



HAL
open science

Etude critique des nouveaux modes “ d’éditorialisation ” de revues scientifiques en accès-ouvert

Pierre-Carl Langlais

► **To cite this version:**

Pierre-Carl Langlais. Etude critique des nouveaux modes “ d’éditorialisation ” de revues scientifiques en accès-ouvert. [Rapport de recherche] Bibliothèque Scientifique Numérique. 2016. hal-01388556

HAL Id: hal-01388556

<https://hal.science/hal-01388556>

Submitted on 27 Oct 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 - ShareAlike 4.0 International License



BIBLIOTHÈQUE SCIENTIFIQUE
NUMÉRIQUE
DIGITAL SCIENTIFIC
LIBRARY

**Etude critique des nouveaux modes « d'éditorialisation »
de revues scientifiques en accès-ouvert**

Rapport synthétisé de l'étude pilotée par BSN 4 et BSN 7

et réalisée par Pierre-Carl Langlais (juin 2016)

comité de pilotage Serge Bauin, Emmanuelle Corne, Jacques Lafait, Pierre Mounier.

Introduction

Lors de la publication de ce rapport, la France vient d'adopter une loi sur le libre accès. L'art. 30 de la « Loi pour une République Numérique » établit que les auteurs d'écrits scientifiques financés au moins par moitié par fonds publics peuvent les « mettre à disposition gratuitement dans un format ouvert, par voie numérique » après une durée exclusive d'exploitation maximale (dite « embargo ») de six mois dans les sciences techniques et médicales et de douze mois dans les sciences humaines et sociales.

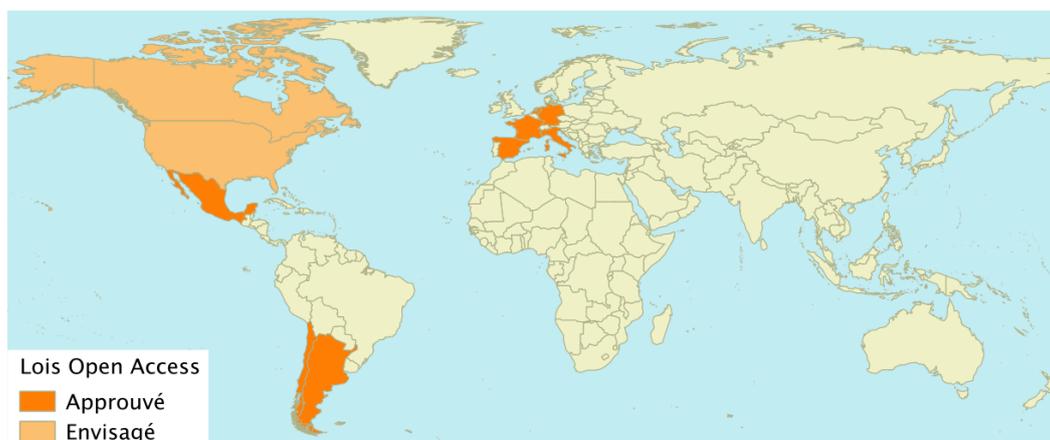
En septembre 2016, ce dispositif a été définitivement approuvé par l'Assemblée et le Sénat. Il vise à lever l'une des principales restrictions limitant la diffusion des publications scientifiques sur les réseaux numériques : la nécessité d'obtenir l'accord de l'éditeur. Les clauses d'exclusivité usuellement prévues dans les contrats d'édition empêchent potentiellement les auteurs de rediffuser leurs contributions scientifiques (par exemple sur un dépôt institutionnel) ou appliquent des restrictions variables¹. Ces arrangements complexes ne facilitent pas la mise en œuvre d'une politique de libre accès cohérente, ni pour les auteurs (qui, d'un éditeur à l'autre, doivent adopter une stratégie de republication radicalement distincte), ni pour les lecteurs. L'inscription du libre accès dans la loi crée un cadre minimal harmonisé apte à simplifier les conditions d'accès à la recherche financée par fonds publics.

La loi française s'inscrit dans la continuité d'un mouvement européen, voire mondial. Le 17 juillet 2012, la Commission européenne formule une recommandation relative à *l'accès aux informations scientifique et à leur préservation*². Partant du principe que l'accès aux recherches financées par fonds publics contribue à « accélérer le progrès scientifique », ce texte préconise de les diffuser « dans les meilleurs délais, de préférence immédiatement et, dans tous les cas, au plus tard six mois après leur date de publication, et au plus tard douze mois pour les publications dans les domaines des sciences sociales

¹ Cf. les conditions répertoriées par Sherpa Romeo : <http://www.sherpa.ac.uk/romeo/journalbrowse.php>

² Recommandation 2012/417/UE : <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32012H0417>

et humaines³ » (art. 1). Quatre ans plus tard, le 17 mai 2016, le Conseil de l'Union Européenne recommande de généraliser ce modèle à l'ensemble des Etats-Membres afin de faire du libre accès une « option par défaut d'ici 2020 ».



Cartographie des lois approuvées et envisagées sur le libre accès

En passant du contrat à la loi, la pratique du libre accès entraîne une redéfinition majeure des modèles économiques existants. L'édition scientifique reste très largement structurée par un système de distribution par abonnement, essentiellement à destination des bibliothèques et des institutions scientifiques. Plusieurs rapports français et européens ont tenté d'évaluer la faisabilité d'une conversion au libre accès. L'enquête récente d'Odile Contat et Anne-Solweig Gremillet, « Publier : à quel prix ? » pour BSN met ainsi en évidence la forte dépendance des revues françaises en Sciences et humaines et sociales à l'égard des financements publics : la « très grande majorité des éditeurs associés aux structures porteuses des revues sont essentiellement financés à partir de fonds publics⁴. »

Le débat tend aujourd'hui à se déplacer d'une évaluation *de principe* du libre accès vers une discussion plus spécifique des modèles possibles. La question n'est plus : faut-il passer au libre accès ? Mais : quel libre accès voulons nous ?

Certaines initiatives majeures du mouvement *open access* se focalisent exclusivement sur les problématiques gestionnaires de la reconversion du modèle de diffusion des éditeurs scientifiques. La douzième conférence de la Déclaration de Berlin organisée en

³ « Conclusions du Conseil concernant la transition vers un système de science ouverte », 17 mai 2016, <http://data.consilium.europa.eu/doc/document/ST-8791-2016-INIT/fr/pdf>

⁴ Odile Contat et Anne-Solweig Gremillet, « Publier : à quel prix ? Étude sur la structuration des coûts de publication pour les revues françaises en SHS », Revue française des sciences de l'information et de la communication, (7), 2015. Adresse : <https://rfsic.revues.org/1716>

décembre 2015 par la Max Planck Gesellschaft débouche ainsi sur une proposition visant à « renverser (*flip*) le modèle économique des journaux diffusé par abonnement » en transférant les fonds alloués aux abonnements vers le paiement de « frais de traitement ».

La conférence avait pour objectif d'élaborer un consensus en faveur d'une coordination internationale incitant les bibliothèques à reconvertir les budgets accordés à la diffusion des revues par abonnement au profit de la couverture des coûts associés à la publication des journaux en libre accès (via les APCs)⁵.

L'attention portée à la reconversion des modèles économiques laisse de côté un point de blocage majeur : l'évaluation. Certaines structures ou communautés universitaires ne reconnaissent pas les publications en libre accès comme des contributions scientifiques. La section 60 du CNU prévoit ainsi que « Les articles dans des journaux uniquement en « open access » ne seront pas pris en compte⁶. » Paradoxalement, si la fiabilité des publications en libre accès est contestée au nom de protocoles d'évaluation normalisés, ces mêmes protocoles font l'objet d'une contestation croissante⁷. Le libre accès pourrait ainsi tirer sa légitimité d'un éventuel contre-modèle d'évaluation, qui ne reposerait plus sur la mise en œuvre de procédures closes dans le cadre d'un échange privé entre éditeurs, auteurs et évaluateurs : l'« évaluation ouverte » ou « *open peer review* ».

Rappel de la commande

La commande initiale du rapport s'inscrit dans le cadre d'un double regard critique sur les nouvelles formes de publications en libre accès et celles qui les ont précédé : quelles formes éditoriales l'État peut-il encourager à l'heure du numérique, de la mutation de

⁵ « Berlin 12 Conference Focuses on Proposal to Flip Subscription Journals to Open Access », arl.org, <http://www.arl.org/news/arl-news/3801-berlin-12-conference-focuses-on-proposal-to-flip-subscription-journals-to-open-access#.VwYsqnrXbvl>, consulté le 7 avril 2016.

⁶ <http://www.cpcnu.fr/web/section-60/conseils-generaux>

⁷ Ce discours de la crise revient fréquemment dans les éditoriaux des revues. Cf. par exemple A. Mulligan, « Is peer review in crisis? », *Oral Oncology*, 41(2), 2005, p. 135-141 qui revient sur le discrédit né de la divulgation de fraudes majeures (« These events have acted as a catalysts within the scholarly community with many questioning the role of peer review ») ou plus récemment Gottfried Schatz, « The faces of Big Science », *Nature Reviews Molecular Cell Biology*, 15(6), 2014, p. 423-426 qui évoque l'effet négatif des taux d'acceptations trop faibles (« Science needs competition, but competition has become so fierce that many fields of science now resemble war zones. Nothing illustrates this better than the current crisis of peer review. »)

l'édition scientifique et de la faillite de l'évaluation scientifique ?

Le passage d'un écosystème de publication à l'autre ne s'accompagne pas seulement d'une succession de « défis » à relever (le numérique, la mutation des pratiques éditoriales). Il rend visible des failles préexistantes, généralement masquées par l'inertie des habitudes : l'accroissement des corpus scientifiques accessibles en ligne autorise une estimation plus fine des déficiences des protocoles de *peer review* ; la refonte des conditions de mise en circulation des textes scientifiques favorise incidemment les expérimentations dans d'autres champs de l'activité éditoriale.

En ce moment critique où les normes établies se fragilisent et les normes à venir restent encore imprécises, les institutions et les communautés scientifiques se trouvent placées en position d'infléchir les recompositions en cours et de favoriser des innovations souhaitées de longue date. Le principe de reproductibilité ou d'ouverture des données de la recherche n'est pas consubstantiel d'une diffusion en libre accès mais son institutionnalisation est facilitée dans un contexte de changement généralisé. Les trois temps (ou « heures ») parallèles spécifiés par la commande initiale dessinent une configuration globale où le développement d'une politique publique d'incitation à certains modèles ou certaines pratiques devient nécessaire : accélération du processus d'informatisation (au-delà de formats, comme le PDF, reproduisant fidèlement le cadre éditorial de l'imprimé), redéfinition des modèles économiques et éditoriaux (faut-il encore parler de revue, ou même d'article ?) et relecture critique de l'efficacité des instances d'évaluation (ce qui concerne non seulement le *peer review* mais aussi toutes les formes d'évaluation de l'évaluation : métriques, listes de revues qualifiantes, etc.).

Comment pourrait se matérialiser cette politique d'incitation ? **Dans un écosystème aussi « interdépendant » que l'édition scientifique numérique, l'encouragement sélectif et exclusif de quelques innovations est difficilement envisageable.** Le terme « innovation » désigne usuellement l'intégration délimitée d'un nouveau produit ou d'une nouvelle pratique, dans un cadre entrepreneurial⁸. L'élaboration de formes d'évaluation

⁸ Voir par exemple la définition proposée par Benoît Godin, dans « "Innovation Studies": The Invention of a Specialty », *Minerva*, 50(4), 2012, p. 397-421.

ouvertes ou de métriques alternatives (ou, finalement, la conversion au libre accès dans son ensemble) ne rentre pas dans cette définition limitative : il ne s'agit pas d'un « produit » isolable mais de la redéfinition d'un aspect fondamental d'une activité.

Plus qu'une politique d'innovation, la généralisation du libre accès appellerait une « politique d'infrastructure », qui, au-delà du soutien d'usages ou d'outils spécifiques définirait des articulations convergentes entre dispositifs, acteurs et pratiques. Un rapport récent de *Knowledge Exchange* recommande ainsi d'« enraciner » le libre accès (*putting down roots*), en l'inscrivant dans un réseau d'acteurs et de politiques convergentes : « La mise en œuvre des politiques de libre accès implique de développer une infrastructure en libre accès à partir de l'actuelle collection disparate de services⁹ ». La publication sur le web se prête à cette démarche : l'écrit s'insère d'emblée dans un réseau intertextuel d'hyperliens, de formats récurrents (définis par les feuilles de style ou les transclusion) et, surtout, de flux d'informations normalisés. Les « Interfaces de programmation » (« Application Programming Interface ») automatisent la récupération des métadonnées et assurent ainsi des croisements entre des activités ou des dimensions distinctes du travail éditorial. Concrètement, des services distincts parviennent à « converser » en s'accordant au préalable sur des standards définissant et délimitant certaines informations.

Le plan du rapport reprend les principaux « champs » évoqués dans la question initiale. La première partie porte sur la métamorphose des systèmes d'édition (*Editorial Manager*, *Open Journal System*) en « plateforme » autonome et connectée à tout un ensemble de services extérieurs. La seconde partie est consacrée à la diversification des formes prises par la contribution scientifique, qui se manifeste par l'éclatement de la publication en plusieurs espaces d'écriture (dépôts de données, de programmes etc.) et l'émergence de dispositifs fondamentalement hybrides (le carnet de code). La troisième partie revient sur les failles actuelles du système classique de *peer review* (en particulier en terme de reproductibilité) et sur les différentes alternatives qui se dessinent sous le nom collectif d'évaluation ouverte (*open peer review*). La quatrième partie aborde les recompositions

⁹ « Putting down Roots », JISC, p. 5 <http://repository.jisc.ac.uk/6269/10/final-KE-Report-V5.1-20JAN2016.pdf>

des modèles économiques : économies d'échelle générées par l'informatisation des procédures éditoriales, diversification des ressources et développement de structures non-commerciales.

Toutes ces transformations sont étroitement liées les unes aux autres. Une cinquième partie conclusive évoque le passage nécessaire d'une politique d'innovation, centrée sur des transformations ciblées, à une politique d'infrastructure : la question n'est plus seulement d'étiqueter des formes d'innovations souhaitables mais d'identifier les cadres les plus propices à une intégration harmonisée des diverses aspirations qui se font jour dans les communautés scientifiques. L'évolution de l'édition scientifique pourrait ainsi passer par la constitution d'écosystèmes structurés, assurant la transmission des textes et des informations d'un acteur à l'autre.

L'étude des modes d'éditorialisation sollicite une approche composite. Les différents aspects d'une production textuelle constituent des champs d'observation commodes plutôt que des entités autonomes. L'étude de l'éditorialisation doit justement s'attacher à décrire leurs points de convergences, leur irréductible concomitance au sein de systèmes d'écriture en redéfinition permanente¹⁰. Appliquée à la publication scientifique, l'étude de l'éditorialisation nous incite à prendre au sérieux la question de l'incidence éditoriale du libre accès : l'altération des conditions de circulation aurait nécessairement une incidence sur l'ensemble du dispositif textuel.

Afin de mettre en œuvre cette approche composite nous avons articulé trois méthodologies distinctes :

- ***La constitution d'une base de données d'une sélection « emblématique » d'organisations éditoriales et de revues.*** Cette sélection comprend des éditeurs de revue (*Elsevier, PLOS, BMJ, EDP, PeerJ, Nature MacMillan, Biomed Central*), des sociétés savantes (*APS, ACS, European Geosciences Union*), des projets expérimentaux (*RIO, Self-Journal of Science, The Winnower, Polymath, F1000, Episciences, ReScience*) mais aussi quelques acteurs « intermédiaires » (les outils

¹⁰ Nous reprenons ici l'essai de définition proposée par Bruno Bachimont dans « Nouvelles tendances applicatives : de l'indexation à l'éditorialisation », in *L'indexation multimédia*, Paris, Hermès, 2007. Consulter à l'adresse http://cours.ebsi.umontreal.ca/sci6116/Ressources_files/BachimontFormatHerme%CC%80s.pdf

d'édition *Editorial Manager*, *Open Journal System* et *Scholar One*). Nous nous sommes attachés à maintenir une certaine diversité entre les différents modèles éditoriaux, au-delà de la distinction entre revue « traditionnelle » et « expérimentale » ou entre accès fermé et accès ouvert. Dans l'ensemble la base de données constitue surtout un outil d'aide à la réflexion. En nous astreignant à remplir systématiquement certaines catégories prédéterminées, nous nous sommes assurés d'observer la plupart des formes prises par le système d'éditorialisation de chaque organisation ou de chaque revue.

- ***Une série d'entretiens avec des personnes travaillant dans ces organisations.*** Nous avons privilégié l'identification d'éléments « invisibles » dans les politiques éditoriales diffusées en ligne, soit, en tout premier lieu, les stratégies et les positionnements des acteurs à plus long terme face aux innovations anticipées. Ce recueil de *verbatim* a permis de faire émerger les dynamiques et les tensions qui les entourent et ainsi d'extrapoler des évolutions et des repositionnements de court terme.
- ***La constitution et la réutilisation de jeux de données complémentaires.*** Notre sélection ne dit en effet rien de l'ampleur du changement, de l'adoption effective de ces normes en gestations au sein des 50 000 revues scientifiques référencées par *Crossref*. Nous avons eu recours à des méthodes de *text mining* pour préciser les contours de certains débats intellectuels (par exemple sur l'évaluation ouverte). Nous avons également repris et analysé plusieurs jeux de données portant, par exemple, sur les usages des outils et des formes de publication dans la recherche scientifique.

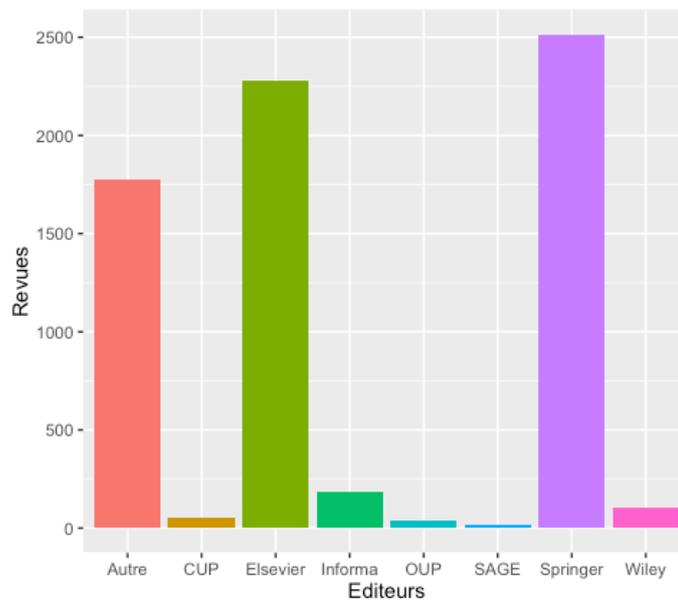
Première partie : Les outils d'édition

Jusqu'aux années 2000, les infrastructures logicielles des revues et des dépôts articles numérisés reposaient généralement sur des produits développés en interne. Au cours de ces quinze dernières années, des outils d'édition normalisés se sont graduellement substitués à ces systèmes « maison ». *Scholar One* est peut-être le plus ancien :

actuellement détenu par *Thomson-Reuters* il équipe plus de 3000 revues. *Editorial Manager* s'impose comme l'acteur dominant sur ce marché (avec près de 7000 publications). *Open Journal System* est un logiciel libre édité par un projet à visée non commerciale, *Public Knowledge Project* (le nombre d'utilisateurs, difficile à connaître, fluctue entre 9000 et 24000). Enfin, le cas des archives ouvertes est particulier : des outils open source comme *DSpace* prédominent (environ six fois plus d'usage par rapport aux solutions propriétaires comme *Digital Commons*).

A l'exception d'OJS, ces outils sont principalement utilisés par les revues détenues par de grandes organisations. Dans *Editorial Manager*, seule une revue sur quatre ne provient pas d'un acteur majeur du marché de l'édition. *Elsevier* et *Springer* prédominent tout particulièrement dans cette répartition, ces deux organisations cumulant plus de revues que tous les autres clients. La faible présence de « petits » éditeurs s'explique par la tarification élevée d'un produit hautement valorisé. Les transactions, conclues de gré à gré, ne sont pas publicisées mais *PLOS* verserait près de 7 millions d'euros chaque année à *Editorial Manager*, soit approximativement 15% de son chiffre d'affaire¹¹. Cet investissement est plus facilement rentabilisé pour des organisations de grande taille ou de taille moyenne, qui peuvent tirer parti d'économies d'échelle en convertissant plusieurs centaines voire de milliers de revues à un système d'édition normalisé.

¹¹ Chiffre avancé par Andrew Kern après recoupement de plusieurs sources comptables : <https://storify.com/jtth/andrew-kern-on-plos>



Répartition des revues électroniques utilisant Editorial Manager par éditeurs¹²

Les outils d'édition comme « plateforme »

Les outils d'édition commerciaux s'apparentent davantage à des systèmes d'éditorialisation plutôt qu'à des outils « exportables » : il ne viennent pas seulement équiper des structures préexistantes mais constituent, pour partie, une structure nouvelle déterminant les conditions d'accès à l'information et de circulation des textes. Nous reprenons ici la définition de la « plateforme » proposée par un récent rapport du *Conseil National du Numérique* :

Une plateforme est un service occupant une fonction d'intermédiaire dans l'accès aux informations, contenus, services ou biens édités ou fournis par des tiers. Au-delà de sa seule interface technique, elle organise et hiérarchise les contenus en vue de leur présentation et leur mise en relation aux utilisateurs finaux. À cette caractéristique commune s'ajoute parfois une dimension écosystémique caractérisée par des interrelations entre services convergents¹³.

Les trois principaux critères identifiés par le *Conseil du Numérique* s'appliquent à

¹² Cette répartition par revue ne rend pas compte du nombre d'articles publiés : une méga-revue comme *PLOS One* compte ainsi autant qu'une revue d'*Elsevier* ou *Springer*.

¹³ Conseil National du Numérique, Rapport Ambition numérique, juin 2015, p. 395. <https://contribuez.cnummerique.fr/sites/default/files/media/CNNum--rapport-ambition-numerique.pdf>

Editorial Manager et *Scholar One* :

- **Un service intermédiaire** : La « plateformes » se traduit concrètement par l'hébergement direct des publications et textes intermédiaires déposés par les auteurs, les évaluateurs et les *éditeurs*. Sur *Editorial Manager*, l'espace d'administration n'est pas localisé sur le site de la revue mais dans un sous-dossier rattaché à la racine du site. *Scholar One*, le principal concurrent d'*Editorial Manager*, ne semble pas avoir sauté le pas de l'hébergement direct. L'outil d'édition dispose cependant d'un substitut : le *Web of Science* publié par sa maison-mère, Thomson Reuters, auquel il a pleinement accès.
- **Une automatisation de la décision éditoriale** : L'activité des outils d'édition repose, de plus en plus, sur une forme d'ingénierie sociale qui automatise notamment les interactions préalables au processus de *peer review*. La stratégie commerciale actuelle d'*Editorial Manager* et *Scholar One* repose en partie sur le développement de fonctionnalités de « suggestion automatisée » des évaluateurs (*reviewer location*). Le vocabulaire et les fréquences des termes employés dans un article soumis sont comparés à un vaste corpus d'articles afin d'identifier des évaluateurs potentiels travaillant sur des sujets similaires. L'*éditeur* effectue un choix final à partir de la sélection proposée. La délégation de la décision éditoriale de la sélection de l'évaluateur à des algorithmes a une incidence sur la morphologie de la publication scientifique : comme sur *Facebook* ou sur *Twitter*, elle repose sur le traitement sélectif de données textuelles et de « graphes » sociaux (réseaux de citation, de collaboration).
- **L'inscription dans un écosystème** : les outils d'édition ne considèrent plus la revue comme une unité de référence mais comme un espace en interconnexion. Pour Josh Dahl la stratégie actuellement menée par *Scholar One* consiste à « s'intégrer dans un écosystème » ce qui passe notamment par une amélioration de la « connectivité des services » (par exemple via l'élaboration d'API normalisées) : « nous ne pouvons plus fonctionner en vase clos, indépendamment de ce que font

d'autres acteurs et des nouveaux standards ouverts¹⁴ ». Dans ces conditions, l'outil d'édition n'est plus seulement un système d'écriture (qui conditionne la morphologie du texte scientifique) mais, aussi, une infrastructure documentaire qui définit les modalités de la circulation des productions textuelles. En tant que partie prenante d'un écosystème, l'outil d'édition est directement impliqué dans une négociation collective des normes : *Editorial Manager* a ainsi redéfini sa classification des rôles éditoriaux en association étroite avec le projet *CREDIT* et l'initiative *ORCID*¹⁵. Si l'outil d'édition pense pouvoir s'émanciper de la revue, la réciproque est également valable. Les principaux éditeurs tendent aujourd'hui à « rapatrier » le processus éditorial sur des outils développés en interne. *PLOS* consacre actuellement une grande partie de ses ressources excédentaires à élaborer un système éditorial en « open source », appelé à se substituer à *Editorial Manager*, *APERTA*.

Cette conversion de l'outil à la plateforme découle des contraintes propres à l'élaboration des contributions scientifiques. L'évaluation prend notamment la forme d'un dispositif de correspondance complexe entre acteurs, ce qui favorise le recours à des processus de régulation algorithmiques ; l'inscription dans un écosystème découle de la « connectivité » intrinsèque des textes de recherche, conçus pour s'inscrire dans des réseaux de citation et d'indexation, potentiellement unifiés par des normes culturelles.

Les logiciels open source : des outils sans plateformes ?

Au cours des quinze dernières années les logiciels *open source* se sont imposés comme une alternative sérieuse aux solutions propriétaires. Avec probablement plus de 10 000 revues équipées, *Open Journal System* devance ses concurrents *Scholar One* et *Editorial Manager* en chiffres absolus. Dans les archives ouvertes, la domination des outils *open source* est encore plus prononcée : ils cumulent plus de 60% des usages (dont 43% pour le leader *DSPACE*) contre moins de 10% pour les outils propriétaires (données d'OpenDOAR).

¹⁴ Entretien avec Josh Dahl

¹⁵ Entretien avec Mark Hester

Les outils *open source* semblent en retrait de la dynamique de transformation de l'outil en plateforme. Dans *Open Journal System*, les mécanismes de récupération des données et des textes sont inexistantes : les concepteurs du logiciels ne connaissent pas précisément le nombre d'utilisateurs. Par compensation, faute d'être intégrée dans une infrastructure uniforme, l'ergonomie du logiciel reste problématique. En demeurant un outil « exportable », OJS doit être directement « customisé » par l'utilisateur au gré de ses besoins. Un effort d'adaptation technique demeure nécessaire. La documentation technique est parfois déficiente : la page d'aide décrivant les possibilités de customisation de l'outil est « très incomplète et parfois peu didactique¹⁶ ».

En l'absence de revenus directs, les outils *open source* dépendent d'un soutien institutionnel : OJS a été initié par le projet *Public Knowledge* et *DSpace* par le MIT. Or, l'amélioration du logiciel repose en effet principalement sur le travail de salariés : « L'essentiel du développement vient de l'équipe professionnelle des développeurs du *Public Knowledge Project*, même si, de par la nature *open source* de l'outil nous pouvons compter sur les apports de quelques intervenants extérieurs. » L'inscription de ce soutien institutionnel dans la durée est parfois problématique. La fiche du DOAJ sur les logiciels libres en répertorie 15 au total, mais la plupart semblent être devenus inactifs voire avoir totalement disparu : les liens donnés pour *topaz*, *peerlibrary*, *hjournal* ou *gapwork* nous ont renvoyés vers des serveurs inexistantes¹⁷.

Les outils *open source* tendent aujourd'hui à diversifier leurs partenariats notamment auprès d'institutions et de projets européens (comme OpenAIRE). En dépit de ces difficultés, les logiciels d'édition *open source* ont contribué à réduire les frais de conception d'un grand nombre de revues, en disséminant des procédés d'automatisation appliqués de longue date au sein de grandes organisations comme *Elsevier* ou *Springer*.

Recommandations

Le travail d'élaboration complexe des outils d'édition scientifique permet de dégager des

¹⁶ Jean-Luc Archimbaud. Document de synthèse sur OJS (Open Journal System), 2015. Déposé sur archivesic à l'adresse http://archivesic.ccsd.cnrs.fr/sic_01074813v2/

¹⁷ http://oad.simmons.edu/oadwiki/Tools_for_OA

économies d'échelle substantielles ; à ce titre, ils contribuent à élargir le champ des modèles éditoriaux et économiques pour les revues scientifiques. Cependant, Les communautés et les institutions scientifiques ne se saisissent pas suffisamment des opportunités (mais aussi des risques) soulevés par ces outils — ce qui se traduit notamment par un volume de contribution encore faible à des logiciels libres comme *OJS*.

Ce constat nous amène à formuler les deux recommandations suivantes :

1. ***Inciter les communautés et les institutions scientifiques à s'impliquer dans l'élaboration de ces outils.*** L'enjeu est essentiel : il en va de la capacité du monde universitaire à faire évoluer par lui-même ses pratiques d'écriture sans être conditionné par des influences externes.
2. ***Prendre en charge directement ou indirectement une partie du travail de développement et de maintien d'outils open source.*** Ces outils contribuent à la démocratisation des économies d'échelle réalisées par l'automatisation et la simplification des procédures. A l'image du rapprochement entre *OJS* et *OpenAIRE*, le soutien peut prendre la forme de collaborations régulières à des projets, institutions et bibliothèques universitaires et, par là, de subventions « en nature » (via la participation de développeurs associés à ces organisation).

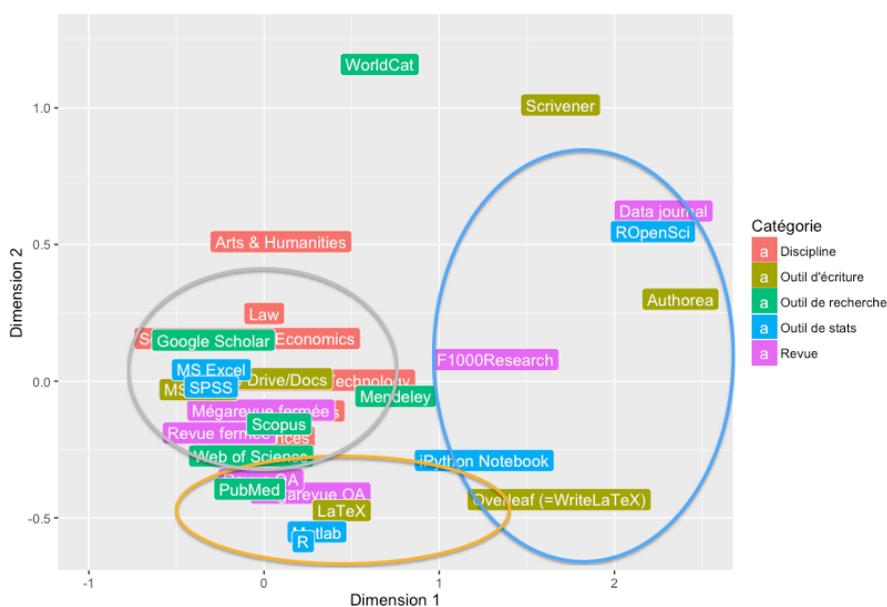
Deuxième partie : les formes de la publication

En 2001, Tim Berners-Lee anticipait la dissolution prochaine de l'article de recherche dans le « web sémantique » : les résultats expérimentaux pourront être directement publiés sur le web, en dehors du contexte d'un article de recherche¹⁸. Quinze ans plus tard, cette révolution promise reste encore à accomplir. Dans leur très grande majorité les textes scientifiques restent publiés sur un format très proche de l'imprimé (le PDF) ou dans une architecture HTML élémentaire (qui, comme pour la plupart des pages web, ne distingue que les structures éditoriales fondamentales : paragraphe, tableau, titre, etc.).

Malgré l'inertie des pratiques, les conceptions culturelles du texte scientifique ne demeurent pas inertes. Nous avons effectué une analyse de correspondance sur les

¹⁸ Tim Berners-Lee et James Hendler, « Publishing on the semantic web », *Nature*, 410(6832), 2001, p. 1023-1024.

résultats d'une enquête menée en 2015 et 2016 sur 20 000 chercheurs dans le monde entier¹⁹. Dans le graphe ci-dessous, plusieurs ensembles cohérents se dessinent : la publication dans les revues « fermées » va de pair avec l'usage d'outils « propriétaires » (*Word, SPSS, Excel*) et/ou détenus par les multinationales de l'édition scientifique (*Web of Science, Scopus, Mendeley*) ; la publication en libre accès s'accorde plutôt avec l'usage de « logiciels libres » (*R, Latex*, même si *Matlab* fait figure d'exception notable) ; un troisième ensemble également dominé par des logiciels libres correspondrait à une exigence nouvelle de « reproductibilité ».



Analyse de correspondances des usages d'outils et d'infrastructures avec les ensembles publication fermée (gris), publication en libre accès (orange) et reproductibilité (bleu)

La cartographie des usages semble ainsi suggérer un éclatement partiel du paradigme actuellement en vigueur. Bien que la révolution promise par Berners-Lee n'ait pas encore eu lieu, de nouveaux « systèmes d'éditorialisation », reposant sur l'interrelation cohérente d'outils, de supports documentaires et de conceptions culturelles du texte émergent.

Publier des données

Les revues scientifiques publient de longue date des séries de données :

¹⁹ Présentation de l'étude : Bianca Kramer et Jeroen Bosman, « Innovations in scholarly communication - global survey on research tool usage », F1000Research, 5, 2016. Adresse : <http://f1000research.com/articles/5-692/v1>

historiquement, dans certaines disciplines comme l'astronomie, les compilations de tableaux précèdent même l'apparition des articles de recherche²⁰. Dans le cadre d'une édition numérique, la reproduction sans altération des tableaux conçus pour une publication imprimée est sous-optimale : elle impose des limites en terme de nombre d'entrée et de colonnes et se prête mal aux réutilisations ultérieures (par contraste avec des formats spécifiques comme le *csv*).

La publication de données passe aujourd'hui de plus en plus par le recours à un intermédiaire : le dépôt de données. Les serveurs des éditeurs ne sont généralement pas optimisés pour gérer et diffuser des compilations volumineuses. Plusieurs acteurs se sont imposés sur ce créneau tels que *Data Dryad* (une organisation à but non-lucratif), *Figshare* (une entreprise) ou *Zenodo* (une infrastructure développée par le *CERN* et *OpenAire*). Si *Zenodo* est gratuit, le modèle économique de *Data Dryad* et de *Figshare* repose sur la perception de *Data processing charges*, qui forment une sorte d'hybride entre coût d'hébergement (il s'agit d'un investissement à renouveler chaque année) et *article processing charges* (l'activité de ces acteurs ne se limite pas au stockage des données mais propose également des fonctionnalités éditoriales telles que l'indexation des jeux de données).

Tout en déléguant l'éditorialisation des données à des intermédiaires, les organisations interrogées ne semblent pas délaisser totalement cette activité. *PLOS* et *PeerJ* ont érigé la publication croisée des données et des textes en principe par défaut, qui ne peut faire que l'objet d'exceptions motivées (par exemple, s'agissant d'informations sensibles). *Elsevier* réoriente sa stratégie éditoriale dans cette direction, en tirant partie de l'expérience acquise par sa maison-mère, *RELX*, dans la gestion des *big data*. Comme pour les outils d'édition, un rapatriement à court ou moyen terme de cette « sous-traitance » n'est pas à exclure.

Qu'elles soient hébergées par un dépôt ou une revue, les données scientifiques conservent un statut périphérique : celui de *documents complémentaires* (*supplementary*

²⁰ Catherine Vassilief, « Les revues françaises d'astronomie. Un cas d'interactivité scientifique. », *Revue de synthèse*, t. 135, n°2-3, p. 239-254.

materials) venus en support d'un article préexistant. Le développement récent de « marques de légitimation » (attribution d'un DOI, décompte des métriques) s'effectue sous les auspices d'une reproduction du dispositif documentaire de l'article. Les jeux de données sont assimilés à des collections closes : les moteurs de recherche de *Figshare* ou de *Zenodo* n'indexent que les métadonnées et non le contenu effectif. Les parcours de lecture institués font de la consultation de l'article une première étape avant d'accéder aux données. Les *data journal* ne s'émancipent pas totalement de ce cadre usuel : ils correspondent plutôt à des *method journals* mettant l'accent sur les conditions de recueil et d'analyse de jeux de données.

Le déficit d'interrelation et d'interopérabilité entre les données pose aujourd'hui tout un ensemble de problématiques qui s'étendent bien au-delà du cas particulier de l'édition scientifique. L'absence d'infrastructures facilitant le repérage et la reprise de données préexistantes se solde par une multiplication des « doublons » où plusieurs services ou entreprises peuvent mener indépendamment un travail similaire. **La « sémantisation » des données remédie à cette dissipation des ressources humaines et techniques.** L'étiquetage standardisé des relations et des catégories permet d'interroger directement les données existantes sans avoir à identifier et télécharger une collection préexistante.

Cette dynamique de sémantisation est aujourd'hui initiée par de nouveaux acteurs : des structures innovantes comme *OpenEdition* et des communautés autogérées à l'image de la « base de connaissance libre » *Wikidata*.

En 2013, *OpenEdition* a généralisé l'emploi des schémas de la *Text Encoding Initiative* (*TEI*). Ce programme de standardisation des formats textuels en ligne inclut également un grand nombre d'attributs permettant de sémantiser « nativement » les productions scientifiques : des balises normalisées indiquent qu'un énoncé porte sur une certaine date, une certaine lieu une certaine personne. Ainsi l'ère hellénistique pourra être définie sous une forme lisible par machine : « <date notBefore="-0323" notAfter="-0031">de la mort d'Alexandre à la conquête romaine</date>²¹ ». Cet effort de sémantisation se prolonge également dans les initiatives institutionnelles telles que la Très Grande

²¹ Exemple extrait de <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/ND.html>

Infrastructure de Recherche *Humanum* et sa plateforme *Isidore* qui se destinent à recueillir et connecter les données de recherche.

Comme *Wikipédia*, *Wikidata* est l'émanation de ses contributeurs qui n'éditent pas seulement son « contenu » mais décident plus généralement de sa morphologie. Les catégories employées font l'objet de propositions et de débats ouverts et continus. Cette rétroaction continue des normes et des usages élargit la couverture de *Wikidata* à l'ensemble des savoirs, en développant et systématisant des champs appropriés à chaque forme de connaissance : il est ainsi possible de spécifier qu'une donnée historique est hypothétique ou qu'elle n'a été valide qu'à une certaine époque.

Faire émerger des formes composites

L'importance nouvelle accordée à des formes distinctes du texte rédigé (données, programme informatique, modélisation interactive...) entraînera-t-elle une transformation profonde des manières d'écrire la science ? L'adoption généralisée du format PDF maintient pour l'instant une continuité totale avec les usages stylistiques hérités des revues imprimés. Le débat sur la « reproductibilité » des méthodes et des résultats favorise cependant l'introduction de dispositifs expérimentaux, articulant étroitement texte, données, code, vidéo et modélisations interactives.

La revue *ReScience* est hébergée sur une plateforme, *Github*, spécialisée dans le recueil de programmes informatiques. Ses fondateurs, Nicolas Rougier et Konrad Hinsén, souhaitent « répondre au déficit de réplique des résultats » en développant un dispositif d'écriture accordant une place centrale à « l'archive du code » :

la chaîne éditoriale est radicalement différente de celle des autres revues traditionnelles. *ReScience* vit sur github où chaque nouvelle implémentation est rendue disponible avec les commentaires, les explications et les tests²².

En apparence, dans ce cadre, la notion classique d'article n'a plus cours. L'effort expérimental de *ReScience* s'étend au-delà d'une simple co-publication du code. Le texte rédigé « éclate » en une multitude de *formes* distinctes : commentaires, interactions,

²² Entretien avec Nicolas Rougier.

codes, données. Cette pratique innovante génère des articles « traditionnels ». Le texte rédigé demeure prépondérant ainsi que les subdivisions consacrées (« Introduction », « Méthode » et « Résultat »). Nicolas Rougier admet que la situation actuelle n'est pas satisfaisante. Maintenant que les procédures éditoriales sont stabilisées, il souhaite faire évoluer le format des articles : « A terme, le standards de publication ne sera plus le PDF mais un format véritablement électronique qui ne s'attachera pas seulement à mimer le papier²³. » Pour s'affranchir complètement des usages établis dans les revues scientifiques depuis près d'un demi-siècle, il songe à un format émergent : le « carnet de code » (*code notebook*).

Les premiers carnets de ce type ont été introduits en 2011 par le projet *IPython* devenu en 2015 le projet *Jupyter*²⁴. Par rapport aux revues traditionnelles tentant de s'ouvrir à la co-publication de code, la démarche du projet *IPython/Jupyter* est exactement inversée : le code constitue la forme d'écriture de référence qui va faire graduellement l'objet d'un enrichissement en intégrant des textes plus rédigés que les simples « commentaires ».

The screenshot shows a Jupyter Notebook interface with the following content:

We will need CrossRef's database of DOI prefixes to identify publishers from the DOIs.
<http://www.crossref.org/06members/50go-live.html>

Don't worry, I scraped all that for you:

```
In [13]: journal_DOIs = pd.read_csv("publisher_Doi_prefixes.csv", index_col = 0)
journal_DOIs[10:15]
```

Out[13]:

	publisher	doi	join	date_1	date_2
10	AAPG/Datapages	10.15530	Nov 02, 2002	Oct 13, 2015	unknown
11	AAPG/Datapages	10.1306	Nov 02, 2002	Mar 02, 2016	unknown
12	AAPG/Datapages	10.15530	Nov 02, 2002	Oct 13, 2015	unknown
13	Aarhus University Library	10.7146	Feb 21, 2012	Mar 07, 2016	unknown
14	AB Svensk Papperstidning	10.3183	Mar 19, 2007	Dec 01, 2015	unknown

Carnet de code ouvert (et modifié) dans l'application Jupyter²⁵

²³ *Ibid.*

²⁴ Cyrille Rossant mentionne une inspiration possible de projets « propriétaires » comme Mathematica ou Maple. Cf. Cyrille Rossant, *Learning IPython for Interactive Computing and Data Visualization*, Packt Publishing Ltd, 2015, p. 2 La création de *Jupyter* marque un élargissement du projet en dehors du langage python (il s'agit d'un mot-valise regroupant les langage **J**ulia, **P**ython et **R**).

²⁵ Carnet de code utilisé pour l'article suivant : John Bohannon, « Who's downloading pirated papers? Everyone », Science, 2016. Adresse : <http://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone>

Par rapport aux usages communs des écritures informatiques, le « commentaire » occupe une fonction prééminente : il n'est pas juste destiné à expliciter les choix mis en œuvre dans chaque fonction mais guide le fil du raisonnement (via des notations en gras). Lorsque le lecteur a installé l'application iPython, **il peut modifier directement le code et altérer ainsi le résultat présenté en sortie**. Dans l'illustration précédente, nous avons altéré le code originel : il affiche non plus l'en-tête du fichier mais les lignes 10 à 15.

Les carnets de code peuvent s'imposer à court terme comme l'un des formats privilégiés des sciences de l'informatique. Et ailleurs ? Dans l'enquête de Kramer et Boesman citée plus haut, l'utilisation des carnets de code n'est pas négligeable, s'agissant d'une technologie créée il y a seulement quelques années et toujours au stade expérimental : 5,18% des sondés en avaient fait une pratique régulière. La proportion d'adoption la plus élevée se trouve non dans les sciences de l'ingénieur et des technologie (qui regroupent l'informatique) mais dans les sciences physiques avec respectivement 8,36 et 9,72% d'utilisateurs ; l'utilisation demeure plus limitée dans les sciences humaines mais pas totalement anecdotique (1,5% en SHS, 1,3% en droit²⁶, 0% en Arts et humanités).

La conversion des formes : du texte aux données et inversement...

Le développement des techniques de *text & data mining* dessine un autre point de convergence entre textes, codes et données. L'enjeu n'est pas de faire cohabiter ces formes composites mais de les « convertir », de faire en sorte que le texte puisse se transformer en une série de données structurées et, inversement, que ces données puissent être de nouveau présentées à l'utilisateur sous la forme d'un texte rédigé.

L'extraction automatisée des textes exploite les récurrences lexicales ou syntaxiques du langage naturel pour transposer les connaissances décrites dans des architectures informationnelles normées et structurées. L'écrit scientifique se prête particulièrement bien à cet exercice : à l'issue de décennies, voire de siècles de négociations collectives les communautés de chercheurs se sont accordés sur des « standards », soit des termes ou

²⁶ Pour le droit la représentativité de l'échantillon demeure incertaine : il n'y a que 7 utilisateurs de carnets de code recensés sur 500 chercheurs (versus 99 sur 6465 chercheurs en SHS).

des expressions désignant avec aussi peu d'ambiguïté que possible un certain objet ou une certaine relation.

Les techniques de *text & data mining* représentent un gain de temps significatif.

Le projet *Text2Genome* a ainsi pu récupérer en quelques mois les données génomiques accumulées par plusieurs millions d'articles scientifiques²⁷. Ces perspectives suscitent depuis peu l'intérêt des grands projets de compilation et de curation des données scientifiques. Une application de *text mining*, *StrepHit*, va prochainement extraire des données des publications scientifiques et les reporter directement sur *Wikidata*²⁸. Dans ce cas, la circulation du texte devenu données ne s'arrête pas là : les fiches de *Wikidata* permettent notamment de générer automatiquement des articles dans des versions linguistiques peu fréquentées de *Wikipédia* (par exemple, la plupart des langues africaines ont moins d'un millier d'articles). L'ubiquité de la connaissance scientifique devient totale : d'énoncé rédigé, elle devient donnée structurée qui peut générer à son tour un autre énoncé rédigé.

Cette mobilité nouvelle des données et des textes soulève plusieurs défis majeurs.

Le cadre légal s'avère inadapté. Alors que la pratique du *text & data mining* ne vise qu'à extraire des « faits bruts » elle implique de diffuser des copies des bases et des corpus étudiés au sein des membres du projet — soit donc, en dehors du « cercle de famille ». Plusieurs exceptions au titre du *text & data mining* à des fins de recherche ont été introduites au cours de ces dernières années au Royaume-Uni et au Japon, en France (dans le cadre du projet de loi sur le numérique) et, par le biais de la jurisprudence, aux États-Unis. L'assouplissement du cadre légal apparaît comme une précondition pour la réalisation de projets pérennes : l'introduction de l'exception au Royaume-Uni a été immédiatement suivi du lancement de l'une des principales initiatives actuelles, le projet *ContentMine* qui ambitionne de collecter plusieurs centaines de millions de « faits » de la littérature scientifique.

En l'état actuel des infrastructures, la circulation extensive des conversions textes-

²⁷ <http://bergmanlab.ls.manchester.ac.uk/text2genome/>

²⁸ https://meta.wikimedia.org/wiki/Grants:IEG/StrepHit:_Wikidata_Statements_Validation_via_References

données et données-textes **peut nuire à la traçabilité de l'information**. Par contraste avec l'usage fréquent des références complètes dans les articles scientifiques, les schémas de bases de données numérisées n'incluent généralement pas de champ spécifique pour indiquer la source d'origine. Cette fonctionnalité existe sur Wikidata mais demeure sous-utilisée²⁹. En mai 2016, une nouvelle initiative, *Wikicite*, ambitionne d'élaborer des outils et des standards adaptés à la citation de source sur *Wikidata* et, plus largement sur les bases de données en ligne³⁰.

Recommandations

Plusieurs évolutions convergentes témoignent d'une transformation structurelle des pratiques d'écriture scientifique. La production de données et de programmes informatiques apparaît de moins en moins comme une activité périphérique et dissimulée : des espaces de publication dédiés (dépôt de données, *data journal*) et des dispositifs hybrides (carnets de code) contribuent à leur éditorialisation. Ces différentes incarnations d'une production scientifique tendent à se rapprocher : les techniques de *text & data mining* assurent ainsi la conversion réciproque des textes et des données.

Les normes en vigueur ne semblent pas adaptées à ces pratiques émergentes. Nous sommes amenés à émettre les recommandations suivantes :

- **Soutenir d'autres formes de publication scientifique** : les textes administratifs encadrant la recherche publique continuent fréquemment de faire de l'article la production scientifique par excellence. La production de bases de données s'est imposée comme un livrable récurrent dans les projets de recherche mais elle ne donne lieu à quasiment aucune reconnaissance par les instances d'évaluation.
- **Introduire une exception au droit d'auteur pour le *text & data mining*** : des réformes visant à sécuriser la pratique du TDM sont actuellement discutées dans plusieurs pays européens. Seule une exception universelle au droit d'auteur à des fins de recherche permettrait de lever efficacement les restrictions.

²⁹ Environ 20% des statements sont référencés à partir de sources externes : <https://tools.wmflabs.org/wikidata-todo/stats.php>

³⁰ Présentation de Wikicite : <https://meta.wikimedia.org/wiki/Wikicite>

- **Initier et accompagner les processus de sémantisation des données.** Les données de la recherche restent trop souvent isolées : en l'état actuel des infrastructures il est difficile d'identifier et de reprendre les travaux accomplis. La sémantisation permet d'étiqueter les données selon des standards interopérables, qui pourraient faire l'objet d'une élaboration collaborative par les communautés scientifique.

Troisième partie : L'évaluation

Le mouvement du libre accès est né d'une crise : la *serial crisis* ou crise des périodiques, marquée par un accroissement exponentiel de la tarification des revues. Il contribue actuellement à en révéler une autre : la *replication crisis* ou crise de reproductibilité.

Cette crise est une conséquence directement de l'accès immédiat et sans restriction des articles scientifiques sur Internet. La dissolution partielle des frontières disciplinaires (jusqu'alors confortées par les politiques d'abonnements préférentiels pratiquées par les bibliothèques) et, plus largement, l'intervention possible d'un public "initié" situé hors du monde scientifique contribue à démultiplier le champ des regards critiques possibles. De nouvelles instances mènent une évaluation *a posteriori* résolument distincte de celle des revues : les contributeurs du forum anonyme *PubPeer* ont ainsi détecté plusieurs cas de fraudes passées inaperçues (dont celles du biologiste français Olivier Voinnet l'année dernière).

Ces vérifications mettent en évidence que de nombreux articles ne sont pas reproductibles : en reprenant strictement la même méthode et des données similaires, les essais de reproduction aboutissent à des résultats différents. Depuis lors, le taux de rétractation (soit le rejet *a posteriori* d'un article préalablement approuvé, généralement suite à une faille du *peer review*) augmente notablement³¹.

Les formes d'évaluation ouvertes connaissent une vogue sans précédent depuis

³¹ R. Grant Steen, Arturo Casadevall et Ferric C. Fang, « Why has the number of scientific retractions increased? », *PLoS One*, 8(7), 2013, p. e68397.

quelques années. Les statistiques de *Google Scholar* illustrent la prégnance croissante de la notion dans le débat scientifique : depuis 2010, le nombre de résultats suit une courbe exponentielle pour atteindre plus de 1000 publications référencées en 2015. Au-delà des essais expérimentaux, les acteurs de référence commencent à s'y ouvrir : la généralisation de l'évaluation ouverte est sérieusement envisagée par *PLOS* à court terme ; *Elsevier* mène également des expériences pilotes en ce sens. Agnès Henri (*EDP Sciences*) nous indique toutefois que l'élaboration de mécanismes appropriés soulève de nombreuses questions : « les comités éditoriaux sont assez enthousiastes sur le principe; dès qu'il s'agit de mettre en œuvre des procédures concrètes il y a beaucoup de frilosités. Faut-il tout publier ? Et sous quelle forme ? (...) La divulgation des énonciateurs pose également problème : faut-il ou non anonymiser les documents³² ? ». Il n'y aurait pas un modèle universel d'*open peer review* mais une famille de modèles reposant de l'ajustement de plusieurs paramètres. Nous allons revenir successivement sur les trois principaux paramètres : la visibilité de l'auteur, la morphologie du dispositif d'écriture et la circulation ultérieure des évaluations.

Les énonciateurs

L'auteur doit-il être public au même titre que sa contribution ? La question répond à un enjeu aussi bien épistémologique que sociologique : l'anonymat semble avoir une incidence sur l'implication de l'évaluateur. *PeerJ* s'est finalement rallié à un système d'évaluation ouvert facultatif, incluant la possibilité d'une contribution anonyme. D'après Jason Hoyt, « nous ne voulions pas effrayer les auteurs ou les évaluateurs potentiels ». De la même manière, l'équipe de *PubPeer* relève que l'anonymat a joué un rôle déterminant dans l'élargissement de la communauté : « Les commentaires sur *Pubpeer* ont été grandement encouragés après la mise en place d'un système d'anonymat contrôlé par l'utilisateur³³. »

Certaines formes d'évaluations ouvertes anonymes ont déjà fait l'objet d'une

³² Entretien avec Agnès Henri.

³³ « Vigilante Scientist », *blog.pubpeer.com*. <http://blog.pubpeer.com/?p=200> En toute logique, ce billet est également anonyme.

expérimentation considérable hors du champ de la publication scientifique. Les 100 000 contributeurs réguliers des diverses communautés wikipédiennes linguistiques privilégient en effet l'emploi de comptes anonymes ou pseudonymes afin de focaliser les débats encyclopédiques sur les idées émises et non sur les personnes. Si elle facilite l'implication des évaluateurs, la pratique de l'anonymat pourrait être également plus facilement détournée³⁴.

Une seconde ligne de démarcation porte sur les modalités du recrutement des contributeurs. Toutes les formes d'*open peer review* que nous avons pu étudier limitaient en pratique l'accès de l'évaluation aux auteurs de contribution scientifique. Dans le cas de l'expérience menée par la revue *VertigO* cette limitation était involontaire : la revue avait « invité ses lecteurs à commenter les textes sans que cela n'apporte de résultat³⁵ ». *Polymath* se voulait également ouvert en principe mais a principalement recueilli des commentaires experts. Le développement d'articulations durables entre sciences participatives et évaluation ouverte reste encore hypothétique.

La barrière à l'entrée est variable. Sur *Pubpeer* ou sur le *Self Journal of Science* l'évaluation de l'ensemble des articles est ouverte à l'ensemble de la communauté scientifique sans faire intervenir un *editor* en amont : il suffit que le chercheur le souhaite pour participer³⁶. D'autres revues maintiennent un mécanisme de sélection des évaluateurs qui peut reposer sur un *editor* (c'est notamment le cas de *PeerJ* ou du *British Medical*) ou sur l'auteur lui-même (qui est fortement incité à suggérer des évaluateurs sur *F1000*) L'expérience *VertigO* a conclu à l'utilité du maintien, du moins temporairement, d'une intermédiation éditoriale : les courriers ciblés, sollicitant en particulier un évaluateur ont rencontré bien plus de succès que les annonces générales³⁷.

³⁴ Il y a ainsi eu plusieurs cas d'évaluations masquées sur PubPeer (par exemple un auteur se faisant passer pour un pair et s'attribuant une bonne évaluation) : <https://forbetterscience.wordpress.com/2015/12/13/post-publication-peer-review-signed-or-anonymous/>

³⁵ Julien Bordier, « Évaluation ouverte par les pairs : de l'expérimentation à la modélisation », 2016, p. 33. Mis en ligne sur HAL à l'adresse : <https://hal.archives-ouvertes.fr/hal-01283582/document>

³⁶ La définition de la "communauté scientifique" varie dans ces deux initiatives : sur *PubPeer*, elle comprend l'ensemble des auteurs indexés dans les archives ouvertes *PubMed* et *ArXiv* ; sur le *Self Journal of Science*, l'adhésion est conditionnée à une affiliation dans une institution scientifique.

³⁷ Julien Bordier, *op. cit.*, p. 33

Les formes de l'évaluation

Il existe aujourd'hui des mises en oeuvre très variées, qui ne visent pas seulement à divulguer les textes échangés mais, potentiellement, à les altérer complètement. Nous retrouvons ici les termes de notre interrogation initiale sur les enjeux éditoriaux du libre accès : dans des systèmes d'éditorialisation toutes les « dimensions » de la production textuelle sont étroitement corrélées.

Les premiers essais de *peer review* ouverts ne s'affranchissent pas totalement des usages stylistiques en vigueur dans les revues : l'un des pionniers de cette pratique, le *British Medical Journal*, se cantonne à un simple acte de transparence en publiant les mails échangés. De nouveaux acteurs tentent de faire émerger des dispositifs inédits.

Le *Self Journal of Science* donne à voir une évaluation en train de se faire : les articles proposés sont immédiatement publiés puis les évaluateurs déposent leurs remarques en commentaires ; un article est officiellement admis lorsqu'il a recueilli l'approbation de cinq membres. Ce dispositif ne requiert aucun *editor* pour faire le travail d'intermédiation entre les auteurs et les évaluateurs : ces derniers se déclarent spontanément.

Le projet *Hypothes.is* a développé un système d'annotation pour le *web* aujourd'hui en voie d'intégration dans plusieurs outils d'édition scientifique (dont *Open Journal System*). Le *peer review* ne prend pas seulement la forme d'un « texte » unifié mais d'une série de commentaires apposés en continu en regard de la contribution scientifique. L'évaluation ne découle plus seulement de l'expertise de l'évaluateur mais d'un échange argumenté et continu³⁸.

Le *Self Journal of Science* et l'expérience *Vertigo* constituent des solutions éditoriales complètes. Plusieurs acteurs de l'*open peer review* se positionnent uniquement sur le segment de l'évaluation. Nous avons déjà mentionné le cas du forum *PubPeer* : *RetractionWatch* assume à peu près la même fonction (vérifier *a posteriori* des articles déjà approuvés). *Publons* et *Rubriq* s'inscrivent plutôt dans le cadre d'une division du travail : ils assurent une évaluation rapide et efficace des propositions soumises (dans le

³⁸ *Ibid.*, p. 30.

cas de *Rubriq*, en rétribuant les évaluateurs) puis les envoient à des revues partenaires.

Faire circuler l'évaluation ouverte

Depuis les années 1960, la mise en circulation des évaluations se réduit au calcul de *métriques*. Au niveau institutionnel, l'estimation de la qualité d'une production ou d'une revue scientifique repose fréquemment sur le calcul du taux de citation ou « facteur d'impact ». Depuis une dizaine d'années, cette pratique est remise en question par plusieurs alternatives : les *altmetrics*. Les *altmetrics* les plus courantes portent ainsi sur le nombre de téléchargements, de visiteurs uniques, de mentions dans les réseaux sociaux... Le développement des indicateurs alternatifs reste cependant pénalisé par l'absence d'un index de citation universel en accès libre : si *Google Scholar* est en libre consultation il n'existe pas d'API pour reprendre automatiquement les informations³⁹. Plusieurs projets en cours visent à réunir les citations des publications en libre accès (puisque rien ne s'oppose à leur libre reprise). *Science Open* travaille depuis le début de l'année 2016 à un *Open Citation Index* qui inclut pour l'instant 2 millions d'articles « principalement extraits de *PubMed* et d'*ArXiv*⁴⁰ ».

Il existe une alternative : en s'ouvrant au public, l'évaluation devient un texte mobile et circulant au même titre que les *métriques*. Tout comme l'article scientifique, elle peut être indexée ou disséminée dans d'autres dispositifs éditoriaux (index bibliographiques comme *crossref*, archives ouvertes, réseaux sociaux académiques). *PeerJ* attribue ainsi depuis peu un identifiant DOI aux évaluations et signale leur taux de citation. La prise en compte des enjeux nouveaux de reproductibilité contribue à valoriser les évaluations et leur conférer quasiment le statut de contribution scientifique « seconde ». Au cours de l'année 2015, le site *pubpeer* a introduit une extension pour le navigateur Firefox permettant de retrouver les commentaires associés à chaque contribution scientifique indexée dans sa base. Sur le *Self Journal of Science* la curation constitue en soi une pratique d'évaluation :

³⁹ Cf. notre tentative de récupération des données de scholar à partir d'un *scraping* (qui a rapidement débouché sur un blocage en consultation) : « Faire du data mining avec Google : comme tromper big brother ? », <http://scoms.hypotheses.org/216>

⁴⁰ Jon Tennant, « The Open Citation Index », *scienceopen.com*, 29 février 2016, <http://blog.scienceopen.com/2016/02/the-open-citation-index/>

chaque chercheur peut créer son propre journal en reprenant des articles jugés emblématiques d'un champ de recherche. Une sélection réfléchie a valeur d'endossement et valide *a posteriori* une démarche scientifique : « Par contraste avec la citation, la reprise est plus difficile à manipuler : elle relève d'un acte éditorial plus fort et est immédiatement visible. »

La mise en circulation des évaluations suscite l'intérêt des archives ouvertes. Le projet Open Aire vient de mettre au point un *Open Peer Review Module* (OPRM) : il s'agit d'une extension associée à une archive ouverte permettant de déposer directement des évaluations, sans avoir à transiter par une revue ou un intermédiaire (comme *PubPeer*).

Toutes ces connexions autorisent l'existence concomitante de plusieurs systèmes d'évaluation. L'examen du même article sur le *Self Journal of Science*, sur le module OPR d'une archive ouverte ou sur un forum ouvert anonyme débouchera probablement sur des conclusions différentes et complémentaires. Par exemple, dans la mesure où il est ouvert à un public généraliste, *PubPeer* semble mieux se prêter à la détection d'erreurs et d'anomalies statistiques qu'à l'évaluation globale de la pertinence d'une étude au sein de son champ de recherche spécialisé. La multiplication des formes d'évaluation ouvre la perspective d'un *méta-peer review* enrichi, apte à répondre aux attentes nouvelles en matière de reproductibilité de la recherche.

Recommandations

Comme pour les formes de publication, les pratiques d'évaluation sont en décalage croissant avec les normes actuellement appliquées. La multiplication des points de vue critiques sur les contributions scientifiques au-delà du seul *peer review* est déjà engagée : des forums interactifs comme *PubPeer* révèlent des failles préexistantes ; les différents modèles d'*open peer review* mis en œuvre au cours de ces dernières années apparaissent comme des alternatives viables au point que des éditeurs établis (*PLOS*, *EDP Sciences*) envisagent de les déployer.

- **Repenser l'évaluation de l'évaluation** : Les textes administratifs encadrant la recherche réduisent usuellement l'ensemble des formes d'évaluation possible au

protocole normalisé de *peer review* mis en œuvre par des revues de référence. Par exemple, l'AERES dressait pour les sciences humaines et sociales des listes de revues qualifiantes. Cette qualification ne permet notamment pas de prendre en compte de nouveaux acteurs intermédiaires (comme *Publons* ou *Episciences*) qui n'assument qu'une partie du travail éditorial de la revue.

- **Signaler les diverses formes d'évaluation dans les archives ouvertes** : de nouveaux dispositifs éditoriaux pourraient préciser que tel article a été certifié par une revue ou par une instance d'évaluation autonome. Cette évolution contribuerait autant à valoriser les plateformes d'archives ouvertes (qui deviendraient le point central où convergent les différentes formes d'évaluation) qu'à faciliter l'avènement de nouveaux acteurs (dont la capacité à certifier une contribution scientifique serait d'emblée reconnue).

Quatrième partie : les modèles économiques

En 2016, l'une des plus grandes bibliothèques scientifiques mondiales est « pirate ». Le site *sci-hub* héberge et diffuse plus de 50 millions d'articles, obtenus pour la plupart illégalement. La taille de ce corpus est comparable aux collections (légales) de méta-données que sont le *Web of Science* (90 millions d'articles indexés) ou *Scopus* (55 millions d'articles). Le développement de *Sci-Hub* repose sur une forme d'industrialisation du piratage marquée par l'absorption des *biblioleaks* préexistants⁴¹ et l'automatisation des « fuites » : lorsqu'un article sollicité n'est pas déjà disponible dans sa base de données, *Sci-Hub* va automatiquement tenter de l'obtenir sur l'une des plateformes de l'éditeur.

L'irruption récente de ce nouvel acteur contribue à « dramatiser » la question du devenir des modèles économiques de l'édition scientifique. Localisé au Kazakhstan, *sci-hub* a peu de chance d'être fermé sur une décision de justice et peut compter sur le soutien tacite ou explicite des « auteurs »⁴². La convergence des besoins des utilisateurs-auteurs de la

⁴¹ Cf. les recherches approfondies de Guillaume Cabanac : « Bibliogifts in LibGen? A study of a text-sharing platform driven by biblioleaks and crowdsourcing », *Journal of the Association for Information Science and Technology*, 67(4), 2016, p. 874-884.

⁴² Cf. ce sondage sur le site de la revue *Science* <http://www.sciencemag.org/news/2016/05/survey-most-give-thumbs-pirated-papers>

recherche et des exigences des institutions implique un basculement général à court terme. Nous examinons ici successivement les principaux aspects de cette transformation : l'évolution des frais d'édition, la diversification des ressources et le développement de structures non-commerciales.

Les frais de l'édition

En 2016, le coût d'une revue est étroitement lié à son modèle éditorial. Sur le *web* le coût de la diffusion devient progressivement trop faible pour pouvoir être facturé (selon l'expression anglaise, il est *too cheap to meter*). Disposant des capacités de stockages colossales du CERN, le dépôt de données Zenodo est entièrement gratuit (par contraste avec ses principaux « concurrents », *Figshare* ou *Data Dryad*). D'après Laurent Romary *HAL* s'inscrit dans la même configuration : « l'hébergement d'un article revient à 5 euros et ce poste ne cesse de baisser avec l'optimisation de la plateforme⁴³. » En somme, la mutualisation des frais revient collectivement moins cher que l'attribution d'un prix (et toute l'infrastructure bureaucratique qu'il faudrait mettre en œuvre pour gérer ces transferts de fonds). Les outils d'édition bénéficient également de cette économie d'échelle : qu'*Open Journal System* soit téléchargé par des centaines ou des dizaines de milliers de revues entraîne un surcoût quasi-nul pour ses concepteurs.

Cette stratégie de réduction des coûts se heurte cependant à une difficulté majeure : la nécessité d'adapter l'équipement et les outils aux besoins propres de la revue et d'une certaine communauté de recherche. Les revues en informatique disposent d'un avantage significatif : elles peuvent compter sur la participation directe de leurs initiateurs. L'équipe du *Journal of Machine Learning Software* ou de *ReScience* assure en continu la maintenance et l'adaptation du dispositif d'édition ; la technique ne se trouve ici jamais dissociée de la pratique. Selon Kevin Stranack (d'*Open Journal System*), le recrutement de développeurs au sein de bibliothèques et d'institutions scientifiques pourrait contribuer à « démocratiser » cette interrelation privilégiée. Ces derniers pourraient notamment concevoir des extensions directement adaptées aux besoins des utilisateurs.

⁴³ Entretien avec Laurent Romary

Dans les entretiens, le *peer review* est fréquemment apparu comme un « coût incompressible ». Si l'évaluation, en tant que telle, est bénévole, la mise en contact des auteurs et des évaluateurs peut représenter plusieurs heures, voire plusieurs journées de travail. Un *editor* de l'*American Physical Society* constate qu'une « une revue vraiment sélective et éditorialisée ne peut être bénéficiaire : elle perd de l'argent sur les articles qu'elle refuse⁴⁴. »

Serait-il possible de réduire les frais associés au *peer review*, par exemple en automatisant une partie des procédures ? Cette piste suscite de nombreuses expérimentations actuellement (reconnaissance automatisée des évaluateurs, optimisation des temps d'interaction sur *PeerJ* ou sur le futur système d'édition de *PLOS*, *APERTA*). Les économies générées par ces innovations ne sont pas négligeables : *PeerJ* peut se permettre de pratiquer une tarification bien inférieure à ses concurrents pour un résultat apparemment comparable (100 euros par auteurs). Certains dispositifs d'évaluation ouverte expérimentaux présentés dans la troisième partie vont au-delà. Ils s'affranchissent totalement des frais du *peer review* : sur *ReScience*, le *Self Journal of Science*, l'*Open-Peer Review Module* ou *PubPeer*, les auteurs et les évaluateurs interagissent sans recourir à un intermédiaire.

Les ressources

Seule une diversification des ressources permet de concilier la publication en libre accès et le maintien *a minima* du modèle de diffusion à visée commerciale. Trois perspectives principales s'ouvrent actuellement. Elles ne constituent pas des modèles clos et semblent même plutôt appelés à s'hybrider.

- **Le modèle auteur-payeur.** Il s'agit, de loin, de la solution actuellement privilégiée. Les négociations les plus récentes entre Elsevier et des consortiums nationaux de bibliothèques et d'institutions scientifiques prévoient ainsi une migration graduelle des financements alloués aux abonnements vers l'achat de droits à publier en libre accès (APCs) dans des revues: selon les termes de l'accord néerlandais, 30 % des

⁴⁴ Entretien avec un editor d'APS.

articles devraient être publiés en libre accès d'ici 2018⁴⁵. La généralisation du modèle auteur-payeur aura probablement pour effet d'entériner une tendance structurelle de l'économie de la publication scientifique : le passage d'un marché de « biens » ou de « services » à un marché purement « symbolique ». L'estimation du prix des APCs sur la durée est très incertaine : en quinze ans, les tarifs pratiqués par *PLOS* n'ont pas baissé mais augmenté (après la dernière hausse ils s'établissent à 1495 \$ pour *PLOSOne*⁴⁶). Les éditeurs traditionnels pratiquent des prix beaucoup plus élevés (environ 5000 \$ pour *Nature Communication*) qui pourraient être amenés à augmenter encore : d'après Mike Taylor, un article diffusé par abonnement rapporte 7000 \$ à Elsevier en moyenne⁴⁷. Dans une économie purement symbolique rien n'interdit que la tarification courante ne finisse par être considérablement plus élevée : « Si les gens sont prêts à payer plus de 230 000 dollars pour un diplôme de Yale (...) pourquoi ne seraient-ils pas prêts à sortir 50 000 \$ pour un article capable de leur obtenir un poste pérenne⁴⁸ ».

- **La vente de données associées** : les grands éditeurs scientifiques s'intéressent de plus en plus aux modèles économiques développés par les réseaux sociaux. Les revenus de *Facebook* ou *Twitter* ne reposent en effet pas sur la perception d'un abonnement mais sur la vente de données personnelles. Les réseaux sociaux académiques comme *ResearchGate*, *Academia* ou *MyScienceWork* sont parvenus à généraliser ce modèle au cas particulier de l'édition scientifique. Elsevier se réoriente largement dans cette direction. L'acquisition d'un service de gestion bibliographique analogue à *Zotero*, *Mendeley*, et l'intégration toute récente, de l'archive ouverte *SSRN* forment l'embryon d'un nouveau département entièrement dédié à la commercialisation de données de consultation (qui se prolonge par des

⁴⁵ Voir la synthèse de l'accord : http://vsnu.nl/files/documenten/Domeinen/Onderzoek/Open%20access/QA_OpenAccess_Akkoord_Elsevier_ENG.pdf

⁴⁶ <http://blogs.plos.org/plos/2015/09/plos-publication-costs-update/>

⁴⁷ <https://svpow.com/2012/07/09/what-does-it-cost-to-publish-a-paper-with-elsevier/>

⁴⁸ Björn Brembs, « How gold open access may make things worse », 7 avril 2016, <http://bjoern.brembs.net/2016/04/how-gold-open-access-may-make-things-worse/>

partenariats exclusifs avec des réseaux sociaux académiques comme *MyScienceWork*⁴⁹). Parallèlement l'éditeur a largement investi dans le développement d'outils d'évaluation et de *monitoring* de la recherche très largement adoptés par des institutions scientifiques : *SciVal*⁵⁰ et *Pure*⁵¹. Nous voyons ainsi se dessiner l'ébauche d'un nouveau marché fructueux : les informations sur les utilisateurs récupérées sur les archives ouvertes, les outils de gestions bibliographique et les réseaux sociaux estampillés par Elsevier convergent pour améliorer l'efficacité (et donc la rentabilité) d'instruments de méta-évaluation.

- **Les services additionnels** : ce modèle économique opère une clarification du modèle auteur-payeur. L'acquisition ne porte en effet plus sur un droit à publier (et donc, sur la réputation présumée de la revue) mais sur des services précis. Le service en question peut être essentiellement symbolique : *The Winnower* fait ainsi facturer l'acquisition d'un DOI. *Open Edition* a développé des services spécifiques à destination des bibliothèques et de leur public. Par exemple, tout en maintenant la diffusion en HTML gratuite, les versions PDF et EPUB sont vendues (sans DRM) aux bibliothèques ou aux particuliers⁵². Le développement des techniques d'extraction pourrait contribuer à accroître la valorisation de ces services. L'élaboration d'une indexation fine en TEI et l'inscription d'emblée des données et des informations dans les écosystèmes d'échange du web sémantique⁵³ pourraient à terme accroître la valorisation et l'attractivité de ces services.

La piste du non-commercial

Des trois modèles économiques compatibles avec une publication en libre accès, seule la vente de services additionnels débouche sur des dépenses maîtrisées : les coûts

⁴⁹ <https://www.elsevier.com/about/press-releases/science-and-technology/sciencedirect-content-now-available-to-mysciencework-users>

⁵⁰ <https://www.elsevier.com/solutions/scival>

⁵¹ <https://www.elsevier.com/solutions/pure>

⁵² Voir par exemple, l'offre pour les livres : <http://www.openedition.org/13052>

⁵³ Par exemple via des outils d'identification automatisés des identifiants de publication (DOI) comme *Bilbo* (<http://maisondesrevues.org/680>)

correspondent à des prestations mesurables. Pour l'heure, il ne s'agit pas de l'option majoritairement retenue. La proposition de *Journal Flipping* de la Max Planck va dans le sens d'une généralisation des APCs ; parallèlement de grands acteurs comme *Elsevier* tentent de réitérer le succès des « jardins clos » du web comme *Facebook* dans le monde scientifique. Ces dérives possibles donnent du crédit à une rupture apparemment radicale : s'affranchir totalement du modèle commercial. Dans ce cadre, l'édition scientifique ne vendrait plus rien (ni articles, ni droits à publier, ni données personnelles) ; sa subsistance reposant exclusivement sur du volontariat et/ou des subventions institutionnelles.

Ce modèle non-commercial est déjà couramment pratiqué. En 2008, un rapport du JISC évaluait la part du travail éditorial à 28 % des frais réels de conception d'un article scientifique (ce qui inclut également la contribution du rédacteur et des évaluateurs) dans une revue électronique en accès fermé⁵⁴. Et ce travail éditorial n'est pas toujours assumé par l'éditeur : l'étude d'Odile Contat et d'Anne-Solweig Gremillet montre que les revues françaises sont usuellement subventionnées « en nature » en déléguant la gestion du *peer review* et la correction des articles à du personnel universitaire pour un coût médian de 1600 euros par article⁵⁵. L'édition scientifique ne devient un « marché » qu'à la toute fin du processus : l'écriture des articles, leur évaluation et, souvent, leur mise en forme sont défrayés par des fonds publics.

En principe, les économies d'échelle générées par les systèmes d'édition électronique autorisent un passage immédiat au libre accès simplement en réaffectant ces subventions dissimulées. La somme médiane de 1600 euros identifiée par Contat et Gremillet est comparable aux coûts pratiqués par des éditeurs de référence (*PLOS* est à 1500 \$) et nettement supérieure à ceux de nouvelles structures comme les épirevues (d'après Laurent Romary, le travail incompressible de secrétariat pour *Episciences* représente environ 200 euros, le reste étant à peu près négligeable ou largement amorti sur la

⁵⁴ Calcul à partir des données p. 184, <http://www.webarchive.org.uk/wayback/archive/20140614211536/http://www.jisc.ac.uk/media/documents/publications/rpconomicpublishing.pdf>

⁵⁵ Odile Contat et Anne-Solweig Gremillet, « Publier : à quel prix ? Étude sur la structuration des coûts de publication pour les revues françaises en SHS », *Revue française des sciences de l'information et de la communication*, 2015, no 7. <https://rfsic.revues.org/1716>

durée⁵⁶).

Le développement du libre accès s'est accompagnée d'une multiplication des revues non-commerciales : environ deux tiers des titres référencés par le *Directory of Open Access Journal* ne font payer ni l'auteur ni le lecteur. Bien que quantitativement importantes ces initiatives commencent tout juste à gagner en visibilité. Le *Public Knowledge Project* (qui élabore également le logiciel libre *Open Journal System*) a lancé depuis peu une étude sur les modèles coopératifs dans l'édition scientifique⁵⁷. Une étude longitudinale publiée dans *PeerJ* en mai 2016 montre que les structures reposant principalement sur du volontariat parviennent à se maintenir sur la durée mais peinent à se développer pleinement : le nombre moyen d'article soumis reste inférieur à la moyenne⁵⁸.

Les revues non-commerciales sont confrontées à une double peine : tout ne disposant pas de fonds propres, elles touchent fréquemment moins de subventions. Des trois structures concernées dans notre échantillon, seules les épirevues d'*Episciences* bénéficient du soutien direct ou indirect d'institutions scientifiques. Les organisations purement bénévoles n'ont paradoxalement pas les moyens de solliciter directement des aides. Cette situation affecte la visibilité et la pérennisation du projet éditorial : pour Michaël Bon la communauté du *Self-Journal of Science* reste restreinte faute d'avoir obtenu une « caution » institutionnelle ; Nicolas Rougier de *ReScience* nous a fait part de ses inquiétudes sur la viabilité à long terme d'une revue hébergée par *Github*.

Plusieurs initiatives récentes s'attachent à « officialiser » le soutien public. Le financement d'*Open Library of Humanities (OLH)* repose sur la perception de souscriptions de bibliothèques : ces dernières n'achètent pas des publications (disponibles en libre accès) ni des droits à publier mais engagent des investissements permettant à terme de faire globalement baisser les frais de publication dans les humanités. D'après l'initiateur d'*OLH*, Martin Eve, disséminer ce coût entre plusieurs institutions leur revient

⁵⁶ Entretien avec Laurent Romary.

⁵⁷ <http://oa-cooperative.org/>

⁵⁸ Bo-Christer Björk, Cenyu Shen et Mikael Laakso, « A longitudinal study of independent scholar-published open access journals », *PeerJ*, 4, 2016, p. e1990. <https://peerj.com/articles/1990/>

individuellement beaucoup moins cher qu'un abonnement (un peu plus de 2 £ pour 190 institutions)⁵⁹.

Des dispositifs plus génériques paraissent nécessaires pour assurer le développement des plus petites structures. En 2014, le rapport Lemoine sur la transformation numérique de l'économie française préconisait d'instaurer un « 1% *open* », soit une somme prélevée d'emblée sur des activités commerciales « pour inciter les entreprises à se consacrer aux projets *open*⁶⁰ » ; de la même manière, le financement des nouveaux modèles de l'édition scientifique pourrait être assuré par un prélèvement systématique sur les montants alloués aux abonnements et APCs. Les structures dépendantes du volontariat seraient également renforcées par l'introduction d'un droit à la contribution : chaque année une petite partie du temps de travail pourrait être dédiée à l'élaboration de projets non-commerciaux.

Recommandations

Sur le plan économique, la généralisation du libre accès est fréquemment assimilée à une généralisation du modèle auteur-payeur. La diversification des modèles éditoriaux constatée dans les parties précédentes se traduit plutôt par une diversification concordante des modèles économiques possibles. Au-delà des APCs, des acteurs émergents élaborent de nouvelles modalités d'organisation : souscriptions, services additionnels, volontariat. Même les éditeurs scientifiques anticipent cet élargissement des perspectives. Sur le long terme, l'exploitation des données sera peut-être plus profitable à *Elsevier* que la perception des APCs.

Dans ce contexte en forte évolution nous émettons les recommandations suivantes :

- **Transformer une partie des dépenses en investissement** : le budget public dédié à l'édition scientifique vise essentiellement à perpétuer le système existant. Le *journal flipping* s'apparente à un dispositif de « bascule » des dépenses actuelles qui seraient au mieux pérennisées. Par contraste, l'émergence et la consolidation de nouveaux modèles économiques et éditoriaux ouvrent la

⁵⁹ Entretien avec Martin Eve.

⁶⁰ Philippe Lemoine, *La Transformation numérique de l'économie française*, p. 21

perspective d'une réduction globale et massive des dépenses engagées. **Cet investissement pourrait prendre notamment la forme d'un 1% innovant** : toute transaction (renouvellement d'abonnements, APCs...) inclurait une petite provision dédiée au financement de nouveaux modèles.

- **Elaborer des formes de subventions adaptées aux entités non-commerciales** : ces structures ne disposent généralement pas des moyens nécessaires pour solliciter les aides directes et indirectes dont bénéficient leurs concurrents commerciaux. L'introduction de dispositifs de micro-financement avec des procédures minimales et/ou la reconnaissance du droit à la contribution constitueraient notamment des réponses appropriées. Cette démarche pourrait s'articuler à la refonte de l'évaluation de l'évaluation préconisée à la fin de la troisième partie, en établissant des incitations positives (bonus...) à publier dans des revues en libre accès non commerciales et à participer à leur élaboration éditoriale.

•

Cinquième partie : de l'innovation à l'infrastructure

Deux facteurs : mobilité et diversité

La réduction drastique du coût de réplique et de republication sur le web et la levée des restrictions légales dans le cadre du libre accès élargissent substantiellement le champ de circulations possibles.

Un texte est amené à avoir plusieurs vies. De document déposé dans une archive ouverte, il peut se muer en article de revue. Il entraîne potentiellement la formulation d'autres textes (tels que les évaluations ouvertes), qui peuvent à leur tour en susciter d'autres. Le contenu même du texte s'émancipe : les techniques d'extraction automatisées autorisent la migration des informations dans des bases structurées et, de là, elles peuvent de nouveau se convertir en texte rédigé.

Toutes ces améliorations successives pourraient être rétroactivement « connectées », les archives compilant les évaluations dispersées, les tentatives de

réplications ou la liste des “faits bruts” extraits par *ContentMine*. L’illustration ci-dessous présente une projection d’interface enrichie d’une archive HAL : des encadrés sur le côté viennent spécifier les évaluations, les réplications ou les diverses extractions de données.

The screenshot displays a HAL interface for a document. At the top left, a red box indicates 'sic_01222973, version 1'. The main title is 'La négociation contre la démocratie : le cas Wikipedia' in French. Below the title, the author 'Pierre-Carl Langlais' is listed, along with the affiliation 'GRIPIIC - Groupe de recherches interdisciplinaires sur les processus d'information et de communication'. The abstract is provided in English, discussing the first pillar of Wikipedia and the concept of democracy. Keywords include 'Computer mediated community', 'Consensual Democracy', and 'Majoritarian Democracy'. The document type is identified as 'Article dans une revue' from 'Négociations, De Boeck Supérieur, 2014'. The domain is 'Sciences de l'Homme et Société / Sciences de l'information et de la communication / Espace public'. On the right side, there are four orange-bordered panels: 'FICHIER' (listing a PDF file), 'EVALUATIONS' (listing a PubPeer review and a Self Journal of Science entry), 'RÉPLICATION' (listing a Jupyter attempt), and 'EXTRACTION DE DONNÉES' (listing Wikidata and StrepHit extractions).

Projection d’une intégration de textes dérivés dans une archive ouverte

Nous voyons ici se dessiner une articulation fondamentale entre mobilité et diversité. La projection d’intégration admet d’emblée la possibilité d’intégrer plusieurs formes d’évaluation, qui n’ont pas forcément le même portée ni la même utilité. Dans la troisième partie nous avons pu voir que les modèles émergents d’évaluation ouverte présentaient des avantages distincts et complémentaires : si le dispositif de *PubPeer* a fait la preuve de son efficacité pour détecter les fraudes, il ne parvient sans doute pas à assumer la totalité des fonctions du *peer review* classique. L’amélioration de la fiabilité de l’évaluation passerait ainsi par la cohabitation de plusieurs systèmes et, donc, par une multiplication des angles et des prises de vues possibles que de nouveaux dispositifs éditoriaux donneraient à lire en concomitance.

Le travail de fond engagé par les outils d’édition depuis une quinzaine d’années facilite cette articulation entre une pluralité de modèles. Les nombreuses formes d’extensions et de customisation possibles correspondent, de plus en plus, à des standards interopérables. Chaque structure éditoriale peut être ainsi amenée à définir son propre « cocktail » adapté à ses besoins et ses ambitions tout en étant en permanence connecté

aux flux de circulation et de dissémination des textes.

Innovation, infrastructure, écosystèmes : une prise de conscience

La structuration actuelle de l'édition scientifique se caractérise par une forme d'entropie de l'effort. Des organisations sont contraintes d'engager des ressources parfois importantes pour assurer certaines activités pour lesquelles elles n'ont pas de compétences ou d'expérience en interne. Les duplications sont monnaie courante : les mêmes articles vont être évalués plusieurs fois (généralement pour parvenir aux mêmes conclusions), des fonctionnalités vont être réinventées indépendamment, des séries de données similaires vont être constituées à plusieurs reprises.

La diminution de ce coût global passe par une mutation conjointe : intégrer l'innovation dans des infrastructures cohérentes et structurées et, à un deuxième niveau, assurer l'interconnexion de ces infrastructures dans des écosystèmes. Les organisations sélectionnées dans notre échantillon nous ont fait part, à différents degrés, de cette prise de conscience. Les co-fondateurs de *ReScience* sont conscients d'une situation « inconfortable » : perpétuer un format classique de l'article (où les fichiers de code et de données n'occupent qu'une place annexe) alors même que le texte rédigé n'occupe plus qu'une place réduite dans le protocole d'évaluation. Il existerait ainsi des agencements (approximativement) optimaux des formes prises par les différents aspects de l'éditorialisation scientifique. Ces agencements sont difficilement perceptibles *a priori* : ils se structurent par la pratique. Une expression revenait ainsi souvent dans les entretiens : *work in progress*. *The Winnower* est « actuellement toujours en phase d'expérimentation⁶¹ » ; chez *PLOS* « la généralisation de l'*open peer review* constitue un *work in progress* ». Aucune organisation nouvelle ne prétend avoir trouvé d'emblée le modèle idéal pour incarner un certain type de changement souhaité.

Le passage de l'innovation à l'infrastructure ne se dessine pas seulement au gré d'une multitude de prises de conscience isolées. Depuis quelques années, l'Union Européenne a entrepris une politique d'infrastructure cohérente. Les politiques de publication mises en

⁶¹ Entretien avec Joshua Nicholson

œuvre par le programme Horizon 2020 requièrent en effet le déploiement de systèmes de republication adaptés ; parallèlement la multiplication des projets de bases de données⁶² soulève d'emblée la question de leur interopérabilité. **La création conjointe de la méta-archive ouverte *OpenAire* et du dépôt de données *Zenodo* pose les bases d'un écosystème européen de la publication scientifique⁶³.** Les outils d'édition sont étroitement associés à ces initiatives : *Open Journal System* travaille « de près avec le projet *OpenAire*⁶⁴ » ; c'est également au travers d'une collaboration avec *OpenAire* que *DSpace* a développé un module d'évaluation pour les archives ouvertes. Plus largement, de nombreux acteurs sont en train d'élaborer des interrelations suivies avec ce noyau en formation : depuis novembre 2015, *OpenEdition* a mis à jour son architecture éditoriale pour être entièrement moissonnable par *OpenAire*⁶⁵ ; *Zenodo* a noué un partenariat avec *GitHub* pour faciliter l'export mutuel des données et des fichiers informatiques⁶⁶.

Les grands éditeurs scientifiques restent peu associés à ce mouvement européen : ils tentent d'élaborer leurs propres écosystèmes. Les politiques d'acquisition et de développement d'Elsevier s'inscrivent pleinement dans ce cadre. Systèmes d'édition (*Pure* et, à terme, *EVISE*), systèmes d'évaluation (*SciVal*), moteurs de recherche (*SCOPUS*) outils de gestion bibliographiques (*Mendeley*), archives ouverte (*SSRN*) et, probablement à terme, réseaux sociaux académiques **forment la base initiale d'un écosystème captif.** Comme sur *Facebook*, *Twitter* ou, finalement, *Google*, l'usage de ces différents services sera au moins partiellement libre. *Elsevier* s'assure par contre une maîtrise complète des flux d'informations et se réserve la possibilité de monnayer ces différentes mesures d'impact auprès des institutions scientifiques.

⁶² A ce jour, plus de 800 infrastructures de recherche ont été financées par des programmes de la commission européenne. Cf. la cartographie http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=mapri

⁶³ D'autres initiatives vont prochainement compléter cet écosystème telles que l'European Science Cloud destiné à "gérer analyser les données issues de sources distinctes" (<http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>)

⁶⁴ Entretien avec Kevin Stranack

⁶⁵ <https://leo.hypotheses.org/12612>

⁶⁶ <https://guides.github.com/activities/citable-code/>

La gouvernance : l'inspiration des communs ?

Le rachat tout récent de l'archive ouverte *SSRN* par *Elsevier*, intervenu alors que l'écriture de ce rapport touchait à sa fin, a révélé une faiblesse intrinsèque du mouvement du libre accès : l'utilisation des licences libres empêche l'appropriation des textes elle ne limite en rien l'appropriation des structures d'éditorialisation et de dissémination. La question qui se pose aujourd'hui n'est plus seulement celle de l'accès mais des modalités de l'accès et du contrôle des voies de communication. Les réseaux sociaux académiques ont posé la base initiale d'une « enclosure » à grande échelle des productions scientifiques en accès ouvert. Sur *ResearchGate*, *Academia* ou *MyScienceWork* il est nécessaire de se connecter pour consulter des articles et donc, tacitement, de conclure cet engagement commercial : céder des données personnelles contre l'accès à la connaissance. Or, la valeur et la précision des données s'accroissent avec la multiplication des sources de captation. *SSRN* ou *ResearchGate* peuvent en tirer beaucoup moins d'enseignement en restant indépendant qu'en s'incorporant dans un conglomérat de la taille d'*Elsevier*.

Dès lors, **la gouvernance devient déterminante**. Dans un contexte fortement évolutif où les innovations ne sont pas encore stabilisées, la pérennité d'une structure éditoriale se mesure d'abord à sa capacité à mettre en place des processus de prise de décision adaptés et réactifs et à impliquer efficacement l'ensemble des parties prenantes. Ainsi, le système de financement d'*Open Library of Humanities* découle de son organisation interne : les souscriptions des 190 bibliothèques et institutions universitaires viennent matérialiser une direction collégiale ; moyennant le versement direct de subventions, les communautés académiques conservent la pleine maîtrise du processus éditorial.

Les formes de gouvernance sont bien moins documentées que les outils ou que les pratiques d'écritures. Le manuel de l'outil d'édition *DSpace* fait près de 1000 pages ; il n'existe pas de manuel des gouvernances possibles pour l'édition scientifique. Chaque institution doit improviser et souvent recréer indépendamment des usages et des méthodes qui existent par ailleurs. Ce déficit de documentation pénalise également les usages. Apparemment, de nombreux contributeurs de l'archive *SSRN* pensaient qu'il

s'agissait d'une association à but non commercial, *a priori* non rachetable par *Elsevier*.

Les nouvelles structures de l'édition scientifique pourraient ici s'inspirer de l'expérience acquise par les *communs* de la connaissance. Le développement du numérique a autorisé l'émergence de très grandes communautés auto-gérées. *Wikipédia*, *OpenStreetMap* ou *Debian* reposent sur les contributions de plusieurs centaines voire plusieurs milliers d'utilisateurs qui décident en permanence du cadre et des normes régissant leur participation⁶⁷. Ces *communs* ne constituent pas des entités absolument nouvelles⁶⁸. L'édition scientifique en général s'apparente à un *commun* hybride (ou *club*), tant cette activité paraît dominée par des *externalités* qui échappent aux grilles d'analyses classiques du marché ou de l'État⁶⁹.

Le passage sur le web a entraîné un saut quantitatif et qualitatif. Tout en reposant sur un système d'autogestion, l'encyclopédie *Wikipédia* atteint aujourd'hui une ampleur comparable à une industrie culturelle (ou, s'agissant d'édition scientifique, à *Elsevier*) : toutes versions linguistiques confondues, près de cent mille contributeurs entretiennent en permanence plusieurs dizaines de millions d'articles. L'automatisation des tâches a joué un rôle déterminant dans l'avènement de ces communs *industrialisés* : la communauté wikipédienne peut déléguer l'application des normes à des *bots* (qui assurent notamment le repérage des modifications inappropriées) et, ainsi, libérer du temps pour participer à des formes de gouvernances collaboratives. L'infrastructure technique émane directement des besoins et des usages de la communauté, au point que nous pourrions parler de *collaborative work by design*.

Ce précédent montre que d'autres modèles sont possibles. **Le rôle joué par *Elsevier* pourrait être assumé par de grandes plateformes auto-gérées** associant toutes les parties prenantes de l'édition scientifique : chercheurs, bibliothécaires, concepteurs d'outil,

⁶⁷ Même s'il ne s'agit pas véritablement d'un manuel, la peer to peer foundation référence de nombreuses pratiques effectives des communautés en ligne : <http://p2pfoundation.net/Category:Peergovernance>

⁶⁸ Dans les travaux de la prix Nobel d'économie Elinor Ostrom les communs désignaient ainsi principalement des sociétés traditionnelles organisées selon d'autres modalités que le marché ou l'État.

⁶⁹ Cette proximité a été suggérée par une étude stimulante de Neyron et al., « Principles for Open Scholarly Infrastructures », *figshare*, février 2015 https://figshare.com/articles/Principles_for_Open_Scholarly_Infrastructures_v1/1314859

voire, dans une démarche de science ouverte, participants extérieurs. Plus que le support technique, ce qui fait aujourd'hui défaut, c'est une dynamique collective internationale : les efforts d'innovation sont trop souvent dispersés sur des initiatives nationales et/ou manquent de moyens pour se développer pleinement.

Recommandations

Alors que le développement du libre accès à court terme paraît acquis, l'élaboration d'infrastructures adaptées devient un enjeu fondamental. Cela présuppose non seulement un travail technique et éditorial (en terme de « connexion » des productions scientifiques) mais aussi un travail social et « politique » : les formes de gouvernance sont déterminantes dans la conduite du changement à long terme et la construction d'un écosystème ouvert et pérennisé. Les trois recommandations suivantes s'inscrivent dans cette perspective :

- **Donner à voir la circulation générale des textes au sein de l'écosystème de l'édition scientifique** : Cette perspective est déjà partiellement mise en œuvre : l'architecture de *HAL* est ainsi optimisée pour disséminer les informations et les métadonnées. A terme, l'enjeu serait de développer des points d'entrées universels pour les différentes incarnations prises par une contribution scientifique et les différentes formes dérivées qui y font référence.
- **Documenter et certifier les formes de gouvernance** : des enquêtes sur le sujet ont été initiées au Canada et en Europe mais le travail à accomplir reste considérable. Cette action peut se concrétiser par le développement de supports documentaires (manuels, retours d'expérience..) et de dispositifs de certification normalisés (à l'image du travail mené par un projet comme *Move Commons*⁷⁰).
- **Encourager une dynamique fédérative internationale**. Actuellement il n'existe que quatre méta-portails universels d'accès aux productions scientifiques : *Web of Science*, *Scopus*, *Google Scholar* et la bibliothèque pirate *Sci-Hub*. La nature fondamentalement globale de la recherche scientifique appelle une réponse

⁷⁰ <http://movecommons.org/fr/preview/>

globale : le rapprochement des principaux acteurs de l'édition ouverte au sens large (éditeurs, archives, institutions, bibliothèques, communautés...) et le développement d'infrastructures communes sous l'égide d'une gouvernance collégiale et collaborative.

Conclusion

Les données recueillies et présentées dans le cadre de ce rapport permettent de répondre à la commande initiale : « quelles formes éditoriales l'État peut-il encourager à l'heure du numérique, de la mutation de l'édition scientifique et de la faillite de l'évaluation scientifique ? » Sauf revirement majeur, la généralisation du libre accès devrait devenir effective à court terme : l'introduction de lois facilitant le dépôt en archive ouverte et, tout dernièrement, les déclarations volontaristes du Conseil européen vont en ce sens.

Les quatre premières parties ont tenté de faire le point sur les évolutions actuelles des principaux aspects du travail d'édition scientifique. Cette dynamique générale de changement, enclenchée par la double conversion des publications universitaires au numérique et au libre accès, ouvre des opportunités nouvelles. Pour chacune de ces dimensions nous avons formulé des recommandations d'actions à engager pour faciliter l'avènement l'émergence de structures innovantes, adaptées aux besoins des communautés scientifiques :

- **Outils d'édition** : le développement de systèmes normalisés contribue à optimiser le processus de publication et à diversifier les modèles éditoriaux. **Les logiciels libres, comme *Open Journal System*, sont propice au déploiement de rétroactions régulières entre les usages et les dispositifs** (via la conception d'extensions *ad hoc*). Pour assurer le plein développement de ces outils non-commerciaux, nous recommandons d'initier une politique active de soutien public direct ou indirect.
- **Formes d'écriture** : sur le web, **l'écriture scientifique se réduit de moins en moins au texte rédigé**. Données, codes et formes hybrides (tels que le carnet de code) occupent une place croissante que traduit le développement d'espaces de

publication et de dissémination dédiés et de marques de légitimations (attribution d'un DOI). Les normes en vigueur restent encore en retrait de cette transformation récente. Nous incitons les administrations de la recherche publique à soutenir d'autres types de publication que l'article et l'État à poursuivre l'élaboration d'un cadre légal adapté.

- **Évaluation** : longtemps cantonnés à des expériences isolées, **les modèles d'évaluation ouverte se disséminent rapidement depuis quelques années**. Ces formes alternatives ne correspondent pas à un *contre-modèle* uniforme mais à un ensemble diversifié qui, selon l'ajustement des principaux paramètres (auteur, dispositif, circulation...) produit des résultats plus ou moins adaptés à telle ou telle situation. Sous l'effet de cette diversification, nous invitons les institutions scientifiques à repenser l'évaluation de l'évaluation et à introduire des dispositifs éditoriaux adaptés pour signaler les différentes formes d'évaluation.
- **Économie** : **le transfert « brut » des abonnements en APCs n'est une issue ni souhaitable ni inéluctable**. Nous appelons les administrations publiques à se saisir des opportunités ouvertes par la mutation générale des modèles économiques pour transformer les dépenses en investissements. La préemption d'une petite partie des dépenses engagées (le « 1% innovant ») et la mutualisation des ressources permettent à terme d'effectuer des économies structurelles et durables.

Au fil de ce développement, nous avons effectué plusieurs constats récurrents qui excèdent le strict cadre des « angles » choisis. La cinquième partie décrit au sens large le passage d'une politique d'innovation à une politique d'infrastructure. Tout en diversifiant, les modèles existants s'inscrivent dans une dynamique d'écosystème : les relations entre les acteurs importent autant que ce qu'ils font. La mobilité des productions textuelles dans le cadre d'un libre accès élargi (intégrant aussi l'ouverture des évaluations, des données, du code...) facilite l'émergence de ces interrelations. Dans la mesure où elle élargit le champ des possibilités techniques, l'informatisation contribue à renforcer l'incidence des modèles organisationnels voire politiques de l'édition scientifique : **la gouvernance**

devient déterminante. L'émergence des *communs numériques* montre qu'il est possible de concevoir de très grands projets, comparables par leur ampleur aux principaux éditeurs scientifiques, tout en étant régis par des formes de gouvernance collaborative et participative. **Le déploiement d'une dynamique internationale associant la plupart des acteurs de l'édition ouverte s'impose aujourd'hui** : le développement de cet écosystème en formation et la prévention de sa captation ne peut vraisemblablement s'effectuer que dans un cadre global.