

Outils et circulation des données en Europe : état des lieux par le consortium euroCRIS

Ce tour d'horizon des systèmes d'information sur la recherche (CRIS) en Europe, met en lumière les outils clés de gestion des données scientifiques et les initiatives en cours pour favoriser leur interopérabilité à l'échelle internationale.



L'INFORMATION SUR LA RECHERCHE

L'information sur la recherche joue aujourd'hui un rôle central dans le paysage scientifique, tant du point de vue de la gouvernance que de celui de la diffusion des connaissances. Selon le Comité pour la science ouverte, cette information regroupe un vaste ensemble de métadonnées qui documentent les différentes étapes de la recherche, depuis sa conduite jusqu'à sa communication^[1]. Ce corpus d'informations, bien que fondamental, est encore trop souvent dispersé, fragmenté, cloisonné entre bases de données hétérogènes issues des services de ressources humaines, des directions financières, des catalogues documentaires ou des archives institutionnelles. Chacune utilise ses propres formats, ses propres standards, ce qui rend difficile leur interconnexion et leur exploitation à grande échelle. Au cœur de cette galaxie de données se trouvent des métadonnées bibliographiques classiques — titres, résumés, auteurs, affiliations — mais aussi des informations plus techniques ou structurelles : logiciels utilisés, instruments mobilisés, jeux de données produits, sources de financement, laboratoires ou institutions impliqués. Autant de composantes qui permettent, une fois réunies et mises en cohérence, de mieux comprendre, évaluer, valoriser et piloter la recherche.

LES SYSTÈMES D'INFORMATION SUR LA RECHERCHE

Pour répondre à ce besoin de cohérence, les systèmes d'information sur la recherche (ou *Research Information Systems*, souvent appelés CRIS pour *Current Research Information Systems*) ont émergé. Ces plateformes sont conçues pour agréger, structurer, stocker et parfois analyser les métadonnées liées à l'activité scientifique. Elles peuvent prendre des formes très diverses : portails nationaux ou régionaux d'agrégation, solutions logicielles commerciales ou open source, dispositifs développés en interne, voire « bricolages » locaux sous forme de tableurs Excel ou autre. Peu importe la forme, le principe reste le même : centraliser l'information

pour en améliorer la qualité, la visibilité et l'utilité. Un mouvement de convergence s'observe également entre ces systèmes CRIS et les archives ouvertes, qui constituent des réservoirs de publications scientifiques. Ces deux univers, autrefois distincts, tendent à se rapprocher. Leur articulation permet de mutualiser les efforts de saisie, de normaliser les processus de dépôt, et surtout de favoriser une évaluation plus fine et plus transparente de la recherche, dans une logique pleinement alignée avec les principes de la science ouverte. Les objectifs assignés à ces systèmes d'information sont multiples : mieux gérer les ressources, produire des indicateurs fiables pour les évaluations, assurer l'interopérabilité entre institutions et plateformes, tout en normalisant les formats de données. C'est dans ce cadre qu'intervient le standard CERIF (*Common European Research Information Format*)^[2], promu par l'organisation euroCRIS^[3]. Ce modèle de données vise précisément à garantir une communication fluide entre systèmes hétérogènes, en définissant une structure commune pour l'information scientifique.

Mais sur le terrain, la diversité reste la règle. Les systèmes CRIS sont aujourd'hui présents à l'échelle mondiale, avec des fonctions similaires — centralisation et gestion de l'information — mais des périmètres, des technologies et des modes de gouvernance très variables. Certains couvrent une seule institution, d'autres une région, un pays entier, voire plusieurs États coopérant entre eux. Les logiciels utilisés peuvent être propriétaires ou libres, développés sur mesure ou adaptés à partir de solutions existantes. Les architectures techniques varient elles aussi : systèmes locaux isolés, plateformes mutualisées entre établissements, hubs d'agrégation au niveau national.

LES DISPOSITIFS NATIONAUX

Au sein de l'espace européen, on observe différentes logiques d'architecture. Certains pays ont opté pour un système centralisé géré au niveau national. D'autres, à l'image de la France, s'appuient sur

[1] CoSO Actualités (16 avril 2024) <https://www.ouvrirlascience.fr/le-comite-pour-la-science-ouverte-est-signataire-de-la-declaration-de-barcelone-sur-louverture-des-informations-sur-la-recherche>

[2] CERIF <https://eurocris.org/services/main-features-cerif>

[3] euroCRIS <https://eurocris.org>



Adobe Stock - YoursTrulyPhotos, identité et gestion d'accès, généré à l'aide de l'IA.

des dispositifs mutualisés intégrant des instances locales interconnectées. D'autres encore préfèrent une architecture d'agrégation, où chaque institution conserve son autonomie tout en transmettant ses données à un portail central.

La gouvernance de ces systèmes diffère également. Dans certains cas, elle est assurée par un ministère, comme en Finlande. Dans d'autres, elle relève d'agences de financement, comme au Portugal. Ces choix ont une influence directe sur la nature des données collectées, leur format, la fréquence des échanges, et les obligations qui pèsent sur les chercheurs.

Prenons le cas de la Finlande. Chaque organisme de recherche y dispose de son propre système, mais les données sont automatiquement transférées vers un système national centralisé⁴, sans saisie manuelle supplémentaire, grâce à des formats XML communs. Ce transfert est obligatoire uniquement si l'institution souhaite recevoir des fonds publics. Le système est piloté par le ministère de l'Éducation et de la Culture, et le CSC (*Centre for Science Information Technology*) en assure la mise en œuvre technique. Cette approche garantit une remontée de données fluide, sans alourdir la charge des chercheurs. Elle permet également d'alimenter efficacement la politique scientifique nationale, d'accroître la réutilisation des données, de renforcer la visibilité des travaux et d'appuyer les processus d'évaluation. De cette expérience, deux leçons majeures émergent : la coopération entre acteurs est une condition indispensable, et il est préférable de ne pas bouleverser les processus internes des institutions, mais plutôt

de construire des ponts et des formats communs à un niveau supérieur.

Au Portugal, l'initiative est portée par l'agence de financement nationale FCT, équivalente de l'ANR. Elle a mis en place un véritable écosystème d'information, structuré autour d'un cadre normatif robuste⁵. Celui-ci inclut un modèle de données cohérent, un dictionnaire standardisé, l'usage d'identifiants internationaux comme ORCID ou ROR, et une nomenclature des disciplines basée sur la classification de l'UNESCO. Cette formalisation favorise l'interopérabilité à tous les niveaux du système, tout en assurant une gouvernance claire. D'autres pays européens ont également développé leurs propres systèmes. En Pologne⁶, la gestion est assurée par le ministère, et les données sont transférées à la fois vers les agences de financement et vers un portail national d'évaluation. En Norvège, le système CRISTIN — bien que désormais intégré dans le nouveau portail NVA⁷ — propose une double alimentation, automatique et manuelle. La plateforme européenne OpenAIRE peut servir de base pour un CRIS, ainsi que c'est le cas des Pays-Bas dont le portail de recherche⁸ s'appuie sur OpenAIRE. En effet, le graphe d'OpenAIRE s'inspire du modèle CERIF (auquel OpenAIRE est conforme, au moins partiellement), et les systèmes CRIS sont interopérables avec OpenAIRE via les *OpenAIRE guidelines*⁹, développées conjointement par euro-CRIS et OpenAIRE. Cependant, certaines métadonnées d'OpenAIRE, notamment sur les instruments (infrastructures) ou les projets de recherche, restent encore insuffisamment couvertes.

[4] Research.fi
<https://research.fi/en>

[5] PT-CRIS <https://ptcris.pt/en>

[6] POL-on <https://polon.nauka.gov.pl/en>

[7] NVA <https://nva.sikt.no>

[8] Netherlands Research Portal
<https://netherlands.openaire.eu>

[9] OpenAIRE Guidelines for CRIS Managers <https://guidelines.openaire.eu/en/latest/cris/index.html>

ENJEUX ET PERSPECTIVES

À travers ces expériences, plusieurs enjeux transversaux se dégagent. Le premier concerne la standardisation des formats, avec un rôle pivot joué par des modèles comme CERIF. Le deuxième enjeu porte sur les identifiants pérennes, devenus indispensables pour relier les entités : chercheurs (via ORCID), publications (DOI), institutions (ROR), financements (GrantID). Le troisième enjeu est celui de la qualité des données : leur nettoyage, le dédoublement, l'harmonisation sont nécessaires pour éviter les biais, les redondances ou les pertes d'information. Cela suppose aussi l'émergence de véritables communautés d'utilisateurs, qui puissent partager des bonnes pratiques et construire ensemble des référentiels communs¹⁰. Enfin, se pose la question d'un cadre éthique pour l'usage de ces identifiants, afin de protéger les données personnelles et garantir leur usage responsable.

Dernier facteur de transformation, mais non des moindres : l'intelligence artificielle. Déjà, elle bouleverse la manière dont les systèmes d'information sur la recherche fonctionnent. Grâce à l'IA, il devient possible d'automatiser l'intégration des données issues de multiples sources, d'en améliorer la qualité par des traitements intelligents, de proposer des recommandations ou des compléments automatiques. Ces technologies offrent des perspectives inédites pour simplifier le travail des chercheurs, fiabiliser les données, enrichir les analyses. Des groupes de travail au sein d'euroCRIS réfléchissent

activement à l'intégration de ces technologies dans les futurs systèmes CRIS, en tenant compte à la fois des possibilités techniques et des enjeux éthiques. En définitive, les systèmes d'information sur la recherche se sont imposés comme des infrastructures stratégiques pour la science contemporaine. Ils soutiennent la transparence et l'accessibilité des travaux scientifiques, facilitent l'évaluation, éclairent les politiques publiques, et accompagnent le mouvement vers une science plus ouverte, plus collaborative, mieux documentée. À l'échelle européenne, une dynamique est à l'œuvre pour construire des écosystèmes cohérents, interopérables, respectueux des diversités locales mais fondés sur des standards partagés. Ce chantier, à la croisée des enjeux techniques, politiques et scientifiques, est plus que jamais d'actualité à l'heure où la donnée devient une richesse centrale pour la recherche et son pilotage.

JOACHIM SCHÖPFEL

Université de Lille, euroCRIS Board member joachim.schoepfel@univ-lille.fr

PABLO DE CASTRO

University of Strathclyde, euroCRIS Technical Secretary pablo.de-castro@strath.ac.uk

[10] A ce propos, la plupart de ces dispositifs mettent à la disposition des fournisseurs de données des recommandations et conseils précis dans le but d'harmoniser l'échange de données, à l'instar du portail flamand FRIS https://www.ewi-vlaanderen.be/sites/default/files/integration_guide_fris_r4_version_2.12.pdf

• Consulter la vidéo de l'intervention : <https://vimeo.com/1088405292>

... QUELQUES RÉFÉRENCES

Bryant, R., Clements, A., Feltes, C., Groenewegen, D., Huggard, S., Mercer, H., ... Wright, J. (2017). *Research Information Management: Defining RIM and the Library's Role*. <https://www.oclc.org/research/publications/2017/oclcresearch-defining-rim.html>

Bryant, R., Clements, A., de Castro, P., Cantrell, J., Dortmund, A., Fransen, J., ... Mennielli, M. (2018). *Practices and Patterns in Research Information Management. Findings from a Global Survey*. <https://www.oclc.org/research/publications/2018/oclcresearch-practices-patterns-research-information-management.html>

Bryant, R., Fransen, J., de Castro, P., Helmstutler, B. & Scherer, D. (2021). *Research Information Management in the United States*. <https://doi.org/10.25333/8hgy-s428>

de Castro, P. (2018). *The role of Current Research Information Systems (CRIS) in supporting Open Science implementation: the case of Strathclyde*. ITLib, (5), 21–30. <https://dx.doi.org/10.25610/itlib-2018-0003>

de Castro, P., Shearer, K., & Summann, F. (2014). *The Gradual Merging of Repository and CRIS Solutions to Meet Institutional Research Information Management Requirements*. Procedia Computer Science, 33, 39–46. <https://doi.org/10.1016/j.procs.2014.06.007>

Ribeiro, L., de Castro, P., & Mennielli, M. (2016). *EUNIS – EUROCRIS joint survey on CRIS and IR. Final report*. <http://www.eunis.org/wp-content/uploads/2016/03/cris-report-ED.pdf>

Schöpfel, J., & Azeroual, O. (à venir). *Les systèmes d'information sur la recherche. Evaluation, performance et responsabilité*. Londres : ISTE.