



LA PUBLICATION  
**SCIENTIFIQUE**  
AUJOURD'HUI:  
**FINANCEMENT,**  
USAGES NUMÉRIQUES

DIST Étude



[www.cnrs.fr](http://www.cnrs.fr)

N° 4 | JUN 2015  
Direction de l'information scientifique et technique



# La publication scientifique aujourd'hui : financement, usages numériques

DIST Etude 4/ juin 2015

**TITRE I**  
**Financer la publication scientifique ..... 4**

**TITRE II**  
**Le halo numérique de l'IST.....25**

# TITRE I

## **Financer la publication scientifique**

**Le "Lecteur » et/ou « l'Auteur » ?**  
Evolutions, Alternatives

Observations de la DIST

## Table des matières

|   |                                    |
|---|------------------------------------|
| <b>Préambule</b> .....  | <b>6</b>                           |
| Un système au milieu du gué .....   | 6                                  |
| <b>Résumé</b> .....   | <b>7</b>                           |
| Le Lecteur et / ou l'Auteur ?.....  | 7                                  |
| <b>Sécuriser le financement de publication scientifique</b> .....   | <b>9</b>                           |
| Un nouvel acteur de la dépense : « l'auteur » .....   | 9                                  |
| L'exposition de la recherche aux APC : une évaluation en cours .....  | 9                                  |
| <b>Le besoin d'une vision stratégique</b> .....   | <b>10</b>                          |
| Le flou : une asymétrie entre abonnements et APC.....   | 10                                 |
| Open Access avec APC : une moitié seulement des journaux OA .....   | 11                                 |
| Paiement des APC : une corrélation médiocre avec la qualité des journaux.....   | <b>Erreur ! Signet non défini.</b> |
| Le difficile consensus entre le lecteur et l'auteur : une corrélation inégale des deux marchés dans les grands accords nationaux..... | 13                                 |
| OCDE : les choix du Royaume Uni et les Pays Bas sont atypiques.....   | 13                                 |
| Une proposition trop rigide: les différences d'APC entre disciplines.....   | 14                                 |
| <b>Alternatives: quelles orientations proposer pour le travail avec Couperin ?</b> .....  | <b>15</b>                          |
| Internet et l'Open Access : paramètres d'évolution de l'édition scientifique numérique .....  | 15                                 |
| Le CNRS est placé en position d'initiative.....   | 17                                 |
| Associer pleinement Couperin aux initiatives .....  | 18                                 |
| <b>Propositions finales</b> .....   | <b>19</b>                          |
| Les solutions à proposer .....  | 19                                 |
| <b>Annexe : Une géopolitique des APC</b> .....  | <b>20</b>                          |
| Position des Etats Unis :.....  | <b>Erreur ! Signet non défini.</b> |
| La situation au Royaume Uni, aux Pays Bas .....   | 21                                 |
| Projets en Allemagne et place de Springer .....   | 22                                 |
| Green et Gold Open : la compétition.....  | 22                                 |

## Préambule

### Un système au milieu du gué

L'édition scientifique internationale est entrée dans une phase d'effervescence. Les prises de position des éditeurs et des communautés de chercheurs se multiplient sur les alternatives offertes aux établissements de recherche pour le financement de la publication scientifique: financement "par l'auteur" (APC) pour certains (anglais, néerlandais, amorce de position allemande), pour d'autres une position conciliant le financement "par le lecteur" (abonnements) et par l'auteur (Etats Unis, Canada, OCDE).

Une clarification globale est à construire.

Sur le fond, la question peut s'énoncer comme suit. **A l'heure numérique, le système de financement de la publication est au milieu du gué**: on ne peut plus laisser financer le système par les seuls abonnements, par le choix des seuls bibliothécaires (paiement par les "lecteurs" - puisque le financement de la publication déborde largement de ce périmètre - mais on ne peut pas d'avantage s'en tenir au financement par les auteurs, par les seuls choix de paiement d'APC par les laboratoires, *puisque ce choix est loin de contenir toute l'alternative aux abonnements*, tout en étant très inégalement réparti entre disciplines et entre pays.

**S'il peut exister une équivalence comptable globale entre ces deux solutions (ce qui reste encore à démontrer au regard du poids des financements externes), celle-ci ne repose ni sur les mêmes critères, ni sur les mêmes circuits de décision, ni sur les choix d'un décideur unique.**

Admettre par hypothèse que l'un équivaut à l'autre sans réflexion préalable sur de nouveaux principes combinant les choix à venir des bibliothèques (financement par les lecteurs) et des communautés de chercheurs (financement par les auteurs là où celui-ci se réalise), c'est faire le jeu des seuls acteurs qui ont intérêt au fragile statu quo actuel: les financiers des grands éditeurs.

Prendre de la hauteur et susciter une discussion globale peut se concevoir en s'appuyant sur certains des principes affirmés dans H2020, et partagés dans le cadre de Science Europe et de l'ERC, **pour appréhender dans sa globalité le service rendu par l'édition scientifique** (financement de l'édition scientifique internationale, service "d'Open process" correspondant aux APC, conditions de la révision par les pairs à l'heure numérique, text mining, propriété intellectuelle...), en réunissant tous les acteurs (autorités publiques, éditeurs, chercheurs...): c'est ce que vient de recommander l'OCDE, dans le sens de points de vue récents de l'ERC.

## Résumé

### Le Lecteur et / ou l'Auteur ?

Comme d'autres négociations nationales qui viennent de s'achever (Pays Bas, Royaume Uni), la négociation nationale proposée par Couperin (négociation Springer) vise à **sécuriser l'évolution de la dépense payée aux grands éditeurs de sciences**, et à la stabiliser. Cette dépense menace actuellement de s'accroître rapidement: d'un côté, en effet, les tarifs d'abonnement (paiement par les lecteurs) se sont accrus fortement, de l'autre, les éditeurs ont mis en place des APC (Article Processing Charges) permettant aux chercheurs la libre disposition immédiate de leurs articles (paiement par les "auteurs") sans s'exposer aux délais "d'embargo" actuellement imposés par les grands éditeurs (de 6 à 18 mois).

En négociant en France une corrélation entre ces deux systèmes de dépenses, paiements par les lecteurs et paiements par les auteurs, en vue de leur plafonnement global, le risque existe de **figer un système inégalement opérant** de gestion bibliothécaire, assez différent de celui des pays voisins (les systèmes anglais et néerlandais étant très particuliers et spécifiques): **les attentes des "lecteurs" et les contraintes des "auteurs" ne coïncident en effet qu'inégalement d'une discipline et / ou d'un pays à l'autre, cependant que l'addition de solutions nationales séparées contient le risque d'une balkanisation de l'IST....**

Pour faire face à l'évolution du modèle technico-économique de l'édition scientifique, qui est rapide, forte, diversifiée vers les services d'analyse numérique, comme l'a récemment montré la DIST<sup>1</sup>, il est essentiel, à la fois de **maitriser l'évolution des APC** dans les stratégies engagées par les communautés de recherche concernées (le CNRS participe aux premières grandes actions internationales dans ce sens, comme SCOAP3 en physique des particules), mais également **d'éviter de lier prématurément ces actions spécifiques à une approche globale des relations avec les éditeurs, et donc avec Springer.**

**L'incertitude subsiste sur la solidité et la justification économique du modèle global** : les études disponibles récentes (2014 et 2015) montrent par ailleurs la grande fragilité des corrélations entre les niveaux actuels d'APC pratiquées par les grands éditeurs, et la « qualité » des publications et articles qui y correspondent. Plus généralement, les « modèles » sont le plus souvent opaques. Peter Gruss, Président de la Société Max Planck, ne dit rien d'autre quand il déclare : « *Instead of the black box of subscription price calculations, we need models based on actual publication costs and characterized by transparency and sustainability* », en finançant de véritables modèles de publication en Open Access.  
(10<sup>e</sup> anniversaire de la Déclaration de Berlin sur l'Open Access)

<sup>1</sup> DIST Etudes, Avril 2015: "L'Edition de sciences à l'heure numérique: dynamiques en cours (2015)"  
<http://www.cnrs.fr/dist/z-outils/documents/Distinfo2/DISTetude2%20%282%29.pdf>

Une stratégie globale, dont Couperin souligne à juste titre qu'elle est indispensable et urgente, doit reposer avant tout sur une **concertation nationale, européenne et internationale entre éditeurs et entités de science publique**, dans les enceintes où cette concertation est en cours (Science Europe, European Research Council (ERC), H2020, mais également au Global Research Council).

**L'objectif à atteindre est celui d'une sécurisation globale sur tous les paramètres d'évolution vers la science ouverte.** Faute de cette sécurisation globale, la juxtaposition des "négociations" nationales, dont les contenus et les résultats ne sont aujourd'hui pas révélés, place les éditeurs en situation d'arbitres de la circulation de l'IST numérique. **Cette situation contient le risque, comme l'observe l'OCDE, de chacun pour soi, de la confusion et du morcellement des collaborations scientifiques internationales, sous le jeu d'intérêts éditoriaux étrangers au partage des résultats de la recherche publique.**



## Sécuriser le financement de publication scientifique

Une hypothèse stratégique répandue est que **les dépenses des lecteurs et les dépenses des auteurs sont substituables et que la transition vers l'Open Access conduira à remplacer la dépense du "lecteur" par celle des "auteurs"**, faute de quoi la dépense des "auteurs" connaîtrait une croissance incontrôlée et menacerait gravement l'accès des disciplines à la publication et / ou, simultanément, les capacités des établissements à faire face aux coûts des APC.

### Un nouvel acteur de la dépense : « l'auteur »

Le débat sur le libre-accès se focalise aujourd'hui de façon insistante sur l'économie des APC. Les APC sont le fondement du modèle économique dit de « Gold OA » et sont exigées des éditeurs pour publier en libre accès, soit dans des revues qui sont en OA « natif », soit dans les revues sur abonnement. Ces dernières permettent toutes désormais au chercheur publiant de voir, moyennant le paiement d'APC élevés et très différenciés (de 1000 à 10 000 euros), son article immédiatement accessible en libre accès sur le site de l'éditeur et / ou une archive ouverte. Ce modèle de publication en OA dans des revues sur abonnements est dénommé « modèle OA hybride ».

### L'exposition de la recherche aux APC : une évaluation en cours

**Le financement général par APC est difficilement envisageable pour le CNRS et pour la recherche en général.** A titre d'exemple, la « généralisation » de l'Open Access Gold aurait pour le CNRS (et plus largement pour les organismes de recherche publics français) des coûts peu soutenables.

Selon les décomptes actuels de l'INIST, les achats de ressources documentaires du CNRS sont aujourd'hui de l'ordre de 15 M€ par an. Par ailleurs, les chercheurs du CNRS publient annuellement en revues un nombre d'articles de l'ordre de 43 000 unités. Si l'on fait l'hypothèse extrême qu'à terme tous ces articles soient publiés en OA sur la base d'un montant d'APC de 2200 € par article (moyenne constatée chez Springer) le coût du Gold Access généralisé supporté par le CNRS serait de 94,6 M€, *soit 6 fois plus que les budgets d'abonnements actuels.*

Si on fait l'hypothèse plus réaliste que seuls un tiers des chercheurs co-publiants du CNRS sont les « auteurs de correspondance » et que seuls ceux-ci sont amenés à s'acquitter des APC, le budget du CNRS serait appelé à se situer autour de 31,5 M€, soit le **double** du budget d'achat de ressources documentaires actuel. **Dans le cadre de sa stratégie IST, la DIST mène actuellement des simulations sur les budgets présents et prévisionnels d'APC : ces résultats détaillés seront communiqués au DGDS avant fin juin.**

Ces premiers chiffres ci-dessus permettent d'approcher les évolutions de moyen terme sur lesquelles un positionnement solide doit être trouvé: premier organisme publiant mondial, le CNRS deviendrait, dans un modèle de Gold Open Access « pur », l'un des tout premiers – voire le premier – contributeurs au financement de la publication scientifique, puisque le Gold OA fait peser sur les chercheurs et leurs

organismes de rattachement les coûts de la mise à disposition en libre accès des résultats de la recherche.

## Le besoin d'une vision stratégique

Avec un complément d'expertise rassemblé à la DIST (direction, INIST, experts extérieurs, instituts...), il apparaît nécessaire de compléter l'appréciation stratégique fournie par Couperin : **cette appréciation paraît en effet devoir être affinée dans les négociations en cours** comme de toute autre négociation du même type, en l'état actuel de l'information et de l'organisation de l'ESR français. Une troisième partie proposera de nouvelles pistes de travail, y compris dans le cadre de Couperin.

La DIST utilisera ci-après les études et articles de référence dans la littérature analytique spécialisée internationale récente (centres de recherche, OCDE, cabinets conseils internationaux...).

### Le flou : une asymétrie entre abonnements et APC

**Si leur développement est aujourd'hui rapide, les APC sont encore loin de faire jeu égal avec les abonnements : négocier l'un et l'autre en corrélation, peut difficilement provoquer une relation de compensation globale et symétrique.**

Le marché de l'édition scientifique de recherche dans son ensemble (y compris l'Open Access) est lui évalué à **12,8 Md.\$** et enregistre une croissance annuelle de 1 à 2%. L'Open Access Gold (marché des APC) ne représente donc aujourd'hui que **2,3%** du marché de la publication scientifique. Même en tenant compte des dynamiques d'évolution différenciées de ces deux segments de marché, **les projections montrent qu'à un horizon de 10 ans (2024) les APC ne représenteraient que 26,5 du marché de l'édition de recherche.**

Malgré les stratégies des éditeurs dans le développement d'offres en Open Access (lire plus loin) l'abonnement restera bien une modalité d'accès importante aux publications scientifiques. Et les budgets d'APC consentis par les organismes de recherche, dont le CNRS, viendront inévitablement s'ajouter à des budgets d'abonnement élevés qu'il faudra maintenir.

Le marché des APC 2014 a été estimé à 300 M\$ mondialement par le cabinet d'études spécialisé américain *Simba Information*<sup>2</sup> et afficherait dans les années à venir une croissance de **30%** par an.

**En toute hypothèse, le modèle de l'abonnement persistera durablement, même dans un contexte de rapide progression de l'OA. Dans ce contexte, le « double dipping » (paiement par les organismes de recherche des abonnements et de volumes d'APC), comme l'observe très justement Couperin, est un risque avéré et aujourd'hui encore difficile à maîtriser.** La DIST travaille sur de nouvelles normes budgétaires et comptables pour clarifier la dépense.

<sup>2</sup> <http://www.simbainformation.com/redirect.asp?progid=86599&productid=8346683>

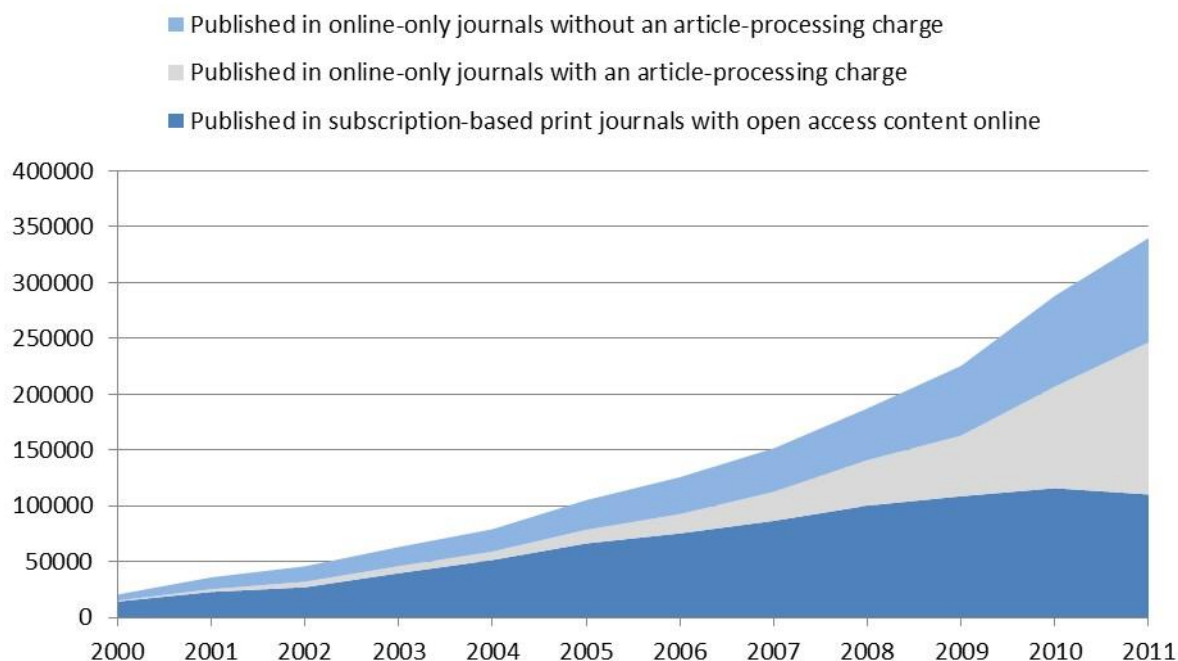
L'idée du couplage entre « négociation abonnements » et « négociation APC » (les crédits des uns se reportant sur les crédits des autres) repose sur le postulat selon lequel il y a pour un éditeur donné, une homothétie entre volume des consommations de ressources documentaires et volume d'APC. Or ce raisonnement est fragile: tel établissement de recherche français s'acquitte, auprès de Springer ou d'Elsevier, de frais d'abonnements de plusieurs dizaines de milliers d'euros par an alors que ses chercheurs n'ont publié en 2014 qu'un très petit nombre d'articles dans une revue Springer.

### Open Access avec APC : une moitié seulement des journaux OA

Considérer les APC comme étant la condition d'un Open Access de qualité serait abusif : **les APC entrent pour moitié seulement dans les dynamiques de la publication en OA**, telles qu'elles ont été récemment mesurées par l'étude globale de référence<sup>3</sup>, en répertoriant toutes les publications d'articles figurant au répertoire global des journaux scientifiques publiés en OA (DOAJ).

**Tableau 1 : Développement de la publication en OA et en OA avec APC**

(7,372 journals listed in the DOAJ on January 1<sup>st</sup>, 2012)



Source: M. Laakso and B.-C. Björk (2012), Anatomy of open access publishing: a study of longitudinal development and internal structure, BMC Medicine 2012, 10:124 - cité par OCDE (2014) "Making Open science a reality" Final report 2015

Importante et vigoureuse, la part portée par l'édition commerciale présente un développement jugé encore aléatoire... S'il est désormais clair que l'OA devient le modèle dominant de la publication, l'OCDE

<sup>3</sup> Björk, Solomon. Helsinki, 2012 et 2014

observe que la part relative de chaque voie est loin d'avoir acquis sa stabilisation, et en l'état actuel « *Most of the major internal shifts in OA journal publishing have only happened fairly recently during the last few years and, judging by the momentum at which things are moving, it is hard to imagine the internal dynamics settling down any time soon* ». <sup>4</sup>

Se lier avec un éditeur dans un tel contexte prend le risque de créer des asymétries entre opportunités de publication pour les chercheurs. Couperin ne s'y trompe d'ailleurs pas en soulignant que toute négociation avec un éditeur privé peut s'accompagner d'un risque, que Couperin écarte au moyen d'un argument qui peut paraître discutable. <sup>5</sup>

### **Paiement des APC : une corrélation médiocre avec la qualité des journaux**

**L'autre difficulté éditoriale majeure à laquelle est confronté le nouveau « marché » du financement par les auteurs, est l'inégale crédibilité commerciale des APC elles-mêmes, au regard des prix pratiqués et de la relation des prix à la « qualité » des publications concernées.** <sup>4</sup>

Une étude exhaustive transversale conduite sur l'ensemble des journaux en OA (DOAJ) et publiée en 2015, fait ressortir un niveau de corrélation inégal et moyen entre la qualité de la publication scientifique et le niveau des APC demandées par les éditeurs. <sup>6</sup> « *The price that publishers sell at to different universities and countries seems to vary a lot, and there is anecdotal evidence of a pattern following the GDP per capita, which more or less sets the budgets universities have for journal acquisitions* ». Les auteurs considèrent par ailleurs que la corrélation des APC à la qualité (mesurée par un indice citationnel robuste) est peu dynamique: « *The correlations between APCs and SNIPS are just slightly lower than what we found on the journal level for full OA journals Article processing charges* ». On observe par ailleurs prudemment que: « ***Preliminary evidence seems to suggest that article processing charges do not strongly correlate with journal impact, especially in the case of hybrid open access*** » (Björk and Solomon, 2014). D'autres études confirment ce dernier point (Romeu et al., 2014 – rapport du CERN).

On voit mal, si cette première grande étude était confirmée par des travaux ultérieurs, comment figer et entériner par la négociation un niveau d'APC donné aujourd'hui, sans avoir défini et stabilisé au niveau international **la nomenclature standard exigible** (nomenclature qui pourrait être qualifiée « **d'Open Process** ») en contrepartie de ces paiements...Des propositions seront formulées en ce sens en troisième partie de ce rapport.

---

<sup>4</sup> Mikael Laakso and Bo-Christer Björk : "Anatomy of open access publishing: a study of longitudinal development and internal structure" Hanken School of Economics, Helsinki, Finland, BMC Medicine 2012, <http://doi.org/10.1186/1741-7015-10-124>

<sup>5</sup> Note Couperin du 27/04: "Le risque de favoriser un éditeur au détriment des autres, en adoptant une politique qui lui donnerait un avantage aux yeux des chercheurs a été évoqué. Nous pensons que la démarche de publication des chercheurs est guidée quasi-uniquement par la qualité et la notoriété de la revue, et que le fait qu'une revue permette de publier en open access sans surcoût n'est pas un motif déterminant dans le choix de publier au sein d'une revue. La politique globale de négociation ne devrait donc pas affecter les logiques de publication des chercheurs."

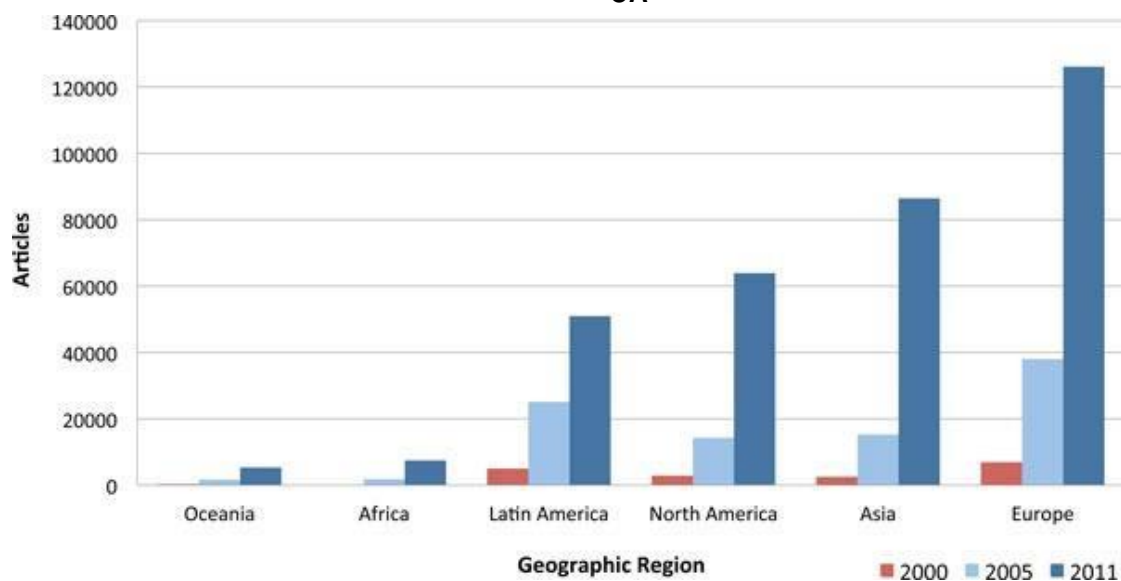
<sup>6</sup> Bo-Christer Björk, David Solomon: "Research on Open Access Publishing, Article processing charges in OA journals –relationship between price and quality" Published online in Scientometrics, March 2015, <http://doi.org/10.1007/s11192-015-1556-z>

### **Le difficile consensus entre le lecteur et l'auteur : une corrélation inégale des deux marchés dans les grands accords nationaux**

Une annexe au présent rapport contient un résumé des principales négociations et / ou positions internationales sur le financement de la publication scientifique. On pourra y constater, à la fois, combien les modèles nationaux diffèrent à l'heure actuelle et combien, en définitive, les négociations aujourd'hui abouties manquent toutes de vision partagée globale des déterminants et des conditions de la publication scientifique : tout est à faire pour asseoir les modèles technico-économiques de la publication scientifique à venir.

**Substituer coûts d'abonnements et APC a pour condition, dans l'hypothèse d'un plafonnement global, que ces financements soient corrélés et que toute modification de l'un soit proportionnelle à celle de l'autre.** Tel n'est pas le cas dans les grands accords internationaux, comme le montrent les données extraites de l'étude de référence précitée. On y voit clairement apparaître, dans le diagramme ci-dessous, l'étendue des différences entre les situations régionales en OA (unité : nombre d'articles dans revues OA par région d'origine des éditeurs ; on fait l'hypothèse que le OA Gold est ici proportionnel à l'OA, ce qui est vrai par ailleurs).

**Tableau 3 : Répartition régionale des pays d'origine des accords pratiquants les publications en OA**



Laakso and Björk *BMC Medicine* 2012 10:124 doi:10.1186/1741-7015-10-124

[Download authors' original image](#)

### **OCDE : les choix du Royaume Uni et les Pays Bas sont atypiques**

Ces différences, en particulier, sont nuisibles à la cohérence globale des choix nationaux dont la publication scientifique a besoin : une accumulation de barrières à la circulation des résultats scientifiques serait évidemment paradoxale et incompréhensible par les chercheurs de l'ère numérique.

Dans ce contexte, loin d'apparaître comme la voie d'avenir vers laquelle se dirigent en priorité les choix les plus éclairés, les accords récemment signés au Royaume Uni et aux Pays Bas portent la marque des situations nationales très particulières qu'ils traduisent, notamment avec la présence sur leur sol d'éditeurs influents de taille globale. L'OCDE, qu'on peut difficilement soupçonner de complaisance en la matière, émet clairement cet avis dans le rapport de 2015 précité :

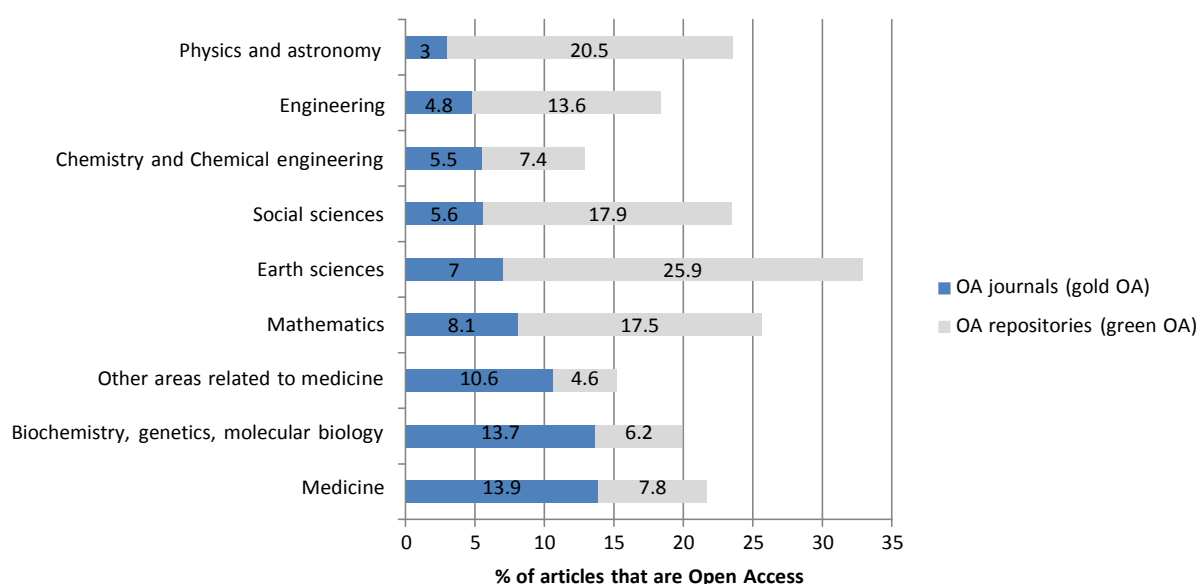
« In a number of OECD countries, major funding agencies have mandated public access to the results of the research they fund. In most countries, the requirement is limited to mandate gratis public or open access to publicly funded research results and let researchers free to choose the way they prefer to disclose research papers, that is whether to follow a gold route (via open access journals) or to self- archive the paper by means of online repositories (the green route).

**Exceptions are the UK and the Netherlands open access policy. In the UK, following the recommendations of the Finch report, the gold route has been preferred to the green one as a way to more effectively mandate and obtain open access to research results. In the Netherlands, the national strategy for open access has a preference for gold open access publishing, although it accepts green.**»

### Une proposition trop rigide: les différences d'APC entre disciplines

Ci-dessous, issu du rapport précité de l'OCDE de 2015, figurent les écarts des pratiques d'APC :

**Tableau 4 : Répartition des APC par disciplines (articles publiés dans les journaux OA - DOAJ)**



Source : OCDE 2015, rapport précité

On peut constater qu'outre les écarts considérables des pratiques, au premier niveau de lecture, il n'y a pas de corrélation apparente entre le recours à l'OA et la pratique des APC, si l'on en juge aux écarts respectifs de la physique, de la médecine, des sciences de la terre....

**On ne saurait faire le tour des constats disponibles, sans examiner les forts écarts de pratiques dans les financements par le lecteur et par l'auteur, qui prévalent entre les disciplines scientifiques.**

Cette dernière objection contient et résume toutes les autres, puisqu'elle conduit à critiquer une approche qui, en solidarissant les moyens globaux dus à un éditeur, néglige fondamentalement d'analyser les conséquences, immédiates et à terme, sur les dynamiques scientifiques sous-jacentes, dans lesquelles une négociation globale doit pourtant d'abord puiser sa raison d'être...

Or, les différences de pratiques des APC de disciplines tournées vers un même éditeur sont spectaculaires, et amènent ainsi à mesurer les risques pour la crédibilité d'une gouvernance scientifique qui seraient associés à la définition unilatérale-léonine, pourrait-on dire, des « droits de tirage » des disciplines associés à la fixation d'un plafond global d'APC, présenté « pour exécution » aux communautés scientifiques...

Aux limites, en effet, un tel accord ne pourrait prendre que deux formes également inacceptables : l'une qui serait le nivellement des APC, distribuées comme des fournitures, l'autre qui serait le statu quo, inacceptable de la même façon par les collègues exprimant de nouveaux besoins et brocardant une logique d'abonnement dévoyée et exacerbée...

Des réserves de ce type ont été d'ores et déjà émises par les professionnels des organismes. **Dans une note en date du 10 avril soumise à l'Assemblée générale de Couperin, 11 membres sur 12 de l'association EPRIST (regroupant des responsables IST de la plupart des organismes publics de recherche) ont exprimé leur scepticisme sur l'éventualité, l'intérêt et la faisabilité d'un couplage a priori entre les financements par le lecteur et par l'auteur.** Cette option, en l'état actuel de l'ESR français, soulève à leurs yeux de nombreuses questions concernant les bases factuelles exactes sur lesquelles engager la négociation ou encore sur les implications négatives d'un tel couplage sur le libre jeu des relations entre chercheurs et éditeurs dans le contexte national et international.

**La répartition des APC doit reposer sur une mise à plat du système, sur des choix dictés par une analyse commune et par une compréhension globale.**

### **Alternatives: quelles orientations proposer pour le travail avec Couperin ?**

Bien qu'avec d'autres réponses, les questions posées par Couperin sont en phase avec le programme de travail du CNRS : c'est l'occasion de proposer, alternativement, des actions nouvelles avec Couperin, dans le cadre international où le CNRS est déjà positionné par sa stratégie IST.

### **Internet et l'Open Access : paramètres d'évolution de l'édition scientifique numérique**

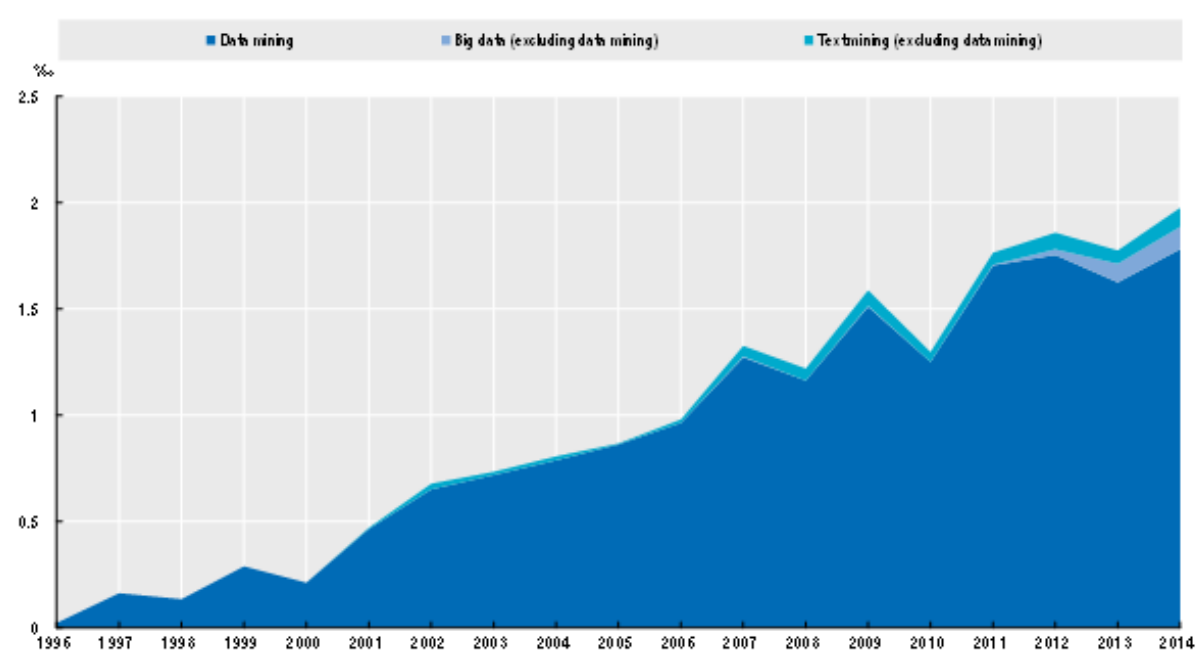
Deux paramètres gouvernent le développement international du partage des connaissances : Internet et l'Open Access.

L'OCDE vient de publier une première image, impressionnante, du développement de l'exploration automatique des corpus de littérature scientifique à toutes les fins de valorisation et / ou de développement de la recherche.

Ces données sont retracées par le diagramme ci-dessous :

**Tableau 5 : Les articles accessibles par Text and Data Mining (TDM)**

Unité : 000 articles



Source: OECD (2014), Measuring the Digital Economy, A New Perspective, OECD publishing

Sources : J. Clark (2013), Text Mining and Scholarly Publishing, Publishing Research Consortium, 2013; European Commission (2014), Standardisation in the area of innovation and technological development, notably in the Field of Text and Data Mining, Report from the Expert Group; Sergey Filippov (2014), Mapping Tech and Data Mining in Academic and Research Communities in Europe, the Lisbon Council, Issue 16/2014; OECD (2015) Inquiries into Intellectual Property's Economic Impact, OECD Paris forthcoming;;

L'incertitude demeure sur les voies et moyens d'un accord, par ailleurs essentiel, entre toutes les parties prenantes. La voie vers l'OA, direction rendue irréversible largement sous l'effet d'Internet, peut être parcourue de multiples façons, comme l'observe l'OCDE, dans son rapport précité de 2015 :

*"As in many other markets where the Internet has thoroughly rewritten the rules of the game, an interesting question is if new entrants, like Public Library of Science and BioMed Central, will take over the market or if the old established actors, commercial and society publishers with subscription-based revenue models, will be able to adapt their business models and regain the ground they have so far lost. Future studies on the internal structure of OA publishing are likely to witness the anatomy transforming yet again. **Most of the major internal shifts in OA journal publishing have only happened fairly***



*recently during the last few years and, judging by the momentum at which things are moving, it is hard to imagine the internal dynamics settling down any time soon”.*

### Le CNRS est placé en position d’initiative

Les débats en cours permettent aux organismes d’associer Couperin aux actions à développer.

Sur les thèmes de la proposition récente de Couperin, le CNRS soutient des positions, conformes à l’analyse ci-dessus, d’abord dans le cadre de **Science Europe**.

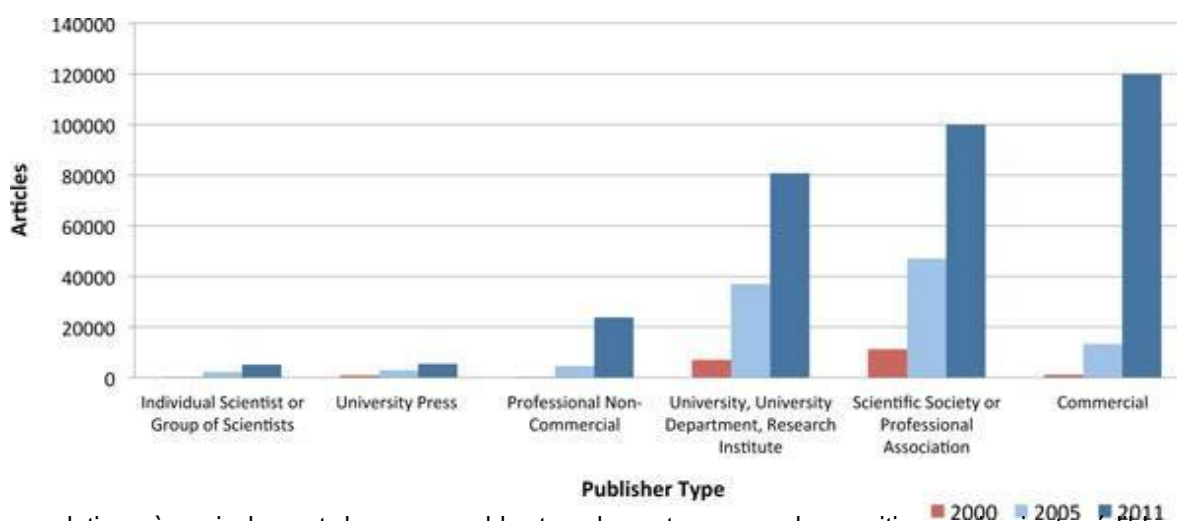
Deux des principes énoncés dans le Position Statement on Open Access (2013), qui engage chacun des membres, sont reproduits ci-après:

- (...) funding of Open Access publication fees is part of a transparent cost structure, incorporating a clear picture of publishers’ service costs;
- (...) stress that the hybrid model, as currently defined and implemented by publishers, is not a working and viable pathway to Open Access.”

Cette position est d’autant plus nécessaire que, sur la voie de l’Open Access, se situent des intervenants nombreux, divers, tous importants, indépendamment des éditeurs commerciaux :

**Tableau 6 : Volume total des articles produits en OA par type d’éditeurs**

Source: Björk and Solomon, op.cit. 2014



Les solutions à venir devront donc rassembler tous les acteurs, sur des positions qui soient crédibles et acceptables par tous. Ce consensus est d’autant plus nécessaire que, comme l’observe également l’OCDE, la concurrence des voies d’accès à l’OA doit être régulée et ne pas faire problème pour les chercheurs.<sup>7</sup>

<sup>7</sup> “Lewis (2012) suggest that gold OA (i.e. when the authors publishes on scientific journals openly available online, commonly referred to as open access journals) could account for 50% of the scholarly journal articles sometime between 2012 and 2017 and 90% of all articles between 2020 and 2025. However, Miguel et al. (2011) show that the percentage of green road journals widely surpasses the percentage of gold road publications. In addition, green open access (i.e. when the author self-archive the article on an online repository) was recently argued to be the most effective and affordable means for funders, institutions and other stakeholders (Houghton and Swan 2013).”

### **Associer pleinement Couperin aux initiatives**

Une stratégie globale, dont Couperin souligne à juste titre qu'elle est indispensable et urgente, doit reposer avant tout sur une **concertation nationale, européenne et internationale entre éditeurs et entités de science publique**, dans les enceintes où cette concertation est en cours (Science Europe, European Research Council (ERC), H2020, mais également au Global Research Council).

**L'objectif à atteindre est celui d'une sécurisation globale sur tous les paramètres d'évolution vers la science ouverte.**

Faute de cette sécurisation globale, la juxtaposition des "négociations" nationales, dont les contenus et les résultats ne sont aujourd'hui pas révélés, place les éditeurs en situation d'arbitres de la circulation de l'IST numérique.

**Cette situation contient le risque, comme l'observe l'OCDE, du chacun pour soi, de la confusion et du morcellement des collaborations scientifiques internationales, sous le jeu d'intérêts éditoriaux étrangers au partage des résultats de la recherche publique.**

Ci-après figurent les conditions que pose le rapport précité de l'OCDE, publié en version provisoire :

#### **Programme de travail de l'OCDE pour l'Open science (2015)**

**Open science is a means and not an end.** Open science strategies and policies are a means to support better quality science, increased collaboration and engagement between research and society that lead to higher social and economic impacts of public research.

**Open science is more than open access to publications or data; it includes many aspects and stages of research processes.** Although this report focuses primarily on open access to publications and research data, it is important to remember that open science is a broader concept that also includes the inter-operability of scientific infrastructure, open and shared research methodologies (such as open applications and informatics code), and machine-friendly tools allowing, for example, text and data mining.

**Clear legal frameworks vis-à-vis the sharing of publications and the re-use of datasets are needed at the national and international level.** A lack of clarity on the interpretation of national and international legal frameworks may prevent the sharing or re-use of research results. In addition, clear guidelines around text and data mining are needed as this tool will become increasingly used by researchers in the future. Some OECD countries are currently discussing or have recently modified national legal frameworks to make them increasingly open-science friendly.

**International collaboration in the area of open science is necessary to address global challenges.** International collaboration is becoming more important than ever, as publications and data in electronic form travel across national frontiers. Shared and inter-operable infrastructure is necessary to disseminate research results and promote scientific collaboration. It can avoid the duplication of efforts and help share the risks or the associated investments.

## Propositions finales

La question de l'économie des APC va, dans les mois et les années qui viennent, être le nœud du débat sur l'Open Access. Parce que le CNRS est le premier organisme de recherche mondial par le nombre de publications que signent ses chercheurs, l'impact financier d'un développement – voulu par les éditeurs – du modèle « hybride » d'Open Access pourrait être problématique, alors même que cette option renforcerait significativement les positions concurrentielles des très grands éditeurs « for profit » et le modèle hybride, au détriment des « pure players » de l'édition scientifique en Gold OA qui pourtant ont introduit une compétition nouvelle en proposant des niveaux d'APC « raisonnables ».

### Les solutions à proposer

- Un « position paper » à destination des Unités de recherche cadrant les éléments de ce débat et rappelant :
  - ✓ que **la voie verte (et en particulier le dépôt dans HAL) est la voie à privilégier pour partager au mieux et au moindre coût les connaissances,**
  - ✓ que l'option « Open choice » permettant à un chercheur de publier en OA dans une revue sur abonnement (modèle hybride) ne doit pas être privilégiée : si un article est accepté dans une revue sur abonnement (tout particulièrement dans les revues d'excellence) c'est le mécanisme de l'abonnement, qui a vocation à perdurer de nombreuses années, qui assurera, au frais de l'éditeur, les coûts de la publication. La publication du chercheur CNRS aura de toute façon vocation à basculer dans le libre accès par dépôt dans une archive ouverte au terme d'un embargo de 12 mois,
  - ✓ que les unités pourront prochainement intégrer dans leurs outils de gestion financière les nouveaux postes de nomenclature qui permettront **d'avoir une vision partagée et consolidée des volumes d'APC .**

**Un projet de guide pour les chercheurs est préparé dans ce sens.**

## Annexe : Une géopolitique des APC

### Position des Etats Unis :

En février 2013, l'OSTP américain (Office of Science & Technology Policy, un organisme chargé de conseiller l'exécutif sur ses choix de politique scientifique) a édicté une directive<sup>8</sup> ayant force de réglementation qui imposait à toutes les agences fédérales disposant d'un budget de R&D supérieur à 100 M\$ par an de définir dans un délai de six mois des politiques à même d'assurer un accès public et une conservation pérenne, dans des archives ouvertes, des outputs de la recherche publique.

Ce mandat de dépôt concernait aussi bien la publication scientifique elle-même que les jeux de données issus de recherches sur fonds publics. Ce dépôt en archive ouverte institutionnelle devait se faire le plus rapidement possible et au plus tard à l'issue d'un embargo de 12 mois s'il existait de ce point de vue des contraintes imposées par les éditeurs ayant publié l'article considéré en revue.

L'objectif implicite de l'OSTP était de disposer, sur le modèle de l'archive PubMed Central gérée par les NIH, d'une archive unique et centralisée permettant de consulter (« Public access »), mais sans forcément que ces matériaux soient en libre accès sous des licences Creative Commons, l'ensemble des connaissances scientifiques produites sur budget fédéral.

Les rivalités entre grandes agences ont eu raison de cette perspective d'« archive universelle ». Chaque agence a développé jusqu'en ce début 2015 sa propre politique d'archive ouverte.

Pour satisfaire aux obligations de dépôts fixées par l'OSTP, deux initiatives ont vu le jour :

- l'archive CHORUS (Clearinghouse for the Open Research of the United States)<sup>9</sup> a été créée dans une logique consortiale par l'ensemble des grands éditeurs scientifiques : ceux-ci proposent de verser directement dans cette archive les articles scientifiques qu'ils ont publié dans leurs revues, et ce à l'issue d'une éventuelle période d'embargo,
- l'archive SHARE (SHared Access Research Ecosystem)<sup>10</sup> a été lancée avec le même but par un consortium de grandes universités américaines.

A ce jour, et en raison de la lenteur des agences finançant des activités de R&D, ces archives n'ont pas vraiment décollé. Mais l'exemple de PubMed Central (qui affiche désormais un taux de 80% de dépôts des articles résultant de recherches financées par les NIH) montre que la montée en puissance des grandes archives ouvertes demande plusieurs années.

**Dans ce contexte la question des APC est remarquablement absente du débat américain sur le libre accès.** Il ne fait pas de doute que beaucoup de chercheurs publiant en Open Access payent des APC (mais probablement peu dans le cadre du modèle hybride). Pourtant l'accent mis par le gouvernement américain sur un développement d'un Green Open Access institutionnel a stérilisé le débat sur les APC.

<sup>8</sup> [https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

<sup>9</sup> <http://www.chorusaccess.org/>

<sup>10</sup> <http://www.share-research.org/>

## La situation au Royaume Uni, aux Pays Bas

**Au Royaume-Uni et aux Pays-Bas, les pouvoirs publics ont clairement choisi la voie du Gold Open Access. Cette option répond à des choix de politiques de soutien à une édition scientifique nationale qui ne peut être imitée sans précautions par la France**

Le gouvernement britannique a en avril 2013 arrêté une politique publique très claire préconisant le « Gold Open Access » comme axe principal de développement du libre accès aux publications scientifiques. Cette politique publique ayant force de loi s'était largement inspirée du « Rapport Finch » paru en juin 2012. Ce rapport soulignait entre autres que l'édition scientifique britannique était une industrie forte, créatrice d'emploi et générant d'importants surplus commerciaux. Le Gold Open Access, porté par les éditeurs scientifiques eux-mêmes, apparaissait comme le seul modèle économique permettant une transition vers le libre accès à la publication scientifique sans fragiliser cette industrie. Cette politique publique claire<sup>11</sup> favorable aux éditeurs défavorisait pourtant clairement la communauté de l'enseignement supérieur et de la recherche (ESR) britannique.

**En effet si les chercheurs britanniques représentent 4% des effectifs mondiaux de chercheurs dans la recherche publique, ils publient 6% des articles au plan mondial et 16% des articles les plus cités. Cette « surreprésentation » des chercheurs britanniques dans la publication scientifique globale se traduit mécaniquement, dans le modèle économique du « Gold Open Access » reposant sur les APC, par un « surcoût ». C'est pour cette raison que le gouvernement britannique a créé et doté de nouvelles lignes de crédit pour les agences britanniques de financement de la recherche spécifiquement destinées au paiement des APC.**

**Quatre des cinq plus importants éditeurs scientifiques mondiaux (Elsevier, Springer, Wolters Kluwer, Macmillan / Nature Publishing Group, Taylor & Francis) sont anglo-néerlandais ou allemands. Le 5<sup>ème</sup>, l'américain Wiley n'a pris son rang (à la troisième place) dans ce classement mondial qu'en rachetant il y a quelques années l'éditeur britannique Blackwell. Au-delà de ces très grands acteurs, tant la Grande-Bretagne que l'Allemagne dispose d'un tissu très serré d'éditeurs scientifiques de taille moyenne mais d'audience internationale. Selon les chiffres fournis par l'association STM ([www.stm-assoc.org](http://www.stm-assoc.org)), l'association internationale des éditeurs scientifiques, l'édition de recherche représente 110 000 emplois directs, dont 40% sont localisés en Europe (essentiellement dans les 3 pays cités) alors que le marché européen ne représente que 28% du marché mondial (en valeur) de l'édition scientifique. En conséquence dans ces trois pays européens les positions des pouvoirs publics relatives aux problématiques d'Open Access ne sont pas dictées par les seules questions de partage des connaissances au sein de la communauté de la recherche, mais aussi par des considérations (légitimes) de soutien à leur édition scientifique nationale. C'est explicite dans le rapport Finch mais cette clé d'analyse vaut aussi pour les Pays-Bas. L'édition scientifique française par contraste est d'un poids limité et risque de se faire laminer si l'essentiel des ressources s'orientent vers les grands éditeurs globaux – dont fait partie Springer.**

---

<sup>11</sup> La Grande-Bretagne reste l'un des rares pays à avoir arrêté une stratégie globale en matière de libre accès

## **Projets en Allemagne et place de Springer**

**L'Allemagne n'a pas défini de politique officiel en matière d'Open Access.** Cependant les responsables de la Max Planck Digital Library (MPDL) viennent de publier un « position paper »<sup>12</sup>, accessible à l'adresse <http://doi.org/10.17617/1.3>, (on ne sait pas si elle a reçu l'imprimatur de la Max-Planck Gesellschaft en tant que telle) dont les présupposés sont très simples voire simplistes : les auteurs constatent que les budgets publics dédiés aux abonnements dans toutes les économies développées permettraient in-abstracto de payer tous les APC de la production scientifique mondiale.

Il suffirait en conséquence, pour atteindre un objectif de libre accès généralisé, de programmer avec les éditeurs, dans le cadre des négociations consortiales, un transfert progressif des budgets des bibliothèques vers des budgets APC, les bibliothèques de recherche devenant la plaque tournante de cette logique de couplage et de transfert préprogrammé. Le couplage des négociations entre budgets d'abonnements deviendraient dans ce schéma la « raison d'être » des consortiums négociant conjointement ces deux postes budgétaires.

**Le poids important de l'édition scientifique et technique allemande d'audience internationale (le N°2 mondial Springer n'est que le plus importants de ces éditeurs allemands** derrière lequel figure une cohorte d'éditeurs moyens) pourrait faire pencher les pouvoirs publics allemands vers des positions similaires à celles du Royaume-Uni ou des Pays-Bas. C'est en ce sens que, même isolée, l'initiative de la MPDL est un signal de veille important.

## **Green et Gold Open : la compétition**

**Le recul dont on dispose sur l'expérience anglaise montre que le Gold Open Access est la stratégie la plus coûteuse pour atteindre l'objectif du libre accès.**

Opérationnelle depuis 2013, la politique publique britannique centrée sur le Gold Open Access vient de faire l'objet d'un premier rapport d'évaluation ([www.rcuk.ac.uk/research/openaccess/2014review/](http://www.rcuk.ac.uk/research/openaccess/2014review/)) qui souligne la complexité administrative qu'a généré pour les unités de recherche britanniques l'accès aux crédits APC et celle leur faisant obligation de documenter la conformité de leur politiques de publications aux exigences formulées par les pouvoirs publics. De son côté le RCUK et le Wellcome Trust, qui fédèrent à elles deux la grande majorité des crédits de recherche publique outre-Manche ont commandité un rapport<sup>13</sup> analysant les forces et les faiblesses des différents scénarios.

**Ce rapport dénonce sans ambages les effets négatifs d'un Gold Open Access Hybrid qui serait la règle.** Citons-le in-extenso sur ce point :

---

<sup>12</sup> Schimmer, R., Geschuhn, K.-K. & Vogler, A. (2015) "Disrupting the subscription journals' business model for the necessary large-scale transformation to open access", <http://doi.org/10.17617/1.3>

<sup>13</sup> L'étude "Developing an Effective Market for Open Access Article Processing Charges", Bo-Christer Björk and David Solomon, mars 2014, rapport remis au Wellcome Trust et au RCUK qui sont les principales agences de financement de la recherche, [http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy\\_communications/documents/web\\_document/wtp055910.pdf](http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp055910.pdf)

« La synthèse de la littérature scientifique et les données empiriques collectées pour la présente étude ont mis en évidence que le problème de l'économie des APC est très différent suivant que l'on considère les revues qui sont en Open Access intégral (« full Open Access ») où les revues en Open Access hybride, c'est-à-dire en pratique toutes les revues sur abonnement puisque les éditeurs proposent toujours pour celles-ci une option « Open Choice ».

Le marché des APC pour publication en « Full Open OA » permet un recul d'une dizaine d'année et peut être considéré comme un marché fonctionnant normalement puisque les éditeurs s'y livrent à une concurrence pour capter leurs « clients » (les chercheurs publiant) en jouant sur différents facteurs : qualité de la revue, services rendus aux auteurs, et niveau de prix des APC. Bien que certains chercheurs voient leurs APC entièrement payées et pourraient ne pas prendre en compte le niveau de ces APC pour choisir la revue dans laquelle ils publieront, diverses sources indiquent que le niveau des APC est un réel critère de choix pour les chercheurs et qu'en conséquence il y a là un facteur qui joue en faveur d'une limitation du prix des APC. La question stratégique essentielle pour les organismes qui financent la recherche est de s'assurer que le marché des APC reste compétitif et innovant. Les éditeurs traditionnels entrent aujourd'hui sur ce marché, accroissant la possibilité que les APC soient liés aux budgets d'abonnement, réduisant de facto la transparence et l'intensité de la compétition sur ce marché. Si les financeurs fournissent des crédits pré-fléchés pour régler les APC dans les revues hybrides cela conduira inmanquablement à une érosion de la pression concurrentielle sur les prix des APC. Un second risque induit est que le marché des APC serait structuré comme celui des abonnements, un nombre limité de très grands éditeurs s'adjudgeant un avantage compétitif massif aux travers de contrats de Big Deals qui incluraient des enveloppes prépayées d'APC. »

### **La voie Green reste la voie la moins coûteuse**

Dans un document récent<sup>14</sup>, David Price, vice-président d'University College London (UCL) souligne l'importance des coûts cachés de la voie Gold : « Il existe des données avérées qui suggèrent que le Gold Open Access reste très coûteux. Une modélisation économique de large ampleur proposée par un rapport<sup>15</sup> financé par Jisc Collections et publié par l'Open Access Implementation Group, suggère que c'est la voie verte, et non la voie dorée, qui est pour les universités de recherche la voie la moins onéreuse. Selon ce rapport le meilleur chemin pour passer d'un modèle économique de l'abonnement au libre accès est la voie verte, non la voie dorée. La plus importante archive institutionnelle britannique est celle de l'UCL, UCL Discovery, qui a généré plus de six millions de téléchargements depuis ses débuts. Un autre rapport montre que la voie dorée nécessite de la part des universités deux fois plus de temps et de coûts que la voie verte sans parler du paiement des APC. Pour la période 2013/2014 les pouvoirs publics ont fléchi 16,9 M.£ vers les APC mais ces coûts directs ont engendré pour

---

<sup>14</sup> David Price se réfère dans son intervention à deux rapports : A. Swan and J. Houghton, *Going for Gold?: The costs and benefits of Gold Open Access for UK research institutions: further economic modelling*, Report to the UK Open Access Implementation Group, June 2012, et le rapport de Research Consulting, *Counting the Costs of Open Access: The estimated cost to UK research organizations of achieving compliance with open access mandates in 2013/14*, November 2014

<sup>15</sup> <http://www.hepi.ac.uk/2015/03/30/open-access-national-licence-answer/>

*les universités des coûts indirects de 9,2 M.£ (...). Jisc Collections, au travers de négociations visant à coupler APC et coûts des abonnements, tente avec succès de contenir la progression globale de ces deux postes de coûts, mais d'une part tous les éditeurs ne se sont pas engagés dans cette voie, et d'autre part les effets de ce couplage sur les budgets d'achats de ressources documentaires ne se sont pas encore fait sentir, ceux-ci n'enregistrant aucune diminution. »*

(NDLR : Les universités britanniques dépensent collectivement 195 M.£, l'équivalent de 271 M€, en achat de ressources documentaires)

**Lectures supplémentaires :**

- Commission européenne: State-of-art analysis of OA strategies to scientific data
- Commission européenne: State-of-art analysis of OA strategies to peer-review publications
- Commission européenne: Comparative analysis of the strengths and weaknesses of existing open access strategies

*Rapports établis par Science-Metrix dans le cadre du contrat  
RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access*



## TITRE II

Méconduites, Lost science, Pertes en ligne...  
**Le "HALO NUMERIQUE" de l'IST**  
Evolutions en cours

Note de travail de la DIST

## Table des matières

|  |           |
|--|-----------|
| <b>Résumé : objectifs d'étude, résultats .....</b>   | <b>27</b> |
| <b>Les méconduites: des formes diverses, un niveau global faible et stable .....</b>               | <b>28</b> |
| Un niveau global faible et stable.....   | 28        |
| Partager l'information: vers une régulation des méconduites....                                    | 28        |
| <b>La "Lost science" : l'évitement du modèle dominant .....</b>                                    | <b>29</b> |
| Evolution de l'offre de publication : un dévoiement évité .....                                    | 29        |
| La "Lost science" et la réaction numérique au H factor .....                                       | 30        |
| Halo numérique et « Lost science » : des solutions émergentes.....                                 | 31        |
| Une métrique plus riche : un enjeu global d'analyse interdisciplinaire.....                        | 32        |
| <b>La perte en ligne d'IST : un halo numérique en surplomb.....</b>                                | <b>33</b> |
| Production des données et production d'articles: un écart croissant ? .....                        | 33        |
| Le halo numérique des données et l'Edition scientifique: la difficile recherche de solutions ..... | 34        |

## Résumé : objectifs d'étude, résultats

A l'heure, numérique, l'IST vit des évolutions mal connues, profondes et ambiguës: d'un côté, l'IST est fortement bénéficiaire des possibilités de diffusion d'Internet, de l'autre, l'IST est confrontée aux aléas qui affectent toutes les autres informations partagées....

Un "**halo numérique**," pourrait-on dire, est ainsi en voie de formation autour de l'IST: mieux en saisir les traits et l'ampleur, permet d'éviter l'alarmisme mais surtout d'anticiper les évolutions en cours du partage des connaissances, et de mieux s'assurer de l'**intégrité** de l'IST: c'est un enjeu global.

Trois directions identifiées par la DIST entreraient ainsi dans la formation du "halo numérique de l'IST, avec dans chaque cas des formes de réponses spécifiques:

### 1. **les méconduites**

Sous diverses formes, elles apparaissent peu "dynamiques". Avec un effet de loupe, les méconduites étant mieux enregistrées aujourd'hui, apparaissent davantage...

### 2. **la "lost science"**

Ce néologisme est depuis peu utilisé pour désigner les formes et conditions de production d'articles, contournant les pratiques dominantes de la publication numérique: ces pratiques sont donc fatalement vouées à de lourdes déperditions...

Vertueuses ou non, les pratiques de "lost science" remettent directement en question l'appréciation de l'article par le biais de la seule fréquence de citation. Ces pratiques de lost science ont leurs effets pervers. Pourtant, en réutilisant mieux ces résultats, une mesure plus systématique permettrait de mieux préciser le développement de l'interdisciplinarité et la qualité des collaborations locales. Une homogénéisation des normes internationale, prônée par l'ONU comme par l'ICSU, constituerait par ailleurs un premier progrès.

### 3. **la perte en ligne d'IST**

La stratégie IST observait un fossé entre d'une part, les volumes de production des données et, d'autre part, les volumes de publication d'articles d'IST. Selon les rares inventaires disponibles, les données inutilisées s'accumulent toujours plus, dans chacun des domaines correspondant de publication... Un "goulet d'étranglement" menacerait ainsi la publication; cet obstacle irait croissant avec la multiplication des capteurs de données.

En collaboration avec la DIST, des réponses émergent, a. pour répertorier et publier les données, b. pour orienter leur passage à la publication, et c. pour ouvrir à de nouvelles formes de publication.

## Les méconduites: des formes diverses, un niveau global faible et stable

### Un niveau global faible et stable

Les sources sont solides, hors SHS, avec le dernier rapport de l'Association STM<sup>16</sup> : sur les 28 100 journaux académiques publiés en langue anglaise en 2014, pour une production totale globale avoisinant les 2,5 millions d'articles, la croissance, 3% l'an actuellement, est en ligne avec celle du nombre des chercheurs. 95% de ces journaux assure la publication de résultats originaux de première main. La croissance de la publication, entre 2003 et 2013, a été de l'ordre de 5% l'an sur les bases WOS et SCOPUS, de 6 à 7% sur Pub Med Central (NSF 2014).

La méconduite ("misconduct" est le vocable en usage) se mesure à divers niveaux : la rétractation, selon un article de Nature en 2014 (eLife 2014) a concerné au cours des 2 dernières années 110 publications d'articles relevant de 6 comités de revue par les pairs, et ceci pour des motifs faisant inégalement relation à la fraude affichée. Certes, le nombre total des rétractations s'est significativement accru, mais, même au regard de cette croissance sur dix ans, la taille totale des méconduites avérées reste située autour de 0,02% (voir sur le détail le rapport cité en note1), dans un contexte où l'unification des systèmes d'information permet de fournir une information toujours plus précise et actuelle.

### Partager l'information: vers une régulation des méconduites....

La croissance observée de la rétractation, s'accompagne d'un partage systématique de l'information, auquel on peut prêter une vertu régulatrice. Créé en 2010, le blog collectif *Retraction Watch* ([www.retractionwatch.com](http://www.retractionwatch.com)) centralise et rend désormais publics, l'ensemble des faits de rétractation signalés par les éditeurs scientifiques.

La rétractation d'articles scientifiques reflète en réalité un éventail ouvert de formes de méconduites: manipulation des données présentées en appui d'une communication scientifique, plagiat, manipulation du process du peer-reviewing.

Les affaires les plus récentes, affectant tous les grands éditeurs<sup>17</sup> sont liées à la production frauduleuse de rapports de peer-reviewing<sup>i</sup>, avalisant une publication en revue. Ces pratiques ont exploité une double faille associée à une insécurité numérique: la possibilité offerte aux chercheurs publiant par de nombreuses revues de suggérer des noms de «reviewers» d'une part, d'autre part une gestion insuffisamment sécurisée des mots de passe dans les logiciels de support des tâches de peer-reviewing. Depuis ces affaires, *tous les éditeurs ont sécurisé leurs procédures pour éviter ce type de fraude, dont l'incidence est appelée à décliner rapidement.*

<sup>16</sup> *The STM Report An overview of scientific and scholarly journal publishing*, Fourth Edition, March 2015, mis en ligne en avril 2015 ([http://www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf)).

<sup>17</sup> <http://www.nature.com/news/publishing-the-peer-review-scam-1.16400>

Source de fraude souvent évoquée dans le passé, le plagiat est largement dissuadé. La conversion numérique des flux éditoriaux permet désormais à des outils logiciels dédiés de repérer les plagiats sur la base de vastes corpus de textes. Le consortium Crossref ([www.crossref.org](http://www.crossref.org)) qui fédère la plupart des éditeurs scientifiques au plan mondial<sup>18</sup> propose le service Crosscheck de détection de plagiat basée sur la technologie iThenticate.

Springer a annoncé en avril dernier l'intégration à son workflow éditorial de l'outil de détection de faux scientifiques SciDetect développé au Laboratoire d'Informatique de Grenoble<sup>19</sup>. Plus facilement décelable le plagiat donne donc lieu à plus de signalements, mais *la puissance des outils de signalement est de plus en plus dissuasive*.

## La "Lost science" : l'évitement du modèle dominant

La "lost science": ce néologisme est parfois utilisé pour désigner les formes et conditions de production d'articles, contournant les pratiques dominantes de la publication numérique et fatalement vouées à de lourdes déperditions...

Vertueuses ou non, les pratiques de "lost science" remettent directement en question l'appréciation de l'article sous l'éclairage de la fréquence de citation. **Ces pratiques de riposte ne vont pas elles-mêmes sans effets pervers : au cœur du halo numérique de l'IST, ces usages gagneraient à être mieux cernés.**

### Evolution de l'offre de publication : un dévoiement évité

La méconduite ne menace pas que les publiants : elle a pu également tenter de faire école chez les éditeurs....

Grâce à la publication numérique, les barrières à l'entrée dans le champ de l'édition scientifique en Open Access ont été considérablement abaissées. Ce qui a été mis à profit par de nombreux nouveaux entrants, dont certains (PLoS, Biomed Central, Hindawi) ont acquis une forte notoriété.

Parmi ces nouveaux entrants, on a pu relever la présence d'« éditeurs » moins scrupuleux proposant de nouvelles revues dont la finalité est avant tout de capter des APC (Article Processing Charges), en fixant celles-ci à des niveaux très faibles (quelques centaines de dollars), dans une sorte de démarche de dumping éditorial... L'émergence de ces « Predatory Journals » (revues prédatrices) a été pointée par différents auteurs<sup>20</sup>. Dans son STM Report (cf. note 1), l'association internationale de publication en STM, relativise très fortement l'idée qu'il y aurait un développement d'une édition scientifique de moindre qualité. Ce rapport note que la profession des « pure players » de l'édition scientifique en OA s'est fédérée au sein de l'Open Access Publishers Association ([www.oasp.org](http://www.oasp.org)) qui a défini des codes de bonnes pratiques exigeants.

<sup>18</sup> La tâche première de Crossref est d'attribuer aux articles scientifiques un DOI)

<sup>19</sup> [www.cnrs.fr/dist/z-outils/documents/Distlefil/distlefil12.pdf](http://www.cnrs.fr/dist/z-outils/documents/Distlefil/distlefil12.pdf)

<sup>20</sup> Beall, J. (2012). Predatory publishers are corrupting open access. *Nature News*, 489

Par ailleurs le STM Report souligne que lors de la soumission de leur projet d'article à un éditeur, les chercheurs publiant sont très attentifs à la notoriété et à l'image de la revue à laquelle ils s'adressent. L'inquiétude sur l'émergence des « Predatory Journals » s'est donc rapidement désamorcée. Parallèlement, les formes de la revue par les pairs se renouvellent, sous l'effet de technologies numériques.<sup>21</sup>

Il ne fait cependant pas de doute que la multiplication – y compris chez les plus grands éditeurs – de méga-revues moins sélectives, la multiplication des nouvelles revues en « Gold Open Access », en élargissant notablement les débouchés de la publication, a contribué dans les 4 dernières années à rendre la publication scientifique plus large dans ses critères.... Ce phénomène doit être analysé à la lumière de la démarche éditoriale qui constitue le cœur de la Lost Science, et non d'une perte de repères et de qualité de la publication scientifique.

### **La "Lost science" et la réaction numérique au H factor**

PLoS One a été l'évènement majeur de ce changement numérique, en affirmant de façon prometteuse : " PLoS ONE features reports of primary research from all disciplines within science and medicine. By not excluding papers on the basis of subject area, PLoS ONE facilitates the discovery of the connections between papers whether within or between disciplines."

Malgré le lancement dans les 12 derniers mois par tous les grands éditeurs scientifiques de méga-revues destinées à concurrencer la revue PLoS One, le système existant de la communication scientifique repose encore sur une logique sélective qui aboutit à l'acceptation d'une minorité d'articles pour publication, **mais élimine bon nombre de publications (chaque revue applique un taux de rejet qui peut aller jusqu'à 90% pour une revue comme Nature, Elsevier ne publie qu'un tiers du million d'articles qui lui sont soumis annuellement).**

Ces rejets ne se font pas en raison d'un non-intérêt scientifique intrinsèque des articles rejetés, mais reflètent des appréciations inégalement profitables à la qualité de la science. Ces appréciations de rejet peuvent être émises, soit parce que ces articles ne sont pas susceptibles d'être ultérieurement très cités, **(et donc nuisent au facteur d'impact de la revue), soit parce qu'ils reflètent des résultats négatifs, soit parce que l'intérêt des résultats présentés n'a pas été perçu par les «reviewers».**

La croissance de la production d'IST ne s'est pas accompagnée ni d'une croissance parallèle du nombre de revues susceptibles d'accueillir cette production, ni d'une augmentation suffisante du nombre d'articles publiés par revue. Un mécanisme a priori restrictif, voire dans certains cas délibérément malthusien, mettait en mouvement le système de publication peer-reviewed, et organisait une rareté des débouchés (les

---

<sup>21</sup> **STM Report cite Note 1:** "There is a significant amount of innovation in peer review, with the more evolutionary approaches gaining more support than the more radical. For example, some variants of open peer review (e.g. disclosure of reviewer names either before or after publication; publication of reviewer reports alongside the article) are becoming more common. Cascade review (transferring articles between journals with reviewer reports) and even journal-independent ("portable") peer review are establishing a small foothold. The most notable change in peer review practice, however, has been the spread of the "soundness not significance" peer review criterion adopted by open access "megajournals" like *PLOS ONE* and its imitators. Post-publication review has little support as a replacement for conventional peer review but there is some interest in its use as a complement to it (for example, the launch of PubMed Commons is notable in lending the credibility of PubMed to post-publication review). There is similar interest in "altmetrics" as a potentially useful complement to review and in other measures of impact."

revues à facteur d'impact), survalorisant les résultats publiés et les revues concernées. Le paradoxe est que cette rareté artificielle de la volumétrie des débouchés en revue reflétait plutôt l'économie du support imprimé (où la limitation du nombre d'articles est liée aux limites du médium physique et de la typographie) alors que la publication électronique permet aujourd'hui de dépasser ces limites du support papier imprimé, à coût variable nul.

La masse de connaissances scientifiques mises en forme (au travers de la rédaction par des chercheurs d'un article pour soumission à une revue) mais qui finalement ne trouvent pas de débouché en raison de la forte sélectivité des revues à facteur d'impact a été conceptualisée, et constitue ce que des auteurs anglo-saxons<sup>22</sup> ont pu désigner sous le vocable de «Lost Science» (que l'on pourrait traduire par la notion de «connaissance oubliée»).

On a donc pu soutenir systématiquement qu'une métrique reposant sur un nombre d'articles publiés, comme celle proposée par les principaux indicateurs bibliométriques, ne reflète donc que partiellement la qualité et l'étendue de l'ensemble de la production scientifique<sup>23</sup>. Divers outils servent de référence et constituent des approximations fréquemment utilisées<sup>24</sup>.

On notera qu'avec l'émergence du thème de la «Lost Science» est pour la première fois posée sur des bases empiriques solides la question de **l'efficacité globale du système de la communication scientifique actuel**. Et ce non pas du point de vue des mérites comparés des modèles «Toll Access» et «Open Access», mais d'un point de vue «épistémologique» qui met en parallèle d'une part le système de publication dans les revues à facteur d'impact, d'autre part la dynamique même de la production de connaissances, en faisant apparaître clairement les «effets pervers» du système actuel de la publication scientifique.

### **Halo numérique et «Lost science» : des solutions émergentes**

La réalité de cette situation n'a pas échappé aux principaux centres d'analyse et de production d'IST aux Etats Unis, dont le Data Science Center de NYU, ce Centre, avec lequel la DIST prépare un séminaire IST pour l'automne 2015, vient précisément de signer avec la DARPA un important contrat de recherches (360 millions de dollars) sur l'extraction de la Lost Science dans le Deep Web à partir d'algorithmes de tri définis par Juliana Freire, notre interlocuteur.

<http://www.baselinemag.com/careers/slideshows/11-tips-for-giving-a-great-speech.html>

---

<sup>22</sup> Neal S. Young, John P. A. Ioannidis, Omar Al-Ubaydii, *Why Current Publication Practices May Distort Science*, PLoS Medicine, 7 octobre 2009. L'article paru dans PLoS Medicine est la synthèse d'un article plus développé également accessible sur le site [www.plosjournals.org](http://www.plosjournals.org). Les auteurs centrent leur propos sur la publication scientifique dans le domaine de la recherche biomédicale (qui représente on l'a vu 30% de la production scientifique mondiale mesurée au nombre d'articles publiés), domaine où les enjeux économiques sont considérables et la compétition entre équipes de recherche très marquée.

<sup>23</sup> La seule exception à ce constat est liée au domaine de la physique des hautes énergies pour laquelle la plate-forme ArXiv collecte systématiquement les articles dans leur version préprint (400 000 articles préprint sont disponibles dans cette base).

<sup>24</sup> <http://googlescholar.blogspot.co.uk/2014/06/2014-scholar-metrics-released.html>

Par ailleurs, les éditeurs de science sont les premiers conscients du phénomène de «Lost science» puisqu'ils sont au cœur du mécanisme très sélectif qui le fonde. Dans les douze derniers mois des réponses partielles ont été données, au travers du lancement de nouvelles revues, pour contrer à l'initiative des éditeurs le phénomène de «Lost Science».

Tous les grands éditeurs sans exception ont créé dans les douze derniers mois des « méga-revues » multidisciplinaires, dont le taux de rejet n'est que de l'ordre de 30% et dont le modèle est PLoS One, première revue en OA créée en 2006. **Ces nouvelles revues publient des matériaux qui jusqu'ici échappait à l'édition scientifique** : articles rendant compte de résultats négatifs, articles fournissant des données méthodologiques détaillées garantissant la reproductibilité des résultats de la recherche, articles de « curation » de jeux de données issus d'une recherche.

**Le dépassement probable du phénomène de «Lost Science» est aussi lié à la maturité des technologies d'exploitation des gisements de connaissances.** Dans un contexte jusqu'à aujourd'hui dominé par les technologies de «search», fournissant en réponse à une requête des listes de résultats pertinents, la limitation des corpus d'articles scientifiques exploitables répondait aussi à la nécessité de ne pas dégrader (par la génération de «bruit» c'est-à-dire de références non pertinentes) la performance de ces outils. Demain les technologies du Text et du Data Mining (TDM), capables de traiter des corpus beaucoup plus volumineux permettront de dépasser le problème de la Lost Science : dans le contexte d'une exploitation en TDM, un résultat négatif, une notice méthodologique, la description d'un jeu de données reprendront leur place de connaissances scientifiques exploitables. La DIST est au cœur de ces logiques avec des projets comme ISTEEX, IA doté de 60 Millions d'euros (chef de projet : R. Fabre). La DIST donne priorité à ces développements, à travers les services d'analyse sémantique proposés par ISTEEX (7 équipes universitaires coordonnées par JM Pierel, PU, UL, Chef de projet Services). C'est une dimension essentielle de services pour les Instituts<sup>25</sup> au cours des années à venir: la stratégie IST prévoit une offre de services en 2016 en la matière.

### **Une métrique plus riche : un enjeu global d'analyse interdisciplinaire**

Une mesure plus systématique apparaît même indispensable dans les nombreux domaines où la métrique n'offre pas aujourd'hui les solutions d'analyse qu'appelle le développement rapide de l'interdisciplinarité.

Une étude longitudinale STM et SHS sur 2800 journaux appartenant à 13 maisons d'Édition de science de premier rang.<sup>26</sup> s'est donné pour tâche de repérer, pour toutes

---

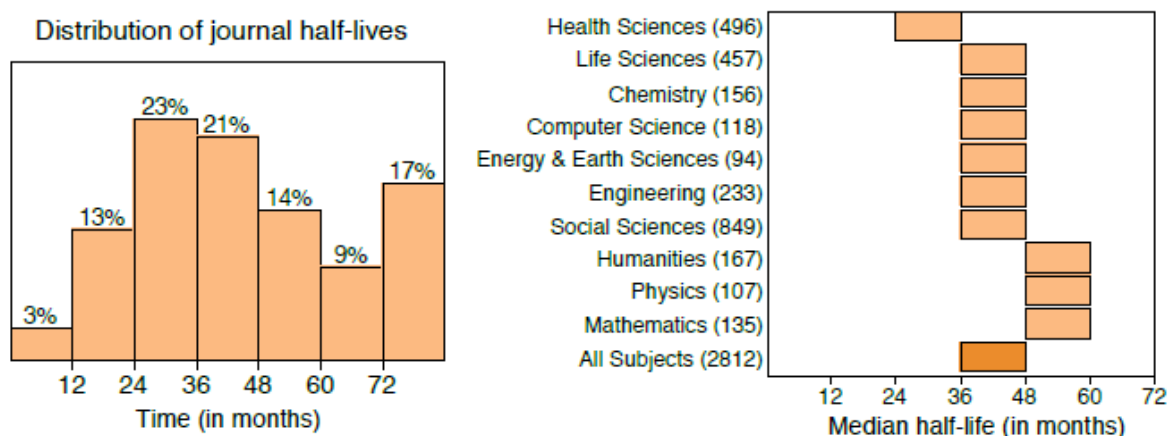
<sup>25</sup> **Voir Rapport STM cite Note 1:** "Semantic technologies have become mainstream within STM journals, at least for the larger publishers and platform vendors. Semantic enrichment of content (typically using software tools for automatic extraction of metadata and identification and linking of entities) is now widely used to improve search and discovery; to enhance the user experience; to enable new products and services; and for internal productivity improvements. The full-blown semantic web remains some way off, but publishers are starting to make use of linked data, a semantic web standard for making content more discoverable and re-usable "

<sup>26</sup> By Philip M. Davis, Ph.D. Phil Davis Consulting, 432 Mitchell St., Ithaca NY 14850. [pmd8@cornell.edu](mailto:pmd8@cornell.edu) | 607 272---5461  
<http://phil---davis.org>



ces revues à facteur d'impact élevé, le temps de "demi-vie" (temps nécessaire à l'obtention de la moitié des résultats entrant dans leur score respectifs de citation). Cette étude originale montre assez clairement ce qu'on savait déjà, à savoir que, à qualité égale des journaux, la fréquence de citation diffère nettement d'un champ disciplinaire à l'autre. Cette différence a pour conséquence de requérir, **pour tous les cas où la publication emprunte ses bases à deux champs de connaissances éloignés dans la demi-vie de citations, la construction d'indices synthétiques d'analyse permettant de faire clairement ressortir les critères d'interdisciplinarité.** Ce n'est pas nécessairement compliqué, et la DIST est prête à travailler là-dessus : encore faut-il le proposer et le faire....

**Tableau:** Demi-vie moyennes des taux de citation globale par durée en mois, et répartition des durées moyennes pour un éventail des principaux journaux de référence des groupes de disciplines évaluées



Source: By Philip M. Davis, Ph.D. Phil Davis Consulting, 432 Mitchell St., Ithaca NY 14850. [pmd8@cornell.edu](mailto:pmd8@cornell.edu) | 607 272---5461 <http://phil---davis.org>

**Une homogénéisation des normes internationales, prônée par l'ONU comme par l'ICSU, constituerait par ailleurs un premier progrès.**

**La perte en ligne d'IST : un halo numérique en surplomb**

**Production des données et production d'articles: un écart croissant ?**

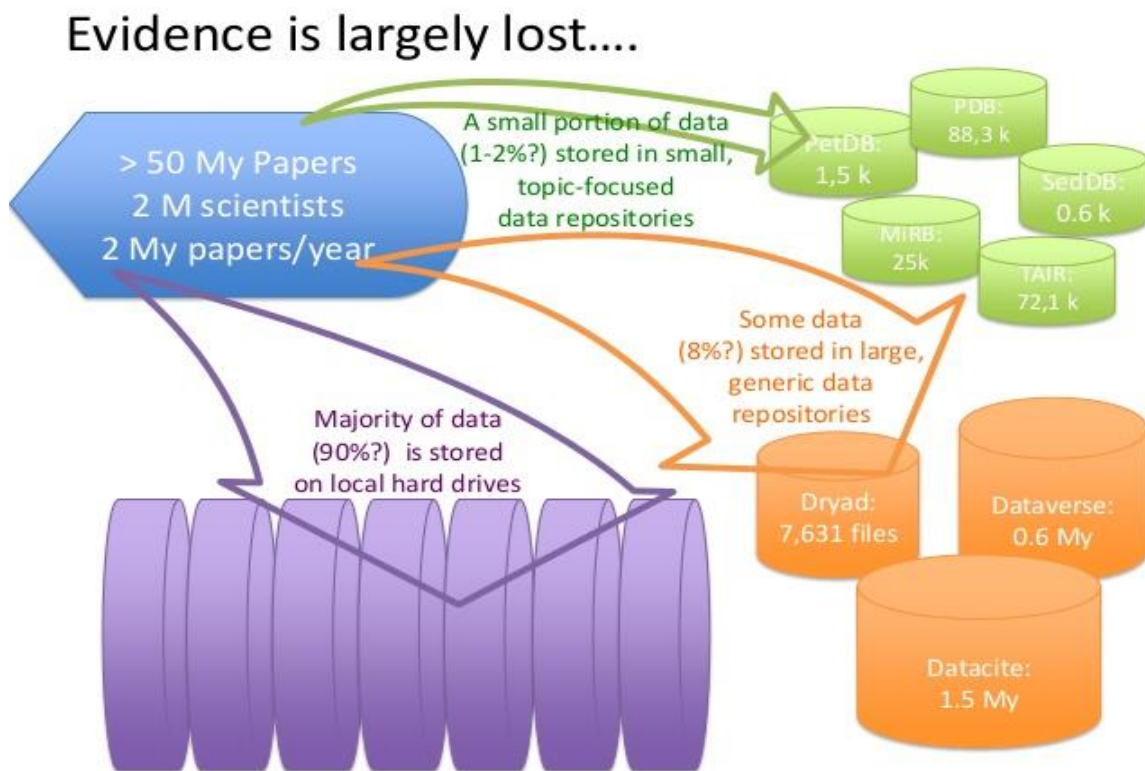
La stratégie IST du CNRS observait que le fossé se creuse entre d'une part, les cycles de production des données et, d'autre part, les cycles de publication des résultats scientifiques. Or, selon les rares grands inventaires disponibles à ce sujet, les données inutilisées s'accroissent toujours plus, dans les domaines correspondant de la publication...

Un "goulet d'étranglement" menacerait ainsi l'efficacité de l'observation et cet obstacle irait croissant avec la multiplication des capteurs de données. Cette analyse, à quantifier et à vérifier, appelle des réponses globales pour prévenir des pertes en ligne croissantes d'IST. En collaboration avec la DIST, des réponses construites émergent, a. pour répertorier et publier les données, b. pour orienter leur passage à la publication, et c. pour ouvrir à de nouvelles formes de la publication.

**Le halo numérique des données et l'Édition scientifique: la difficile recherche de solutions**

Il y a peu d'images globales des difficultés actuelles associées à cet écart, pourtant situé au cœur des futures logiques de publication de bon nombre de disciplines. Le Schéma suivant, assez éloquent, est systématiquement utilisé par le Disruptive Technologies Department d'Elsevier en 2014 et en 2015 : on peut donc lui prêter au moins de chercher à convaincre.... Si cette présentation globale était clairement confirmée, la déperdition des données (stockage sur les disques durs des laboratoires....) atteindrait 90% des données produites, chiffre qui doit évidemment être affiné et analysé...

**Tableau:** Production de données et production de publication: les pertes en ligne d'IST



**Source :** Elsevier, Disruptive Technologies Department 2014

La question posée à l'Édition scientifique est claire, quel que soit l'ordre de grandeur précis, et c'est une question éditoriale majeure, qui conduira à préciser le contour du halo numérique actuel de l'IST. Le STM Report précise pour 2015 observe ce qui suit sur le sujet :

The explosion of data-intensive research is challenging publishers to create new solutions to link publications to research data (and *vice versa*), to facilitate data mining and to manage the dataset as a potential unit of publication. Change continues to be rapid, with new leadership and coordination from the Research Data Alliance (launched 2013): most research funders have introduced or tightened policies requiring deposit and sharing of data; data repositories have grown in number and type (including repositories for “orphan” data); and DataCite was launched to help make research data cited, visible and accessible.

Meanwhile publishers have responded by working closely with many of the community-led projects; by developing data deposit and sharing policies for journals, and introducing data citation policies; by linking or incorporating data; by launching some pioneering data journals and services; by the development of data discovery services such as Thomson Reuters' Data Citation Index. Text and data mining are starting to emerge from niche use in the life sciences industry, with the potential to transform the way scientists use the literature. It is expected to grow in importance, driven by greater availability of digital corpuses, increasing computer capabilities and easier-to-use software, and wider access to content.

---

*The STM Report An overview of scientific and scholarly journal publishing*, Fourth Edition, March 2015, mis en ligne en avril 2015