

INTERVIEW

« DANS L'ESR FRANÇAIS, C'EST L'ABES QUI, À TRAVERS IDREF, INCARNE LE LIEN ENTRE RÉFÉRENTIELS, DOCUMENTATION ET RECHERCHE »

Entretien avec David Reymond et Henri Bretel, membres du consortium CRISalid

Membres du consortium CRISalid, qui porte les projets « SoVisu+ » et « Idya », David Reymond et Henri Bretel¹ détaillent pour *Arabesques* les liens entre référentiels, documentation et système d'information recherche.

Pourriez-vous, tout d'abord, expliquer ce qu'on entend par « système d'information recherche », ses grands principes, fonctions et objectifs, et sa place dans un système d'information d'établissement de recherche ?

Un système d'information recherche ou SI recherche, ou encore *Current Research Information Systems* (CRIS), est défini par Guillaumet, García et Cuadrón comme un dispositif dédié à la gestion et au suivi des informations liées aux activités de recherche, offrant une variété de fonctionnalités qui s'adaptent aux besoins spécifiques des institutions de recherche.

Ce grand principe se décline en plusieurs fonctionnalités : créer et administrer des profils pour des chercheurs et des structures ; agréger les publications, brevets, thèses, en mesurer l'impact, identifier les tendances ; rassembler les informations sur les projets de recherche (ceci peut inclure les financements, collaborations, documents de suivi administratif) ; faciliter la collaboration entre chercheurs, par exemple par la gestion d'appels à projets ou la génération et la visualisation de réseaux à partir des profils ; enregistrer les accords de licence, embargos et restrictions d'accès, dépôts dans des archives ouvertes, protocoles, approbations éthiques, déclarations de conflit d'intérêts) ; et enfin, relier les informations concernant la recherche avec les autres systèmes institutionnels ou externes pour automatiser les flux de données et réduire la saisie manuelle : bases de données bibliographiques, archives ouvertes et entrepôts de données, systèmes de gestion des subventions etc.

En s'appuyant sur le web sémantique, l'outil que nous construisons permettra aux acteurs de l'écosystème, chercheurs, gestionnaires de la recherche, bibliothécaires, de constituer et valider un « graphe de connaissances institutionnel »² de leur institution.

Parmi les fonctionnalités évoquées, certains CRIS, comme le projet « SI Labo », priorisent la gestion administrative. Notre projet met davantage l'accent sur les acteurs et résultats de la recherche.

Le lien entre référentiels, documentation et SI Recherche n'est pas forcément immédiat. Pourriez-vous détailler les interactions qui les lient ?

Le lien est indirect. Les référentiels servent à s'assurer que les entités soient bien reconnues dans les différents contextes où elles sont utilisées : les structures, personnes, productions, concepts ont chacun un identifiant unique, issu d'un référentiel. L'absence d'ambiguïté bénéficie par exemple aux structures de recherche, dont la visibilité et la crédibilité dépendent d'une caractérisation exacte des compétences. Quant aux classements d'institutions, en plus de toutes les précautions qu'ils impliquent, ils nécessitent également des référentiels de production scientifique et des classifications disciplinaires partagées. Dans l'ESR français, c'est l'Abes qui, à travers IdRef, maintient et expose ces référentiels, incarnant le lien entre référentiels, documentation et recherche. Un chantier semblable existe à l'échelle européenne autour du référentiel CERIF.

Au regard de ces deux pôles, référentiel et SI recherche, et de la situation dans d'autres pays, quel est l'état d'avancement de l'ESR français et quelles sont ses spécificités ?

En France, la gestion locale est entravée par l'organisation complexe de la recherche, marquée par l'affiliation par « tutelle », par opposition au standard international de l'affiliation par employeur. Cette situation favorise la multiplication des « silos » de données. Elle nécessite des

outils résilients, ce qui peut devenir un atout, mais elle a aussi des inconvénients que nous voulons pallier. Tout d'abord, elle décourage l'enrichissement des données. SoVisu+ accompagnera la communauté pour améliorer la qualité de l'archivage de la production scientifique, en utilisant au besoin l'IA. Ensuite, elle réduit la lisibilité de la science pour la société. L'*Expert Finder System* permettra de mieux valoriser les forces des institutions. Enfin, elle est un obstacle au partage de données. BiblioLabs adapte aux spécificités françaises, EPE, UMR, la gestion des données bibliographiques et la curation des données.

Les référentiels sont donc des piliers fondateurs. Pouvez-vous exposer les enjeux que véhiculent les données et les exigences qu'elles portent en conséquence ?

Le premier enjeu des données est la qualité, dont dépend le service rendu à tous les acteurs de la donnée, chercheurs, services administratifs, documentalistes. L'enjeu de l'interopérabilité est directement lié aux référentiels. Les données visibles à l'international doivent être normalisées selon les formats CERIF ou VIVO. Mais l'identification des données françaises passe par l'utilisation des référentiels correspondants en France. C'est pourquoi nous travaillons avec les services de l'Abes, mais aussi les référentiels d'AurÉHAL, que l'usage de HAL rend incontournables. Enfin la certification par l'instance d'autorité concernée, l'Abes pour les alignements d'identifiants ou une direction de la recherche pour le lien entre unité et tutelles par exemple, est nécessaire pour valider la correspondance entre la réalité et la donnée normalisée.

Pourriez-vous nommer les acteurs, agents et institutions en charge de ces données, expliquer leurs rôles, et la manière dont ils s'articulent ?

Les directions administratives veillent à la mise à jour des référentiels de structures internes, qui ont vocation à être répercutés à l'échelle nationale, par exemple dans le RNSR. Les chercheurs tiennent à jour les données concernant leurs travaux, avec le trio ROR, ORCID, DOI pour l'international, et les informations spécifiques à leur champ. Les documentalistes maintiennent les référentiels bibliographiques et accompagnent la normalisation des données. La création des données est le fait de chaque agent, mais l'enrichissement et la certification sont le fait d'interactions. Le chercheur est seul à même de définir ses champs d'expertise, mais la normalisation passe par l'interaction chercheur-documentaliste ou chercheur-autorité administrative. Les institutions nationales ou internationales interagissent entre elles pour assurer la compatibilité des référentiels.

Telle une course de relais, chaque acteur doit porter un bout de la donnée, la qualifier et la passer enrichie au coureur suivant. Sans froisser ou distribuer de bons points, quelles sont, actuellement, les articulations les plus sûres et celles qui sont « à risque » ?

L'architecture décentralisée s'écarte de l'image de la course de relais car chaque agent du circuit est en même temps producteur et utilisateur de la donnée, interagissant potentiellement avec tous les autres. Dans ce réseau, les maillons d'autorité et de certifications sont critiques car ils ont une responsabilité sur la production de la donnée pour tous les autres acteurs.

Vous formez, autour de SoVisu+, un collectif d'institutions engagées dans le développement d'un outil communautaire apte à doter les établissements français de recherche d'un SI recherche. Pourriez-vous exposer votre démarche, son point de départ et sa cible, et en dresser les clés du succès ?

La communauté s'est réunie autour de trois projets initiaux, SoVisu de l'université de Toulon, l'*Expert Finding System* de l'université Paris 1 Panthéon-Sorbonne, et BiblioLabs de l'université Paris-Saclay, avec pour objectif d'urbaniser et industrialiser ces éléments sous forme de modules logiciels libres et techniquement à jour. Elle demeure ouverte aux nouvelles participations. Elle est dirigée par un comité de pilotage et répartit le travail dans des groupes d'expertises et de conception. L'objectif est de développer des briques applicatives indépendantes, conformes à l'état de l'art technologique, favorisant la durabilité, l'écoresponsabilité et la réduction des coûts. Le projet s'inscrit dans les préconisations de rapports internationaux qui encouragent l'investissement dans la gestion institutionnelle des données, le soutien des identifiants pérennes, la collaboration entre fonctions, l'investissement dans du personnel dédié, et l'intégration des informations de recherche dans la gouvernance. La démarche implique de définir une architecture décentralisée, de développer les services de manière agile, d'arrêter les choix de formats et de technologies. La communauté communique vers l'extérieur (FU, Esup Days, webinaire du GT Recherche ESR), et consulte des experts nationaux ou internationaux dont les communautés SemApps et VIVO. Le déploiement des installations pilotes est envisagé pour la rentrée prochaine. Les travaux visent à homogénéiser les attentes, à anticiper les impacts sur les activités existantes, et à identifier les points techniques critiques et les besoins de formation. Des moyens seront sollicités pour orchestrer et animer la communauté, documenter les réalisations, et développer ou intégrer des modules répondant aux exigences de qualité et d'interopérabilité.

Propos recueillis par

FRANÇOIS MISTRAL ET VÉRONIQUE HEURTEMATTE

[1] David Reymond est maître de conférences à l'université de Toulon, et Henri Bretel, chargé de bibliométrie à l'université Paris-Saclay.

[2] Les graphes de connaissances représentent des concepts (des personnes, des lieux, des événements) et leurs relations sémantiques. Ce sont des structures de données utiles à des fins d'exploration, de cartographie et de visualisation. Les données ouvertes concernant la recherche forment désormais un réseau mondial de connaissances interconnecté et décentralisé, qui peut être modifié et enrichi par la communauté.