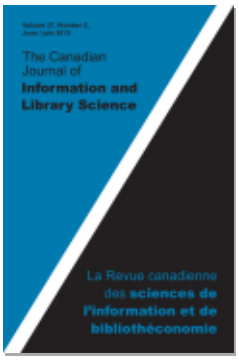
PROJECT MUSE®

Archives institutionnelles : Observations sur un nouveau mode d'information scientifique / Institutional Repositories: Observations on a New Form of Scientific Information



Joachim Schöpfel, Hélène Prost

Canadian Journal of Information and Library Science, Volume 37, Number 2, June/juin 2013, pp. 122-136 (Article)

Published by University of Toronto Press
DOI: *https://doi.org/10.1353/ils.2013.0008*

For additional information about this article
https://muse.jhu.edu/article/510865

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

Archives institutionnelles : Observations sur un nouveau mode d'information scientifique

Institutional Repositories: Observations on a New Form of Scientific Information

Joachim Schöpfel
Laboratoire GERiiCO, Université de Lille 3
joachim.schopfel@univ-lille3.fr

Hélène Prost
CNRS, membre associé au laboratoire GERiiCO, Université de Lille 3
helene.prost007@gmail.com

Résumé : Les archives institutionnelles sont devenues un vecteur significatif de la communication scientifique. Aujourd'hui, quatre archives ouvertes sur cinq sont des archives institutionnelles. L'une de leurs caractéristiques réside dans leur grande diversité. Il n'existe pas un modèle unique, mais une grande variété, pas une seule procédure mais de multiples options. Notre article décrit quatre modèles de base ou « archétypes » (plateforme de publication, container, site de patrimoine, dépôt), leurs facteurs de succès et quelques perspectives d'évolution. La légitimité des institutions, leurs choix stratégiques mais aussi, leurs infrastructures numériques impactent l'avenir de ces sites. S'ils tournent le dos aux communautés scientifiques, leur destin sera celui des catalogues et autres bases de données : utiles certes, mais pas indispensables.

Mots-clés : archives institutionnelles, archives ouvertes, libre accès, communication scientifique, information scientifique

Abstract: Institutional repositories have become an important vector for scientific communication. Today, four out of five open archives are institutional archives. One of their characteristics is their great diversity. There is no one single model, but a variety; not a single procedure but multiple options. Our article describes four basic patterns or "archetypes" (publishing platform, container, heritage site, deposit), their success factors and some prospects. The legitimacy of institutions, but also their strategic choices and their digital infrastructure impact the future of these sites. If they neglect the needs of the scientific communities, their fate will be that of catalogs and other databases: useful indeed, but not essential.

Keywords: institutional repositories, open archives, open access, scientific communication, scientific information

© 2013 The Canadian Journal of Information and Library Science

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

Le libre accès à l'information scientifique

La "Open Access Week 2012"[1] fut l'occasion de témoigner de la vitalité du mouvement pour le libre accès à l'information scientifique et de faire le bilan du chemin parcouru. Depuis la déclaration de Berlin de 2003[2] qui avait défini le libre accès comme "une source universelle de la connaissance humaine et du patrimoine culturel ayant recueilli l'approbation de la communauté scientifique", le nombre des archives ouvertes a dépassé 2 000 sites[3], on compte plus de 8 000 revues en libre accès[4], et il est impossible de dire combien d'articles, communications, thèses etc. sont désormais librement accessibles. Une partie de la communication scientifique est ainsi devenu "bien public" avec une diffusion plus vaste qu'auparavant (Willinsky 2005).

Le succès indéniable de ce mouvement en terme de chiffres ne doit cependant pas donner à penser qu'il s'agit là d'un ensemble cohérent, uniforme, coordonné. Au contraire, la diversité est la règle, aussi bien par rapport aux réalisations techniques que pour les modèles économiques.

Il y a vingt ans, l'idée du libre accès correspondait surtout au concept de la communication directe, sur le modèle de la plateforme arXiv de la communauté des *High Energy Physics* de Los Alamos, considérée comme première archive ouverte. Sont apparues ensuite les premières revues en libre accès, telles que *Nucleic Acids Research*, *New Journal of Physics* et *Atmospheric Chemistry and Physics* ou encore *ALSIC* en France.

Stevan Harnad définissait ces deux approches comme "green road" et "gold road" (Harnad et al. 2008). La voie verte désigne l'auto-archivage des publications dans les archives ouvertes, tandis que la voie dorée correspond à la publication d'articles dans les revues en libre accès (Baruch 2007, Guédon 2008).

La voie dorée s'est transformée en quelques années en un nouveau marché où cohabitent des initiatives locales avec de nouveaux modèles économiques[5] et l'offre commerciale des éditeurs qui désormais font payer les auteurs (ou leurs institutions) pour la libre diffusion de leurs publications.

La voie verte s'est également développée en plusieurs directions. Armbruster et Romary (2009) distinguent quatre différents types d'archives : des archives thématiques pour un domaine ou une discipline spécifique (comme ArchiveSIC), des archives à vocation nationale comme par exemple HAL, des archives institutionnelles pour les publications des chercheurs d'un établissement (comme Archimer de l'IFREMER), puis d'autres sites à caractère scientifique qui n'appartiennent à aucune de ces catégories (comme par exemple Symposcience ou les Bibliothèques Virtuelles Humanistes). Mais il est parfois difficile de faire la distinction entre ces quatre catégories et aussi par rapport à d'autres sites scientifiques (bibliothèques numériques, portails etc.).

Après une étude sur le devéloppement des archives ouvertes en France dont les résultats ont été publiés sur ArchiveSIC (Schöpfel et Prost 2010), nous nous sommes intéressés de plus près aux archives institutionnelles qui constituent, d'après notre enquête, de loin la catégorie la plus dynamique et importante de la voie verte en France. Tandis que Armbruster et Romary (2009) soulignent l'importance des archives à vocation nationale, telle que HAL, c'est en fait la catégorie

des archives institutionnelles et en particulier celles des universités qui se développe le plus rapidement.

En fait, en analysant les données de notre enquête et en discutant avec des collègues, nous nous sommes rendus compte de différences typiques entre les sites, parfois flagrantes, parfois moins perceptibles. Comment décrire ces différences ? Existent-ils des tendances d'évolution ? Peut-on dégager une nouvelle typologie d'archives institutionnelles ? Quels sont les facteurs de succès ?

Dans la suite de l'article, nous allons d'abord préciser le concept d'archive institutionnelle puis expliquer la méthodologie de notre analyse. Ensuite, nous allons proposer quatre types d'archives institutionnelles et un modèle bidimensionnel. Dans un deuxième temps, nous allons décrire quelques stratégies de développement et facteurs de succès avant de conclure avec des recommandations du point de vue d'un chercheur.

Pour éviter tout malentendu, précisons dès maintenant: Même si les données empiriques et plusieurs de nos exemples sont issus du paysage français, notre proposition typologique se veut plus large dans la mesure où nous avons intégré des analyses et des études d'autres pays. Même si la spécificité française — une centralisation des ressources technologiques autour de la plateforme HAL[6] — impacte les résultats de notre enquête, les différents types d'archives institutionnelles se retrouvent facilement au plan international.

La diversité d'un modèle

Les archives institutionnelles sont devenues un vecteur significatif de la communication scientifique. Leur ambition est de servir les intérêts de l'institution (université, organisme de recherche, laboratoire. . .) et des chercheurs. Elles servent à la collecte de leur production scientifique sur le long terme (Carr et al. 2008) et reflètent l'engagement de l'institution pour la gestion et la préservation de cette production (Lynch 2003).

Aujourd'hui, quatre sur cinq archives ouvertes sont des archives institutionnelles (source : OpenDOAR). La condition primordiale pour leur succès repose sur la volonté clairement affichée d'une institution de développer une plateforme d'édition et de diffusion libre pour sa production scientifique, en conformité avec sa propre stratégie et son environnement. Ou comme le formule une enquête de COUPERIN sur les archives ouvertes en France[7] : « L'entrepôt institutionnel est largement considéré comme faisant partie de l'identité propre de l'établissement ».

Néanmoins, l'une des caractéristiques des archives institutionnelles réside dans leur grande diversité. Il n'existe pas un modèle unique, mais une grande variété, pas une seule procédure mais de multiples options, parfois isolées, parfois en réseau (Foulonneau et al. 2007).

Chacune des raisons pour la mise en place d'une archive institutionnelle entraîne des implications sur son contenu, sa conception et son financement. L'institution doit définir de façon explicite l'implication et les rôles de chacun, tout en étant capable de s'adapter au changement, dans le contexte mouvant de la communication scientifique (Friend 2011).

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

Les archives institutionnelles sont régies par des politiques distinctes, présentent différentes procédures et fonctionnalités et offrent une variété de services et métadonnées. Gérées sous différents modèles économiques et modes de financement (Swan et Awre 2006), elles peuvent contenir bien plus que la production courante des chercheurs. Ainsi, Smith (2008) détaille un grand nombre de contenus numériques - articles de recherche, pré et post-publications, thèses et mémoires, documents administratifs, ressources pédagogiques mais aussi des données de la recherche, des produits multimédia ou des documents patrimoniaux.

Méthodologie

Notre étude s'appuie sur une double approche qui réunit un état de l'art et la nouvelle analyse des résultats d'une enquête empirique.

1. Dans un premier temps, nous avons procédé à un état de l'art de la recherche sur les archives institutionnelles. À partir d'une recherche bibliographique, nous avons établi une synthèse d'études et d'analyses des contenus, stratégies de développement et facteurs de succès de ces sites. En particulier, nous avons mis l'accent sur les travaux du JISC ou de l'équipe d'Eprints de l'université de Southampton (Friend ; Carr etc.).[8]
2. D'autre part nous avons repris les résultats de notre enquête de 2008–2009 sur le développement des archives ouvertes en France (Schöpfel et Prost 2010) pour revoir l'analyse et l'interprétation des données statistiques, afin de dégager quelques grandes tendances ou types de développement. Cette enquête avait identifié 151 archives ouvertes en France. La sélection de ces sites avait été faite à partir de l'analyse et de la comparaison de 19 répertoires, annuaires et autres listes d'archives ouvertes.[9] Chacun des sites répertoriés et sélectionnés a été décrit selon 58 critères.[10] En vue de proposer une nouvelle approche de analyse de cette enquête, nous avons surtout exploité l'information sur la typologie des sites, sur leurs contenus, leur politique de dépôt, leur développement et la validation des contenus, avec une attention particulière portée sur les contenus « atypiques » comme des collections patrimoniales ou des résultats de la recherche (*datasets*).

Voici le résultat de cette double approche, présenté d'une manière plus conceptuelle qu'empirique.

Quatre modèles de base (« archétypes »)

Comme indiqué plus haut, la création d'une archive institutionnelle reflète un choix institutionnel par rapport aux objectifs et contenus. Nous nous sommes donc demandé quelles étaient les options pour ce choix. D'après la littérature scientifique et nos propres observations, il se dégage quatre modèles de base ou « archétypes ». Ces modèles diffèrent non seulement en fonction de leur contenu ou finalité (mission ou vocation) mais aussi selon leurs options et leur déploiement au sein de l'institution.

Si tous les sites poursuivent l'objectif de favoriser la visibilité de la production scientifique de leur établissement, ils ne se déploient pas tous de la même manière. Aussi, ils ne remplissent pas tous la même fonction. Nous exposons ici quatre modèles distincts, suivant entre autre l'analyse stratégique de Friend (2011). Cependant, il s'agit de modèles types, alors que sur le terrain, les archives institutionnelles sont souvent plutôt une combinaison particulière de ces différents modèles.

Plateforme de publication

Dans ce premier modèle se retrouvent les deux fonctions principales d'une archive ouverte pour les communautés de recherche : communiquer et publier. L'accent est mis sur l'accès rapide et direct au texte intégral. L'intérêt stratégique de l'institution peut être double :

- Accroître l'impact de la recherche ou de l'enseignement en exposant les différents résultats dans la vitrine du libre accès.
- Réduire les coûts et accroître les bénéfices de la diffusion de la recherche de l'institution.

Sur le plan financier l'alimentation et la préservation du référentiel sont prises en charge par l'institution, peut-être aussi avec une contribution d'une société savante.

Le but est de diffuser l'ensemble des documents (*preprints*, *postprints*, thèses, actes de conférences, rapports non publiés...) produits par les membres de l'établissement. Une attention particulière est portée sur la qualité scientifique des documents, alors que les incitations appuyées au dépôt sont reléguées au second plan. Communiquer est, d'après l'enquête de COUPERIN, l'objectif principal des chercheurs. Les métadonnées servent surtout à faciliter la recherche et l'accès aux documents. Cela signifie qu'elles sont probablement de qualité médiocre et assez imprécises pour la plupart des documents, sauf peut-être pour les thèses.

La principale fonction de ces sites est la communication et l'accès au texte intégral, via les moteurs de recherche et/ou via l'interface de navigation et de recherche proposée à l'intérieur des sites. Les éléments clés sont un taux élevé de documents avec texte intégral, une qualité scientifique acceptable et garantie, l'accès sans restriction, suivie par un nombre élevé et représentatif de dépôts.

D'autres services, moins importants, peuvent apporter une valeur ajoutée :

- Les statistiques d'utilisation.
- Les services de conservation.
- Les services d'édition (dépôt).

Container

Dans le deuxième modèle, l'archive institutionnelle peut être comparée à un container pour toutes sortes de matériel produit par les chercheurs, enseignants et étudiants. Thèses, mémoires, rapports et actes de congrès côtoient images, supports de cours, articles, données de la recherche, présentations, posters, etc.

L'attention se porte vers la disponibilité et la visibilité des documents déposés, sans sélection ou contrôle *stricto senso* du caractère ou de la qualité scientifique.

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

L'objectif stratégique peut être de rassembler tous les publications et autres documents de recherche et d'enseignement comme un enregistrement permanent des réalisations de l'institution, mais sans aucune idée préconçue de leur utilisation (Friend 2011).

Ici encore le modèle économique le plus approprié est un soutien institutionnel et l'établissement peut mettre en place une politique de dépôt obligatoire ou inciter fortement les chercheurs à déposer leurs publications sur la base du volontariat.

En ce qui concerne le contrôle de qualité ou de rédaction, les métadonnées ne sont certainement pas une question importante. Très probablement, les services sont limités à la préservation, l'édition, la découverte des ressources, peut-être aussi à l'évaluation de la recherche et à la veille. Il est également possible d'ajouter l'indexation sociale et le *data mining*. Il n'y a pas de vision claire sur la collecte et l'acquisition. Mais la perspective la plus prometteuse sera peut être le lien avec les réservoirs des données de recherche (*datasets*).

Site de patrimoine

Dans ce scénario, l'archive est la vitrine de la production passée et actuelle de l'établissement, qu'il s'agisse des publications, de la littérature grise ou des ressources pédagogiques. L'objectif est triple :

- Accroitre l'impact de l'enseignement et de la recherche par la voie du libre accès.
- Réduire les coûts et augmenter les bénéfices d'une large diffusion.
- Contribuer au mouvement du libre accès à la recherche publique.

La différence avec les autres archétypes est que ce modèle inclut les anciennes collections de l'établissement, comme par exemple des thèses, des séries de comptes-rendus, actes ou rapports, des photos, cartes, monographies ou revues éditées et publiées par une université.

Contrairement au modèle du container, la sélection du contenu et l'alimentation du site font l'objet d'une réflexion attentive qui est souvent prolongée par une stratégie de numérisation d'anciens ouvrages, thèses etc. De même, l'indexation et les métadonnées sont généralement de meilleure qualité. Dans la mesure où les anciens documents sont dans le domaine public ou supposés orphelins, il n'y aura pas d'entrave d'un embargo, d'un accès réservé ou d'une absence de texte intégral.

Le modèle économique approprié est le soutien institutionnel. Mais il peut y avoir d'autres modes de financement, comme par exemple une subvention publique pour numériser le patrimoine scientifique ou des collections particulières.

L'existence d'un centre de numérisation local permet à ces sites de voir leur contenu croître plus rapidement, en particulier grâce à la littérature grise (Westell 2006).

L'idée sous-jacente est la préservation numérique des collections patrimoniales et leur diffusion aux communautés scientifiques et à toutes personnes intéressées. Cela peut être complémentaire aux archives commerciales (*backfiles*) des éditeurs.

Ce modèle est probablement le plus proche des activités traditionnelles d'une bibliothèque, avec des questions liées à la qualité, l'indexation et la classification des contenus ; en revanche l'évaluation ou la scientométrie ne sont pas nécessairement au cœur du projet.

Dépôt

Promu par Stevan Harnad comme la « voie verte » du libre accès, ce quatrième modèle correspond aux sites dont l'alimentation est régie par une obligation de dépôt (mandat) notamment des articles publiés dans des revues scientifiques pour obtenir l'adhésion de près de 100% des auteurs-chercheurs. L'objectif est de démontrer la valeur de l'institution par le biais d'un cadre de dépôt quasi légal et de maîtriser ou contrôler la production scientifique ainsi que son évaluation. Ce modèle reflète l'une ou plusieurs des stratégies suivantes (Friend 2011) :

- Rendre compte aux agences de moyens des publications et autres produits de recherche et d'enseignement, afin d'appuyer d'autres demandes de subvention.
- Signaler les publications et autres produits de recherche et d'enseignement aux agences de moyens dans le cadre d'un audit des dépenses
- Démontrer aux gouvernements et contribuables l'impact sociétal de l'institution, aussi à l'aide d'indicateurs scientométriques.
- Accroître l'impact de chaque membre du personnel de l'institution en mettant en lumière les retombées académiques et commerciales des publications et autres documents mis en libre accès.

Le modèle économique est le soutien institutionnel ; il peut inclure des services tels que les statistiques d'utilisation, l'évaluation et le suivi des projets scientifique, la cartographie et l'aide au transfert de technologie (valorisation). En outre, la connexion à un système d'information sur la recherche (CRIS) devrait être possible.

Le principal intérêt de ces sites n'est ni la publication de documents en ligne (comme pour le modèle I) ni la constitution d'une collection (comme pour le modèle III) mais l'évaluation de la recherche. Pour le contenu, l'impact est double. D'une part, les publications avec comité de lecture y jouent un rôle majeur, par rapport à d'autres types de documents moins valorisés par les agences d'évaluation. D'autre part, l'obligation de dépôt généra un taux relativement élevé de métadonnées sans texte intégral en raison de la politique d'embargo des éditeurs, de contenus sensibles, ou de l'absence d'autorisation par les co-auteurs.

Par ailleurs, le type de mandat peut varier : il existe des mandats institutionnels ou départementaux, et des mandats liés à une agence de moyen ou à un type de document (les thèses le plus souvent). Le site ROARMAP de l'université de Southampton donne une idée de ces différentes réalisations au plan international[11]. La situation en France a été décrite par L'Hostis et Aventurier (2006).

Cartographie et exemples

Comme nous avons indiqué plus haut, les quatre archétypes sont des potentialités, des pistes possibles (scénarios) pour le choix stratégique d'une institution.

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

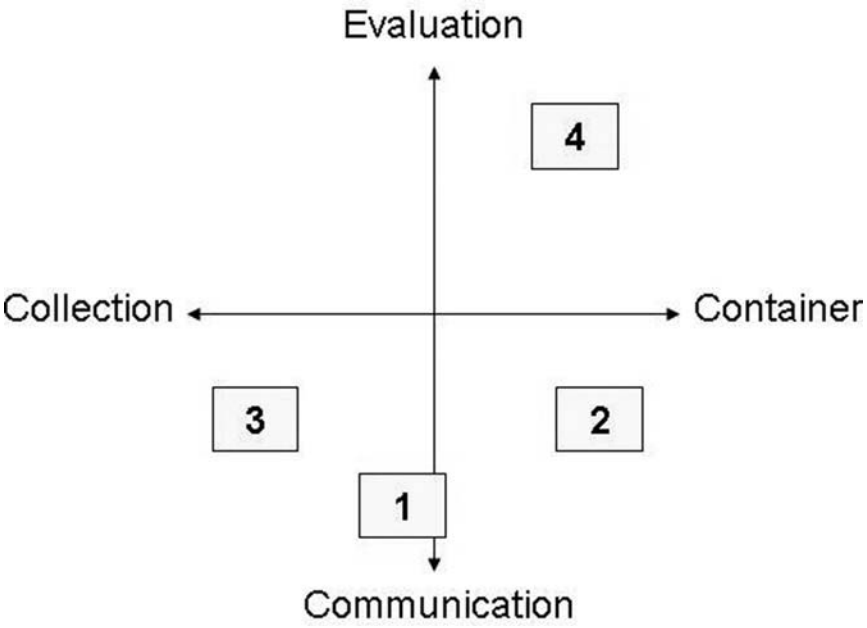


Figure 1: Cartographie des quatre modèles

Les modèles correspondent à différentes positions selon deux dimensions (cf. figure 1).

Le premier axe définit la priorité du site, s'il met l'accent sur l'évaluation de la recherche (évaluation d'une institution, d'un laboratoire ou d'un chercheur) ou sur la communication des résultats via la diffusion des publications en libre accès.

La deuxième axe définit le caractère du contenu du site, s'il reflète un vrai choix et une sélection réfléchie des contenus (collection), ou si son contenu est plus ou moins aléatoire (container).

Trois modèles (1, 2 et 3) mettent clairement l'accent sur la communication de la production scientifique (diffusion) tandis que la priorité du modèle 4 est de favoriser et maîtriser l'évaluation de la production scientifique de l'institution.

Le modèle 3 fonctionne comme une bibliothèque et construit des collections. Cet aspect ne joue pas un rôle pour les modèles 2 et 4 qui mettent en avant le nombre et l'appartenance institutionnelle (affiliation) des auteurs. Le modèle 1 se situe entre les deux extrêmes.

Sur le terrain des archives institutionnelles, les réalisations des sites mélangent souvent deux ou trois modèles et leur position sur cette carte est moins claire. Pour cette raison, nous n'avons pas essayé de chiffrer les quatre modèles d'après les résultats de notre enquête. Néanmoins, à titre indicatif, voici quelques sites représentatifs afin de mieux comprendre les différences :

I – Plateforme de publication : *Archive électronique*[12] de l'Institut Jean Nicod — toutes les publications de l'institut sont déposées en texte intégral. Néanmoins, ce site a aussi un aspect patrimonial dont la mesure où il contient quelques publications anciennes. Un autre exemple est représenté par les *Thèses en ligne de ParisTech* (PASTEL)[13] de douze grandes écoles de Paris avec presque 4 000 thèses, toutes en texte intégral. Pour finir, un exemple d'un autre pays, le *Dokumenten- und Publikationsserver* (edoc-server)[14] de l'université Humboldt de Berlin, une archive institutionnelle doté d'une politique éditoriale et d'un service d'aide pour les publications en ligne des chercheurs.

II – Container : *Archive numérique des publications scientifiques et techniques de l'IN2P3* (HAL-IN2P3)[15] — une archive ouverte proposant une très grande variété de documents, y compris des mémoires d'étudiants, brevets, cours, HDR, références sans documents et documents non classés. Un autre exemple est l'*Archive Ouverte INSEP*[16] de l'INSEP — cette archive contient des articles, communications et ouvrages mais aussi des cours, des études techniques, des rapports et des mémoires d'étudiants. D'autres exemples : le site *Institutional Repository* (IRep)[17] de l'université de Nottingham Trent, avec presque 20 000 références d'articles, images, présentations, cours, rapports, thèses etc., dont beaucoup sans texte intégral, et *Academic Commons*[18] de Columbia University dont voici l'appel au dépôt : « Current faculty, staff, and students of Columbia and its affiliates are invited to deposit research materials in any digital format, including: articles, monographs, theses and dissertations, working papers, technical reports, conference papers and presentations, datasets, images, video and other multimedia creations, and software code ». Autrement dit, il s'agit d'une invitation à tout déposer, sans sélection.

III – Site de patrimoine : *Publications et Travaux Académiques de Lorraine* (PETALE)[19] de l'université de Lorraine — plateforme de diffusion en ligne des thèses et mémoires en accès ouvert, avec plus de 1 500 thèses de doctorat et environ 200 thèses d'exercice auxquels s'ajouteront plusieurs centaines d'anciennes thèses qui font l'objet de programmes de numérisation. Un autre exemple est l'archive institutionnelle *Dspace@MIT*[20] du Massachusetts Institute of Technology proposant une sélection de thèses anciennes qui remontent au XIXe siècle (*MIT Thesis Collection*[21]).

IV – Dépôt : *Open Repository and Bibliography* (ORBi)[22] de l'université de Liège – dépôt obligatoire et 61% de références avec texte intégral (octobre 2012). Un autre exemple est la plateforme de dépôt et de consultation des articles scientifiques produits par des auteurs du Laboratoire « Biogéochimie et écologie des milieux continentaux » *HAL-Bioemco*[23] avec presque 90% de références sans texte intégral. Un exemple écossais : l'archive institutionnelle *Enlighten*[24] de l'université de Glasgow avec un mandat institutionnel et beaucoup de références sans texte intégral[25].

Facteurs de succès

L'appartenance des archives institutionnelles à l'une ou l'autre de ces catégories n'est pas liée à leur développement ou leur qualité. Aussi, leur succès ou échec ne dépendent pas nécessairement de ces choix stratégiques mais d'autres facteurs,

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

opportunités et risques. Nous avons identifié plusieurs aspects qui paraissent importants pour la réussite, l'impact et le développement d'une archive institutionnelle (cf. aussi Westell 2006).

- Un soutien institutionnel qui donne toute sa légitimité au projet.
- Une stratégie institutionnelle et une politique clairement affichée dans le domaine des archives ouvertes et de mandat de dépôt.
- Des ressources humaines suffisantes avec de bonnes compétences en informatique et en sciences de l'information.
- Des métadonnées normalisées (Dublin Core, OAI, TEF, CERIF etc.).
- L'interopérabilité avec les systèmes locaux et, le cas échéant, l'infrastructure régionale ou nationale, en particulier via le protocole OAI-PMH qui permet le moissonnage et l'agrégation des métadonnées de plusieurs archives ouvertes (Bester et Chartron 2010).
- Une solution pour l'archivage pérenne.
- Des services à valeur ajoutée, tels qu'ils ont été décrits par Bester (2010).

Tous ces facteurs contribuent à la réussite d'une archive institutionnelle. Ils sont des conditions nécessaires. Cependant, d'une manière plus générale, d'autres aspects conditionneront le développement durable d'un tel site pour les années à venir :

La communauté : Dans leur analyse d'une archive institutionnelle de l'Université Cornell, Rupp et LaFleur (2004) soulignent l'importance d'associer les chercheurs et l'administration dès le début au projet et au *workflow* du site, par des actions de communication et des contacts directs (*marketing one-to-one*), afin de susciter une prise de conscience, d'identifier des publications très en amont du site, et d'inciter leur dépôt. Leur verdict : sans communauté, pas d'archive institutionnelle. Pour le personnel de la bibliothèque, ceci nécessite une très bonne connaissance des communautés scientifiques et l'entretien d'une proximité avec les chercheurs. Cependant, il faut être conscient du décalage entre les intérêts et attentes des professionnels et du manque de motivation des chercheurs (Seaman 2011).

La qualité : Westell (2006) fait remarquer qu'une archive fourre-tout « qui contient tout pour tout le monde » n'a pas de sens et manque d'intérêt. Une action spécifique doit être engagée dès le début du processus, afin de garantir une qualité minimum du contenu, des données et des services. Des procédures de contrôle et des spécifications techniques doivent assurer la qualité de la sélection des documents et de l'ensemble du projet (Luzi et al. 2004). La démarche de certification entamée par le réseau allemand DINI depuis dix ans va dans ce sens — créer un référentiel pour évaluer puis conforter la qualité des archives institutionnelles.

L'infrastructure : Certains auteurs insistent sur l'intégration des archives institutionnelles dans l'infrastructure des systèmes d'information sur la recherche (CRIS). Les publications font partie des entités élémentaires d'un CRIS, et les archives institutionnelles peuvent proposer d'une part leurs métadonnées, leur

lien d'accès d'autre part, devenant ainsi un maillon important de l'infrastructure, avec des avantages à la fois pour l'organisation elle-même et pour ses chercheurs (Lambert et al. 2005). Dans un tel environnement, format (standard) et qualité (richesse) des métadonnées deviennent primordiaux.

Les nouveaux contenus : Jeffery (2007) défend l'idée d'une passerelle entre CRIS et archives institutionnelles et conseille de connecter les publications des archives à des données de recherche (*datasets*) d'autres réservoirs via les CRIS. Dans une étude sur l'évolution des publications scientifiques, Doorenbosch et Sierman (2011) ont décrit comment le lien aux données et résultats de la recherche crée une version augmentée et enrichie puis comment leur conservation à long terme présente un défi inédit aux archives institutionnelles. Ils prédisent que ces environnements virtuels de recherche collaborative vont devenir les futurs espaces de travail des chercheurs.

Perspectives

En théorie, tout paraît simple. La médiation, les pratiques numériques, les modèles économiques et le libre accès à l'information scientifique : les concepts sont simples, et c'est justement leur force. C'est sur le terrain que tout se complique. Quand nous avons voulu faire un bilan du développement des archives ouvertes en France, nous avons adopté une approche déductive, du concept vers le terrain. Au lieu d'imposer *a priori* une définition et une terminologie à l'observation, nous avons compilé et comparé le contenu des répertoires, annuaires et autres listes sur les sites du libre accès, tel que ROAR, OpenDOAR, CCSD ou COUPERIN (GTAO). Le résultat de cette approche était surprenant — un nombre élevé d'archives mais des réalisations et projets très variés, parfois non conformes aux théories et concepts. Dans certains cas, les différences prévalent sur les points communs.

En même temps, force est de constater qu'il n'y a que peu de consensus sur ce qu'est une archive institutionnelle. Seulement à peine la moitié des archives institutionnelles de notre étude a été répertoriée par l'ensemble des sites significatifs (tel que CCSD/HAL, OAIster etc.). Ceci pose un problème supplémentaire, celui de la terminologie. Comment appeler des sites hétérogènes, référencés de manière incohérente ? Y a-t-il un terme commun ? Nous avons adopté le terme d'archive institutionnelle sans perdre de vue les alternatives, telles que réservoir, plateforme, dépôt, serveur de publication etc. Le réseau allemand DINI par exemple a remplacé ce terme par *Dokumenten- und Publikationsservice*, « service de documents et publication » ou plus précisément « service de dépôts et de publication en libre accès », afin de souligner la convergence des archives ouvertes avec des plateformes d'édition en ligne comme par exemple revues.org en France.[26] Cela se discute.

À plusieurs reprises, diversité et choix terminologique ont fait l'objet de critique. Ici, nous avons adopté l'approche inverse, inductive, du terrain des réalisations vers le concept. Nous avons essayé de repérer à travers des études et projets sur le terrain les modèles émergents et cohérents, ainsi que les aspects et facteurs qui impactent réellement ou potentiellement leur développement.

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

Nous avons décrit quatre modèles « archétypes » : l'archive institutionnelle comme plateforme de publication, container, site patrimonial ou dépôt. Nous avons décrit les caractéristiques pour chaque modèle, conscients que la réalité est plus complexe, métisse. Pour tous les modèles, nous avons indiqué des facteurs qui contribuent à leur succès ou échec.

Ici, la connaissance des communautés scientifiques et leur soutien est primordial[27]. Tant qu'une archive institutionnelle, aux yeux de ces communautés, reste l'initiative d'informaticiens et/ou de bibliothécaires, parfois imposée de surcroît par un conseil d'administration ou scientifique, son succès et son intérêt pour les chercheurs seront tout relatifs et de portée limitée.

Or, placer les communautés scientifiques au cœur des projets n'est pas anodin et ne se réduira pas à des campagnes de sensibilisation, communication ou de formation. Concrètement cela implique deux constats : D'abord admettre que le même concept et modèle ne convient pas à l'ensemble des disciplines, organismes, universités, laboratoires ou types de documents et fichiers. Puis tenir compte des besoins et priorités des chercheurs, enseignants et étudiants qui ne sont pas nécessairement les mêmes que ceux des métiers ou administrations. Citons-en trois, à titre d'exemple :

- **L'accès au texte intégral :** Une archive ouverte qui contient majoritairement des métadonnées sans texte intégral est aussi attractive qu'une bibliothèque dont les rayonnages sont vides. De même, les documents confidentiels, classifiés ou sous embargo n'ont pas leur place dans une archive ouverte mais sont à gérer via des catalogues ou des bases de données.
- **Une qualité minimale :** L'idéal serait une archive contenant des *post-prints*, comme Stevan Harnad ne cesse de clamer. Mais même sans *peer review* ou autre forme de sélection, les documents déposés doivent répondre à certains critères de qualité, par validation (thèses soutenues), labelling (tampon institutionnel) etc. Accepter des dépôts sans intérêt revient à garder des papiers inutiles sur le bureau. L'auto-archivage ne crée pas automatiquement une collection qui pourrait intéresser une communauté scientifique.
- **Pérennité :** Les archives institutionnelles doivent offrir une solution fiable pour garantir la conservation et la diffusion pérenne des documents déposés (format et interprétation des données). La pérennité est une notion toute relative ; le DINI (*op. cit.*) demande une garantie de cinq ans minimum — c'est peu en égard à d'autres supports mais probablement réaliste.

Les archives institutionnelles sont devenues un vecteur significatif de la communication scientifique. Cependant, leur destin est aussi lié au développement des institutions. La légitimité des institutions, leurs choix stratégiques mais aussi, leurs infrastructures numériques impactent l'avenir de ces sites. Il n'est pas certain qu'ils survivent dans leurs formes et fonctions actuelles. Il n'est pas certain non plus que les archives institutionnelles réussissent le grand écart entre les intérêts, attentes et besoins de toutes les parties prenantes. Mais il est certain que si elles tournent le dos aux communautés scientifiques, leur destin sera celui des catalogues et autres bases de données : utiles certes, mais pas indispensables.

Notes

1. *Open Access Week 2012*, 22-28 octobre 2012 http://www.openaccessweek.org/.
2. http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre.
3. Source : OpenDOAR http://www.opendoar.org.
4. Source : DOAJ http://www.doaj.org.
5. Comme le modèle *OpenEdition Freemium* du CLEO http://www.openedition.org/8873?lang=fr.
6. Pour éviter aussi un autre malentendu : HAL n'est pas une archive institutionnelle du CNRS mais sert de plateforme pour les archives d'autres institutions.
7. Synthèse sur les résultats de l'enquête Archives ouvertes Couperin-ADBU-INRA-INRIA 2012 http://www.couperin.org/fr/news/261/1004.
8. Pour la liste totale des références, cf. http://www.citeulike.org/user/Schopfel/tag/institutional_repository.
9. La liste complète des répertoires et archives ouvertes se trouve en libre accès dans le rapport cité (Schöpfel et Prost 2010). Y figurent notamment les sites français répertoriés par ROAR, OpenDOAR, CCSD, COUPERIN, Scientific Commons et BASE.
10. La grille d'analyse est décrite dans le rapport Schöpfel et Prost (2010). Voici les grandes lignes :
    - *Information générale : 8 critères (nom, acronyme, URL, institution etc.)*
    - *Information spécifique : 13 critères (type d'archive, contenu, logiciel, taille etc.)*
    - *Contenu : 26 critères (littérature grise, rapports, articles etc.)*
    - *Données qualitatives : 7 critères (politique, métadonnées, validation etc.)*
    - *Commentaires : 4 critères (date de l'enquête, responsable de la collecte, commentaires)*
11. *Registry of Open Access Repositories Mandatory Archiving Policies* http://roarmap.eprints.org/.
12. http://jeannicod.ccsd.cnrs.fr.
13. http://pastel.archives-ouvertes.fr/.
14. http://edoc.hu-berlin.de/.
15. http://hal.in2p3.fr.
16. http://www.sportdocs.insep.fr/.
17. http://irep.ntu.ac.uk.
18. http://academiccommons.columbia.edu/.
19. http://www.uhp-nancy.fr/universite/actualites/petale_archive_ouverte_des_theses_et_memoires_numeriques.
20. http://dspace.mit.edu/.
21. http://dspace.mit.edu/handle/1721.1/7582.
22. http://orbi.ulg.ac.be/.
23. http://hal-bioemco.ccsd.cnrs.fr.
24. http://eprints.gla.ac.uk/.
25. Voir par exemple les dernières références http://eprints.gla.ac.uk/cgi/latest.
26. http://edoc.hu-berlin.de/series/dini-schriften/2010-3-fr/PDF/dini-zertifikat-2010-3-fr.pdf.
27. Voir aussi les résultats de l'enquête de COUPERIN qui vont dans le même sens.

Bibliographie

Baruch, Pierre. 2007. « La diffusion libre du savoir Accès libre et Archives ouvertes ». *L'Archicube* (3), 77–95. http://archivesic.ccsd.cnrs.fr/sic_00169330.

Bester, Emma. 2010. « Les services pour les archives ouvertes : de la référence à l'expertise ». *Documentaliste – Sciences de l'information* 47 (4): 4–15. http://archivesic.ccsd.cnrs.fr/sic_00550933.

[3.145.60.149] Project MUSE (2024-04-26 15:23 GMT)

Bester, Emma, et Ghislaine Chartron. 2010. *Difficile convergence des archives ouvertes en SIC*. http://archivesic.ccsd.cnrs.fr/sic_00627729.

Carr, Leslie, Wendy White, Susan Miles, et Bill Mortimer. 2008. "Institutional Repository Checklist for Serving Institutional Management." In *Third International Conference on Open Repositories 2008*, 1–4 April 2008, Southampton, United Kingdom. http://pubs.or08.ecs.soton.ac.uk/138/1/IRChecklist.pdf.

Doorenbosch, Paul, et Barbara Sierman. 2011. "Institutional Repositories, Long Term Preservation and the Changing Nature of Scholarly Publications." *Journal of Digital Information* 12 (2). http://journals.tdl.org/jodi/index.php/jodi/article/view/1764/1851.

Foulonneau, Muriel, Anne-Marie Badolato, Wolfram Horstmann, Karen Van Godtsenhoven, Mary Robinson, Sophia Jones, Martin Feijen, et Kasja Weenink. 2007. *Réseaux d'archives institutionnelles en Europe : logiques de développement et convergences*. Rapport. http://archivesic.ccsd.cnrs.fr/sic_00205049.

Friend, Frederick. 2011. *Open Access Business Models for Research Funders and Universities*. Rapport. Copenhagen: Knowledge Exchange. http://www.knowledge-exchange.info/Files/Filer/downloads/Open Access/KE Briefing paper_BUSINESS_MODELS_FOR_RESEARCH_FUNDERS_AND_UNIVERSITIES-08-09-11.pdf.

Guédon, Jean-Claude. 2008. *Accès libre, archives ouvertes et États-nations : les stratégies du possible*. Rapport. Université de Montréal. http://archivesic.ccsd.cnrs.fr/sic_00277755.

Harnad, Stevan, Tim Brody, François Vallières, Les Carr, Steve Hitchcock, Yves Gingras, Charles Oppenheim, Chawki Hajjem, and Eberhard R. Hilf. 2008. "The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update." *Serials Review* 34 (1): 36–40. http://dx.doi.org/10.1016/j.serrev.2007.12.005.

Jeffery, Keith G. 2007. "Greyscape." In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape*, 10–11 December 2007, Antwerp, Belgium. http://www.opengrey.eu/item/display/10068/697881.

L'Hostis, Dominique, et Pascal Aventurier. 2006. *Archives ouvertes – Vers une obligation de dépôt? Synthèse sur les réalisations existantes, les pratiques des chercheurs et le rôle des institutions*. Rapport. Paris : INRA. http://archivesic.ccsd.cnrs.fr/sic_00115513.

Lambert, Simon, et al. 2005. "Grey literature, institutional repositories and the organisational context." In *Seventh International Conference on Grey Literature: Open Access to Grey Resources*, 5–6 December 2005, Nancy, France. http://www.opengrey.eu/item/display/10068/697820.

Luzi, Daniela, et al. 2004. "The integration of GL documents with a research information system on occupation safety and health." In *Sixth International Conference on Grey Literature: Work on Grey in Progress*, 6–7 December 2004, New York, NY.

Lynch, Clifford A. 2003. *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*. Rapport no. 226. Washington : ARL Association of Research Libraries. http://www.arl.org/resources/pubs/br/br226/br226ir.shtml. http://dx.doi.org/10.1353/pla.2003.0039.

Rupp, Nathan, et Lee J. LaFleur. 2004. "Making Grey Literature Available through Institutional Repositories." In *Sixth International Conference on Grey Literature: Work on Grey in Progress*, 6–7 December 2004, New York, NY. http://www.opengrey.eu/item/display/10068/697849.

Schöpfel, Joachim, et Hélène Prost. 2010. *Développement et usage des archives ouvertes en France. 1re partie : Développement*. Rapport. Lille : Université Charles-de-Gaulle Lille 3. http://archivesic.ccsd.cnrs.fr/sic_00497389/fr/.

Smith, Kathlin. 2008. "Institutional Repositories and E-Journal Archiving: What Are We Learning?" *Journal of Electronic Publishing* 11 (1). http://dx.doi.org/10.3998/3336451.0011.107.

Swan, Alma, et Chris Awre. 2006. *Linking UK Repositories: Technical & Organisational Models to Support User-Oriented Services Across Institutional & Other Digital Repositories*. Rapport. London: JISC. http://www.jisc.ac.uk/publications/reports/2006/linkingukrepositoriesreport.aspx.

Westell, Mary. 2006. "Institutional repositories: proposed indicators of success." *Library Hi Tech* 24 (2): 211–26. http://dx.doi.org/10.1108/07378830610669583.

Willinsky, John. 2005. *The Access Principle: The Case for Open Access to Research and Scholarship*. Cambridge: MIT Press.