Ces préoccupations se concrétisent désormais par l'adoption de plusieurs lois et réglementations au sein de l'Union européenne, comme le Digital Services Act (2022), ainsi que par des mesures nationales et un renforcement des investissements dans les infrastructures locales.

Ces initiatives en matière de souveraineté numérique varient tant par leur forme que par leur ambition, mais elles partagent une préoccupation commune : si les systèmes qui stockent les résultats de la recherche d'un pays et en permettent l'accès sont hébergés et gérés ailleurs, que se passe-t-il lorsque les conditions d'accès changent, que les relations politiques se détériorent ou que les services sont tout simplement interrompus ?

Le rôle des dépôts dans la mise en oeuvre de la souveraineté numérique

La souveraineté numérique dans la recherche désigne la capacité des chercheurs, des institutions et des nations à contrôler, accéder et préserver leurs productions scientifiques sans dépendre d'intermédiaires commerciaux ou de juridictions étrangères. Bien qu'il soit extrêmement difficile pour de nombreux pays d'atteindre une indépendance technologique totale (les infrastructures numériques sous-jacentes étant souvent réparties à l'échelle mondiale et hors du contrôle de toute entité unique), ils peuvent contrôler l'endroit où leurs productions sont stockées. C'est précisément la fonction des archives : veiller à ce que les ressources nationales précieuses issues de la recherche soient conservées au sein d'une infrastructure fiable et gérée localement.

Les dépôts institutionnels font partie du paysage scientifique depuis plus de deux décennies, avec les premières plateformes open source apparues au début des années 2000, en même temps que l'Initiative de Budapest pour l'accès ouvert, qui a identifié les archives comme une infrastructure essentielle pour rendre la recherche librement accessible (Initiative de Budapest pour l'accès ouvert, 2002 ; Rothfritz et al., 2025). Depuis, les dépôts se sont étendus dans toutes les régions du monde, formant un réseau large et distribué d'infrastructures principalement hébergées par des institutions. Pour les pays cherchant à accroître leur souveraineté en matière de recherche, elles constituent un outil important qui peut être exploité pour soutenir ces objectifs.

Dans le contexte de la souveraineté numérique, la fonction de conservation des dépôts prend une signification particulière. Les dépôts sont conçus pour gérer les productions de recherche sous la responsabilité des institutions et des communautés qui les ont produites, offrant une alternative beaucoup plus sûre et responsable que de confier le patrimoine de recherche d'une nation à des entités commerciales étrangères. De plus, la nature décentralisée du réseau des

dépôts implique que cette conservation ne dépend d'aucune institution ni d'aucune plateforme en particulier.

Bien qu'une solution tentante pour la gestion nationale des productions de recherche consiste à tout centraliser sur une seule plateforme nationale, ce type de concentration comporte un risque inhérent : un point de défaillance unique et une architecture "taille unique" qui risque de ne pas répondre aux besoins variés des différentes disciplines, institutions et communautés de recherche. Les réseaux de dépôts offrent un modèle fondamentalement différent : fédéré plutôt que centralisé, chaque nœud conservant sa propre gouvernance, son propre périmètre et ses propres fonctionnalités, tout en restant connecté aux réseaux nationaux et mondiaux grâce à des protocoles et normes partagés. Dans cette optique, la souveraineté n'exige pas l'uniformité, mais plutôt l'interopérabilité.

L'interopérabilité au service de la résilience et la visibilité mondiale

Une préoccupation légitime concernant la tendance vers la souveraineté numérique est que les infrastructures construites localement pourraient devenir inaccessibles au-delà des frontières nationales ou parce qu'elles n'ont pas été développées en tenant compte des normes internationales, contribuant ainsi à une fragmentation accrue du domaine public mondial de la connaissance (Silva, 2025). Cependant, une souveraineté recherchée par l'isolement est contre productive : si les pays et les institutions se replient dans des « jardins clos », la connectivité dont dépend la recherche internationale s'en trouvera compromise. L'interopérabilité internationale constitue donc une exigence fondamentale pour ces services nationaux et institutionnels. C'est le mécanisme qui permet aux nations de conserver un contrôle local tout en contribuant au système mondial de l'information.

L'interopérabilité est déjà un principe de conception inhérent à l'architecture des dépôts. La plupart des plateformes largement utilisées prennent en charge des API et protocoles communs, tels que le protocole OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), qui permettent la collecte et la découvrabilité des contenus. Mais cette compatibilité technique de base est insuffisante à elle seule. Pour qu'un dépôt soit interopérable et découvrable au-delà de son contexte local ou national, il doit répondre à un ensemble d'exigences à plusieurs niveaux : un étiquetage multilingue, des licences ouvertes, des métadonnées de haute qualité soutenues par des identifiants pérennes (PIDs), et des protocoles et API émergents qui permettent le lien (COAR Notify) entre les ressources et un accès fiable par machine. Il s'agit là de facteurs essentiels pour garantir que les ressources d'un dépôt soient visibles au-delà des utilisateurs directs de la plateforme, grâce à leur indexation, leur collecte et leur exploration par les agrégateurs et les moteurs de recherche.

Renforcer les dépôts pour relever le défi

La contribution des dépôts à la souveraineté numérique est bien établie, mais un certain nombre d'améliorations structurelles et opérationnelles pourraient renforcer leur efficacité à cet égard. La capacité de conservation est peut-être l'une des lacunes les plus significatives. La conservation numérique englobe la gestion active et continue des objets numériques afin

de les protéger contre la dégradation des bits, l'obsolescence des formats et la discontinuité institutionnelle. Cela implique des processus techniques complexes et gourmands en ressources. La plupart des dépôts en libre accès répondent à certaines exigences liées à la conservation numérique, telles que des copies de sauvegarde et le contrôle de l'intégrité des fichiers ; toutefois, rares sont ceux qui parviennent à satisfaire aux cadres de conservation les plus complets, comme l'OAIS (Open Archival Information System), qui définit la norme professionnelle en matière de gestion archivistique à long terme.

La plupart des dépôts ne répondent pas aux exigences strictes en matière de conservation numérique, et ce pour diverses raisons d'ordre structurel, technique et de gouvernance. Le financement est peut-être la contrainte la plus fondamentale : la conservation nécessite des investissements continus dans le stockage, la migration et le personnel spécialisé, que beaucoup d'archives ne peuvent soutenir (Barrueco & Termens, 2020). De plus, de nombreuses plateformes de dépôts ont été conçues principalement pour l'accès et la découverte, plutôt que pour la gestion à long terme. Elles manquent donc souvent de l'infrastructure sous-jacente nécessaire pour la vérification de l'intégrité des documents (*fixity checking*), la migration des formats, ou le stockage redondant réparti géographiquement, comme l'exigent les normes de conservation (Rieger et al., 2022). Au niveau de la gouvernance, de nombreuses archives fonctionnent sans politiques de conservation formelles, laissant l'engagement institutionnel informel et vulnérable aux changements de direction ou de priorités (Dowding, 2016).

Face à ces défis, une approche plus réaliste pour une conservation rigoureuse des productions de recherche dans les dépôts pourrait consister à intégrer les archives institutionnelles à des initiatives nationales de conservation plus complètes, qui fournissent l'expertise et les ressources nécessaires. Les lacunes actuelles en matière de capacités de conservation numérique des dépôts ne remettent toutefois pas en cause l'importance de leur rôle, qui consiste à assurer un niveau fondamental et irremplaçable de gestion locale des productions de recherche d'une institution et d'un pays.

L'interopérabilité reste également un défi permanent. De nombreux dépôts n'ont pas encore pleinement adopté les pratiques techniques et en matière de métadonnées nécessaires pour s'interconnecter avec d'autres systèmes, un problème qui touche à la fois l'infrastructure et les capacités institutionnelles (Rothfritz et al., 2025). Le manque de ressources et les effectifs insuffisants entraînent des pratiques incohérentes en matière de métadonnées d'un dépôt à l'autre, ce qui compromet la capacité des agrégateurs et des moteurs de recherche à indexer leur contenu de manière fiable. Ce problème n'est pas nouveau : dès 2010, Park et Tosaka avaient constaté que la précision et la cohérence des métadonnées étaient considérées comme les indicateurs de qualité les plus importants, mais aussi les plus difficiles à maintenir, un constat qui, plus d'une décennie plus tard, reste largement d'actualité.

L'un des obstacles les plus tenaces à l'optimisation de l'exploitation des dépôts réside dans la combinaison d'un manque de personnel et d'une infrastructure logicielle vieillissante. Une revue systématique réalisée en 2025 a révélé que 82 % des études identifiaient la professionnalisation comme un problème majeur, la pénurie de personnel qualifié étant constatée dans les dépôts de données de pays situés dans toutes les régions.(Shearer et al., 2023 ; Rothfritz et al., 2025 ; Macgregor & Davidson, 2026). Les conséquences sont visibles

au sein même de l'infrastructure : des systèmes obsolètes, des interfaces inutilisées et des dépôts laissés à l'abandon.

La fragilité de nombreux dépôts reflète le faible niveau de priorité que les bibliothèques et les institutions ont traditionnellement accordé à leurs opérations. Les archives ont souvent occupé une position marginale au sein de leur institution, une situation qui reflète une convergence de facteurs structurels, culturels et financiers. Dans la plupart des institutions, les dépôts ont été créés comme des mécanismes de conformité pour les politiques d'accès ouvert, plutôt que comme des infrastructures stratégiquement valorisées en tant que telles (Rothfritz et al., 2026). En conséquence, ils ont généralement reçu le minimum d'investissements nécessaires pour rester opérationnels, les questions de personnel, de développement technique et de communication étant considérés comme des dépenses secondaires plutôt que essentielles. Ce positionnement est renforcé par les mécanismes de récompense en vigueur dans les établissements universitaires, qui ont historiquement privilégié les résultats de la recherche — publications, subventions et citations — au détriment des infrastructures qui permettent leur gestion et leur diffusion (Hahn, 2008). L'effet cumulatif conduit à un modèle chronique de sous-investissement qui a laissé de nombreux dépôts techniquement obsolètes, sous-dotés en personnel, et mal intégrés dans l'écosystème plus large de la recherche et de l'édition. Ces défis devront être relevés à mesure que nous développerons nos dépôts afin de répondre aux exigences croissantes en matière de souveraineté numérique.

Conclusion

La concentration de l'infrastructure scientifique entre les mains d'un petit nombre d'éditeurs commerciaux et de plateformes étrangères a créé des dépendances dont de nombreux pays commencent seulement aujourd'hui à prendre pleinement la mesure. À mesure que le paysage géopolitique évolue et que la fragilité de ces dépendances devient de plus en plus difficile à ignorer, la question de savoir qui contrôle les systèmes qui stockent, gèrent et donnent accès aux résultats de la recherche nationale passe du statut de préoccupation de niche à celui de priorité stratégique.

Les dépôts ne sont pas une nouvelle réponse à ce défi. Ils font partie du paysage scientifique depuis plus de deux décennies, remplissant discrètement des fonctions qui ne sont que maintenant reconnues comme géopolitiquement significatives. Leur architecture décentralisée, régie par des institutions, les rend particulièrement bien adaptés à un modèle fédéré de souveraineté : un modèle qui préserve le contrôle local sans sacrifier la connectivité mondiale, et qui peut s'adapter aux besoins variés des différentes disciplines, langues et traditions de recherche. En d'autres termes, l'infrastructure existe déjà. Ce qui a fait défaut jusqu'à présent, c'est la reconnaissance de sa valeur stratégique et les investissements nécessaires pour exploiter pleinement son potentiel.

Cet investissement est en retard. Le manque chronique de personnel, le vieillissement des systèmes, les pratiques de métadonnées incohérentes et la capacité de conservation limitée ne sont pas des faiblesses inhérentes au modèle des dépôts. Ce sont les conséquences accumulées du fait de considérer les dépôts comme des outils de second plan plutôt que comme un atout national essentiel. Pour y remédier, il faudra un financement durable, des infrastructures partagées, une politique coordonnée et la volonté d'accorder à l'exploitation

des dépôts le même sérieux que celui dont les gouvernements commencent aujourd'hui à faire preuve à l'égard de la souveraineté numérique en général.

Les enjeux sont clairs : lorsque les productions de recherche sont hébergées sur des plateformes gouvernées par des institutions étrangères ou des intérêts commerciaux, une décision politique lointaine ou un changement de modèle économique peut rompre l’accès du jour au lendemain, sans avertissement ni recours. Les archives, correctement financées et interconnectées, offrent quelque chose de différent : une fondation durable et gouvernée localement sur laquelle une véritable souveraineté de la recherche peut être construite. L’élan est là. Ce qui est nécessaire maintenant, c’est l’engagement pour le soutenir.