



HAL
open science

De la décentralisation à la vision collective: une architecture pour moderniser la gestion de la recherche, ses applications puis l'utilisation par l'écosystème scientifique

David Reymond

► To cite this version:

David Reymond. De la décentralisation à la vision collective: une architecture pour moderniser la gestion de la recherche, ses applications puis l'utilisation par l'écosystème scientifique. 2023. hal-04502667v1

HAL Id: hal-04502667

<https://hal.science/hal-04502667v1>

Preprint submitted on 13 Mar 2024 (v1), last revised 5 Sep 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

De la décentralisation à la vision collective : une architecture pour moderniser la gestion de la recherche, ses applications puis l'utilisation par l'écosystème scientifique.

Résumé : Les avancées indéniables du traitement de données textuelles ont ouvert de nombreux outils cartographiques fréquemment utilisés pour le pilotage, le suivi et l'évaluation de la Recherche. Les données étant manquantes, biaisées ou partielles, le plus souvent ces outils se révèlent inutiles et/ou de simples leurres, leur utilisation stigmatise l'écosystème, induit des clivages et conduit à des dérives de la pratique scientifique. La fiabilisation de la donnée est la première étape *sine qua non* à mettre en place et ne peut se solutionner par l'apposition de simples tuyaux informatiques connectés aux bases existantes qui embarquent, elles, un lourd historique emmêlé depuis les premiers pas de l'informatique documentaire. Sur un autre plan, le recours à la donnée issue des silos centralisés induit fatalement une obsolescence sans une action de maintenance permanente de la donnée. Reste un principe simple : la donnée est fiable seulement si elle a été vérifiée par son auteur. Elle restera fiable si elle est contrôlée, sécurisée ou mise à jour si nécessaire. Dès lors se pose la question de l'autorité : sur la partie production de la recherche qui d'autre que l'expert peut opérer la première étape de fiabilisation ? Des certificateurs peuvent de surcroît en attester. Un mécanisme qui peut a priori paraître inatteignable sauf à raisonner « décentralisé » et rendre utile à chacun de maintenir sa donnée tout en capitalisant ce même travail pour le collectif. Ce mode opératoire permet de surcroît d'abstraire la complexité du mille-feuilles institutionnel et d'adapter les principes de Leiden en repensant l'évaluation scientifique. Dans cet article, nous positionnons les limites des applications de bibliométrie sur des données non fiables (applications qui dépassent largement le cadre de l'évaluation auquel nous nous référons cependant), précisons le contexte national et soulignons la charge actuelle que fait peser son application *in fine* aveugle. L'objet est ensuite de développer les principes d'une architecture décentralisée simplifiant les processus de collaboration internes par des procédés automatisés et conduisant mécaniquement à une réduction de la charge de chacun et à la fiabilisation collective de la donnée. Celle-ci est alors formalisée en un *graphe de connaissance institutionnel* utilisable en conséquent pour des applications qui réduisent l'évaluation à un gadget amusant sans toutefois entraver son utilité et qui étendent significativement la portée de cette fiabilisation.

Mots-clés : bibliométrie, système d'information, ESR, décentralisation, données liées, web sémantique, web social

Introduction

Les outils logiciels d'évaluation et les systèmes d'information sur la recherche (Guillaumet, García, et Cuadrón 2019) élargissent constamment leurs fonctionnalités pour exploiter ces données et les fiabiliser manuellement pour en extraire du sens (Krüger et Petersohn 2022). La bibliométrie, en tant qu'outil d'évaluation et de pilotage de la science, se positionne comme une source informationnelle essentielle dans le paysage de la recherche contemporaine. Fondée sur l'analyse quantitative à partir des métadonnées des publications scientifiques, elle offre une méthodologie rigoureuse permettant de mesurer l'impact, la visibilité, et la productivité des chercheurs, des institutions, voire des domaines de recherche entiers ou des pays (Chaignon et Egret 2022). Cette approche, reposant sur l'examen des publications, des citations, des collaborations, d'impact, s'avère cruciale pour synthétiser et comprendre les dynamiques de la production scientifique, identifier les tendances émergentes à partir des contenus, et soutenir la prise de décision éclairée (Borgman et Furner 2002; Bornmann 2020).

Depuis les travaux fondateurs de Scientométrie par Price (de Solla Price 1986) et Garfield (1955), les indicateurs évoluent tout comme les bases de données qui se prêtent à ces métriques (Birkle et al. 2020). Toutefois il est de notoriété (Gingras 2008; 2014) qu'une démarche politique fondée sur ces indicateurs ne peut être satisfaisante à de nombreux niveaux : d'une part mesurer la production scientifique sur le seul objet que sont les publications est une limite de la représentation de l'action académique dans la société, et d'autre part la couverture même des bases bibliographiques (Mart'in-Mart'in et al. 2020; Visser, van Eck, et Waltman 2020) ou encore leur consistance (Singh et al. 2020; Wilder et Walters 2021) est très discutable. Les indicateurs sont ainsi biaisés par nature selon les disciplines, les revues rendant les comparaisons exogènes hasardeuses tout comme le pilotage direct (Bornmann 2020). Des métriques alternatives « altmetrics » viennent compléter les indicateurs de bibliométrie sur les productions de la recherche et leur réception mais souffrent tout autant d'insuffisance des données : dispersées dans des bases plus ou moins correctement remplies ou rattachés aux organismes soutenant la production. Pour tenter de nettoyer ces dernières des identifiants persistants en format machine sont conçus pour identifier de manière non équivoque les chercheurs, les organisations de recherche et les objets de recherche mais sont encore loin d'être uniformément répandus (ref) et ne suffiront pas seuls à établir les principes de Leiden.

La tentation malgré tout d'utiliser ces indicateurs a été entretenue par les éditeurs (Thomson...) et la mise en place des tableaux d'honneur type Shanghai. La course à la production massive, à l'impact le plus fort ou le rayonnement international dure depuis de nombreuses années¹ à tous les niveaux (national, institution, laboratoire, individu), et a montré des effets pervers d'introduction de biais (des disciplines avec des revues à facteur d'impact démesuré et inversement) et de flou dans la production scientifique même depuis les pratiques malines (un résultat est atomisé pour être dispersé dans une ou plusieurs publications) aux plus déviantes (faux résultats, fausses données, plagiat, utilisation de l'IA générative, etc.). La vague Science ouverte a proposé de sonner le glas du système *Publish or Perish* (Boukacem 2023) par ce constat de la mise en pratique perverse de ces logiques d'évaluation via ces indicateurs quantitatifs qui au-delà des écueils précédemment esquissés se réalisent encore à l'insu des principaux intéressés. La tendance actuelle à focaliser le recueil de données à partir de données sur le CV apparaît comme une première issue et peut se fonder sur des pratiques internationale déjà éprouvées². Dans cet article, nous explorerons la constitution d'un graphe de connaissance institutionnel à partir des données des chercheurs assemblées par eux-même via une interface de gestion de type environnement numérique de travail : garantissant la sécurisation, la certification et la mise en cohérence avec l'écosystème organisationnel. Nous montrerons d'une part l'intérêt de cette architecture pour la simplification des procédures et des collaborations avec les autres corps de métiers de l'ESR, ainsi que l'enjeu dans l'évaluation de la recherche, mettant en lumière son potentiel en phase avec les préconisations actuelles (Boukacem 2023) et les principes de Leiden incontournables pour guider les politiques scientifiques, à orienter les financements, et à promouvoir une science de qualité et transparente.

Contexte

En France, les données nécessaires pour le fonctionnement, le pilotage et l'évaluation de la Recherche sont, par nature complexes, réparties dans différentes bases et de niveau inégal entre les institutions : d'un côté les descriptions de laboratoires, de l'autre la gestion financière, les ressources humaines, la gestion des productions, etc. Autant de silos de données orientés métier d'accompagnement de la

¹ <https://www.ouvrirlascience.fr/le-manifeste-de-leiden-pour-la-mesure-de-la-recherche/>

² C'est le cas par ex. du Système d'Information Recherche BR-CRIS (<https://brcris.ibict.br/>) qui en appuie sur la base des CV Lattes que chaque chercheur doit renseigner depuis le master, reconstitue le graphe institutionnel du Brésil. La base de CV est réputée d'autorité faisant preuve d'attestation de diplômes ou de travaux.

recherche et communication qu'il faut alimenter et maintenir. Ainsi, au sein d'une institution, pour obtenir une lecture de la production scientifique, la stratégie est d'en recourir aux bases bibliographiques (le plus souvent externes et onéreuses), de démêler l'historique et les variantes d'affiliations, de réidentifier les chercheurs pour confirmer la liste des productions ainsi reconstituée. L'action est réalisée chaque année, confirmée tous les quatre ans dans les universités par consultation individuelle de la part des laboratoires. Il en va de même pour de nombreuses opérations utiles à différentes étapes du cycle de vie de la Recherche comme la liste des membres d'une entité de recherche ou l'identification d'experts, le suivi de projets... Ce sont de multiples opérations répétitives et coûteuses à l'ensemble de l'écosystème qui se couronnent par la mise en place d'habitudes de fonctionnement et de résignation des chercheurs à ne pas regarder la qualité des données obtenues (ref).

Par le mode opératoire du quotidien des chercheurs et pour leurs relations avec l'écosystème et les autres corps de métier, les données de la recherche (contrats, produits, travaux, publication...) sont réparties dans des bases (s'il en est) et, dans le temps, sans aucun autre lien entre elles que le chercheur et sa mémoire. Leur agrégation dans les systèmes projetés sur une modélisation centralisée s'est avérée très insatisfaisante. Et pour cause : consigner une information par nature volatile dans différentes bases induit non seulement une charge permanente de maintenance pour les chercheurs mais aussi un travail qui peut être perçu comme inutile dans sa durée. Par essence, le recours aux silos de données métier qui verrouillent une information descriptive d'un travail de recherche à une date t ne peut conduire qu'à l'obsolescence de celle-ci : le travail de recherche est dynamique et sa description évolue par nature. Ainsi, le processus de gestion des données de recherche commence idéalement au tout début, dès la conception du projet et doit se poursuivre tout au long du cycle de vie de la recherche, le recours aux silos centralisés (Pin 2021) induit fatalement une obsolescence sans une action de maintenance permanente de la donnée.

Vers la gestion décentralisée des données

SoVisu a été conçu pour accompagner les chercheurs à l'auto-archivage (Reymond 2022; 2021; Reymond, Tabariès, et Bara 2022). En montrant à chacun le reflet de sa production associée à ses identifiants (Figure 1), le chercheur est responsabilisé par rapport à l'état de sa représentation dans les index. Chacun peut alors en devenir maître et décider de soigner la lisibilité de son profil scientifique. Côté institution, par capitalisation de cette information agréée par les experts (soignée ou non), nous produisons des indexations de qualité, que ce soit du point de vue de la fiabilité ou de la pertinence, utiles au pilotage, à l'évaluation ou à la valorisation. SoVisu ouvre ainsi une voie de dialogue instrumenté avec les chercheurs pour opérer à une curation collective des données, obtenir des informations qualitatives et des informations « internes ». En contrepartie de l'outillage ainsi fourni et sous couvert des autorités de certification, la communauté contribue à reconstituer et à vérifier le **graphe de connaissance** des entités d'affiliation et, récursivement, jusqu'à l'échelon de l'établissement. C'est à la constitution de ce graphe que nous allons donner une implémentation robuste, en nous appuyant sur les formats et protocoles du Web Sémantique et du Web Social, et en exploitant le potentiel de l'IA et des Grands Modèles de Langage (GML) afin de faciliter le travail de chacun, éliminer les multiples saisies et dépasser l'état de l'art des *Current Research Information System*. Les flux d'informations relatives à la production (publique ou pas) sont gérés au niveau de l'établissement en cohérence avec l'écosystème de la SO (identifiants, référentiels, services) dans **une boucle vertueuse visant la qualité des dépôts et l'adhésion de la communauté**. Par incrémentation progressive du périmètre fonctionnel, un mode de développement agile et participatif, nous étendrons progressivement à l'activité de recherche en général pour constituer un dispositif participatif de pilotage institutionnel (Boukacem 2023).

SoVisu+ va amplifier³ d'une part en étendant progressivement aux autres produits de la recherche (thèse, brevet, prépublications...) et progressivement reconstituer le CV, en mettant à disposition des outils renforcés par l'IA pour aider en premier lieu à l'application de métadonnées issues de vocabulaires contrôlés et en recueillir les usages pour renforcer les modèles. SoVisu+ proposera un ensemble de services aux chercheurs (Reymond 2023; Reymond et Lapôtre 2023), intégrant les flux nécessaires à tout l'écosystème, en aidant à améliorer la qualité de diffusion de leurs travaux et, en conséquent, leur lisibilité ou découvrabilité. En utilisant ces services, les chercheurs gagnent du temps, sécurisent et fiabilisent leurs données et la communauté reconstitue et maintient collectivement le "capital connaissance" des entités auxquelles ils sont affiliés et, récursivement, jusqu'à leur institution. En s'appuyant, de façon transparente pour les usagers, sur les technologies du Web Sémantique, la communauté reconstitue naturellement le "graphe de connaissance institutionnel" et, de surcroît, en certifie la qualité pour le consolider en fiabilisant une fois pour toutes la topologie des points de fiabilisation et de certification des données : ce graphe peut alors servir à alimenter les portails publics, sites web, ou outils de recherche d'expertise de l'institution etc. avec une précision mesurable et, par évidence les rapports d'évaluations, suivi de la production et pilotage.

Formats et interopérabilité

L'interface chercheurs de SoVisu+ constitue le lieu de « concertation » ou de médiation indirect avec les experts : le prédicat de base est qu'ils connaissent leurs affiliations, leurs productions, leurs expertises, leur(s) discipline(s), etc. et qu'ils peuvent en attester ou, de préférence, valider simplement ces informations, puis, peut-être, mettre en valeur certaines. Dans cette approche, les chercheurs sont, par essence, l'autorité de leur profil scientifique. Ces données sont alors déposées dans un container qui peut décliner des vues différentes (ce qui est privé, public, relevant d'une autorité institutionnelle pour sa certification, etc...), des informations appréciées par l'expert, voire complétées et le cas échéant, expliquées. La description de ces contenus, autant que faire se peut, suivra une des normes en vigueur à l'international (CERIF ou VIVO (Krafft *et al.* 2010 ; Ilik *et al.* 2018) avec une garantie de continuité car d'une part ces modèles sont extensibles et, d'autre part, ils sont structurés pour couvrir l'activité de recherche au sens large (cf. Figure 1) dans un format évolutif. L'utilisation d'un de ces formats et d'ontologies spécialisées pour les métadonnées descriptives est d'ailleurs une condition nécessaire pour garantir l'interopérabilité non seulement au plan national mais aussi international. La mise en lumière de ce profil sur les portails institutionnels, les systèmes d'identification d'experts, les rapports bibliométriques ou même les dispositifs d'évaluations garantissent qu'ils soient très probablement soignés au niveau d'exigence de l'expert qu'il représente qui, de fait, reste maître de ce qui est diffusé. Le dispositif devra permettre aussi d'accompagner progressivement les activités de recherche : un nouveau projet de recherche depuis sa définition, sa soumission, son évolution, sera sous l'autorité du profil. Certaines parties relèveront de l'autorité de certification quant à la validation des données sur le même principe que précédemment. Sa description, compatible avec le référentiel

³ A la date de rédaction, un consortium d'établissements (INRIA, Nantes Université, Université Claude Bernard Lyon 1, Université de Montpellier, Université Paris 1 Panthéon-Sorbonne, Université Paris-Saclay, Université de Toulon, Université Paris-Est Créteil) ont engagé des moyens et des personnels à des degrés variés. Ce collectif s'est nommé CRISalid et a pour vocation la création collective, ouverte, modulaire et dans l'état de l'art technologique d'outils, procédés, nomenclatures, référentiels pour la création de Graphe de Connaissance Institutionnel et leur partage faisant défaut à tout l'ESR. La communauté s'organise et se structure pour capitaliser, maintenir et entretenir depuis les fondements génératifs à la mise en place et son accompagnement institutionnel en suivant les recommandations de l'OCLC (Bryant *et al.* 2021). Le suivi des différents projets et événements de la communauté est là : <https://www.esup-portail.org/wiki/display/ESUPCRISalid/Accueil+ESUP-CRISalid>

institutionnel, permettra ainsi non seulement le suivi mais en garantira la fiabilité en collant à sa dynamique évolutive.

La Figure 1, ci-dessous montre le périmètre du modèle CERIF (Jörg, Höllrigl, et Baker 2014) (REF), maintenu en permanence pour évoluer avec les besoins de la gestion de la recherche. Au cœur de ce modèle, les Unités Organisationnelles (Laboratoires, Groupes, ...) se représentent tout comme les personnes ou les projets et sont rattachés si nécessaire, les profils, productions, infrastructures, activités, et résultats, tout comme les indicateurs. Cette représentation formalisées dans des normes interopérables et lisibles par les machines, s'appuie sur des identifiants pérennes autant que possible pour leur suivi.

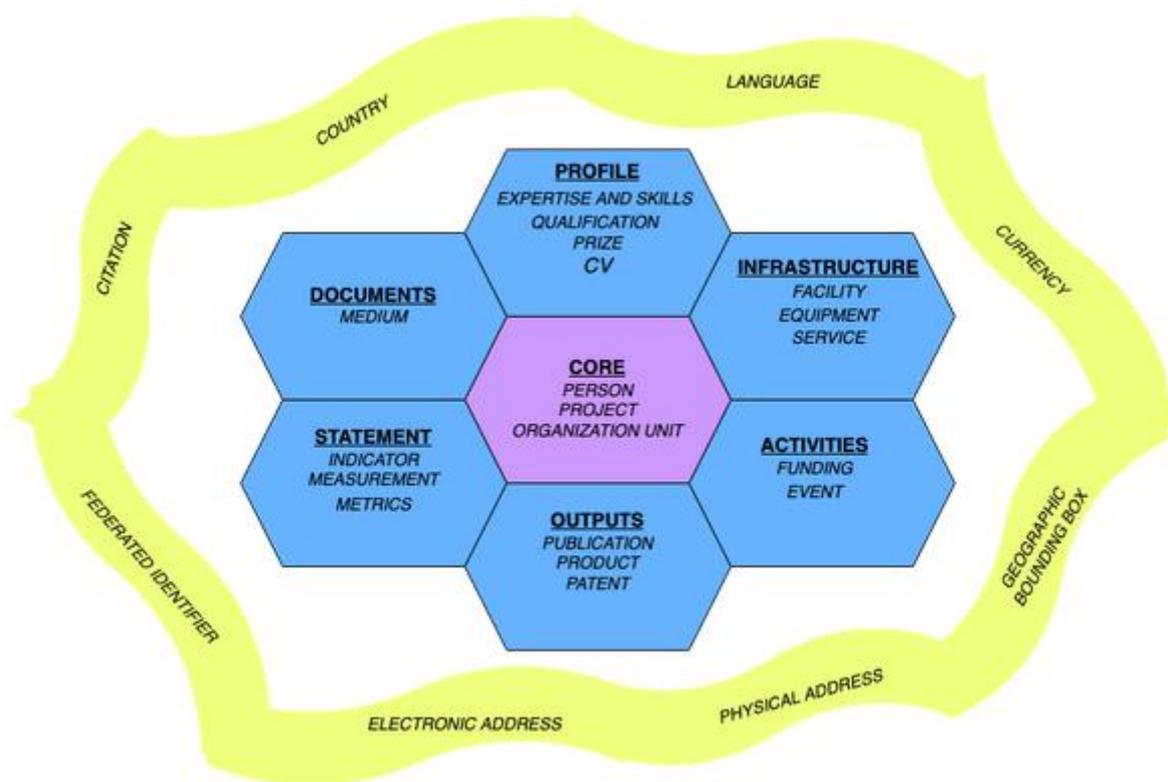


Figure 1: Périmètre du modèle CERIF couvrant les informations de la gestion de la recherche

Les technologies décentralisées du Web

Nous ne faisons pas ici un état des lieux exhaustif mais mettons l'accent sur les travaux récents de Tim Berners Lee sur la technologie Solid (« Solid-Based Approach for Research Information Interoperability - General Discussion » 2023).

La communication scientifique est à repenser (Sompel et al. 2004; Sponberg et al. 2023) et les travaux restent très actifs sur le domaine. Parmi d'autres projets ciblant la décentralisation dans la production scientifique (par. ex. OnePub⁴), le projet *Researcher Pod*, financé par la Fondation Mellon, explore un système de collaboration appuyé sur des protocoles de type web social pour instrumenter les différentes étapes. Le système est fondé sur la vision du Web décentralisé de Tim Berners-Lee et les différents protocoles afférents de la technologie Solid (sécurité, authentifié, identifié pérennisent, ...) et tel qu'envisagé par Capadisli (2020) et Van de Sompel (Sompel et al. 2004) pour les usagers. Dans le

⁴ <https://www.lri.fr/~mbl/ONE/onepub.html>

contexte du projet Solid, un Conteneur⁵ Solid vise à donner aux individus le contrôle sur leurs propres données sur le web. Il fournit un espace personnel de stockage en ligne où les utilisateurs peuvent stocker et gérer leurs données, puis accorder des autorisations aux applications ou à d'autres utilisateurs pour y accéder. Les serveurs de données liées (*Linked Data Platform*), dans ce contexte, sont utilisés pour mettre en œuvre la partie serveur de ces Conteneurs Solid, permettant aux utilisateurs de gérer et de partager leurs données conformément aux principes de Solid et des données liées.

Les Conteneurs Solid peuvent ainsi être utilisés dans un large éventail d'applications (Verstraete, Verbrugge, et Colle 2022) où les individus veulent maintenir la propriété et le contrôle sur leurs données personnelles (Pandit 2023; Solanki 2021). Cela inclut les réseaux sociaux, le partage de données, les espaces de travail collaboratifs, et bien plus encore. Transposé au domaine de la recherche, les travaux de Langer et al. décrivent les spécifications de Conteneur type « projet de recherche » (Langer, Vu Nguyen Hai, et Gaedke 2020).

Un réseau de Conteneurs Chercheur (*Researcher Pod*) est décentralisé : chaque chercheur référence des résultats de recherche dans son container de données personnelles (Mansour et al. 2016; Sambra et al. 2016) hébergé sur un domaine web personnel. Cette action, et ses mises à jour, alertent des services automatisés ou des acteurs du réseau de ces modifications. De la valeur est ainsi ajoutée sans autre action de l'utilisateur initial, à ces résultats de recherche par interaction avec des services découplés appropriés (acteurs ou robots)⁶. Les spécifications définies dans les protocoles *Activity Stream* (« Activity Streams 2.0 » 2017) solidifient l'infrastructure pour ce mode de collaboration hybride tout en laissant ouverte l'extension à d'autres domaines. Dans ce cadre, nous étendons ce principe actuellement spécifié pour la publication scientifique aux autres activités de la recherche depuis les élémentaires de l'affiliation aux autres détails de productions scientifiques et leur écosystème : projets, thèses, événements. Un ensemble d'éléments qui, à terme, couvrira l'activité scientifique faisant converger l'interface à la fois en une gestion de CV mais également en une interface d'échange et de médiation avec l'écosystème (documentation, RH, finances, ...). Les silos métiers pourront pointer sur ces données pour éviter doublons, redites et décohérence des données. Au-delà de ces détails techniques qui feront l'objet de spécifications ouvertes répondant aux exigences du W3C, nous centrons ici sur la stratégie d'intégration dans un établissement qui conditionne la mise en cohérence du système via les identifiants chercheurs en particulier l'alignement du quadruplet (WebId, idhal, OrcId, Idref) et l'affiliation à une entité de recherche (groupe, laboratoire, ...) après le passage d'une autorité de certification. La combinaison des trois technologies Pod Solid, Activity Stream et Données liées (LDN) ouvrent la voie à la capture de données de la recherche (productions, activités, etc.) ouvertes par le chercheur au ayants-droits et en temps réel : les Container informent des robots de tout changement (ajouts, suppression et mises à jour), qui en réfèrent si nécessaire aux nœuds d'autorité (responsable de laboratoire, services de documentation...), pour certification ou validation sans aucune autre action nécessaire. Le Graphe de Connaissances réalisé par agrégation récursive est maintenu collectivement en cohérence avec le déroulement de l'activité scientifique.

⁵ Dénommé Pod (*Personal Object Data*) dans la littérature. Le terme Pod faisant de longue date écho dans l'ESR à des supports vidéo pour la pédagogie...

⁶ Ce peut être simplement la mise à jour du graphe institutionnel, pour intégrer la présence de ce nouveau document (et donc inutile de disposer de collecteurs qui scrutent en permanence les référentiels pour se mettre à jour, charge conséquente et lourde au plan développement durable), ou d'autres actions plus complexes telle l'adjonction de métadonnées, la mise en relation avec des jeu de données antérieurs, etc. etc.

Stratégie de reconstruction du graphe de connaissance : des nœuds aux branches

Le Schéma 1 montre le processus de la première connexion pour la création du Container suite à authentification dans un Système d'Information⁷.

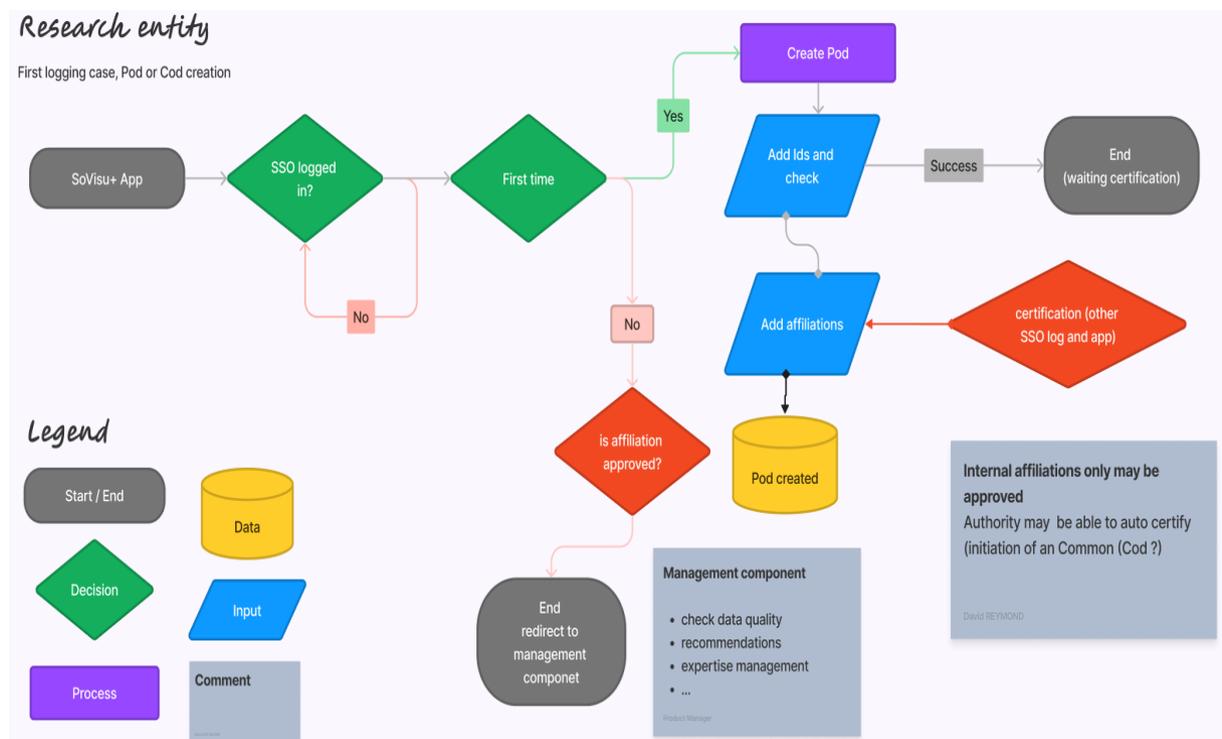


Schéma 1 : Première connexion et création du Container chercheur

Dans ce processus, le Container est créé, alimenté par les données afférentes du système d'information (id, noms, affiliations, signature) et de l'écosystème de publication (la production scientifique limitée dans un premier temps aux publications). Une zone du Conteneur est sujette à certifications (affiliation d'appartenance à un laboratoire par ex.) qui informe alors le nœud d'autorité pour validation de cette information : un tampon apposé ou vérifié. Dès l'approbation le Container du chercheur est marqué comme affilié à celui du laboratoire en y faisant référence. Réciproquement, le Container du laboratoire référence à son tour le Container du chercheur et associe immédiatement les données déclarées utiles à ce niveau (comme la production scientifique par ex.).

L'interface chercheurs

L'application de gestion de Conteneur individuel doit fournir des services utiles pour le chercheur, simplifiant la production des données qui peuvent se réutiliser aisément dans toutes circonstances. A nouveau, dans une vision Agile de cette conception nous reprenons un mode cumulatif. Le point de départ est inspiré de l'interface de gestion de CV et de la production scientifique de SoVisu intégrant la dimension qualitative (Reymond 2022). Afin de faciliter l'appropriation, un support d'aide à la production de métadonnées (extraction automatique à partir d'un texte type *preprint*, utilisation de vocabulaires normalisés et IA pour accompagner l'exercice). La poussée en archive est aussi facilitée par un simple clic. A terme, d'autres fonctions complémentaires viendront s'ajouter : création de

⁷ Le cas des individus disposant d'entrée dans des institutions différentes n'est pas traité. Un système type fédération est à prévoir ? De même le procédé pour l'initiation des autres acteurs de l'institution reste à définir selon les différents rôles, des Containers doivent être générés pour tous les éléments cœur du modèle CERIF (tous les rôles d'autorité de certification, tous les projets à terme). En mode Agile et développement continu les éléments arriveront progressivement .

projet (mise en lien avec les services de la recherche), création d'évènements : application de la charte de signature, liens avec les laboratoires et autres chercheurs tout est maintenu sur une seule et même « zone ». Avec ces fonctions s'ajoutent sur le réseau à valeur ajoutée les processus automatiques de certification et de notification de l'écosystème : l'action du chercheur est facilitée, l'écosystème informé et les dispositifs pointent définitivement sur le Conteneur dont le chercheur a la responsabilité. Ce même lieu de données agrège progressivement tous les témoins de l'activité scientifique et peut servir de référentiel et zone de médiation pour pouvoir repenser l'évaluation et le pilotage de façon ouverte et transparente.

Evaluation et pilotage 3.0

La bibliométrie, en tant qu'outil d'évaluation et de pilotage de la science, se positionne comme une discipline essentielle dans le paysage de la recherche contemporaine. Fondée sur l'analyse quantitative des publications scientifiques, elle offre une méthodologie rigoureuse permettant de mesurer l'impact, la visibilité, et la productivité des chercheurs, des institutions, voire des domaines de recherche entiers. Cette approche, reposant sur l'examen des citations, des collaborations, et d'autres indicateurs bibliographiques, s'avère cruciale pour comprendre les dynamiques de la production scientifique, identifier les tendances émergentes, et soutenir la prise de décision éclairée. Depuis les travaux fondateurs de Garfield (1955), les indicateurs évoluent tout comme les bases de données qui se prêtent à ces métriques (Birkle et al. 2020). Toutefois il est de notoriété (Gingras 2008; 2014) qu'une démarche politique fondée sur ces indicateurs ne peut être satisfaisante à de nombreux niveaux : d'une part mesurer la production scientifique sur le seul objet que sont les publications est une limite de la représentation de l'action académique dans la société, et d'autre part la couverture même des bases bibliographiques (Mart'in-Mart'in et al. 2020; Visser, van Eck, et Waltman 2020) ou encore leur consistance (Singh et al. 2020; Wilder et Walters 2021) est discutable. Les indicateurs sont biaisés par nature selon les disciplines, les revues et les comparaisons peuvent être hasardeuses tout comme le pilotage direct (Bornmann 2020). Des métriques alternatives « altmetrics » viennent compléter les indicateurs de bibliométrie sur les productions de la recherche et leur réception. Les identifiants persistants en format machine sont conçus pour identifier de manière non équivoque les chercheurs, les organisations de recherche et les objets de recherche. Les outils logiciels d'évaluation et les systèmes d'information sur la recherche (Guillaumet, García, et Cuadrón 2019) élargissent constamment leurs fonctionnalités pour exploiter ces données et en extraire du sens (Krüger et Petersohn 2022). DORA, CoARA recommandent une évaluation transparente, exploitant d'autres données que la simple quantification de la production scientifique et plus qualitative⁸...

Que dire sur la production d'un chercheur qui s'est exilé pendant une grande période pour collecter des données dont l'analyse pourrait lui prendre encore énormément de temps avant de publier quoi que ce soit ? Que dire aussi des enseignants-chercheurs qui accompagnent l'art et la culture. Par ex. l'accompagnement à la réalisation de supports média (radio, télévision, etc.). Ce ne sont pas des productions scientifiques anoblies à être admises dans les indicateurs. Il est de nombreux exemples entre ces deux extrêmes où l'activité de l'enseignant chercheur peut ne pas satisfaire les exigences de l'évaluation de la recherche (REf). Pourtant cette activité préalable ou parallèle va se traduire par un manque énorme pour les indicateurs de suivi de production qu'ils soient individuels ou collectifs. Un trou qui ne peut se justifier qu'à partir d'une interaction avec le chercheur ou réciproquement une interdiction morale, éthique de produire et d'utiliser ce type d'indicateurs. Sauf à disposer d'un moyen d'inclusion ou pas de l'individu dans la production mesurée. Sous un autre angle, les indicateurs thématiques de la production (suivre les tendances, les émergences, etc.), les Objectifs de

⁸ Cf. <https://www.ouvrirlascience.fr/un-accord-pour-la-reforme-de-levaluation-de-la-recherche/>

Développement Durable utiles pour les rapports aux différentes instances, les cartographies de rayonnement international, sont de puissants outils pour qualifier et analyser synthétiquement les travaux mais ils dépendent par évidence de la présence de métadonnées dans les fiches bibliographiques. Certaines sont produites par les machines qui ne sont pas exemptes d'erreurs et nous avons montré que plus de 60% des données déposées en archive ne disposent ni de résumés, ni de mots-clés. De fait, tous ces indicateurs sont grossièrement des leurres s'ils ne sont pas accompagnés d'indicateurs de consistance des métadonnées sur lesquels ils s'appuient. Sur le plan sémantique d'autant plus qu'ils sont générés automatiquement mais, à notre connaissance, non vérifiés et, qui plus est, par l'autorité d'expertise de ces données : les chercheurs l'ayant produite.

Dans le cadre de SoVisu+ en reliant la fonction d'accompagnement à l'autoarchivage, la gestion de son CV à celui de l'évaluation nous pouvons implémenter des modalités transparentes qui ouvrent une dimension médiatisée et qualitative :

- Distincte de la production des laboratoires,

En raisonnant avec la notion de « collection » d'une Organisation de recherche qui retrace, depuis sa création, malgré ses dénominations différentes, les différentes productions de publications scientifiques. Cette notion porte la trace de l'historique scientifique, des stratégies des différents directeurs, des apports de connaissances au développement de la Science, permet d'en apprécier leur lien, leur impact et peut-être même leur génétique. En revanche, si l'on prend à un instant t, l'agrégation de la production scientifique historique des chercheurs qui sont affiliés à cette organisation nous obtenons une structure de connaissance totalement différente avec probablement quelques communs. En disposant de ces deux types de connaissance de l'Organisation on peut aisément comprendre qu'il s'agit là de données complémentaires à différencier pour le pilotage stratégique. La première, d'historisation, porte les acquis et les résultats, la seconde, vue partielle du capital humain en vigueur, dirige et préfigure le potentiel de l'organisation.

- Impliquant directement les acteurs,

L'indicateur simpliste de mesure la production dénombre celle-ci et en fait la moyenne par les acteurs membres de l'organisation. Les indicateurs de suivi sont directement déclinés de cet indicateur (que l'on identifie l'impact sur le critère du facteur d'impact ou du h-index, que l'on identifie les collaborations internationales en inhérent à partir des noms d'auteurs (un procédé très flou), ou d'autres enrichissements (thématiques etc.) en découle naturellement. Sans compter le problème de la collaboration (auteur multiples, doit-on compter plusieurs fois), il s'agit de déclinaisons de l'indicateur simpliste. Le simple fait de pouvoir compter ou pas dans le calcul de ces moyennes donne au chercheur la possibilité d'agir sur le processus d'évaluation, et préfigure une première possibilité de médiation.

- Permettant de préciser, spécifier les volets d'activité dominants du chercheur,

Un second degré de potentiel d'action du chercheur dans son évaluation (et celle de son organisation de tutelle), peut être ouvert par la possibilité de choisir des volets d'activité de prédominance (organisation de manifestations, activités pédagogiques, activités administratives, collecte de données...) qui inclurait ou exclurait automatiquement des indicateurs associés.

- Transparente car démontrée : en rendant lisible les indicateurs précédent, le chercheur peut se situer et apprécier par avance quel sera la représentation de son travail sur des indicateurs.

Peuvent être apposés à ce stade les niveaux de référence, et une apposition de remarques qualitative peut être apposée.

- Ouverte à de nouveaux critères (diffusion et qualité de par ex.) et évolutive,

Le dispositif décrit globalement est une trame laissant place à des variations singulières quant aux choix des indicateurs, et laissant place à d'autres. Dans les fondements cette procédure est ouverte et transparente en ouvrant un espace de médiation au niveau individuel et en intégrant des possibles singularités dans les modes de calcul des indicateurs.

Conclusion

Fondé sur une architecture ouverte et décentralisée, le système répond à une logique de restitution de la souveraineté des données de chacun, permet d'inclure les principes d'autorité et de certification institutionnel, et d'opérer le principe « dites-le nous une fois » tout en se structurant et s'adaptant indépendamment de la complexité alors déportée aux différents agents, aux logiques historiques des institutions de l'ESR malgré leurs variantes et singularités. Un principe d'économie appréciable tant en gain de temps dans les transactions administratives ou au niveau individuel, ainsi qu'un gain sur les coûts de la maintenance induite alors coupée de court. A la souveraineté s'ajoute la fiabilisation des données : la production des indicateurs représentatifs de chacun en est une forme d'incitation, la seconde en arrière-plan est celle de l'assurance qualité et de la découvrabilité des travaux de chacun performant ainsi la mission de Diffusion du Chercheur et la qualité des données ouvertes. La fonctionnalité de la mise en archive simplifiée et facilitée participe à la sauvegarde et la sécurisation. Dans ce principe gagnant-gagnant, le mille-feuille institutionnel est collectivement reconstitué, fiabilisé et maintenu collectivement par la simple action de la maintenance individuelle correspondant à son rôle dans l'institution. Progressivement, de nombreuses autres applications sont possibles, et pourront être connectées ou apportées à la communauté dans le temps. Nous avons raisonné ici sur la gestion des publications scientifiques, mais la notion des données ouvertes laisse place à l'introduction d'autres informations essentielles de l'activité de recherche et fondée sur son soutien pour pouvoir l'accompagner sans la freiner et automatiquement collecter des données fiables. Celles-ci seront alors utiles et pertinentes pour le suivi, le pilotage ou l'évaluation en général. L'introduction de cette architecture décentralisée ouvre ainsi à une modernisation de tout l'écosystème et soulève la remise en question des pratiques collaboratives entre les différents métiers, des fonctionnements historiques de certaines procédures deviennent transparentes, ou encore de la possibilité de disposer de référentiels chercheurs fiables dont la consistance (granularité des données, date de dernière vérification ?) peut de surcroît s'apprécier ou se décliner selon des points de vue applicatifs (suivi, maintenance, automatisation complémentaires...).

Bibliographie

- « Activity Streams 2.0 ». 2017. 23 mai 2017. <https://www.w3.org/TR/activitystreams-core/>.
- Birkle, Caroline, David A. Pendlebury, Joshua Schnell, et Jonathan Adams. 2020. « Web of Science as a Data Source for Research on Scientific and Scholarly Activity ». *Quantitative Science Studies* 1 (1): 363-76. https://doi.org/10.1162/qss_a_00018.
- Borgman, Christine, et Jonathan Furner. 2002. « Scholarly communication and bibliometrics ». *Annual Review Information Science And Technology* 36: 3-72.
- Bornmann, Lutz. 2020. « Bibliometrics-based decision trees (BBDTs) based on bibliometrics-based heuristics (BBHs): Visualized guidelines for the use of bibliometrics in research evaluation ». *Quantitative Science Studies* 1 (1): 171-82.
- Boukacem, C. 2023. « La réforme de l'évaluation de la recherche ». In . <https://mediaserveur.u-bourgogne.fr/videos/5-la-reforme-de-levaluation-de-la-recherche/>.

- Capadisli, Sarven. 2020. « Linked research on the decentralised Web ». Phd, Universitäts- und Landesbibliothek Bonn. <https://csarven.ca/linked-research-decentralised-web>.
- Chaignon, Lauranne, et Daniel Egret. 2022. « Identifying Scientific Publications Countrywide and Measuring Their Open Access: The Case of the French Open Science Barometer (BSO) ». *Quantitative Science Studies* 3 (1): 18-36. https://doi.org/10.1162/qss_a_00179.
- Garfield, Eugene. 1955. « Citation indexes for science : a new dimension in documentation through association of ideas » 122 (juillet): 108-11.
- Gingras, Yves. 2008. « La Fièvre de l'évaluation de la recherche du mauvais usage de faux indicateurs ». *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 100 (1): 41-44.
- . 2014. *Les dérives de l'évaluation de la recherche: du bon usage de la bibliométrie*. Raisons d'agir éditions.
- Guillaumet, Anna, Francesc García, et Oscar Cuadrón. 2019. « Analyzing a CRIS: From Data to Insight in University Research ». *Procedia Computer Science* 146: 230-40. <https://doi.org/10.1016/j.procs.2019.01.097>.
- Jörg, Brigitte, Thorsten Höllrigl, et David Baker. 2014. « Harmonising and Formalising Research Administration Profiles CASRAI / CERIF ». *Procedia Computer Science* 33: 95-102. <https://doi.org/10.1016/j.procs.2014.06.016>.
- Krüger, Anne K., et Sabrina Petersohn. 2022. « From research evaluation to research analytics. The digitization of academic performance measurement ». *Valuation Studies*. <https://api.semanticscholar.org/CorpusID:255653808>.
- Langer, André, Dang Vu Nguyen Hai, et Martin Gaedke. 2020. « SolidRDP: Applying Solid Data Containers for Research Data Publishing ». In *Web Engineering*, édité par Maria Bielikova, Tommi Mikkonen, et Cesare Pautasso, 12128:399-415. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-50578-3_27.
- Mansour, Essam, Andrei Vlad Sambra, Sandro Hawke, Maged Zereba, Sarven Capadisli, Abdurrahman Ghanem, Ashraf Aboulnaga, et Tim Berners-Lee. 2016. « A demonstration of the solid platform for social web applications ». In *Proceedings of the 25th international conference companion on world wide web*, 223-26.
- Mart'in-Mart'in, Alberto, Mike A Thelwall, Enrique Orduña-Malea, et Emilio Delgado López-Cózar. 2020. « Google scholar, microsoft academic, scopus, dimensions, web of science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations ». *Scientometrics* 126: 871-906.
- Pandit, Harshvardhan J. 2023. « Making Sense of Solid for Data Governance and GDPR ». *Information* 14 (2): 114. <https://doi.org/10.3390/info14020114>.
- Pin, Clément. 2021. « Évaluation d'un plan national de simplification administrative. Le cas de l'enseignement supérieur et de la recherche ». *LIEPP Policy Brief*, n° 50 (mars). <https://doi.org/10.25647/liepp.pb.50>.
- Reymond, David. 2021. « SoVisu, terrains de recherches SIC en cours ». Présenté à Réunion IMSIC AXE 1 Enjeux et usages des dispositifs sociotechniques numériques et des mutations informationnelles Captas, documentation et numérique, juin 18.
- . 2022. « SoViSu : visibilité et lisibilité en SO ». Présenté à Séminaire IMSIC AXE 1 Brevets, captas et Science Ouverte, février 24.
- . 2023. « SoVisu+ (SVP) : optimisation de la qualité des données dans l'Enseignement Supérieur et la Recherche grâce à l'Intelligence Artificielle ». In *Journée d'étude AFNOR/BNF 2023*. Paris.
- Reymond, David, et Raphaëlle Lapôtre. 2023. « SoVisu+: starting point and foundations of a national CRIS ». In *Semantic Web in Libraries*. Berlin / Germany.
- Reymond, David, Alaric Tabariès, et Esteban Bara. 2022. « SO-VISU ». Linux, docker. Python, HTML, JS, KQL, elasticsearch, kibana. Université de Toulon. <http://sovisu.univ-tln.fr>.
- Sambra, Andrei Vlad, Essam Mansour, Sandro Hawke, Maged Zereba, Nicola Greco, Abdurrahman Ghanem, Dmitri Zagidulin, Ashraf Aboulnaga, et Tim Berners-Lee. 2016. « Solid: a platform for decentralized social applications based on linked data ». *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep.*

- Singh, Vivek Kumar, Prashasti Singh, Mousumi Karmakar, Jacqueline Leta, et Philipp Mayr. 2020. « The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis ». *Scientometrics* 126: 5113-42.
- Solanki, Manishkumar R. 2021. « Solid: A web system to restore the control of users' personal data ». In *ICT Systems and Sustainability: Proceedings of ICT4SD 2020, Volume 1*, 257-67. Springer.
- « Solid-Based Approach for Research Information Interoperability - General Discussion ». 2023. Solid Community Forum. 21 juillet 2023. <https://forum.solidproject.org/t/solid-based-approach-for-research-information-interoperability/6774>.
- Solla Price, Derek John de. 1986. *Little science, big science... and beyond*. Columbia University Press New York.
- Sompel, Herbert van de, Sandy Payette, John Erickson, Carl Lagoze, et Simeon Warner. 2004. « Rethinking scholarly communication: Building the system that scholars deserve ». *D-Lib Magazine*; 2004 [10] 9.
- Sponberg, Adrienne, Matthew A McCary, Catherine O'Riordan, Pamela Padilla, et Richard L Wallace. 2023. « Rethinking Scholarly Communication ». *Frontiers in Ecology and the Environment* 21 (7): 307-307. <https://doi.org/10.1002/fee.2670>.
- Verstraete, Melanie, Sofie Verbrugge, et Didier Colle. 2022. « Solid: Enabler of decentralized, digital platforms ecosystems ». In *31st ITS european conference*, 1-19.
- Visser, Martijn S., Nees Jan van Eck, et Ludo Waltman. 2020. « Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic ». *Quantitative Science Studies*, 1-22.
- Wilder, Esther Isabelle, et William H. Walters. 2021. « Using conventional bibliographic databases for social science research: Web of science and scopus are not the only options ». *Scholarly Assessment Reports*. <https://api.semanticscholar.org/CorpusID:238851254>.