

Les mutations du paysage de l'information scientifique

Joachim Schöpfel

1. Introduction

En quelques années, les nouvelles technologies de l'information et de la communication (NTIC) ont modifié le paysage de l'information scientifique et technique (IST) en profondeur. Toute la chaîne de valeur de la communication scientifique est concernée, depuis la production et la diffusion de l'information jusqu'à l'acquisition et la consommation des produits et services.

Cette modification a fait l'objet d'un grand nombre d'études et de publications. Notre chapitre tente de dégager les grandes lignes de l'évolution en cours pour donner aux professionnels de l'information les moyens de mieux comprendre les enjeux et répondre aux besoins des jeunes chercheurs.

Nous dressons ici un panorama : l'objectif n'est pas de fournir de nouvelles informations mais de proposer une structure et aussi, une perspective, c'est-à-dire une certaine manière de voir, une approche.

2. Pourquoi l'information scientifique et technique ?

Depuis le lancement des premières revues scientifiques au 17^e siècle, les finalités et fonctions d'une publication scientifique ont peu évolué. Comment protéger les droits de l'auteur d'une idée ou d'une découverte scientifique ? Comment diffuser les résultats de la recherche ? Comment les exposer à la critique et au débat ? Et comment en préserver une trace pérenne ?

Aujourd'hui encore, les chercheurs publient essentiellement pour les mêmes raisons. Et les vecteurs de l'IST, en premier lieu les revues mais aussi les livres, rapports, conférences ou thèses, remplissent toujours les fonctions élémentaires définies depuis bientôt 350 ans : l'enregistrement (c'est-à-dire, la protection de la propriété intellectuelle), la diffusion, l'évaluation par les pairs (*peer review*) et l'archivage.

Bien sûr, les NTIC ont changé la donne. Le numérique affecte toutes les fonctions de la publication scientifique, modifie la demande des chercheurs et le contenu des publications. La communication scientifique est également tributaire des modes de pilotage, de financement et d'évaluation de la recherche (Duval & Schöpfel, 2009). C'est la loi du *Publish or perish* : les universités, organismes de recherche, agences de moyens et d'évaluation comptent désormais les « publiants » et s'organisent pour suivre la « production scientifique » à la trace, afin

d'avoir une idée plus précise du nombre d'articles, livres, communications et brevets. Publier est devenu une condition nécessaire pour la carrière scientifique et le travail de recherche. Les chercheurs passent désormais 50% de leur temps sur la recherche et 50% pour l'évaluation, et on entend dire qu'ils publient « comme des poules en batterie »...

Mais changeons de perspective et demandons pourquoi un (jeune) chercheur a besoin de l'information scientifique. Hans E. Roosendaal, professeur en gestion stratégique à l'université de Twente (Pays-Bas) fournit quatre éléments de réponses (Roosendaal et al., 2010).

Accès aux résultats : Un chercheur a besoin de se tenir au courant des résultats de la recherche dans son domaine, des données scientifiques, théories, méthodes etc. L'information scientifique constitue pour lui une matière première indispensable, et le temps qu'il passe à la recherche et la consommation de l'information est un investissement nécessaire. Avec l'avènement d'Internet, ce temps qu'un chercheur consacre à la veille, à la recherche et à la lecture d'information a considérablement augmenté.

Information propre à la communauté : Pour ses choix stratégiques (thématiques, collaborations, subventions...), un chercheur a besoin d'information sur ses réseaux mais aussi, sur ses concurrents. Qui fait quoi ? Quels sont les nouveaux projets de recherche ? Quelles sont les nominations, les nouvelles équipes ou institutions qui comptent ? Qui travaille et publie avec qui ? Où trouver des subventions ? Tout cela est essentiel pour la production scientifique, et l'information scientifique fournit des réponses.

Information relative à la société : Mais pour ses choix stratégiques, le chercheur a aussi besoin d'information sur l'environnement sociétal, sur la politique scientifique du moment, sur les besoins de la société civile etc. Prenant quelques exemples de thèmes à l'ordre du jour de la société civile : le vieillissement de la population et la prévention et le traitement de la maladie d'Alzheimer, le changement climatique, les nouvelles énergies, les moyens de transport urbain non polluants, les nouvelles technologies de l'information et la protection des données personnelles etc. Ces thématiques impactent les orientations scientifiques d'un chercheur, et il doit se tenir au courant.

Préparation d'une publication : Lors du choix d'un vecteur de divulgation de ses résultats, un chercheur a besoin d'information sur les supports de publication, les titres et périmètres de revues, les comités de rédaction, leur politiques éditoriales et contrats de publication, les éditeurs et collections de livres, les conférences etc. Où trouver cette information indispensable par rapport à l'évaluation de sa production scientifique ?

Le chercheur ne vit pas au-dessus de la mêlée ou dans une tour d'ivoire. L'information scientifique lui sert aussi à garder le contact avec la société, dans les deux sens. Pour toutes les raisons citées, un chercheur a besoin de l'accès à l'information scientifique – autrefois via « sa » bibliothèque, aujourd'hui de plus en plus grâce à Internet.

3. Contenus

En parlant d'information scientifique, on a parfois tendance à confondre information et contenu, comme si l'information se réduisait aux seuls documents et données, aux revues, livres, rapports etc., sans tenir compte des produits et services d'accès, de sélection etc.

Nous allons d'abord décrire les mutations qu'on peut constater sur le plan des contenus : nous procéderons en deux temps, en donnant d'abord quelques chiffres avant d'étudier l'impact du numérique sur les caractéristiques fonctionnelles et techniques.

Pour retracer l'impact du numérique sur les contenus de l'IST, nous avons fait un choix. Loin d'être exhaustif, nous avons centré l'étude sur plusieurs catégories documentaires, les revues, articles, livres et la littérature grise, avant d'aborder les archives de données scientifiques et le « 4^e paradigme » de la cyberinfrastructure.

Le réseau thématique et pluridisciplinaire RTP-DOC du CNRS a conceptualisé la notion du document dans son passage au numérique, en montrant les transformations radicales du document sous ses trois aspects forme, contenu et médium (cf. Pédaque, 2003).

Notre ambition est bien plus modeste. Nous allons évoquer quelques aspects ayant un impact direct sur le comportement du consommateur, c'est-à-dire du chercheur en tant qu'utilisateur final, en distinguant les caractéristiques quantitatives de l'offre des fonctionnalités et spécifications techniques.

3.1. Revues

La revue scientifique est une publication périodique relevant d'un ou de plusieurs disciplines (STM et/ou SHS) dont la finalité principale est de publier des résultats de recherche originaux sous forme d'articles.

Le nombre des titres vivants à caractère scientifique et soumis à sélection (*peer review* etc.) est estimé à environ 27 000 en 2010, par rapport à 15 000 en 2001. Leur nombre a régulièrement augmenté depuis sur une longue période (figure 1).

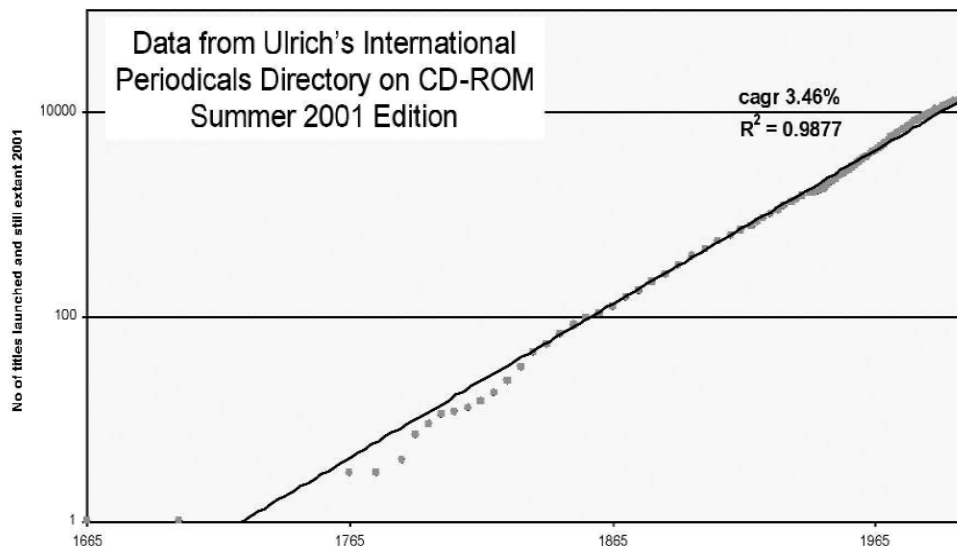


Figure 1 : Augmentation du nombre des revues (Mabe, 2003)

D'après Mabe (2003), cette augmentation a été en moyenne de 3,5% par an avant 2001. Au regard des chiffres plus récents, cette évolution s'est fortement accélérée depuis pour atteindre environ 7% par an.

Pourquoi cette augmentation que certains décrivent comme une explosion en regrettant de voir aujourd'hui beaucoup trop de titres ? En fait, il n'y a pas trop de titres. Mabe (2003) a démontré le rapport direct entre le nombre des revues, le nombre des chercheurs et le budget global de la recherche, avec une corrélation positive significative entre ces trois chiffres. Autrement dit : le nombre des revues augmente au même rythme que celui des chercheurs et des dépenses pour la recherche.

Traditionnellement, une revue imprimée comptait en moyenne 100 auteurs et 500 à 1 000 lecteurs (dont les auteurs). C'était sa communauté. Avec le passage au numérique, ces relations ont changé mais uniquement du côté des lecteurs – aujourd'hui, une revue atteint potentiellement bien plus de lecteurs. Par contre, le nombre d'auteurs par revue n'a pas vraiment changé ce qui explique l'augmentation du nombre des titres : toujours plus de chercheurs ont besoin de publier et cherchent un « produit » pour leurs articles, et les éditeurs et sociétés savantes répondent à cette demande par le lancement de nouveaux titres. A ceci s'ajoutent d'autres facteurs, tels que le développement interdisciplinaire et l'émergence rapide de nouveaux domaines, ou encore la percée de grands pays comme la Chine, l'Inde ou le Brésil sur le marché de l'information scientifique.

D'après la base *Ulrichsweb*, 70% de ces titres existent 2010 en format numérique, avec ou sans version imprimée. Si on ne regarde que les revues avec facteur d'impact, ce pourcentage est encore bien plus élevé et dépasse les 90%.

Du point de vue utilisateur, on peut décrire les quatre aspects novateurs d'une revue électronique :

1. Un ou plusieurs formats numériques, le plus souvent HTML et/ou PDF.
2. Un accès en ligne sur Internet, le plus souvent sur le Web ; donc virtuellement accessible à partir de n'importe quel moment, endroit et poste.
3. Une navigation entre citations, références et documents via des liens hypertexte internes et/ou externes ou des recherches (DOI, CrossRef, OpenURL).
4. De nouvelles fonctionnalités à valeur ajoutée telles que McKiernan (2002) les a listées depuis bientôt dix ans. Il s'agit par exemple d'ajouter un workflow d'édition en amont de la publication en ligne, de rendre possible d'interagir avec le texte, les illustrations, tableaux, données etc. (annotations, manipulations etc.), ou de pouvoir personnaliser l'accès aux documents (environnement personnalisé, alertes, profils de recherche etc.).

3.2. Articles

Le nombre d'articles publiés dans les revues scientifiques est difficile à estimer. Mabe (2008) indique 1,9 millions d'articles pour 2008, avec une augmentation annuelle de 3-4%. En 1996, ils n'étaient « que » 1,1 millions, selon Mabe.

Ce chiffre est probablement sous-estimé. En extrapolant plusieurs sources (*Web of Science*, *Scopus*, *Ulrichsweb* etc.), on obtiendrait plutôt 2,5 millions articles pour 2010. La plupart de ces articles sont *peer reviewed*, indexés dans une ou plusieurs base(s) de données, et publiés dans une revue à audience internationale.

Cette masse d'articles serait rédigée par plus d'un million d'auteurs et lue par 10-15 millions de lecteurs.

D'une manière générale, on peut décrire la transformation numérique de l'article sous deux aspects :

1. Le trait caractéristique d'un article numérique a été décrit par RTP-DOC et d'autres (cf. Pédaque, 2007 ; Chaudiron et al., 2008). Il s'agit de la déconstruction et fragmentation de l'unité du texte ou, selon les termes de RTP-DOC, de « l'érosion de la notion monolithique du 'document' dans le paradigme hypertextuel ». Récemment, Elsevier a lancé le prototype du « Article of the Future » sur le site de *Cell Press*. Ce prototype reflète l'éclatement de l'article en plusieurs éléments distincts (titre et résumé, illustrations, résumé graphique, références etc.).
2. Le 2^e trait caractéristique de l'article numérique est en partie lié à cette fragmentation. Il s'agit de l'enrichissement par de nouveaux contenus et fonctionnalités (multimédia, données etc.) et de l'interactivité, notamment avec les outils du Web 2.0 (annotations, *peer review* de la version publiée, *tagging* ou marquage etc.).

Comment cette transformation va-t-elle modifier la rhétorique de la publication scientifique ? Il est peut-être trop tôt pour le dire. Des études récentes démontrent clairement le potentiel de changement mais indiquent aussi que l'écriture des articles reste pour l'instant assez classique (cf. Delmotte, 2009).

3.3. Livres

Les livres numériques, les e-books ont déjà fait couler beaucoup d'encre. A titre d'exemple récent, on peut citer l'excellent mémoire de Conjaud

(2009) déposé dans MemSIC, sur la place des e-books dans la documentation médicale.

Au lieu de dresser un panorama exhaustif et de contribuer à la conceptualisation du phénomène – livre numérisé ? livre électronique ? liseuse ? (cf. Dacos & Mounier, 2010) – nous nous contenterons de quelques chiffres et remarques d'ordre général.

En juin 2010, Google Books annonçait avoir numérisé plus de 12 millions de livres dont un certain nombre à caractère scientifique. Voici au hasard, trois disciplines :

- Mathématiques : 12 100 livres dont 4 700 du domaine public.
- Sciences politiques : 29 000 livres dont 3 700 du domaine public.
- Psychologie : 25 900 livres dont 2 000 du domaine public.

Pour ces disciplines, Google Books compte environ 65 000 livres dont plus de 10 000 immédiatement accessibles. A titre comparatif, la collection *eBooks* d'Elsevier contient 10 500 livres dont 2 444 titres en science de la vie et médecine et 434 titres en chimie. Quant à Springer, sa plate-forme *SpringerLink eBooks* donne accès à 37 454 livres dont 2 944 titres en science de la vie et médecine et 1 962 titres en chimie.

Ces chiffres sont impressionnants et confrontés avec quelques études récentes sur l'usage (Nicholas et al., 2010), ils annoncent un changement du comportement des chercheurs par rapport à la lecture. Ce qui manque par contre, c'est une analyse plus fine de l'impact de cette évolution au niveau des fonctionnalités et services, par exemple pour un produit tel que le *Brain Navigator* qui a remplacé l'ancien *Atlas of the Human Brain*. Comment définir cet objet numérique non identifié qu'est devenu le livre ? Il faudrait probablement une nouvelle typologie.

3.4. Littérature grise

A côté de l'offre des éditeurs scientifiques, le chercheur a accès à d'autres documents à caractère non commercial, produits à tous les niveaux par des administrations, organismes scientifiques et laboratoires R&D et diffusés en dehors des circuits habituels. Il s'agit de thèses, mémoires, rapports, conférences, *working papers* etc. qu'on désigne sous le terme de littérature grise (Farace & Schöpfel, 2010).

Pour les bibliothèques et centres de documentation, ces documents sont souvent plus difficiles à acquérir, et leur signalement est moins bien contrôlé que pour les revues, articles et livres des éditeurs. Par contre, leur contenu est riche, ils contiennent des résultats et informations qui parfois ne sont jamais publiés ailleurs, ou bien plusieurs mois ou années plus tard. Leur importance pour la recherche est indéniable mais des études scientométriques (analyses des citations par exemple) montrent des dissemblances significatives entre les différentes disciplines et domaines scientifiques.

Chiffrer la production « grise » est difficile car il n’y a pas de base bibliographique, catalogue ou d’index exhaustifs. Pour donner une idée : en 2009 la littérature grise représente plus de 300 000 ou 18% des documents déposés dans les archives ouvertes françaises. Il s’agit surtout de thèses et de conférences (Schöpfel & Prost 2010). Le SUDOC signale environ 17 000 thèses scientifiques par an. Un autre chiffre : en tant que bibliothèque nationale scientifique de l’Allemagne, la TIB Hannover reçoit tous les rapports issus de la recherche publique ; leur nombre s’élève à 8 000 par an.

Le passage au numérique change la nature de ces documents et de leur distribution. D’une part, Internet facilite la diffusion de la littérature grise, par exemple grâce aux dépôts dans les archives ouvertes. De l’autre côté, le signalement de ces documents reste souvent un problème ; les trouver ne devient pas nécessairement plus facile. L’absence d’enjeux économiques n’incite pas à la normalisation des données ni à l’optimisation des circuits de diffusion.

L’exception qui confirme la règle concerne les thèses électroniques. Depuis plusieurs années, des réseaux nationaux et internationaux pour le signalement, la diffusion et la conservation des thèses se mettent en place. En France, après quelques initiatives locales l’ABES coordonne aujourd’hui un dispositif autour de l’application STAR et un archivage pérenne sur les serveurs du CINES, avec intégration des métadonnées normalisées dans le SUDOC.

3.5. Archives de données

Jusqu’ici nous avons uniquement parlé de publications. Mais l’offre ou plutôt la mise en ligne des résultats scientifiques se développe rapidement, en parallèle, parfois en lien, parfois aussi en concurrence avec les publications. Une partie est librement accessible sur le Web, en particulier sur des serveurs correspondant aux critères du libre accès. D’autres résultats – des *dark data* - appartiennent au *deep web* et sont virtuellement inaccessibles (cf. Heidorn, 2008). Dans les archives ouvertes françaises, leur proportion aux dépôts atteint 5-10%.

L’enjeu de ces *datasets* est double, l’accès et l’exploitation. Il faut des métadonnées normalisées (dont un identifiant comme par exemple le DOI) pour décrire, gérer et signaler les résultats. Et il faut des services pour faciliter la recherche et l’accès aux données. Certains sites et/ou revues créent des liens entre publications et résultats. Cela n’est pas toujours très efficace (cf. Savage & Vickers, 2009).

En France, on peut citer plusieurs réservoirs de données, dont la base de cristallographie de l’université du Maine ou les différentes bases de l’observatoire de Paris à Meudon. On peut aussi mentionner le projet européen *DataCite* pour attribuer et gérer des identifiants DOI aux ensembles de données scientifiques.

3.6. Le « 4^e paradigme »

L'information scientifique se transforme ici en Cyberinfrastructure ou eScience, constituée non plus de littérature et documents mais de données. Jim Gray, chercheur chez Microsoft, a décrit cette évolution comme changement de paradigme¹ :

- 1^e paradigme (il y a 1000 ans) : recherche empirique pour décrire des phénomènes naturels.
- 2^e paradigme (depuis 100 ans) : recherche théorique, à partir de modèles et de généralisations.
- 3^e paradigme (depuis quelques décennies) : recherche informatisée basée sur des simulations de phénomènes complexes.
- 4^e paradigme (aujourd'hui) : exploration des données, intégration des théories, simulations et expérimentations.

Les résultats ne sont plus simple support pour des publications mais y sont intégrés s'ils ne les remplacent pas tout simplement, par exemple sous forme d'épi-journaux (*overlay journals*). Certains prédisent qu'à terme, les publications comme vecteur de communication scientifique disparaîtront au profit d'une communication directe machine/machine (Hey et al., 2009). Après la révolution, la fin des publications scientifiques ? On n'y est pas encore.

Mais c'est un fait qu'aujourd'hui un chercheur est confronté à une masse formidable d'information scientifique, une nébuleuse de contenus interconnectés autour des disciplines et domaines de recherche. Cette structure a été rendue visible par les travaux du projet MESUR (Bollen et al., 2009a), sous forme d'une *map of science* sur la base des statistiques d'utilisation de plusieurs milliers revues en ligne.

L'offre toujours plus large et l'accès en ligne ont créé un autre monde ou plutôt, une autre façon de voir le monde de l'information scientifique.

¹ <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>

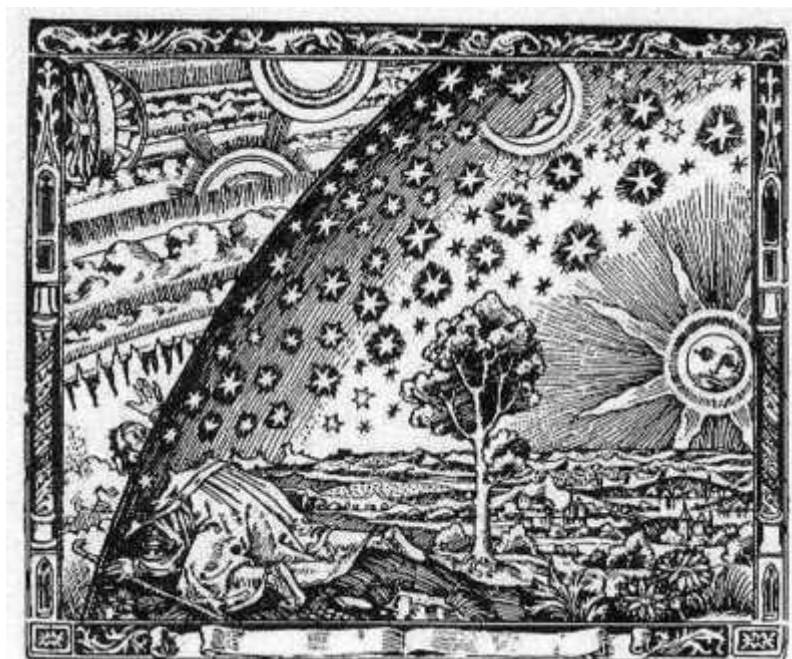


Figure 2 : Un nouveau monde... (C. Flammarion, *L'atmosphère*. Paris, 1888).

Alors, comment voyons-nous ce monde ? Quels sont nos yeux, nos outils pour voir et comprendre ? Les études d'usage le confirment : en premier lieu, il s'agit de Google... Mais l'offre de service à disposition du chercheur est bien plus riche et plus large ? Le prochain chapitre tente d'en proposer un aperçu structuré.

4. Services

Pour donner une idée de l'importance de cette masse d'information, nous avons extrait le nombre de revues et d'articles de la base SCOPUS², pour plusieurs disciplines :

Discipline	Revues	Articles	Articles/jour
Sc économiques	483	21659	60
Chimie	495	122573	335
Sc de l'information	116	3733	10
Géologie	141	9190	25
Environnement	725	67905	190
Mathématiques	735	78546	215
Cancérologie	186	26174	70

Figure 3 : Revues et articles par discipline (SCImago, chiffres pour 2007)

Un chercheur en chimie, s'il voulait faire de la veille sur les meilleurs titres de sa discipline, devrait suivre environ 500 revues avec plus de 120 000

² Via SCImago (2007). SJR — SCImago Journal & Country Rank.
Retrieved April 27, 2010, from <http://www.scimagojr.com>

articles par an, c'est-à-dire plus de 300 articles par jour. C'est impossible. Pour d'autres domaines, ce n'est guère mieux. Et il est tout simplement irréaliste d'imaginer un chercheur en sciences de l'information feuilleter une dizaine d'articles par jour. Alors, comment faire ? Comment faire face à cette masse d'information ?

Une des maximes du Web est « content is king ». Confronté à cette surabondance d'information, on peut douter de cette affirmation. Et si le roi était nu ? Il est certain que le contenu est une condition nécessaire pour une communication scientifique réussie ; mais à lui seul il ne suffit pas. Sans un environnement de service à forte valeur ajoutée, cette masse d'information produirait plutôt l'effet inverse.

A titre d'exemple nous avons choisi plusieurs services à valeur ajoutée, à utiliser par ou pour le chercheur.

4.1. Veille

Quelles sont les sources d'information utilisées par les doctorants ? Nous avons consulté plusieurs analyses de citation³. D'après ces études, il s'agit surtout d'articles de revues scientifiques, même si la part varie d'un domaine à l'autre, entre 40 et 95%. Les livres représentent entre 5 et 35% des sources consultées, suivis par la littérature grise (thèses, conférences etc.) avec 5 à 15% et les pages sur le Web avec 5% des citations.

Comment assurer un suivi efficace et rationnel d'autant de sources ? Sans mettre en place un dispositif sophistiqué de veille, le chercheur (mais aussi le professionnel de l'information) peut s'appuyer sur des agrégateurs qui signalent la mise à jour pour un ensemble de sites Web et importent le nouveau contenu via la syndication de flux RSS.

Des outils comme Google Reader, Netvibes ou Symbaloo permettent de suivre des sites Web (archives ouvertes, éditeurs etc.) des blogs, des actualités sur les réseaux ou sur Twitter mais aussi les sommaires d'une sélection de revues. A terme, un tel dispositif personnalisé est un gain de temps inestimable.

Mais il existe aussi des agrégateurs publics, mis en place et partagés par des établissements et professionnels de l'information. Ainsi, le site Bibliopédia propose plusieurs dizaines de bibliothèques, portails et univers Netvibes, certains généralistes ou liés à un service documentaire, d'autres thématiques et assez spécialisés.⁴

Pour la syndication des sommaires, le JISC britannique a financé un projet qui offre des flux RSS pour un millier de revues scientifiques.⁵

La plupart des bibliothèques numériques, plates-formes et archives offrent aussi un service d'alerte, sur domaine, mot-clé ou profil.

³ Gooden (2001), Chikate & Patil (2008), Heil (2009)

⁴ http://www.bibliopedia.fr/index.php/Bibliothèques_sur_netvibes

⁵ <http://www.tictocs.ac.uk/>

4.2. Evaluation

D'autres outils aident le chercheur à évaluer la qualité et la pertinence des sources d'information (publications, sites), mais aussi la production et l'impact des institutions et des chercheurs.

Ces outils exploitent des données quantifiables, telles que les citations, consultations, liens ou encore la taille du contenu.

Citons ici des outils payants comme le Web of Science de Thomson et son concurrent direct, SCOPUS d'Elsevier qui proposent des services scientométriques de plus en plus sophistiqués comme par exemple la toute nouvelle suite de produits *SciVerse* d'Elsevier qui extrait des données de SCOPUS sur les publications, domaines, institutions et chercheurs.

D'autres outils sont librement accessibles sur Internet et proposent des indicateurs plus ou moins précis et pertinents : *SCImago* (revues, pays, institutions), *Harzing's Publish or Perish* avec son h-index (chercheurs), *Webometrics* (sites, archives), *Google Scholar* (publications, auteurs), *PLoS statistics* (revues et articles).

L'évaluation scientifique est un sujet délicat, et les outils et méthodes d'évaluation font l'objet d'analyse et de critique, notamment dans le domaine des SHS (Zarka, 2009). Mais il n'est pas lieu ici de mener cette discussion qui, sur le terrain de la recherche, semble souvent vécue et acceptée comme contrainte professionnelle.

4.3. Sélection

Comment trier, filtrer, limiter, affiner la sélection des informations et notamment, des articles ? En premier lieu avec des moteurs de recherche spécialisés dans le domaine des publications scientifiques, comme Scirus et Google Scholar. A l'aide des fonctionnalités avancées, le chercheur peut filtrer l'information par l'année de publication, la langue du texte, le format du fichier ou de la page, le nom de l'auteur etc. Scirus permet aussi une limitation à partir de sites présélectionnés (archives ouvertes etc.).

En fonction du domaine, d'autres outils peuvent être utiles, notamment des bases de données bibliographiques et, plus récent, des « avatars » autour d'un site comme PubMed⁶.

Les fonctionnalités des bases de données scientifiques offrent également ce genre de services.

4.4. Accès (payant)

L'information scientifique est en grande partie diffusée par des producteurs, éditeurs et intermédiaires commerciaux, avec un accès payant, restrictif et limité. Des portails, bibliothèques numériques, listes

⁶ Cf. GoPubMed® <http://gopubmed.com/>

a-to-z, base de données et collections de revues et e-books sont mis à disposition des chercheurs selon leurs droits d'accès. Hormis les SHS, les plus importants produits d'IST couvrent toute la gamme (ou une grande partie) des disciplines scientifiques et sont largement connus (ScienceDirect d'Elsevier, SpringerLink, Wiley InterScience, Journals@Ovid etc.) mais il existe également une offre plus spécialisée et/ou de moindre taille (Oxford Journals, EmeraldInsight, JSTOR, CAIRN etc.).

L'acquisition, l'agrégation et la gestion d'une telle offre posent des questions spécifiques d'ordre juridique (licence, gestion des droits, définition des utilisateurs) et techniques (accès direct ou via un serveur proxy, environnement) que nous avons abordées ailleurs (Drouard et al., 2009). Insistons ici uniquement sur l'intérêt d'une analyse fine des besoins des communautés scientifiques à servir, élaborée par exemple par une bibliothèque scientifique de proximité ; ce service sera complété par la mise à disposition d'une offre transversale plus large, dans le cadre d'une licence nationale.

4.5. Accès (libre)

Depuis bientôt 20 ans se développe, en marge d'abord et désormais en concurrence, l'accès libre à l'information scientifique sur Internet, essentiellement sous deux formes : les archives ouvertes et les revues en accès libre dont le nombre accroit rapidement.

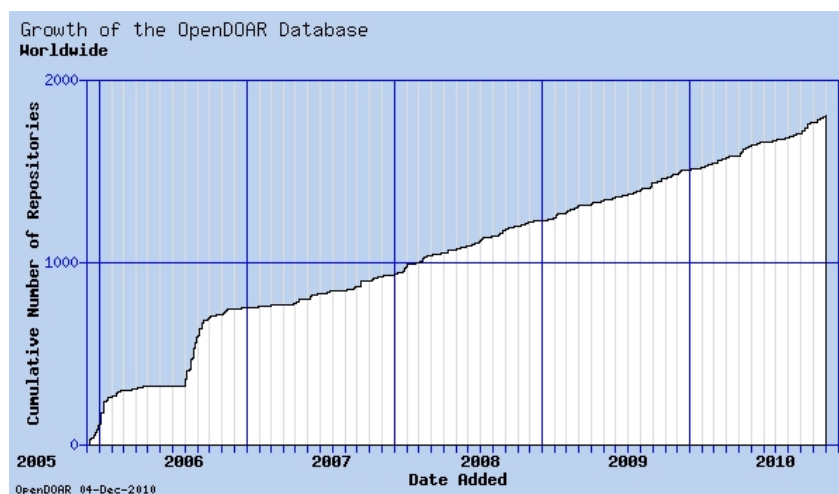


Figure 4 : Le nombre des archives ouvertes (OpenDOAR, décembre 2010)

Début décembre 2010, le répertoire d'archives ouvertes le plus connu, OpenDOAR⁷, comptait 1 805 sites au monde dont un tiers aux Etats-Unis et au Royaume-Uni. Quatre sites sur cinq contiennent les publications des chercheurs d'une université, d'un laboratoire ou département ou d'un organisme de recherche.

⁷ <http://www.andoar.org>

Quant aux revues en accès libre, le répertoire DOAJ⁸ contient 5 803 titres avec presque 500 000 articles. A cela s'ajoute un nombre toujours croissant d'autres revues dites « hybrides » publiées par des éditeurs commerciaux et partiellement libres.

L'évolution en France, notamment des archives institutionnelles, est rapide (cf. Schöpfel & Prost, 2010). Les sites les plus connus sont l'archive HAL, créée comme miroir de la première archive ouverte arXiv, la plateforme revues.org ou encore les sites d'archives PERSEE (revues SHS) ou NUMDAM (revues mathématiques).

Les archives ouvertes en France contiennent environ 10% des articles scientifiques publiés par des chercheurs. Comparé à l'édition commerciale, c'est un pourcentage significatif, certes, mais pas représentatif et ne suffit pas à satisfaire les besoins d'information des communautés scientifiques qui demandent l'accès le plus large.

Nous avons déjà évoqué l'émergence d'une nouvelle forme d'IST sous forme de *datasets*. Une partie de ces données sont en libre accès. En France, ces *datasets* constituent déjà environ 7% du contenu des archives ouvertes (2009).

4.6. Réseaux

Aujourd'hui, les réseaux sociaux proposent des services en lien avec l'information scientifique. Des réseaux professionnels, scientifiques ou privés comme Facebook, LinkedIn, Viadeo ou ResearchGate contiennent des informations utiles pour la carrière scientifique (offres d'emploi, recommandations etc.), pour la veille (alertes), pour l'élaboration de projets de recherche (contacts, carnets d'adresses, compétences, références, propositions de valeur) ou d'événements, pour le débat et l'échange (forums, groupes) ; mais ils contiennent aussi des documents, images, vidéos etc.

D'une certaine manière, un réseau comme ResearchGate ou LinkedIn se substitue partiellement à un *Current Research Information System* (CRIS). Utilisé de façon pertinente, un réseau social peut devenir une source d'information incontournable pour un chercheur ou une communauté scientifique.

4.7. Publications, éditeurs

Un chercheur a régulièrement besoin d'information sur les publications (revues, livres) et les éditeurs afin de sélectionner un vecteur de communication pour ses résultats. Souvent, il cherche des informations précises ; parmi ces informations, citons le facteur d'impact d'une revue, l'existence et la composition d'un comité de lecture, la procédure de soumission et la sélection de manuscrit (*peer review* etc.), la politique éditoriale et le *scope* de la revue, le nombre d'articles, les éventuels

⁸ <http://www.doaj.org/>

projets de numéro spécial, ou bien la politique de l'éditeur en matière de droit d'auteur (cession de droit, autorisation du dépôt dans une archive ouverte).

Où trouvera-t-il cette information ? Soit auprès de « son » bibliothécaire ou documentaliste ; soit en consultant les produits et services spécialisés, comme le Web of Science, Scopus, SCImago, Ulrichsweb ou encore SHERPA/RoMEO pour la politique des éditeurs en matière de libre accès. Ce dernier site aura prochainement un équivalent en France (Héloïse).

5. Conclusion : consommation et production scientifique

Jusqu'ici nous avons parlé des mutations du paysage de l'information scientifique et technique uniquement sous l'aspect de l'offre. Or, que savons-nous de l'utilisation réelle de cette offre, qu'il s'agisse de contenus ou de services ?

Un corpus croissant d'études et de recherches permettent aujourd'hui de mieux connaître et comprendre l'usage fait par les chercheurs et communautés scientifiques⁹.

Premier constat : « Scholars like it simple... they bounce » (Nicholas et al., 2009). Souvent, un chercheur ne passera pas beaucoup de temps pour rechercher de l'information sur Internet mais « zappera » d'une page à l'autre : « les usagers transposent leurs activités de navigation sur le Web sur les outils de recherche d'information en général, et les plates-formes d'éditeur en particulier (...) L'efficacité de cette stratégie (est jugée) à la quantité d'informations qu'ils brassent tout au long de cette navigation et à la pluralité des espaces qu'ils visitent (académiques, professionnels, commerciaux, associatifs,...) et des tâches qu'ils accomplissent simultanément. » (Boukacem-Zeghmouri, 2010b).

Deuxième constat : Les premières études ont probablement surestimé l'effet générationnel (« génération Google ») comme facteur-clé pour comprendre le comportement des chercheurs. Une relecture des résultats fait apparaître des différences au moins aussi importantes entre disciplines scientifiques. Il paraît tout à fait possible que l'usage des ressources en ligne ne fait que transposer les divergences disciplinaires dans l'environnement des NTIC et que l'effet générationnel et homogénéisant d'Internet n'est qu'apparent et transitoire. D'autres analyses confirmeront oui ou non cette hypothèse.

D'après ces études, une chose est certaine : Génération Google ou pas, les chercheurs ne sont pas toujours très soucieux de la légalité de leur comportement sur le Web, et ils ne sont pas vraiment fidèles non plus dans leur préférence pour tel ou tel site.

Depuis quelques mois, on accède aussi aux premiers résultats empiriques sur l'impact réel de l'utilisation des ressources en ligne.

⁹ Cf. Boukacem-Zeghmouri (2010a), Williams et al. (2010), Jamali et al. (2009), Jamali & Nicholas (2010)

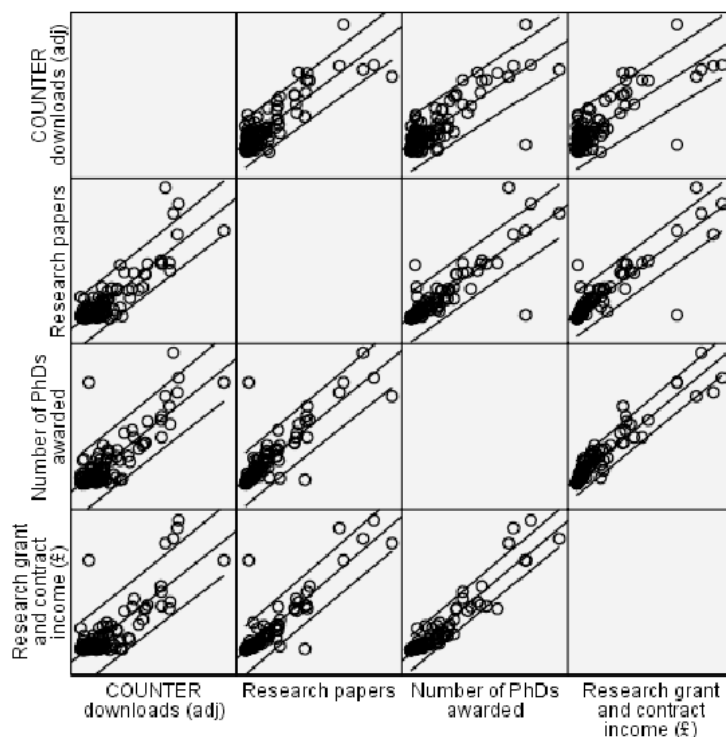


Figure 5 : Relation entre statistiques d'utilisation et production scientifique (RIN, 2009)

Récemment, le Research Information Network de Londres a publié un rapport qui, pour la première fois analyse les corrélations entre les statistiques d'utilisation (téléchargements) et la production scientifique d'un campus, mesurée par le nombre de publications, le nombre de thèses et le montant des subventions de recherche (RIN, 2009).

Le résultat est positif ; visiblement, il existe un lien significatif entre l'usage des revues électroniques et la production scientifique. Afin d'étudier la nature exacte de ce lien, l'équipe CIBER autour de D. Nicholas a lancé une étude longitudinale sur cinq ans – est-ce que la consommation de l'information scientifique incite et facilite la production, ou est-ce que la production scientifique stimule la consommation ? Il est tout aussi possible que l'effet agit dans les deux sens. Dans tous les cas de figure, cette recherche souligne – si besoin – le rôle stratégique de l'information scientifique. -

En guise de conclusion, répétons nos trois idées forces sur les mutations du paysage de l'information scientifique et technique :

1. L'information scientifique et technique remplit des fonctions précises.
2. Elle intègre les contenus scientifiques (publications, données etc.) aussi bien que les produits et services professionnels, à destination des professionnels de l'information ou, de plus en plus, des chercheurs eux-mêmes.
3. La mise en ligne de l'information scientifique a eu des effets aussi bien quantitatifs que qualitatifs sur la nature de cette information.

Ce chapitre a essayé d'apporter quelques éléments pour mieux comprendre ces changements. Pour autant, quatre questions essentielles n'ont pas (ou pas assez) été abordées :

1. L'analyse des usages et pratiques des chercheurs pour comprendre comment ils utilisent l'offre à leur disposition.
2. Les modèles économiques et la chaîne de valeur de l'édition scientifique, en particulier leur évolution dans l'environnement du commerce électronique et du libre accès.
3. L'environnement légal et juridique, notamment après le vote des lois récentes (DADVSI, HADOPI) mais aussi lié au code des marchés publics et au droit des contrats privés (pour les licences).
4. La question des institutions et des acteurs, c'est-à-dire qui fait quoi, avec qui, pourquoi et pour qui etc.

6. Bibliographie

Bollen Johan, Van de Sompel Herbert, Hagberg Aric, Bettencourt Luis, Chute Ryan, Rodriguez Marko A. Balakireva Lyudmila. Clickstream Data Yields High-Resolution Maps of Science. *PLoS ONE*, 2009, vol. 4, no. 3, e4803+.

Bollen Johan, Van de Sompel Herbert, Hagberg Aric Chute Ryan. A principal component analysis of 39 scientific impact measures. *PloS one*, 2009, vol. 4, no. 6, e6022+.

Boukacem-Zeghmouri Chérifa (dir.). *L'information scientifique et technique dans l'univers numérique. Mesures et usages*. Paris, ADBS Editions, 2010.

Boukacem-Zeghmouri Chérifa. Pratiques de consultation des revues électroniques par les enseignants chercheurs : les STM en France. *Documentaliste - Sciences de l'Information*, 2010, vol. 47, no. 2, 4-13.

Chaudiron Stéphane, Ihadjadene Madjid Maredj Azzeddine. La fragmentation et l'unité documentaire en question. In : *16eme Congrès de la SFSIC, 11-13 juin 2008, Compiègne*.

Chikate R. V. Patil S. K. Citation Analysis of Theses in Library and Information Science Submitted to University of Pune: A Pilot Study. *Library Philosophy and Practice*, 2008.

Conjaud Charlotte. *La place du livre numérique dans un centre de documentation médicale. Le cas du centre de documentation de l'hôpital Jeanne de Flandre*. Mémoire de Master, Université Charles-de-Gaulle Lille 3, 2009.

Dacos Marin, Mounier Pierre. *L'édition électronique*. Paris, La Découverte, 2010.

Delmotte Stéphanie. Publications scientifiques en sciences humaines. *Les Cahiers du Numérique*, 2009, vol. 5, no. 2, 53-84.

Drouard Françoise, Petitjean Sophie Schöpfel Joachim. Les portails d'information du CNRS. *Communication & Langage*, 2009, no. 161, 75-92.

Duval Raymond, Schöpfel Joachim. « Nouvelles formes de communication scientifique ». In: Roux Magalie (dir.). *Biologie: l'ère numérique*. Paris, CNRS Editions, 2009.

Farace Dominic J., Schöpfel Joachim (dir.). *Grey Literature in Library and Information Studies*. Berlin, De Gruyter Saur, 2010.

Gooden Angela M. Citation Analysis of Chemistry Doctoral Dissertations: An Ohio State University Case Study. *Issues in Science and Technology Librarianship*, 2001, no. 32.

- Heidorn P. Bryan. Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends*, 2008, vol. 57, no. 2, 280-299.
- Heil Kathleen A. Citation Analysis of Theses/Dissertations for a Marine Biological Laboratory. In: *2009 SAIL Conference: From the River to the Sea, 11-14 May 2009, Wilmington, North Carolina*.
- Hey Tony, Tansley Stewart, Tolle Kristin (dir.). *The fourth paradigm. Data-intensive scientific discovery*. Redmond, Microsoft Corporation, 2009.
- Hull Duncan, Pettifer Steve R. Kell Douglas B. Defrosting the digital library: bibliographic tools for the next generation web. *PLoS computational biology*, 2008, vol. 4, no. 10, e1000204+.
- Jamali Hamid R., Nicholas David Rowlands Ian. Scholarly e-books: the views of 16,000 academics: Results from the JISC National E-Book Observatory. *Aslib Proceedings: new information perspectives*, 2009, pp. 33-47.
- Jamali Hamid R. Nicholas David. Intradisciplinary differences in reading behaviour of scientists: Case study of physics and astronomy. *The Electronic Library*, 2010, vol. 28, no. 1, 54-68
- Mabe Michael. The growth and number of journals. *Serials*, 2003, vol. 16, no. 2, 191-197.
- Mabe Michael. « L'édition des revues scientifiques ». In : Schöpfel Joachim (dir.). *La Publication Scientifique*. Paris, Lavoisier, 2008.
- McKiernan Gerry. « E is for Everything: The Extra-Ordinary, Evolutionary [E-]Journal ». *The Serials Librarian*, 2002, vol. 41, n° 3-4, 293-321.
- Nicholas David, Rowlands Ian, Clark David, Jamali Hamid. Digital consumers: Virtual Scholars. In : *Ressources électroniques académiques: mesures et usages. Colloque international*. Lille, 26-27 novembre 2009.
- Nicholas, David, Rowlands, Ian Jamali, Hamid R. E-textbook use, information seeking behaviour and its impact: Case study business and management. *Journal of Information Science*, 2010, vol. 36, no. 2, 263-280.
- Pédauque Roger T. Document : forme, signe et médium, les re-formulations du numérique. Rapport, CNRS RTP-DOC, 2003.
- Pédauque Roger T. *La redocumentarisation du monde*. Paris, Cepaduès, 2007.
- RIN. E-journals: their use, value and impact. Rapport, Research Information Network, 2009.
- Savage Caroline J., Vickers Andrew J. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLoS ONE*, 2009, vol. 4, no. 9, e7078+.
- Schöpfel Joachim (dir.). *La publication scientifique. Analyses et perspectives*. Paris, Lavoisier Hermes Science, 2008.
- Schöpfel Joachim (dir.). Communication scientifique: les nouveaux enjeux. Numéro spécial. *Les Cahiers du Numérique*, 2009, vol. 5, no. 2.
- Schöpfel Joachim, Prost Hélène. Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie : Développement, Rapport, Université Charles-de-Gaulle Lille 3, 2010.
- Vajou Michel, Martinez Ruth Chaudiron Stéphane. Les enjeux économiques de l'édition scientifique, technique et médicale : Analyses et questions clés. *Les Cahiers du Numérique*, 2009, vol. 5, no. 2, 143-172.

Williams Peter, Nicholas David Rowlands Ian. E-Journal Usage and Impact in Scholarly Research: A Review of the Literature. *New Review of Academic Librarianship*, vol. 16, no. 2, 192-207.

Zarka Jean-Yves. L'évaluation : un pouvoir supposé savoir. *Cités*, 2009, vol. 37, no. 1, 113-123.