Eléments d'une communication scientifique ouverte et publique

Laurent Romary

▶ **To cite this version:**

Laurent Romary. Eléments d'une communication scientifique ouverte et publique. Lisette Calderan; Pascale Laurent; Hélène Lowinger; Jacques Millet. Publier, éditer, éditorialiser. Nouveaux enjeux de la production numérique, De Boek, 2016, Information & Stratégie, 9782807306653. hal-01328192v4

HAL Id: hal-01328192
https://inria.hal.science/hal-01328192v4

Submitted on 15 Jul 2016

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 4.0 International License

Éléments d'une communication scientifique ouverte et publique

Laurent Romary
Inria, projet Alpage

Résumé : Le paysage actuel de la communication scientifique est en pleine tension, entre d'une part une communauté scientifique qui aspire à des mécanismes plus fluides et moins coûteux de diffusion des résultats de recherche et d'autre part, un tissu privé d'édition scientifique qui vise à préserver au maximum un chiffre d'affaires construit à partir de modèles d'un autre âge. Nous essayerons dans ce cadre d'identifier différents éléments qui permettraient de concevoir des environnements de diffusion de l'information scientifique plus adaptés aux possibilités techniques actuelles, mais aussi aux capacités financières et éditoriales de la recherche publique. De façon très pragmatique, nous repérerons des pistes d'évolution liées aux infrastructures techniques, aux licences, à la relecture par les pairs ou à la diffusion de données, en partant de l'expérience acquise au sein d'Inria depuis plusieurs années, mais nous essayerons aussi d'esquisser une vision intégrée qui puisse fournir une base cohérente pour les actions à mener à l'avenir.

Laurent Romary est directeur de recherche au sein du projet Alpage d'Inria, où il mène des recherches dans le domaine des humanités numériques et plus particulièrement sur la modélisation et la représentation de données en sciences humaines et sociales. Depuis plusieurs années, il a contribué à la définition des politiques d'information scientifique du CNRS, de la société Max Planck et d'Inria, où il a participé notamment à la mise en place d'une obligation de dépôt en archives ouvertes dans HAL. Il a aussi participé de longue date à la spécification et à la maintenance des directives de la TEI, notamment comme membre, mais aussi comme président du conseil technique de la TEI, et préside actuellement le comité 37 de l'ISO.Il dirige l'infrastructure Européenne DARIAH pour le développement de méthodes numériques en sciences humaines et sociales.
https://cv.archives-ouvertes.fr/laurentromary

1. Introduction

Cet article a pour objectif de présenter ce que peut (ou doit) être une politique institutionnelle d'information scientifique et technique (IST) au sein d'un grand organisme de recherche, en partant de l'exemple d'Inria. Cette présentation est motivée par la place toute particulière qu'a prise Inria ces dernières années dans le paysage français en instaurant une « obligation de dépôt » des publications dans l'archive nationale HAL[1]. En particulier, cette obligation s'accompagne, ou plutôt est précédée, d'une réflexion prospective sur le modèle cible que notre organisme souhaite viser dans toutes les composantes liées à l'accès et à la diffusion des connaissances scientifiques. Il est notamment essentiel de comprendre comment une politique affichée peut se traduire par des décisions élémentaires pragmatiques qui, mises bout à bout, peuvent contribuer à faire évoluer le paysage global de l'IST.
Pour cela, nous aborderons successivement un certain nombre de grands thèmes, politiques, éditoriaux ou techniques qui doivent faire partie de la définition au jour le jour d'une politique

[1] Hyper Archives en Ligne, développé et géré au sein de l'unité mixte de service CCSD (CNRS, Inria et Université de Lyon). Lien vers le portail Inria : hal.inria.fr

d'IST, en explorant occasionnellement des perspectives à moyen terme vers lesquelles nous aimerions tendre.
Ces réflexions sont le fruit d'une double expérience de chercheur en informatique fortement ancrédans la diffusion de données scientifiques [Romary, 2012] et de responsable ou conseiller en IST pour trois organismes de recherche, le CNRS [Romary, 2005], la Société Max Planck [Romary, 2007] et Inria.Elles n'engagent bien sûr pas les établissements mentionnés, ce qui garantit de ma part une assez grande liberté de positionnement. J'ai essayé de traduire au mieux la différence entre une position nettement individuelle (le « je ») et une réflexion de nature plus collective (le « nous »), en lien notamment avec mes collègues du réseau Information et Édition Scientifiques(IES) d'Inria.

2. Veillée d'armes

2.1. Il est l'or, Monseignor

Alors que cet article se veut une synthèse de différentes actions menées sur une période plutôt longue, sa rédaction est presque polluée par l'actualité. En quelques mois, voire quelques semaines, de nombreux sujets ont pris une importance toute particulière à l'échelon national ou européen :

- Le calendrier législatif a vu se recouvrir une initiative européenne pour la révision de la directive sur le droit d'auteur (Reda, 2015) et l'étude du projet de loi pour une République Numérique[2]. Dans les deux cas, les développements législatifs risquent d'influer le cadre dans lequel nous allons évoluer à l'avenir, qu'il s'agisse d'avoir une base pour des embargos à partir desquels le dépôt en archives ouvertes sera dans tous les cas autorisé, ou d'identifier un cadre permettant la libre fouille de contenus scientifiques ;
- Le thème de l'"*open science"* est très présent dans l'agenda de la présidence néerlandaise de l'Union Européenne (janvier-juin 2016) et un communiqué de presse[3] annonce, au moment où j'écris, une ambition claire pour 2020 d'ouverture intégrale des articles scientifiques, quelques mois après que la Société Max Planck a lancé une proposition particulièrement conservatrice pour le domaine de l'édition scientifique, nommée OA2020 en toute modestie ;
- On annonce le rachat de l'archive de publications en sciences sociales SSRN ou de l'outil de gestion de carnets de laboratoire Hivebench par Elsevier, quelques mois après que le groupe a acquis Mendeley, un des plus gros sites de gestion de bibliographies et de contenus scientifiques.

Il n'est pas possible de résumer ici l'ensemble des propositions, des différentes opinions et des arguments associés, mais nous pouvons identifier quelques points de repère généraux :

- La prise de conscience européenne concernant l'*open access* est croissante depuis les tensions du printemps 2006 où une pétition[4] recueillant près de 28 000 signatures dont de nombreuses institutions venait répondre à un communiqué très ferme de l'association STM[5] concernant son rejet *a priori* de toute politique favorisant la libre diffusion des articles scientifiques. Les financements des projets PEER et OpenAire, ainsi que le déploiement d'une politique de mise à disposition systématique des

[2] http://www.gouvernement.fr/action/pour-une-republique-numerique
[3] http://francais.eu2016.nl/documents/persberichten/2016/05/27/communique-de-presse---tous-les-articles-scientifiques-europeens-en-libre-acces-a-partir-de-2020

[4] http://legacy.earlham.edu/~peters/fos/2007/02/20000-signatures-for-oa-presented-to-ec.html
[5] http://legacy.earlham.edu/~peters/fos/2007/02/publishers-issue-brussels-declaration.html

articles associés à des financements européens dans le programme H2020 (EC, 2012) ont dynamisé la prise de conscience européenne en la matière, avec en 2016 le lancement d'une plate-forme sur la science ouverte[6] ;

- Parallèlement, les éditeurs privés se sont effectivement approprié le thème de l'*open access* et se positionnent comme moteurs dans le cadre d'un modèle économique auteur-payeur qu'ils ont su intelligemment prôner. Ce modèle a malheureusement servi de base de réflexion (!) à un certain nombre de rapports et initiatives plus que maladroits allant du rapport Finch (cf. Shieber, 2013) à OA2020 déjà cité, en passant par des campagnes de négociation avec des éditeurs pour en figer les conditions d'application en Grande Bretagne ou aux Pays Bas. Même la récente déclaration de la LERU, qui, malgré son titre agressif, reprend à son compte le terme de « transition »[7], issu du modèle OA2020, ne permet pas d'être complètement optimiste quant au niveau de vigilance des institutions universitaires européennes ;
- Enfin, l'épisode lié à SSRN n'est qu'un élément dans un paysage où de nombreux acteurs privés se sont engouffrés dans le secteur de l'information scientifique pour offrir eux-mêmes leur plate-forme d'archivage (Academia, Research Gate) ou de certification (F1000, ScienceOpen, My ScienceWorks, peer.us). Je vois avec effroi mes collègues chercheurs déposer leurs productions sur ces plateformes dont le modèle économique, la pérennité et la confidentialité vis-à-vis des contenus sont loin d'être assurés.

La faible présence institutionnelle au niveau de l'expression de nouvelles idées, de la prise d'initiative, le peu de capacité d'innovation publique doivent-ils nous laisser penser que la bataille pour une science ouverte et publique est déjà perdue ? Avons-nous la possibilité, en tant qu'institution, de réaffecter nos efforts à la définition de modèles alternatifs et surtout à l'implémentation de ceux-ci ? Nous allons essayer de reconstruire ici un peu d'optimisme en commençant par un petit retour en arrière.

2.2. Un rêve déçu ?

J'ai fait partie de ceux qui, il y a maintenant une dizaine d'années, ont pensé qu'il était possible de faire évoluer le paysage de l'édition scientifique de l'intérieur en participant à une redéfinition des mécanismes qui pouvaient lier chercheurs, bibliothèques, journaux et édition scientifique privée. C'était une époque d'innovation importante où de nouveaux acteurs se mettaient à proposer un modèle auteur-payeur natif tels que BMC[8] ou PLOS, accompagné parfois de réelles innovations éditoriales et partenariales.

L'une des initiatives les plus exemplaires de l'époque est de mon point de vue le travail mené autour du portefeuille de journaux de l'Union Européenne des sciences de la terre (EGU) en collaboration avec l'éditeur privé Copernicus[9]. La démarche de collaboration entre ces deux acteurs public et privé a permis d'explorer de nombreux domaines qui restent aujourd'hui encore très innovants :

- Les articles publiés étaient librement accessibles en ligne ;

---

[6] http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform

[7] Le terme de « transition » est principalement utilisé des les communications des tenants d'une bascule intégrale des journaux sous abonnement dans le modèle auteur-payeur.Il est malencontreusement (ou intentionnellement ?) repris aussi dans le titre même des conclusions du conseil : "Council conclusions on the transition towards an open science system", http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf

[8] Où un certain Jan Velterop était à l'œuvre pendant quelques années avant de rejoindre Springer.

[9] Fondé par un ancien chercheur de la société Max Planck, Arne Richter.

- Le modèle de financement reposait pour plusieurs journaux non pas sur des contributions auteur-payeur individuelles, mais bien sur un financement institutionnel direct de l'EGU ou de grosses institutions partenaires (CNRS, Max-Planck) ;
- Un modèle de coût auteur-payeur étagé permettait aux auteurs de choisir s'ils désiraient des services éditoriaux supplémentaires, notamment de *copy-editing* ;
- Enfin, une expérience extrêmement précurseur d'*open peer review* (relecture par les pairs ouverte) fut développée, qui permettait à la fois de rendre les rapports de relecture accessibles en ligne avec les articles acceptés, mais aussi à la communauté d'exprimer un retour sur les pré-publications[10] (Poeschl, 2010).

La démarche ouverte, dès la phase de soumission, a eu des conséquences extrêmement vertueuses sur l'ensemble du dispositif. La visibilité précoce des manuscrits a encouragé les auteurs à ne soumettre que des documents relativement solides, ce qui a considérablement réduit le nombre de soumissions. La pression sur les relecteurs devenant ainsi moins forte, la qualité des relectures s'est améliorée, tout en s'associant à un taux d'acceptation de plus en plus élevé. Les rapports étant rendus publics, leur qualité a aussi crû et la participation à la relecture des journaux correspondants est devenue un élément de notoriété pour leurs auteurs. Les journaux de ce programme ont fortement bénéficié de ces développements : moins de charge administrative de gestion de mauvais manuscrits, facilité accrue à trouver des rapporteurs et au final un indice de citation en croissance stratosphérique sur une période de temps très courte. Alors que des bénéfices similaires ont été observés dans d'autres communautés, comme la biologie avec EMBO (Pulverer 2010), on peut réellement s'étonner que si peu d'autres initiatives aient vu le jour, notamment en SHS.

Une deuxième expérience particulièrement intéressante fut de suivre la mise en place de l'initiative SCOAP3 au sein du consortium initié par le CERN à la fin de 2005. Cette initiative vise à faire passer un groupe de journaux en physique des hautes énergies d'un modèle par abonnement à un modèle en accès ouvert, directement financé par les bibliothèques d'un large groupe d'établissements d'enseignement supérieur et de recherche. Même si la phase opérationnelle, qui n'a débuté qu'en 2014, et se voit maintenant renouvelée jusqu'en 2019,se présente comme un modèle Gold classique avec paiement de frais de publications[11], il est important de rappeler les principes fondateurs qui ont guidé les développements initiaux de SCOAP3 :

- Création d'un groupe de journaux dont on découple la propriété de l'entité éditrice qui va en gérer la publication ;
- Mise en concurrence des éditeurs sur l'attribution d'un journal, notamment en ce qui concerne l'offre tarifaire ;
- Rédaction d'une charte qui détermine le degré de service minimal attendu des éditeurs retenus pour gérer un journal ;
- Financement centralisé de ces journaux afin qu'ils soient tous disponibles en accès libre par création d'un consortium d'entités contributrices ;
- Intégration de l'ensemble du corpus des articles ainsi financés au sein d'un serveur centralisé géré au CERN : repo.scoap3.org

Au final, avec 3000 bibliothèques impliquées, on peut estimer que SCOAP3 est plutôt un succès, même si le portefeuille de 10 journaux est somme toute assez réduit. Les APC par article restent limités, de 500 à 2000 $, et toutes les publications sont associées à une licence

[10] Nous utiliserons ce terme de pré-publication pour désigner les manuscrits auteurs avant leur publication dans un journal ou une conférence, comme équivalent du terme *preprint* en anglais.

[11] Article Processing Charges (APC)

Creative Commons avec attribution (CC-BY[12]) garantissant une large variété de réutilisations possibles (comme la fouille de données). L'élément le plus important est pour moi la mise en place d'une sorte de groupement de commande centré sur le CERN qui a pu donner au fil des ans non seulement une crédibilité à l'initiative, mais surtout un réel pouvoir de négociation. On s'approche ici d'un mécanisme par lequel c'est effectivement la force publique qui définit les conditions de mise en œuvre de journaux scientifiques par des tiers privés. Même si le résultat reste très conservateur vis-à-vis des journaux existants et du fonctionnement de ceux-ci, c'est un exemple à connaître quand on cherche à établir la cartographie des modèles possibles d'édition scientifique.

Le dernier exemple que je voudrais présenter ici est bien sûr la négociation menée en 2007 avec Springer alors que je dirigeais la bibliothèque numérique de la société Max Planck. À l'époque, le modèle auteur-payeur n'était que très peu discuté et l'essentiel de l'attention portait sur les grandes négociations institutionnelles et nationales qui étaient menées avec les éditeurs ("Big deals"). Springer avait cependant introduit depuis quelques mois un mécanisme innovant, appelé *Open Choice* [13], qui permettait à des auteurs, moyennant un forfait de3000dollars[14], d'avoir leur article disponible en accès libre au moment de la publication (ce que nous appelons maintenant le modèle hybride).
Fin 2007, la société Max Planck devait renouveler son contrat avec Springer, et comme d'habitude, l'éditeur avait laissé traîner la négociation pour pouvoir forcer la main à l'organisme de recherche, sachant que si l'accord n'était pas renouvelé, les accès aux articles seraient bien sûr coupés[15]. La négociation tourna mal. La Max Planck s'était donné un objectif économique : diminuer d'un quart le budget d'abonnement. Springer s'y refusait. Fin 2007, la Max Planck annonça la rupture du contrat et dut gérer une situation de crise non anticipée : achat des articles à l'unité, abonnements locaux pour des journaux ou des collections essentiels, et surtout d'énormes pressions sur la présidence de l'organisme pour que l'on rétablisse le dialogue avec l'éditeur.Du côté du directoire de Springer, la pression était aussi forte de peur de perdre un client aussi exemplaire que la société Max Planck.
Jan Velterop et moi avons alors obtenu le feu vert de nos directions pour trouver un compromis acceptable sans que les deux parties perdent la face, et nous en profitâmes pour pousser le bouchon de l'accès libre. Nous nous sommes vite arrêtés sur le principe que nous voulions expérimenter : un mécanisme offrant un accès libre complet pour les auteurs de la Max Planck. Le résultat, dans le contexte de crise, alla au-delà de nos espérances. L'accord obtenu intégrait les éléments suivants dans un cadre expérimental planifié pour deux ans :

- Diminution de 10% (de moitié de ce qu'exigeait la direction de la société Max Planck) du budget et augmentation du portefeuille de titres couvert par l'accord ;
- Mise en ligne sur le site de l'éditeur de tout article pour tout auteur[16]issu de la société Max Planck ;

[12]http://creativecommons.org/licenses/by/4.0/
[13] Le programme a été initié par Jan Velterop, passé entre-temps chez Springer comme Monsieur Open Access. Cf. (Velterop, 2007), et la réaction de Peter Murray Rust : https://blogs.ch.cam.ac.uk/pmr/2007/07/08/springer-i-resign-from-your-journal/
[14] Ce coût très élevé s'explique par l'extrême réticence du directoire de Springer à adopter ce mécanisme et ce fût Derk J. Haank, le PDG de Springer, qui pris lui-même la décision de donner suite à la proposition de Jan Velterop auquel il croyait.
[15] Dans un monde où les chercheurs n'ont pas toujours conscience des enjeux de l'information scientifique, ils sont souvent des instruments de pression efficaces sur les organismes de recherche.
[16] L'accord ne stipulait pas qu'il s'agissait de l'auteur principal !

- Reversement des articles correspondants dans l'archive de publication de la Max Planck.

Cet accord, que j'ai toujours présenté comme une réelle négociation de *services d'édition*, est arrivé beaucoup trop tôt techniquement et conceptuellement. Il fût extrêmement critiqué en interne de part et d'autre et au final ne fût pas renouvelé. Les responsables documentaires traditionnels de la société Max Planck trouvaient que cet accord de diffusion, dont les tenants n'avaient pas, à ma connaissance, été expérimentés à cette échelle avant, venait polluer des négociations qui avaient trait à l'abonnement. La direction de Springer, à l'exception de son PDG, craignait que l'éditeur s'engage dans un modèle qui ne lui serait pas nécessairement bénéfique (sic).

Il y a une certaine ironie à faire le point sur cette initiative quand on observe que quelques années plus tard, le modèle expérimenté à l'époque est devenu la référence pour les grands éditeurs commerciaux et que la société Max Planck s'engage maintenant dans une initiative similaire à grande échelle, mais dans un contexte qui est devenu beaucoup plus hasardeux.
Le modèle auteur-payeur est maintenant un outil pour préserver le chiffre d'affaire de l'édition. Les coûts exorbitants, jusqu'à 7500 € pour un article dans The Lancet, ne permettent pas plus qu'auparavant aux institutions de maîtriser leur budget, et surtout, une économie spécifique d'éditeurs prédateurs[17] s'est mise en place pour faire publier facilement, mais moyennant finance, des articles sans réel contrôle scientifique. On a surtout perdu, au fil des discussions sur les modèles économiques, la nécessaire réflexion sur les services attendus d'un journal scientifique. Peut-on penser les choses autrement ?

3. Défendre une vision

3.1. Obligation de dépôt ?

Revenons maintenant à la politique, affichée et implémentée, d'Inria en matière d'information et d'édition scientifique. Vu de loin, le réflexe est en général de qualifier cette politique de "green". « Inria a un mandat de dépôt ! »[18], entend-on généralement. Il est donc important de rappeler qu'Inria n'a que très récemment instauré cette obligation de dépôt que nous sommes par ailleurs encore très peu à avoir mis en place en France[19], mais a plutôt défendu une politique d'ensemble de diffusion de ses productions scientifiques qui s'est récemment cristallisée autour d'une vision que nous pouvons résumer ainsi :

- Favoriser la dissémination maximale de nos productions, afin d'offrir une excellente visibilité à nos travaux, mais aussi et surtout de contribuer à la diffusion rapide des connaissances et des savoirs ;
- Constituer un corpus institutionnel fiable qui reflète la production de l'institut et défendre par cela un certain principe de souveraineté numérique ;
- S'affranchir des contraintes extérieures et/ou privées dans la préservation de ce patrimoine ;
- Contribuer à l'évolution du paysage de la communication scientifique, tant dans les processus éditoriaux que dans l'usage qui est fait des productions scientifiques.

Une telle vision ne peut cependant se concevoir sans tenir compte des contraintes et des possibilités offertes par les environnements techniques disponibles et surtout sans garder un

[17] https://scholarlyoa.com/publishers/
[18] J'adresse un salut amical à Serge Bauin qui rappelle régulièrement qu'avec un mandat de dépôt, plutôt qu'une « obligation », on ne se fait pas nécessairement ouvrir toutes les portes. Cf. https://fr.wikipedia.org/wiki/Mandat_de_dépôt_en_France
[19] dont notamment IFREMER.

œil attentif, quasiment minute par minute, sur les évolutions rapides que subit le paysage de l'information scientifique. C'est dans ce cadre que nous avons dû mettre en place très tôt, avant même d'instaurer notre fameuse obligation de dépôt, une politique concernant le modèle émergeant auteur-payeur (alias "gold", avec paiement de frais de publications)[20]. C'est donc naturellement par ce volet que nous commençons ici.

3.2. Gestion des frais de publication

Comme nous l'avons vu, l'apparition de journaux en auteur-payeur s'est faite dans un contexte plutôt accueillant de la part des différents acteurs du système de recherche, chercheurs, bibliothèques ou institutions. Des journaux tels que BMC, PLOS ou encore ceux de l'EGU (via Copernicus) offraient la possibilité d'expérimenter de nouveaux modes de publication à des coûts qui pouvaient paraître raisonnables à l'époque. L'émergence de propositions hybrides, comme celle de Springer,s'est accompagnée d'une prise de conscience du risque de payer deux fois pour le même contenu[21] et de voir émerger des propositions qui contrecarreraient une politique de diffusion des publications reposant sur le dépôt en archives ouvertes.

La définition d'une politique précise et prospective en la matière était particulièrement compliquée. Les grands éditeurs[22] communiquaient peu sur ce thème et, dans la période 2006-2012, centraient leur action sur une opposition ferme à toute forme de position politique de la commission européenne, notamment, en matière d'*open access*. De leur côté, les chercheurs étaient statistiquement trop peu en situation de fréquenter le modèle auteur-payeur pour que des débats puissent se cristalliser en leur sein. Pour une institution telle qu'Inria, il était de fait difficile de connaître l'étendue des frais de publication payés en direct par les chercheurs, et seuls des questionnements sporadiquesremontaientvia la direction ou le réseau documentaire.Nous n'avions donc, à la fin des années 2000, que peu d'éléments pour identifier une politique claire, avec surtout l'impression qu'il était difficile de prévoir sous quelle forme l'ensemble du système éditorial était susceptible d'évoluer. Et pourtant nous souhaitions avoir un impact sur cette évolution…

C'est sur cette base qu'Inria a défini sa position de *contrôle circonspect* du modèle auteur-payeur en conduisant l'analyse suivante, enrichie a posteriori, il faut bien l'avouer.

La mise en ligne d'un article de journal directement par les éditeurs privés contre paiement de frais de publication pose trois problèmes principaux :

1. Elle ne règle en rien le problème des coûts documentaires puisque les acteurs commerciaux concernés sont essentiellement les mêmes (cf. infra), et qu'ils ont pour objectif global au mieux de maintenir et si possible d'augmenter leur chiffre d'affaire ;
2. Elle ne permet pas non plus de négocier les offres de services associés, alors même que certains d'entre nous avaient identifié des moyens d'aller dans cette direction (cf.supra) ;
3. Elle entraîne surtout un fractionnement des budgets documentaires, ramenant le paiement des frais de publications au niveau des chercheurs eux-mêmes, avec l'impossibilité d'en contrôler d'éventuelles dérives.

[20] Un autre salut amical envers Jean-Claude Guédon, qui me rétorquera que la voie dorée ne se limite pas au paiement de frais de publications, ce à quoi je réponds un peu plus loin (section *Arrêtons d'avaler des couleurs*).

[21] Les abonnements et les frais de publications se cumulant : l'expression consacrée en anglais est *double dipping*.

[22] En incluant ici les actions de communication et de lobbying de la très active association STM.

Ce dernier point a servi de base au développement de la politique d'Inria en la matière.Nous avons identifié en effet que la priorité à court terme était d'endiguer le risque d'hémorragie budgétaire si les chercheurs de l'institut se mettaient à adhérer massivement au modèle auteur-payeur (natif ou hybride). Nous avons donc mis en place un budget centralisé, géré par le réseau IES de l'institut, et par lequel les chercheurs doivent en théorie obligatoirement passer pour payer leur frais de publication.
La mise en place de ce budget centralisé s'est accompagnée d'une politique très ciblée concernant ce qui peut être financé ou non dans le cadre de ce budget. La décision politique la plus forte fut de rejeter le paiement de frais de publication dans le cadre du modèle hybride, c'est-à-dire pour des journaux par ailleurs financés sur abonnement. Cette politique fut décidée très tôt, dès avril 2007, mais fut peu communiquée et déployée en interne car elle ne s'intégrait pas complétement dans une vision cohérente telle que nous pouvons la formaliser maintenant. Le rejet de l'hybride repose pour nous sur deux arguments principaux :

- Le paiement de frais de publication afin d'obtenir la mise en ligne d'articles acceptés dans un journal scientifique est par essence inutiledans le contexte de la politique Inria puisque nous exigeons que tout chercheur dépose son manuscrit final accepté dans HAL ;
- Le modèle hybride est probablement l'un des mécanismes les plus conservateurs que l'on puisse imaginer, puisqu'il ne s'applique qu'à des journaux existants et ce majoritairement au sein de grands groupes internationaux qui ont su déployer des mécanismes transactionnels pour gérer de tels paiements (comme c'est le cas pour Springer par exemple).

Très clairement, nous pensons que le budget correspondant doit rester disponible pour les actions de recherche des équipes et, dans le cas de budgets centraux d'information scientifique, ne pas affecter notre capacité à innover.Bien que la démarche soit particulièrement facile à comprendre, et de fait largement partagée par les chercheurs Inria, nous ne sommes pas en mesure d'effectuer un contrôle administratif étroit de tels paiements. Il arrive ainsi régulièrement que nous apprenions que tel ou tel chercheur a utilisé ses crédits de recherche pour financer la mise en ligne d'un de ses articles dans un cadre hybride. Je pense à titre personnel qu'il ne s'agit que d'écarts sporadiques qui doivent juste nous encourager à communiquer toujours plus sur la logique et les ambitions de notre démarche.

La politique mise en œuvre chez Inria a reçu deux types de commentaires ou critiques auxquels nous essayons de répondre ici.
La première objection est souvent venue de nos partenaires d'autres établissements qui ont eux-mêmes rejeté l'idée de mettre en place un mécanisme similaire de centralisation des frais de publication.Au-delà d'une réaction immédiate sur la complexité administrative inhérente à l'organisme (facultés, départements ou instituts, étanchéité des lignes budgétaires), les réticences reposent principalement sur les arguments suivant :

- La liberté du chercheur : celui-ci doit pouvoir continuer à publier où il veut, et l'institution dont il relève ne doit pas le forcer à des choix particuliers en la matière ;
- L'absence de fonds disponibles pour financer les APC, souvent associée au fait qu'on ne sait pas exactement combien de frais de publications sont effectivement dépensés dans l'institution correspondante ;
- L'évocation d'un risque d'explosion budgétaire qui pourrait résulter de l'existenced'un budget non plafonné et disponible à loisir ;
- Enfin, le risque de « découpler le donneur d'ordre et le payeur, comme pour les abonnements, ce qui déresponsabilise le donneur d'ordre »[23].

[23] Serge Bauin, communication personnelle.

Reprenons les arguments un à un.
La liberté du chercheur de publier où il veut ne doit effectivement pas être remise en cause si on veut que la politique de centralisation budgétaire soit efficace. En gros, nous devons prendre en charge dans le budget commun *tout* ce que le chercheur aurait de lui-même pris sur son budget d'équipe de recherche. Le chercheur se retrouve malgré tout influencé puisqu'il apprend à dialoguer avec les responsables de la ligne budgétaire, qui vont par exemple lui signaler l'existence d'éditeurs prédateurs ou fournir des points de comparaison sur le niveau de frais demandé par tel ou tel journal.
À l'échelle d'une institution, l'absence de fonds est évidemment une fausse question qui relève plus de l'organisation interne. Si on considère que l'on prend en charge, au niveau central, des coûts qui existeraient de toute façon au niveau des individus ou des équipes de recherche, il s'agit seulement de déterminer comment en assurer systématiquement la couverture. La réelle question stratégique pour une institution est ici de savoir s'il s'agit de frais de recherche, pris par exemple sur les prélèvements effectués sur les contrats, ou de frais d'infrastructure et donc intégrés au budget dédié nativement à l'information scientifique.
Le débat porte sur l'impact d'une telle politique sur une maîtrise, ou au contraire une expansion incontrôlable, des coûts de publication en auteur-payeur. La réflexion repose sur la mise en perspective de deux types de mécanismes de régulation :

- Quand les frais de publication sont pris sur les budgets individuels, une régulation naturelle s'opère au niveau du chercheur qui peut voir à tout moment son niveau de dépense. Cet argument est en partie affaibli par la pression naturelle qu'a le chercheur à devoir publier (« à tout prix ») et qui va donc considérer facilement que cela justifie des sacrifices sur d'autres postes, mais aussi au fait que dès qu'une équipe dispose d'un contrat de recherche conséquent (ANR ou EU), ce qui n'est pas rare chez Inria, le budget de l'équipe devient rapidement « infini » pour l'individu concerné ;
- Quand les frais de publications sont pris sur un budget centralisé, on passe d'une logique d'un acte anodin de dépense à celle d'une demande de prise en charge par un tiers, ce qui entraîne forcément une réflexion sur le bien-fondé de faire une telle demande. Là aussi, cet argument est contrebalancé par le fait que si le budget central n'est pas plafonné et que le chercheur identifie que toute demande sera automatiquement acceptée, il risque effectivement d'abuser de la possibilité correspondante.

La liaison entre le donneur d'ordre et le payeur dans la mise en place d'une telle politique est essentielle. Alors que nous avons fait le choix d'un budget centralisé, le déploiement de celui-ci repose sur un dialogue de proximité avec les chercheurs afin qu'ils soient eux-mêmes dans la situation de demander un soutien aux personnels documentaire, sans que ce soutien soit perçu comme automatique. Les chercheurs prennent maintenant régulièrement l'habitude de demander en amont, avant soumission de leur article, s'il est raisonnable d'envisager le paiement de certains APC, où même, dans certains cas, si la revue correspondante est vraiment sérieuse.
La pertinence de la politique adoptée chez Inria dans ce cadre repose essentiellement sur le pari que nous avons fait de nous doter petit à petit d'un réel instrument d'observation en continu de ce qui est dépensé en frais de publication et, par ce biais, de fournir des élément objectifs qui contribuent à une culture partagée par l'ensemble des acteurs, chercheurs, documentalistes et direction de l'institut sur la façon dont le modèle auteur-payeur est perçu et évolue.Inversement, je pense qu'il y a un risque non nul que certaines institutions, à ne pas vouloir prendre le phénomène en main, se fassent dépasser par la base et voient un budget – invisible – de frais de publication exploser en leur sein.

Une deuxième objection, qui me fait toujours chaud au cœur, vient de certains de nos chercheurs au sein même d'Inria qui s'étonnent que nous puissions afficher une réticence vis-à-vis du modèle auteur-payeur et offrir malgré tout la possibilité de payer intégralement les frais correspondants. Pourquoi dès lors ne pas tout simplement interdire toute contribution financière à ce modèle que nous devrions simplement rejeter ? Bien que sympathique, puisqu'elle indique que les chercheurs Inria peuvent être très engagés dans les réflexions liées à l'information scientifique, cette objection appelle trois types de réponse :

- Bien que le modèle auteur-payeur soit globalement peu répandu au sein de la communauté informatique, nous ne pouvons pas négliger le fait que certaines communautés, par exemple en bio-informatique, publient régulièrement dans des journaux reposant sur ce modèle ;
- Une réponse qui pourrait aussi s'appliquer à la première objection issue des autres institutions est que si le phénomène auteur-payeur s'étend, indépendamment de nous, nous risquons effectivement de ne pas avoir été prêts et d'avoir péché par inertie ou inactivité. Le risque de n'avoir aucun moyen de ralentir un échauffement budgétaire dans le domaine du paiement des frais de publication est réel ;
- Nous, globalement le secteur de l'enseignement supérieur et de la recherche, ne sommes pas encore opérationnelspour offrir des alternatives viables et déployées à grande échelle qui puissent offrir une troisième voie pour l'édition scientifique. Nous ne pouvons pas nier la réalité telle qu'elle se présente à nous en l'état.

Au bilan, j'estime que la politique de centralisation budgétaire a plutôt porté ses fruits. Même s'il est difficile de comparer avec d'autres institutions qui n'ont par définition pas les mêmes outils pour suivre le développement du modèle auteur-payeur, nous maintenons un budget extrêmement réduit de frais de publication. En 2015, nous avons ainsi dépensé 24 000€ pour 20 demandes de prise en charge, en estimant que moins d'une dizaine de paiements nous ont échappé. Il est particulièrement intéressant de constater que l'évolution au fil des années (cf. figure 1) est relativement stable, avec même un tassement ces deux dernières années. L'élément « objectif » le plus satisfaisant est effectivement le sentiment que les tenants de la politique sont largement partagés par les chercheurs Inria qui comprennent massivement les dangers du modèle hybride, et sont très prudents, voire résolument contre le modèle auteur-payeur en général.

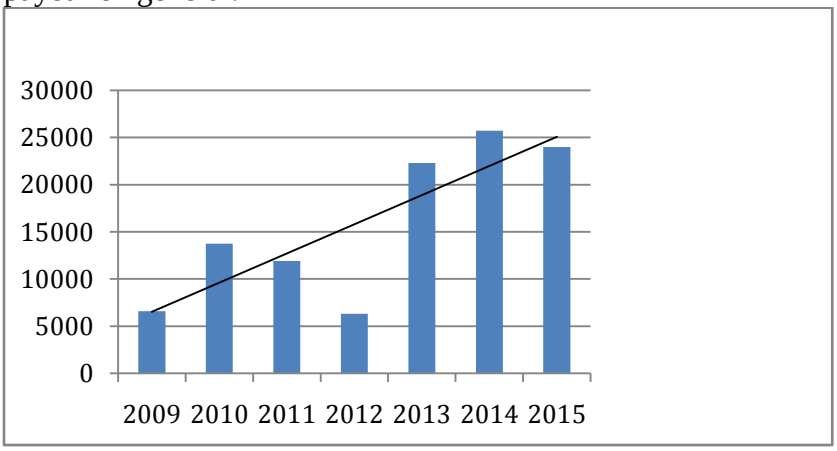

Figure 1 : Evolution des coûts centralisés de frais de publication chez Inria (en euros, par année)

3.3. Diffusioninstitutionnelle des productions scientifiques(was : Green)

Inria a très tôt identifié la nécessité de disposer d'un outil de large diffusion de ses productions scientifiques. Relevant d'une certaine fierté scientifique institutionnelle, justifiée, la démarche fut initiée par Jean-Pierre Verjus quand il était directeur de l'Information

Scientifique et de la Communication, et donc par construction en position de pousser cet agenda de communication scientifique. Après une phase d'exploration assez courte de quelques mois, il prit la décision d'ouvrir en avril 2005 un portail sous HAL[24] et d'y encourager les chercheurs de l'institut à y déposer leurs articles. En parallèle, il initia une politique à long terme d'accompagnement du portail en y impliquant le réseau de services IST présents dans les différents centres d'Inria. L'idée était de développer des services favorisant l'appropriation de la plateforme par les chercheurs : interface corporate, communication ciblée dans les centres de recherche, développement d'outils d'import et d'export, ainsi que de pages web d'équipe, et enfin mise en œuvre de mécanismes simplifiés pour la production du rapport des équipes qui déposaient leur production dans HAL. Bien que le terme d'obligation fut tabou, la perspective était bien d'aboutir au plus vite à une situation qui permette de couvrir la production de l'institut le plus largement possible.

C'est en effet la réalisation d'un deuxième objectif, en plus de la visibilité, qui a influencé la réflexion au sein d'Inria : défragmenter le corpus de publication. En effet, le système de communication scientifiquesur support imprimé tel qu'il s'est développé au fil des années reposait sur deux principes complémentaires :

- Délégation de la production des contenus à des éditeurs scientifiques qui gèrent la relecture par les pairs ainsi que la mise en forme précise rendue nécessaire par le caractère figé du support ;
- Achat massif et en de multiples exemplaires des objets imprimés qui sont réunis en corpus largement disponibles au sein des bibliothèques.

Le passage à l'électronique a remis en cause l'équilibre entre ces deux aspects puisque le support physique disparaissant, on pouvait imaginer que la diffusion des contenus pourrait devenir beaucoup plus facile, sans que l'on ait déterminé ce qui pourrait garantir la préservation du patrimoine associé. Ce glissement explique la tension qui s'est créée entre d'uncôté des éditeurs qui ont voulu préserver leur prérogative de diffusion et promettaient d'assurer la préservation patrimoniale des contenus, et de l'autre le monde de la recherche qui d'une part comprenait que la diffusion n'avait plus nécessairement à passer par des tiers et qui voyaitpar ailleurs, avec la disparition du rôle des fonds imprimés, le risque d'une perte de maîtrise sur un corpus cohérent. Les mécanismes de concentration du secteur de l'édition privée n'ont fait qu'augmenter les craintes.

Cette analyse a joué un rôle essentiel dans la démarche d'Inria et permet d'expliquer, presque dans le détail, la forme sous laquelle elle a été mise en œuvre. En effet, le passage d'une forte incitation à une obligation de dépôt, qui a fait initialement grincer des dents de nombreux chercheurs qui pensaient qu'on en resterait là, a résulté de l'observation qu'il fallait au plus vite reprendre notre souveraineté sur *l'ensemble* des productions de l'institut à la fois du point de vue de leur *diffusion*, mais aussi de leur *préservation*, et par la même de tout processus ultérieur d'*analyse*. Il s'agissait alors de ne plus confier à des éditeurs le soin d'assurer ces fonctions, mais bien de les prendre à notre charge.

C'est donc assez naturellement que la perspective d'une obligation de dépôt a pris corps dès lors qu'au niveau de la direction, des réseaux documentaires, mais aussi de nombreuses équipes de recherche, la perception des enjeux était croissante.

L'obligation de dépôt, annoncée en janvier 2013, et formalisée à l'automne 2015 par une double note d'implémentation, intègre de nombreux éléments que nous avons discutés jusqu'à présent.

[24] On voit ici la différence entre la fierté institutionnelle et une forme d'ego mal placé qui aurait pu nous pousser à mettre en œuvre notre propre archive de publications en dehors de toute démarche de mutualisation.

Tout d'abord, l'obligation de dépôt couvre l'ensemble de la production scientifique de l'institut telle que reflétée au sein du rapport annuel d'activité des équipes de recherche. Inversement, le principe d'obligation repose sur le fait que seuls[25] les articles déposés en texte intégral dans HAL sont pris en compte pour la production des rapports annuels.
La politique Inria, définie en fonction d'un objectif propre à l'institut, s'affranchit par ailleurs de nombreuses contraintes exogènes, notamment celles imposées par le monde de l'édition privée. Le devoir de dépôt n'est donc pas subordonné à une politique affichée officiellement dans des serveurs internationaux tels que SHERPA/ROMEO[26] et elle s'applique quelles que soient les exigences imposées par les éditeurs de signer un accord de transfert exclusif des droits d'auteurs[27]. Dans cette démarche, il y donc bien sûr peu de place pour la notion d'embargo qui me semble de toute façon incompatibleavec une vision plus large de dépôt en continu des manuscrits auteurstelle que nous allons ladécriredans la section suivante.

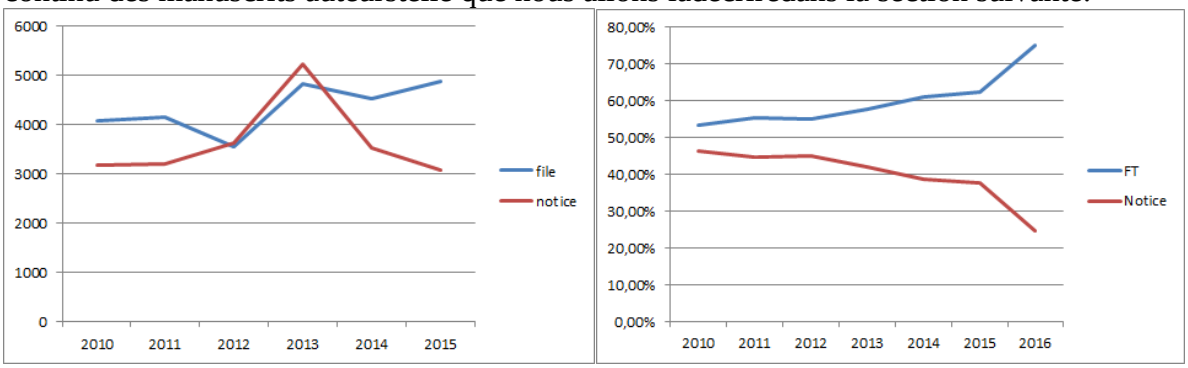

Figure 2 : Evolution du taux dépôt (notices et textes intégraux) chez Inria dans HAL

L'obligation de dépôt est aussi indépendante du modèle économique par lequel un article a été publié dans unjournal scientifique. Ainsi, quand un article est diffusé dans un journal en *open access* natif (avec ou sans paiement de frais de publication d'ailleurs), il doit tout aussi bien être déposé dans HAL. La situation est facilitée par le fait que l'article, dans sa version diffusée sur le site de l'éditeur, est en général associé à une licence Creative Commons. Nous pouvons donc déposer cette version publiée, en prenant bien soin d'y appliquer la même licence que celle affichée par l'éditeur.
Ce point nous amène à préciser qu'il ne nous semble pas judicieux d'aller trop loin dans le déplacement des lignes de forces entre institutions de recherche et éditeurs privés, au moins pour ne pas trop mettre les chercheurs en première ligne, et donc ne pas encourager le dépôt des versions éditeurs des articles scientifiques. L'obligation de dépôt Inria porte donc sur la version auteur après relecture par les pairs, même si de fait, comme nous allons le voir, la vision que nous défendons nécessite d'avoir une ambition un peu plus élevée.

[25] Dans une phase se transition, et pour tenir compte de la spécificité de certaines communauté (par exemple en bio-informatique), l'obligation s'applique avec une certaine souplesse du point de vue du texte intégral (cf. Figure 2). Par ailleurs, les articles déposés dans arXiv ou PMC sont bien sûr éligibles.
[26] J'insiste souvent sur le fait qu'on ne peut déployer une politique de dépôt institutionnel en gardant un œil sur ces serveurs et j'encouragerais plutôt de ne jamais regarder ce qui s'y trouve.
[27] Le présent article n'est pas le lieu de discuter de la légalité de tels accords de transfert de droits dans le droit français, mais un indicateur fort est l'absence d'un dépôt de plainte depuis autant d'années. La perspective d'un article spécifique dans la future Loi pour une République Numérique cadrant au maximum ce que peut exiger un éditeur nous facilite aussi le travail de conviction.

Au bilan, et après trois ans de mise en œuvre, nous pouvons considérer que l'obligation de dépôt, avec plus de 100% de couverture descriptive[28] et environ 80% de textes intégraux (chiffre qui ne semble pas vouloir se tasser), est une vraie réussite.
Une enquête effectuée auprès des chercheurs n'ayant déposé qu'une notice sans texte intégral a montré que le dépôt obligatoire générait encore des incertitudes chez certains chercheurs, qu'il nous faudra mieux comprendre et gérer. Nous n'avons aussi que peu déployé une politique uniforme du point de vue des licences à associer à chaquedépôt, même s'il y a unanimité pour identifier la licence Creative Commons avec attribution comme la meilleure base de travail.

3.4. Du vert clair au vert foncé

L'obligation de dépôt n'est en fait qu'un point de passage, certes obligé, vers la mise en œuvre d'une vraie reprise en main tant individuelle qu'institutionnelle de nos productions scientifiques. Il s'agit, comme nous l'avons esquissé dans la deuxième partie de (Romary, 2012b) d'aller vers un espace fiable de gestion, d'archivage et de diffusion des contenus scientifiques sous toutes leurs formes.
De ce point de vue, la politique Inria comporte une réelle incitation envers tous ses chercheurs à considérer le dépôt des documents scientifiques au stade le plus précoce possible de leur production. L'idée est ici de considérer HAL comme un canal de diffusion des idées, dès leur mise par écrit en encourageant le dépôt de manuscrits à toutes les étapes et dans toutes leurs versions.
Il y a de fait une réelle logique derrière cette démarche, dont les avantages sont multiples.
Le premier avantage est de pouvoir cibler la diffusion d'une première version d'un article vers une poignée de collègues ou plus largement vers la communauté de recherche concernée, notamment via les réseaux sociaux. Cette diffusion, certes classique en physique des hautes énergies avec le serveur arXiv (Gentil-Beccot et alii, 2009), s'avère être particulièrement bénéfique pour le dialogue scientifique,puisqu'elle offre la possibilité d'un retour constructif de la part de collègues avant la diffusion d'une version plus stable et sa soumission à un comité éditorial.
Un des avantages à une diffusion précoce des manuscrits est bien sûr de rendre la visibilité d'un article indépendante de possibles délais de publication par un éditeur et donc de pouvoir être lu, et par conséquence cité, bien avant qu'une version formelle soit disponible.C'est le cas également pour les chapitres d'ouvrages collectifs, qui sont soumis à des délais de publication encore plus importants que les articles.
Une autre conséquence, certes moins connue, est la possibilité pour le manuscrit concerné d'être utilisé comme document d'antériorité dans une recherche de brevet. Dans ce cadre, la temporalité du document est essentielle, car l'acceptationd'une demande de brevet va dépendre de l'existence ou non de travaux antérieurs « rendus publics » portant sur la même invention. Une archive de publications telle que HAL, dont la stabilité et surtout les mécanismes de datation sont reconnus par les organisations concernées, peut dès lors servir de référence pour identifier de tels documents. On trouve ainsi régulièrement dans les bases de brevets des citations de documents queles examinateurs sont allés trouver dans HAL et qu'ils ont référencé en lien à leur datation précise. À titre d'exemple, on peut voir que le manuscrit inria-00326784[29] du portail HAL-Inria a été utilisé comme document X (antériorité bloquant

[28] Sic. Les chercheurs déposent plus d'objets dans HAL que ce qui apparaît effectivement dans le rapport d'activité annuel des équipes.
[29] https://hal.inria.fr/inria-00326784

l'acceptation de la demande) pour la demande faite à l'OEB sous la référence EP2547095 A1[30].
Au-delà de la diffusion précoce et de la pérennisation des étapes d'un document dont une version est finalement publiée sous la forme d'un chapitre de livre, d'un article dans un journal scientifique ou de conférence, l'intérêt d'une gestion cohérente de toutes les versions d'un document permet aussi de considérer qu'il peut y avoir une vie après la publication pour le document en question. Rien n'empêche en effet de considérer qu'on peut vouloir publier (au sens de rendre public) des versions ultérieures qui vont bien sûr corriger des erreurs, comme on le fait couramment dans les journaux de mathématiques, mais aussi pour en compléter l'argument en fonction des retours de certains lecteurs. C'est typiquement le cas du présent chapitre, objet de commande, et qui est donc écrit sous la pression du temps. Archivée sous HAL, sa version publiée dans un ouvrage ne sera pas nécessairement la dernière version disponible en ligne.
Une autre raison qui peut faire évoluer un article au-delà de sa publication est le fait de vouloir ajouter des éléments complémentaires au document lui-même : il peut s'agir des transparents utilisés pour le présenter à une conférence, la vidéo de cette même conférence, un poster, ou éventuellement des données de recherche qui viennent compléter l'argumentaire de la publication.
On voit bien que la vision à moyen terme est de savoir gérer des documents vivants ("living documents") dont on identifiera occasionnellement des versions stables, par exemple parce qu'elles ont fait l'objet d'une relecture par les pairs. On peut ainsi imaginer qu'une version particulière en cite une autre plus ancienne pour marquer explicitement la différence. Dans ce cadre, l'objectif d'une plateforme telle que HAL n'est plus d'être un miroir d'un monde de publications qui lui est extérieur, fait de journaux, de livres etc. publiés par des éditeurs bien souvent privés, mais bien d'être un espace documentaire où le chercheur va gérer la continuité de sa production. Les notions de pré-publication, *post-print*(manuscrit auteur après relecture par les pairs) ou version éditeur deviennent ici rapidement obsolètes, puisqu'elles ne sont que le reflet d' « accidents » extérieurs au processus de production documentaire.

3.5. Au loup !

Une telle démarche volontariste de mise en ligne systématique adossée à une présence institutionnelle forte peut aussi être source de craintes multiples de la part des chercheurs, mais aussi parfois des personnels documentaires. Nous en reprenons ici les plus courantes.
Le premier niveau de crainte est lié à la relation avec l'éditeur privé chez qui l'article est publié et s'exprime de deux façons possibles :

- La peur de ne pas être publié : la présence d'une pré-publicationdans une archive ouverte pourrait empêcher l'acceptation d'une soumission à unjournal. Les cas effectifs de rejet sont rares et mon expérience personnelle sur deux cas où des rapporteurs exprimaient des scrupules à ce sujet a montré que les éditeurs en chef connaissent de plus en plus la notion de pré-publication ;
- La peur de subir des poursuites de la part des éditeurs s'ils déposent leur manuscrit en ligne après publication.

Un autre niveau de crainte souvent partagé par les institutions concernées porte sur la propriété intellectuelle et s'articule autour de deux risques principaux :

- La peur du plagiat, à savoir que la mise en ligne précoce de manuscrits puisse permettre plus facilement à d'autres d'en reprendre les contenus, tend à disparaître. Il semble que l'accroissement de la culture numérique a bien fait comprendre qu'au

[30]http://www.google.com/patents/EP2547095A1

contraire, la mise en ligne d'une information permet plus facilement d'en détecter d'éventuels plagiats ;

- Il reste parfois une crainte concernant la possible valorisation des résultats sous la forme de brevet par exemple. Là encore, de plus en plus d'acteurs sont conscient que la présence en archives ouvertes d'un article publié par ailleurs ne change rien sur le fond. Toute publication, quelle qu'elle soit, y compris une présentation orale à une audience un tant soit peu publique, est susceptible d'être opposée à une demande de brevet.

Le champ le plus sensible où s'expriment les craintes est au final surtout lié à la possible exploitation institutionnelle des contenus de l'archive ouverte. On entend crier au loup que les institutions, de par leur investissement croissant dans HAL et, pour certaines, par l'établissement d'une obligation de dépôt, cherchent à contrôler étroitement le travail des chercheurs. L'argument me paraît toujours étrange quand, d'une part, on accepte, en tant qu' employé d'une institution de recherche, d'avoir comme obligation minimale de justifier nos activités, et d'autre part, que les institutions concernées, en l'absence de données plus fines (je dirais plus intelligentes) se sont parfois fourvoyées à faire reposer leurs analyses sur des bases privées très chères telles que le Web of Science (Thomson Reuters) ou Scopus (Elsevier). J'estime que là encore il est indispensable de reprendre notre souveraineté informationnelle sur la façon dont on veut analyser la production scientifique des individus et des établissements, avec une priorité probablement plus grande donnée aux outils de pilotage stratégique.

Afin de lever tout doute en la matière, il est essentiel de défendre et de mettre en œuvre une démarche où la responsabilité des dépôts et de leurs usages incombe au chercheur avec l'entier soutien de son institution. L'acte de dépôt, même s'il est occasionnellement délégué à un assistant ou un documentaliste, doit rester de l'ordre de la responsabilité individuelle. Le cas de l'obligation de dépôt Inria est exemplaire. Couplé à la production du rapport annuel, il permet au chercheur de choisir ce qu'il veut y voir apparaître et donc ce qu'il va déposer[31].

De la même façon, le chercheur doit pouvoir lui-même définir les indicateurs qui le caractérisent. L'utilisation de requêtes permettant de compter tout type de données ou de combinaisons de celles-ci dans HAL ne doit pas être réservée à un groupe fermé d'experts en bibliométrie, mais faire partie du tableau de bord offert au chercheur. Il s'agit là d'une forte priorité pour les développements à venir de HAL pour accroître la confiance vis-à-vis de la plateforme. Un travail en ce sens, en flattant occasionnellement la fierté personnelle des chercheurs devrait aussi contribuer à renforcer la base des utilisateurs.

3.6. Nouveaux modèles éditoriaux

Une politique de diffusion des contenus scientifiques doit reposer sur trois piliers : la maîtrise des flux exogènes et de leurs conséquences économiques, la réappropriation des contenus sur des plateformes ouvertes et publiques et enfin l'éditorialisation des contenus afin de mieux gérer leurs conditions de production et d'évaluation. Ce dernier aspect est ainsi l'un des développements récents de la politique Inria d'information scientifique et repose sur deux directions complémentaires :

- L'expérimentation du développement d'épi-journaux dans le cadre de la plateforme Episciences (Berthaud, 2014). Il s'agit ici de comprendre comment des journaux peuvent se développer comme des espaces de certification de contenus déposés préalablement en archives ouvertes, de sorte à différentier complètement, dans le

[31] On doit savoir aller plus loin en permettant au chercheur de sélectionner, parmi ses dépôts, ceux qui seront visibles dans un rapport particulier. Ce doit être un des éléments du futur espace de travail du chercheur.

processus de diffusion d'information scientifique, la publication proprement dite d'une validation par les pairs ;

- Le soutien à l'utilisation de SciencesConf comme environnement public de gestion de conférences. L'articulation avec l'archive ouverte intervient ici en amont puisque la sélection des articles s'effectue avant qu'une collection soit effectivement créée sur HAL.

Les deux initiatives vont dans le même sens, il s'agit non seulement de disposer de plateformes qui implémentent les fonctions de base de la communication scientifique telle que vécue par le chercheur et ainsi en maîtriser les coûts, mais aussi au passage d'ouvrir une boîte de Pandore supplémentaire concernant la propriété et les droits afférents aux actes de relectures des chercheurs. Ainsi, les évaluations, critiques ou commentaires, souvent accompagnés de suggestions de nature scientifique sont actuellement absorbés par les plateformes commerciales d'hébergement de journaux, mais aussi des environnements privés de gestion de conférences (easychair, softconf).

Nous avons tous un devoir d'expérimentation en la matière, mais nous devons aussi offrir de réels services qui à terme, bien sûr, doivent être mutualisés à l'échelle nationale et internationale. Maîtriser un outil tel qu'Episciences est aussi l'opportunité d'envisager d'autres applications qui correspondent à de réelles attentes des communautés scientifiques, notamment dans le domaine des données ou des logiciels. On peut ainsi imaginer facilement (Romary et alii, 2016) qu'Episciences puisse servir de base à la mise en place de *data journals* comme certification de jeux de données présentes sur des plateformes d'hébergement telles qu'Ortolang (pour les ressources linguistiques).

Pour conclure rapidement sur cet aspect par ailleurs essentiel dans notre démarche institutionnelle, il est important de rappeler que nous ne considérons pas les épi-journaux comme la seule réponse à la recherche d'alternatives pour une communication scientifique ouverte, mais comme une opportunité d'aller jusqu'au bout d'un modèle qui, nous l'espérons, viendra compléter une offre publique plus large.

3.7. Mettre tout cela en musique – arrêtons d'avaler des couleurs

Le travail mené chez Inria à tous les niveaux, chercheurs, réseau documentaire, direction, a conduit à définir tout un ensemble de mesures parfois extrêmement précises destinées à mettre en œuvre une politique globale de diffusion des savoirs qui corresponde au mieux aux intérêts de l'institution et de la mission qu'elle doit servir : excellence dans la création des connaissances et transfert (et donc communication) de celles-ci vers la société économique et civile. Cette politique n'a jamais vraiment été définie sur la base de la nécessité de faire de l' « open access », ni même réellement d'adopter une voie verte ou dorée. Mon expérience au cours de ces années, tant chez Inria que dans d'autres organismes, m'a plutôt montré que ces grandes voies colorées favorisent plus une vision conservatrice du système de l'édition scientifique qu'elles n'ouvrent de réelles possibilités d'exprimer des transformations en profondeur de ce système.

Ainsi la voie dite verte, qui correspond principalement à l'idée que l'on donne accès aux *publications scientifiques* en en déposant une version dans une archive ouverte, repose par définition sur le principe qu'un système extérieur d'édition scientifique existe et qu'il faut un mécanisme parallèle de diffusion. Des acteurs comme Steven Harnad[32] ont ainsi souvent répété qu'il ne fallait pas se préoccuper de définir d'autres modèles, mais que le système basculerait de lui-même, comme par magie,quand tous les articles, publiés par ailleurs, seraient aussi disponibles en archives ouvertes, sans que l'on sache vers quel(s) modèle(s) cible(s) la bascule s'effectuerait. C'est ainsi que la voie verte porte avec elle par exemple la

[32] "Green is the self-archiving of *published, peer-reviewed articles*" (Harnad, 2005)

notion d'embargo qui n'est qu'un élément de synchronisation avec le système existant et immuable d'édition scientifique, et qui d'ailleurs a été introduit par les éditeurs scientifiques eux-mêmes. L'embargo[33] est une manière de rappeler que ce qui rythme la temporalité des publications, c'est bien le temps de la diffusion dans des journaux scientifiques. Une autre conséquence est souvent aussi un rétrécissement malheureux de la portée de ce qui peut ou doit être déposé dans une archive ouverte, puisque la référence en termes de documents est constituée par les articles de journaux. On marginalise ainsi souvent les manuscrits-auteurs avant publication, ou encore les productions « périphériques » telles que les présentations en transparents, les posters, ou tout autre matériau un peu différent de la structure d'un article.

La voix dorée n'offre guère plus de marge créative. C'est même pire, si l'on considère qu'il s'agit principalement de définir des modes de financement alternatifs pour les journaux scientifiques dans leur configuration actuelle. De fait, quels que soient les appels à concevoir cette voie dorée de façon extrêmement large ([Guédon, 2008]), on ne peut que reconnaître l'extrême habilité de l'édition privée en la matière et en particulier de l'association STM, qui s'est appropriée le concept pour en faire un élément de la démonstration que les éditeurs eux-mêmes pouvaient offrir l'*open access* au monde de la recherche.

Y a-t-il un espace d'innovation dans ce monde bicolore ? Des tentatives sont bien faites d'ajouter des couleurs à ce paysage[34], mais le problème de fond reste entier, cette simplification en grandes voies ne permet pas d'aborder la description d'initiatives récentes au bon niveau de granularité.

Quel est ainsi le statut d'un modèle d'épi-journaux tel que nous l'avons décrit avec la plate-forme Episciences ? Est-ce un modèle de voie verte puisque la diffusion se fait en archives ouvertes ? Est-ce une voie dorée,puisque la mise en ligne de l'article final est un acte de publication résultant d'une décision éditoriale du journal, de façon presque traditionnelle ?

Qu'en est-il aussi de modèles économique tels que le Freemium qui accompagne le déploiement de nombreuxjournaux de la plate-forme revues.org ? Une vision dorée ou même diamant (sans APC) est extrêmement réductrice pour décrire un mode de financement externe par les bibliothèques (et donc par les lecteurs), du fonctionnement des journaux (alors que la plateforme est vue comme une infrastructure financée par des fonds publics), qui elles-mêmes ne sont que partiellement en accès libre, puisque seule la version web (HTML) est librement accessible et que l'accès à d'autres formats ou services (par exemple le catalogage) est subordonné au paiement de la licence Freemium par les bibliothèques universitaires ou de recherche.

On peut aussi observer combien la focalisation sur des voies que l'on présente souvent comme alternatives a pu cristalliser des débats qui n'avaient pas lieu d'être. Des discussions récentes concernant les licences libres qui pourraient être associées aux articles diffusés en libre accès[35]ont ainsi été polluées par le fait que de nombreux éditeurs modulent la licence associée à un article publié en fonction du niveau d'APC effectivement payé. Il est ainsi devenu quasiment impossible d'ouvrir un débat sur des listes internationales telles que COAR sur le principe même qu'une licence simple avec attribution (CC-BY) soit optimale pour la diffusion d'un document scientifique, quel que soit le modèle effectif de diffusion adopté.

Pour conclure ici, la vision de l'accès ouvert en grandes voies colorées est selon moi plus un frein à la nécessaire imagination dont nous devons faire preuve dans la période actuelle de rapide évolution du paysage qu'un système simplifié de classification qui serait utile à la réflexion. Elle empêche d'une partl'expression d'une politique globale et cohérente, dès lors que l'on est forcé de se positionner vis-à-vis d'un modèle vert ou doré et, d'autre part, nuit à

[33] On lira l'intéressante analyse historique de (Altman, 1996).

[34] Comme le modèle « Diamant » (Farge, 2012)

[35] On notera la très bonne synthèse d'arguments à ce sujet dans (Couture, 2015)

l'expression d'une grande finesse dans les détails d'implémentation de cette politique, quand il s'agit de définir la couverture effective du dépôt dans une archive de publication, l'identification de mesures d'accompagnement (obligation, services), les usages des contenus (chercheur ou institution), la détermination de licences de diffusion ou de modèles pérennes de financement. Comme l'ont montré des rapports récents comme celui de Pierre Carl Langlais (2016) ou plus anciens (Salençon, 2008), une démarche de prospective holistique est devenue nécessaire pour explorer une variété, multicolore, de modèles éditoriaux, d'environnement de dissémination des savoirs etde financement de toute une palette possible de services associés. Peut-on, doit-on même encore parler d'*open access* ? Rien n'est moins évident et ce d'autant moins que nous devons maintenant creuser les fondements technique et documentaire du paysage éditorial actuel ou futur.

4. Rôle d'une infrastructure technique et éditoriale

4.1. Partir de ce dont nous disposons

Une politique de diffusion de l'information scientifique, aussi cohérente puisse-t-elle être, doit aussi être en phase avec les moyens humains et techniques dont l'institution dispose. De fait, Inria a su faire fructifier, au sein d'un réseau (encore) très actif de professionnels de la documentation, un réservoir de compétences permettant d'accompagner ses chercheurs dans leurs différentes activités d'accès, de création et de diffusion de leurs publications et, dans une moindre mesure, des autres productions de recherche. L'existence d'un groupe motivé et actif a été au fil des années un élément essentiel pour le positionnement fort d'Inria dans le domaine de l'IST. Cependant, la politique Inria s'est aussi appuyée sur une composante longtemps restée externe à sa propre gouvernance, l'archive de publication HAL, créée et maintenue au CCSD[36]. Avant d'en devenir partenaire début 2015, nous nous sommes engagés dans une démarche qui a consisté d'une part à considérer HAL comme notre plateforme de référence pour tout ce qui touche à la diffusion institutionnelle de nos articles, et d'autre part à contribuer directement à différentes actions de dissémination, de développements techniques ou d'accompagnement éditorial, dans une démarche résolument mutualisée. Bien que non dénué de risque, ce choix s'est avéré optimal à bien des points de vue, notamment en termes de budget et de pérennité.

Il est donc bon à ce stade d'effectuer un bilan de notre perception de la plateforme, notamment pour repérer les différents aspects qui nous semblent essentiels en tant qu'infrastructure d'archivage et de diffusion des publications scientifiques. Dans le cadre de ce bilan nous allons à la fois évaluer ce qui pour nous, a postériori, justifie pleinement notre choix de ne pas avoir mis en œuvre notre propre plateforme institutionnelle, mais aussi d'ouvrir des pistes possibles d'évolution susceptibles de contribuer encore plus à nos objectifs à moyen et long terme.

4.2. Éléments techniques

Un certain nombre d'éléments techniques sont indispensables pour garantir l'accès et l'enregistrement pérenne du contenu d'un entrepôt de données en général et d'un entrepôt numérique de publications en particulier : identification précise et stable des contenus (avec liaison directe sur les objets numériques eux-mêmes), archivage à long terme et stabilité continue de l'offre de service. Sans élaborer ici ces différents aspects, nous pouvons signaler qu'ils reposent beaucoup sur la notion de transparence et de confiance. Il ne suffit pas en effet de savoir qu'un objet est archivé, mais dans quelles conditions et avec quel niveau de sécurité.

[36] Centre pour la communication scientifique directe, unité propre de service du CNRS jusqu'en 2015.

L'existence d'institutions dédiées à ces tâches comme le CINES permet en grande partie de répondre à ces attentes. Disposer d'une telle infrastructure permet de complètement s'affranchir des services équivalents qui peuvent être proposés par l'édition privée.
Un deuxième élément essentiel est de concevoir l'archivage d'un document dans un cadre plus large que le simple dépôt d'un texte de publication. Il faut savoir en effet l'associer à tout un ensemble d'objets complémentaires qui peuvent venir l'illustrer ou l'enrichir, qu'il s'agisse de graphiques, de jeux de données oude la vidéo d'une présentation faite lors d'une conférence. Il faut aussi relier l'entrée d'une archive ouverte à une écologie plus large par le biais de références à d'autres sites de publication, qu'il s'agisse d'autres archives (arXiv ou PMC) ou des sites de gestion de journaux ou de conférences. Il faut enfin relier l'entrée à des jeux de données externes quand il n'est pas possible de les grouper directement avec une publication.
Enfin, l'accès aux informations de l'entrepôt par des tiers, indépendamment de la consultation sur le portail HAL lui-même, peut être envisagé sous de nombreuses formes : indexation, moissonnage, interrogation en direct. HAL offre dans ce cadre une palette complète de possibilités qui contribuent à la grande visibilité de la plateforme et de ses contenus. Alors que l'accent a été mis depuis des années par les communautés des archives ouvertes sur la disponibilité d'une interface de moissonnage OAI/PMH, couplée à un profil de métadonnées de type Dublin Core, l'expérience prouve que la pauvreté des informations ainsi transmises ne donne finalement pas une réelle valeur ajoutée à un portail tel que celui d'OpenAire, construit sur cette base. Données de publication précises et structurées, affiliations, intégration dans des collections : tout cela est perdu quand on compare une entrée dans HAL et son équivalent moissonné par OpenAire, mais aussi par les serveurs privés sur lesquels nous reviendrons.
De fait, il faut effectivement distinguer une visibilité de surface, qui permet en particulier aux articles d'être bien référencés dans des moteurs génériques tels que Google, et la fourniture d'une relation fine aux données qui permet, par le biais d'interfaces de fourniture (SWORD) ou de téléchargement (API d'interrogation), d'offrir de véritables formats « intégraux », c'est-à-dire couvrant l'ensemble des métadonnées disponibles dans la base. Cette démarche est essentielle, comme nous le verrons un peu plus loin, pour intégrer HAL dans un écosystème de services plus riches.

4.3. Services

Le milieu de la recherche n'est jamais trop prompt à utiliser un outil complexe si celui-ci ne démontre pas qu'il minimise le différentiel entre effort d'appropriation et avantages effectifs. C'est dans cette perspective que de nombreux efforts ont porté ces dernières années, notamment avec la version 3 de HAL, pour offrir des services concrets associés au dépôt et à l'utilisation des contenus. De mon point de vue, les éléments suivants sont la clef actuelle de la qualité de l'environnement sous HAL :

- La gestionde son identité en lien avec les grands référentiels nationaux et internationaux (ORCID, Idref, VIAF, etc.), ainsi que la possibilité de se créer une page personnelle où tout ou partie des publications contenues dans HAL peuvent être présentées ;
- L'organisation du corpus des publications en collections pouvant être générées manuellement ou automatiquement. Cette fonctionnalité permet une gestion souple de corpus aussi variés que des volumes de journaux ou de conférences, des publications associées à un laboratoire ou un projet, ou encore l'édification de collections thématiques ;
- Une large palette d'outils visant à faciliter le dépôt d'un article avec bien sûr Grobid (Lopez et Romary, 2015) pour l'extraction de métadonnées, de nombreuses

possibilités de complétion automatique ainsi que la possibilité de reprendre des dépôts existants comme modèles.

Il reste encore malgré tout de nombreux progrès à faire dans le domaine des services. Nous reviendrons sur certains qui ne peuvent se concevoir sans une perspective plus large de l'usage possible d'une plateforme telle que HAL. Qu'il s'agisse du moteur de recherche, des outils statistiques ou d'outils plus fins de mise en réseau des articles (antériorité), il ne sert à rien d'avancer pour l'instant dans cette direction sans avoir une idée claire du public que l'on veut servir et des usages que l'on en attend. On entre dans le domaine du prospectif…

4.4. Affiliations – référentiels

Au-delà des services immédiats qu'offre HAL à l'utilisateur, il est un service relativement invisible, ou au contraire vu comme une contrainte par les déposants, à savoir la nécessité d'associer chaque auteur d'un article avec une affiliation précise.

Bien que pouvant être vu comme fastidieuse, cette caractéristique est un élément essentiel de l'organisation d'une archive de publication dans une perspective plus institutionnelle. Elle permet pour une équipe, un laboratoire ou une institution donnée de repérer automatiquement les publications qui lui correspondent, ce qui permet de simplifier la production de bilans ou la génération de pages web, mais surtout de pouvoir disposer d'une vue d'ensemble des activités scientifiques d'une structure de recherche, comme nousl'avons exploré dans l'action de développement d'Inria anHALytics[37].

L'affiliation d'un auteur est une image instantanée des circonstances dans lesquelles celui-ci a écrit son article, circonstances qui peuvent naturellement changer d'une production à une autre. Le chercheur peut changer de laboratoire, il peut vouloir signaler qu'un travail a été fait lors d'un séjour dans une autre structure, et donc d'indiquer une affiliation multiple ; il peut aussi vouloir signaler qu'un article relève d'une production plus privée, sur un sujet qui ne concerne pas son affiliation principale.

Le modèle de données d'une archive de publications souhaitant offrir des mécanismes d'affiliation riche à ses auteurs doit donc être en mesure de représenter chaque auteur sous la forme d'une combinaison individu + affiliation(s). Les affiliations elles-mêmes doivent refléter l'organisation en plusieurs niveaux de dépendance qui subordonne une équipe à un ou plusieurs laboratoires et un laboratoire à une ou plusieurs institutions de rattachement, avec d'éventuelles possibilités de combinaisons plus ou moins complexes. Il s'agit de fait d'une structure en graphe acyclique orienté qui permet, dès lors qu'une publication est indiquée comme étant rattachés à un niveau donné, de connaître toutes les structures de niveau supérieur pour lesquelles la publication est pertinente.

La richesse d'un tel mécanisme provient essentiellement des possibilités de maintenir un niveau élevé de pertinence et de qualité au sein d'un référentiel sous-jacent qui va garantir à terme une identification unique des structures de recherche concernées. Bien qu'il existe un répertoire des structures nationales de recherche (RSNR), le référentiel des structures intégré à AUReHAL apporte une réelle plus-value liée à son mode de gestion à deux niveaux :

- Alimentation en continu, à l'entrée de chaque publication, des structures qui n'existent pas encore dans une forme qui s'apparente au *crowdsourcing* ;
- Validation et contrôle par un large réseau documentaire, complémentaire du réseau (encore) plus limité des modérateurs, et garantissant là aussi une relation étroite avec la réalité du terrain.
- La mise en place effective d'AURéHAL n'a eu lieu qu'en octobre 2014 et la bascule des anciennes structures de HAL V2 dans ce référentiel a créé un fonds de référence

[37] https://github.com/anHALytics, voir aussi une démonstration sous : http://traces1.saclay.inria.fr/anHALyticsv1/search/

initial extrêmement bruité (informations parcellaires, doublons) ; un important travail a donc été effectué au fil des mois dans certains établissements, dont Inria, pour nettoyer le référentiel, y intégrer le lien avec les identifiants du RNSR, et ainsi augmenter sa fiabilité globale.

Au final, l'importance d'AURéHAL dans la gestion des affiliations des publications de nos communautés de recherche ne doit pas être négligée : plus fin que la plupart des référentiels administratifs, plus large aussi, puisque couvrant tous les laboratoires et les établissements avec lesquels les chercheurs collaborent dans des publications communes, il permet d'offrir un paysage extrêmement précis des modes de production de recherche en France, comme aucune autre communauté internationale n'est véritablement capable de le faire. Bien que s'accompagnant d'un travail non négligeable des collègues en charge des contenus, la richesse des affiliations dans HAL doit conduire l'ensemble des acteurs de notre système de recherche à se questionner sur la pertinence de devoir consacrer encore du temps à définir dessystèmes adéquats d'affiliation papier, dans l'espoir incertain d'influencer un repérage par un tiers privé (Thomson Reuters) de la visibilité (ou « impact ») de nos publications, que nous-mêmes devons à nouveau filtrer manuellement au niveau de l'OST et de nos établissements. L'inutilité de ce travail de Sisyphe m'a toujours étonné et attristé pour ceux qui s'y consacrent et la présence d'un environnement aussi riche que celui que nous avons mis en place avec les affiliations dans HAL devrait nous encourager à changer complètement nos priorités d'investissement à cet égard.

4.5. Modération

Tout dépôt dans HAL est soumis à une *modération*, c'est-à-dire à un contrôle préalable visant à garantir que le document relève bien de la politique éditoriale de la plateforme, à savoir de contenir des documents de nature scientifique. Une telle modération est bien sûr indispensable à toute plateforme de contenu en ligne, mais il est intéressant de constater que le modèle choisi ici, une modération dite *scientifique* en amont de la publication, reproduit de façon simplifiée le schéma de la relecture par les pairs. Dans les premières années de HAL, la modération fut même conduite par l'un des initiateurs de la plate-forme, Frank Laloë, très attaché qu'il était à la qualité scientifique de ce qui devait s'y trouver. La modération est maintenant principalement effectuée par des personnels documentaires qui se retrouvent parfois dans la situation difficile d'appliquer des règles de nature scientifique à des contenus qu'ils n'ont pas vocation à maîtriser.

Modérer des articles dans HAL couvre de fait un large spectre de questions pour lesquelles il n'est pas toujours facile d'apporter des réponses :

- Comment identifier un document scientifique, existe-t-il des critères objectifs de forme (linéarité du texte écrit) ou de taille (une démonstration mathématique peut-elle être vue comme un article) ?
- Quels sont les droits associés à l'objet déposé ? Sait-on bien identifier le ou les auteur(s) ? Le texte n'est il pas soumis à un régime légal contraignant sa diffusion ?
- Peut-on juger de la qualité scientifique d'un document, sans interférer avec le mécanisme plus large du dialogue scientifique et de l'évaluation par les pairs ? C'est probablement là que se trouve le point de tension le plus fort avec une modération proprement documentaire.

La démarche d'Inria pose, il me semble, des jalons pour une vision à moyen terme de cette activité à la fois essentielle et sensible. L'obligation de dépôt, comme on l'a déjà signalé, repose sur une responsabilisation du chercheur qui va lui-même déterminer ce qu'il met en ligne. Dès lors, la modération doit intrinsèquement faire confiance au chercheur et éventuellement l'accompagner pour améliorer la qualité de ses dépôts. [développer ?]

La modération, vue comme un acte centralisé de décision d'acceptation ou non d'un document dans HAL, possède par ailleurs ses limites. Dans de nombreuses situations, il faut envisager de dé-modérer le processus, à partir du moment où l'on sait que le document, ou le flux de document, a déjà fait l'objet d'une validation scientifique, ou entre, au moment même du dépôt, dans un tel processus de validation. C'est le cas notamment de toutes les démarches d'intégration de corpus d'articles comme nous l'avons expérimenté dans le cadre du projet PEER ou de la bibliothèque numérique de l'IFIP. C'est aussi le problème que l'on rencontre quand des dépôts sont explicitement destinés à être soumis à un épi-journal. Sans avoir toutes les réponses ici, nous devons trouver des moyens de fluidifier ces processus tout en maintenant un accompagnement dans le maintien de la qualité des contenus sous HAL. Affranchissement de modération, modération conditionnelle ou modération a posteriori : il nous faudra collectivement être flexibles pour rendre le processus le plus transparent possible pour les chercheurs.

4.6. Les conditions d'un espace pérenne

Les cas d'Academia et de ResearchGate, qui attirent souvent de nombreux collègues soucieux de plus de visibilité,peuvent servir d'exemples sur au moins un point : par la centralisation de la plateforme, la dévolution aux utilisateurs de l'alimentation massive de contenus et l'accumulation d'outils de fouille de données, ces plateformes agrègent inexorablement informations et abonnés. Ce dont ne disposent pas ces environnements, et c'est pour cela qu'on les voit essayer de nous démarcher, c'est d'un réseau d'institutions qui soutiennent ce développement, à la fois en encourageant leur utilisation dans le cadre de leur politique de diffusion des connaissances, mais aussi, et c'est probablement très négligé dans les analyses, en apportant un soutien en personnels d'accompagnement qui vont garantir la qualité des contenus.

Dans le cas d'une plateforme telle que HAL, nous, institutions d'enseignement et de recherche au sens large, devons tirer les leçons de ce qui se passe et être plus lucides dans l'affirmation d'une démarche cohérente au niveau national qui sache combiner une *plateforme centralisée* avec une *gestion distribuée des services*. Bien qu'allant à contre-courant d'un instinct quasi primitif de mise en avant de son ego institutionnel dans un contexte de forte compétition sur des moyens limités (mais la raison peut savoir équilibrer cela), cette démarche doit avoir plusieurs conséquences pour le développement futur de HAL, de ses contenus et de ses services :

- Nous devons créer les conditions d'une mutualisation du développement de la plateforme en généralisant une approche centrée sur des services ouverts et en créant des espaces (techniques et humains) où il est naturel pour les développeurs d'expérimenter, d'échanger et d'intégrer de nouvelles fonctionnalités. Le rôle du CCSD peut ainsi évoluer de façon sensible de celui de développeur primaire à celui d'intégrateur ;
- La gestion éditoriale des contenus, y compris les référentiels, doit pouvoir être largement décentralisée en direction des établissements, mais aussi des communautés scientifiques et des réseaux documentaires. Chacun doit se sentir impliqué dans l'alimentation d'un fonds qui peut devenir une richesse unique pour notre espace de recherche future et là encore, le rôle du CCSD doit pouvoir évoluer vers celui de coordinateur de cette variété potentielle de contributeurs ;
- La gouvernance globale de HAL doit davantage s'ouvrir pour refléter cette évolution et faire en sorte que les établissements se sentent plus à l'intérieur comme participants qu'à l'extérieur comme clients. De nombreuses modalités sont possibles, par exemple en renforçant le rôle de BSN, ou en passant par un conseil scientifique solide proche de celui de l'ABES actuel ;

- Enfin, nous devons, collectivement, savoir adosser à la richesse des fonds de HAL un espace de recherche qui en exploite tous les aspects. La démarche que nous proposons doit ainsi contribuer à redynamiser notre capacité en fouille de données scientifiques.

Serge Bauin (2014) a mené une analyse très fine de la situation actuelle de HAL dans le paysage académique français global et dresse une feuille de route d'un certain nombre de priorités à court terme qui nous semble aller dans le sens du schéma cible que j'essaye d'esquisser ci-dessus.

En particulier, il semble que nous soyons en phase sur le fait que la plateforme HAL doit évoluer petit à petit pour devenir un véritable espace de travail documentaire pour le chercheur. Cet espace doit permettre au chercheur de gérer tout type de document nécessaire à son travail scientifique et en déterminer les modes de diffusion, allant d'une sphère strictement privée à différents modes de diffusion ciblée (autres chercheurs, groupes ou projets, ouverture intégrale). Son espace doit lui permettre de gérer non seulement ses propres documents, mais aussi lui permettre d'organiser sa propre bibliothèque de recherche en y intégrant documents et références externes.

Le tableau de bord du chercheur doit pouvoir intégrer toute une palette d'outils dont il a la maîtrise et qui doivent l'aider à qualifier et présenter son activité. Nous devons donc aller de plus en plus vers un espace fluide associant :

- Des outils de fouille de données qui permettront au chercheur de structurer et de parcourir sa bibliothèque numérique personnelle. Ces outils pourront relier un document donné à un graphe de citations antérieures ou postérieures,en lien avec des fonds externes tels que celui qui se met en place au sein du projet Istex ;
- Une version plus élaborée de ces outils qui pourra intégrer des recherches d'antériorité permettant d'associer un document avec d'autres documents similaires pour suggérer des associations (identification de collègues ayant des intérêts similaires, proposition de références)[38] ;
- Cet espace de travail sera le point d'entrée naturel avec des outils de certification dans la lignée de ce que nous avons commencé à expérimenter avec Episciences (et qui tend à former un continuum avec des plateformes de gestion de congrès telles que SciencesConf) ;
- Un des éléments essentiels sera pour le chercheur de pouvoir gérer son mode de communication externe à partir de son espace personnel. En plus d'avoir toute latitude sur le degré de dissémination de ses contenus, le chercheur doit disposer de tous les moyens techniques nécessaires pour mettre en avant les indicateurs qu'il estime pertinents ;
- Enfin, l'espace documentaire du chercheur doit intégrer de façon transparente un accès réciproque avec d'éventuelles archives de données qui commencent à émerger (Nakala, Ortolang), et qui elles-mêmes doivent tendre à partager de nombreuses fonctionnalités avec HAL (identification et référentiels en particulier).

Un tel espace ne peut fonctionner à terme que si le chercheur se sent proche de sa communauté, plutôt que de se retrouver isolé dans un environnement générique abstrait. Le travail initié au niveau des portails de HAL et la plus grande décentralisation éditoriale que j'ai appelé de mes vœux doit se traduire par une liaison étroite entre l'espace du chercheur et un support de proximité des réseaux de recherche, institutions et réseaux documentaires

[38] Il s'agit ici de techniques similaires à celles développées dans (CLEF-IP) et qui à l'occasion sont déployés pour la reconnaissance de plagiats. Ce dernier thème n'est de fait qu'un épiphénomène au regard de la palette des usages qu'il est possible d'envisager et ne doit en aucun cas focaliser toute notre attention.

concernés. Cette démarche doit permettre d'accroître la culture des chercheurs dans le domaine de l'information scientifique dans toutes ses dimensions : communicative, patrimoniale et éditoriale.
Un des moyens de progresser dans cette direction est que HAL intègre dans sa feuille de route le soutien au développement de bibliothèques numériques institutionnelles. Typiquement, c'est pour Inria un enjeu stratégique d'identifier et d'ouvrir le patrimoine de ses productions depuis sa création. Nous expérimentons actuellement une telle démarche dans le cadre de notre collaboration avec l'IFIP avec laquelle nous intégrons l'ensemble de leurs publications sous la forme d'un ensemble de collections dans HAL, mais aussi dans le cadre de notre réflexion sur la reprise possible de notre antériorité documentaire, notamment pour nos collègues chercheurs qui, dans la perspective d'un prochain départ à la retraite, cherchent à pérenniser sous forme numérique l'ensemble de leurs publications.

5. Aller plus loin encore ?

La définition d'une politique d'information scientifique d'un établissement est, comme on l'a vu au fil des pages de cet article, un exercice constant d'équilibre entre de grands principes, de petites avancées et tout un ensemble de contraintes scientifiques, techniques ou budgétaires. J'ai esquissé à plusieurs endroits des pistes d'actions ou d'ambitions pour contribuer un peu plus à rendre le système de diffusion des productions scientifiques plus ouvert et fluide, dans un cadre que je fixe clairement comme étant celui d'une infrastructure numérique dont l'indépendance et la pérennité ne peuvent qu'être assurées au sein de l'espace public. Je ne vais donc pas en conclusion chercher à effectuer une synthèse de ce que pourraient être des priorités d'action, mais rappeler un élément essentiel d'une action efficace dans le domaine de l'information scientifique : le personnel documentaire. La plupart des actions décrites dans cet article n'auraient pu avoir lieu si, notamment chez Inria, il n'y avait pas un réseau de personnels compétents à même de mettre en œuvre et d'accompagner tout ce qui nécessite une perception fine des contenus, de leur description et de leur représentation. S'il n'y a qu'un message aux responsables des établissements d'enseignement supérieur et de recherche qui seraient tentés de suivre certaines des directions de cet article, c'est en priorité de stabiliser et renforcer les équipes qui permettront à tout cela de fonctionner dans la durée.

6. Références

Altman, Lawrence K (1996). "The Ingelfinger rule, embargoes, and journal peer review–Part 1". *The Lancet*. 347(9012), pp. 1382-1386.doi:10.1016/S0140-6736(96)91016-8.

ArmbrusterChris, Laurent Romary (2010). Comparing Repository Types - Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. *International Journal of Digital Library Systems*, IGI Global, 2010, 1 (4), pp.61-73. <inria-00465197v2>

Bauin Serge (2014),L'open access à moyen terme : une feuille de route pour HAL, rapport CNRS. http://corist-shs.cnrs.fr/sites/default/files/billets/cnrs_dist_rapport_bauin_sur_ccsd_et_hal_septembre_2014.pdf

BerthaudChristine, Laurent Capelli, Jens Gustedt, Claude Kirchner, Kevin Loiseau, et al.. EPISCIENCES - an overlay publication platform. Dr. Panayiota Polydoratou. *ELPUB2014 - International Conference on Electronic Publishing*, Jun 2014, Thessalonique, Greece. IOS Press, pp.78-87, 2014, <http://www.ebooks.iospress.nl/publication/36552>. <10.3233/978-1-61499-409-1-78>. <hal-01002815v2>

Couture, Marc (2015), Les licences Creative Commons et l'accès libre , http://r-libre.teluq.ca/389/1/LicencesCC-AL.pdf
European Commission (2012). *European Commission background note on open access to publications and data in Horizon 2020* : http://ec.europa.eu/research/science-society/document_library/pdf_06/background-paper-open-access-october-2012_en.pdf
Fargue, Marie (2012), À propos des publications en accès libre, http://wavelets.ens.fr/BOYCOTT_ELSEVIER/MARIE_FARGE/2012_NOTE_POUR_LA_MINISTRE_DE_LA_RECHERCHE/2012_06_29_Publications_en_acces_libre.pdf
Gentil-BeccotAnne, Salvatore Mele, Travis Brooks (2009).Citing and reading behaviours in high-energy physics, *Scientometrics*, 84(2), DOI: 10.1007/s11192-009-0111-1
Guédon Jean-Claude (2008)Mixing and Matching the Green and Gold Roads to Open Access—Take 2, *Serials Review*, Volume 34, Issue 1, March 2008, Pages 41–51.
Harnad Stevan (2005), Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold, *Ariadne*, 42.
Lopez Patrice et Laurent Romary (2015). GROBID - Information Extraction from Scientific Publications. *ERCIM News*, 100. http://ercim-news.ercim.eu/en100/r-i/grobid-information-extraction-from-scientific-publications
Pöschl,Ulrich (2010), Interactive open access publishing and public peer review: The effectiveness of transparency and self-regulation in scientific quality assurance, IFLA Journal, vol. 36 no. 1, pp. 40-46. doi: 10.1177/0340035209359573
Pulverer, Bernd (2010),Transparency showcases strength of peer review, Nature, vol. 468.
Reda, Julia (2015). EU copyright evaluation report. https://juliareda.eu/copyright-evaluation-report/
Romary, Laurent (2005) Préface. In: Aubry, C., and Janik, J., eds. Les archives ouvertes . Enjeux et pratiques. Guide à l'usage des professionnels de l'information. Paris: ADBS, 2005. pp. 5-8.
RomaryLaurent (2007). OA@MPS - a colourful view. *Zeitschrift für Bibliothekswesen und Bibliographie*, Vittorio Klostermann, 2007, 7 p. <hal-00164041>
RomaryLaurent (2009). Communication scientifique : Pour le meilleur et pour le PEER. *Hermes*, CNRS-Editions, 2009, pp.131-136. <inria-00424254>
RomaryLaurent, Chris Armbruster (2010). Beyond institutional repositories. *International Journal of Digital Library Systems*, IGI Global, 2010, 1 (1), pp.44-61. <hal-00399881>
RomaryLaurent (2010). Archives scientifiques en ligne : entretien avec Laurent Romary, propos recueillis par Dominique Chouchan. *Les Cahiers de l'INRIA - La Recherche*, INRIA, 2010, La nouvelle physiologie du goût. <inria-00511279>
Romary Laurent (2012a). Data management in the humanities. *ERCIM News*, ERCIM, 2012. <hal-00680193>
RomaryLaurent (2012b). Scholarly Communication. Mehler, Alexander and Romary, Laurent. *Handbook of Technical Communication*, de Gruyter, 2012, 978-3-11-022494-8. <inria-00593677>
RomaryLaurent, Mike Mertens et Anne Baillot (2016). Data fluidity in DARIAH – pushing the agenda forward. *BIBLIOTHEK Forschung und Praxis*, De Gruyter, 2016, 39 (3), pp.350-357. <hal-01285917v2>
RomaryLaurent et Sally Chambers (2014). DARIAH: Advancing a digital revolution in the arts and humanities across Europe. *e-data&research*, Data Archiving and Networked Services (DANS), 2014. <hal-00913691>

Romary Laurent (2015). TEI challenges in an accelerating digital world. *DiXiT Convention week*, Sep 2015, The Hague, Netherlands. 2015, <http://dixit.huygens.knaw.nl>. <hal-01254365>

Schöpfel Joachim (2015). Open access - the rise and fall of a community-driven model of scientific communication. *Learned Publishing*, Association of Learned and Professional Society Publishers, 2015, 28 (4), pp.321-325. <10.1087/20150413>. <hal-01282744>

Salençon Jean (2008), Rapport du comite Information scientifique et technique, Rapporteur : Alexandre Moatti, Remis le 19 mai 2008 au Directeur général de la recherche et de l'innovation et au Directeur général de l'enseignement supérieur

ShieberStuart M. (2013), Ecumenical open access and the Finch Report principles July 10th, 2013, *The Occasional Pamphlet on scholarly communication*, http://blogs.harvard.edu/pamphlet/2013/07/10/ecumenical-open-access-and-the-finch-report-principles/

Velterop Jan (2007). Open Access is a Choice. ZfBB 54 (4-5), pp.268-272.

7. Remerciements

Je tiens à remercier ici chaudement toute l'équipe IES d'Inria avec laquelle j'ai pu interagir depuis des années sur les sujets évoqués dans cet article. Au niveau national, les échanges nombreux avec différents collègues de l'enseignement supérieur et de la recherche, que je pourrais facilement agréger sous l'étendard de BSN ont fortement contribué à élaborer mes réflexions et mes positions. Enfin un grand merci aux relecteurs de première ligne de cet article, dont Serge Bauin, Naomi Truan, Jacques Millet, Hélène Lowinger et Pascal Laurent.

1. INTRODUCTION 1

2. VEILLEE D'ARMES 2

2.1. IL EST L'OR, MONSEIGNOR 2
2.2. UN REVE DEÇU ? 3

3. DEFENDRE UNE VISION 6

3.1. OBLIGATION DE DEPOT ? 6
3.2. GESTION DES FRAIS DE PUBLICATION 7
3.3. DIFFUSION INSTITUTIONNELLE DES PRODUCTIONS SCIENTIFIQUES (WAS : GREEN) 10
3.4. DU VERT CLAIR AU VERT FONCE 13
3.5. AU LOUP ! 14
3.6. NOUVEAUX MODELES EDITORIAUX 15
3.7. METTRE TOUT CELA EN MUSIQUE – ARRETONS D'AVALER DES COULEURS 16

4. ROLE D'UNE INFRASTRUCTURE TECHNIQUE ET EDITORIALE 18

4.1. PARTIR DE CE DONT NOUS DISPOSONS 18
4.2. ÉLEMENTS TECHNIQUES 18
4.3. SERVICES 19
4.4. AFFILIATIONS – REFERENTIELS 20
4.5. MODERATION 21
4.6. LES CONDITIONS D'UN ESPACE PERENNE 22

5. ALLER PLUS LOIN ENCORE ? 24

6. REFERENCES 24

7. REMERCIEMENTS 26