

# QUELLES TECHNOLOGIES INNOVANTES POUR LA VEILLE EN SCIENCES HUMAINES ET SOCIALES ?

LAURE ENDRIZZI

## INTRODUCTION

Dans le contexte qui nous intéresse aujourd'hui – la veille scientifique en sciences humaines et sociales – la dématérialisation croissante des accès à l'information scientifique, couplée à des formes de publication de moins en moins coûteuses et des modes de distribution de plus en plus centrifuges, met à l'épreuve les approches documentaires traditionnelles. Dans cet article, nous explorerons les usages en émergence autour de la syndication de contenus (RSS) et des services bibliographiques mutualisés. Nous évoquerons les nouveaux outils de mesure de la réputation, notamment via les *blogs*, qui apportent un éclairage inédit sur les logiques de réseaux qui se mettent en œuvre. L'apparition de ces technologies innovantes, quelquefois qualifiées de « *disruptive* », bouleverse nos pratiques informationnelles et nos repères traditionnels en termes d'évaluation de l'information scientifique. En illustrant notre propos d'exemples, nous tenterons ici d'offrir quelques pistes

pour appréhender ces usages et développer de nouvelles stratégies à l'égard de l'information scientifique.

## 1. SYNDICATION ET FLUX RSS : DE NOUVELLES FORMES POUR LA CIRCULATION DE L'INFORMATION

### LES ENJEUX, DE LA DIFFUSION À LA RÉCEPTION

La technologie RSS – dont l'acception la plus communément admise est « *really simple syndication* » – permet la diffusion et la réception de flux ou de fils d'informations textuelles, audios ou vidéos entre des ordinateurs connectés à Internet. Basée sur le langage XML, elle est signalée par une icône généralement orange **XML** ou **RSS**, ou par le terme de « syndication ». Elle a été popularisée par les blocs-notes (*blogs*) qui, dans une grande majorité, disposent de cette fonctionnalité, mais n'est pas réductible à cette forme de publication.

En matière de réception, le RSS permet à l'utilisateur d'automatiser les tâches répétitives de recherche d'informations sur un serveur : il peut s'abonner aux mises à jour d'un site Internet (ou de certaines de ses pages), sans qu'il lui soit nécessaire de s'identifier nominativement auprès du serveur, ni de fournir son adresse électronique. Cette procédure s'avère donc moins intrusive (pour le moment ?) que les alertes par courriel et listes de diffusion, qui souffrent de la recrudescence des *spams* et autres virus.

Ces deux modes de diffusion / réception que sont d'une part le courrier électronique et d'autre part le RSS ne sont pas d'ailleurs sans présenter une certaine porosité, et la question de la concurrence ne manquera pas de se poser de manière croissante à moyen terme. Il n'est sûrement pas anodin, en effet, que la dernière version du logiciel *Sympa*, qui gère l'ensemble des listes de diffusion du comité réseau des universités (CRU), intègre cette fonctionnalité d'abonnement RSS. Ainsi, les messages diffusés sur des listes comme *Tematice*, par exemple, peuvent être collectés soit via un fil RSS, soit plus traditionnellement via un message électronique. D'une manière générale, il est important de comprendre que c'est l'utilisateur qui garde à tout moment la main sur ses abonnements : on ne passe plus son temps à rechercher le message qui contient le lien à cliquer pour se désabonner d'une liste ; ni même à supprimer les messages sans intérêt : en effet, le flux est (normalement) paramétré en amont pour proposer une dizaine ou une quinzaine d'« items », chaque nouvelle actualité écrasant en quelque sorte la dernière. Avec le RSS, on s'inscrit pleinement dans une logique de consommation de l'information, basée sur une lecture par balayage.

Le RSS offre par ailleurs une flexibilité forte en termes de diffusion orientée usager : les contenus issus du serveur source sont non seulement mis à jour en temps réel, mais sont aussi manipulables, c'est-à-dire intégrables dans un site Internet ou intranet, sans mention juridique spécifiant explicitement leur réutilisabilité. Tout se passe comme si la mise à disposition via le RSS constituait une condition suffisante pour l'exploitation externe des contenus. Les revues de presse personnalisées sont donc en plein essor : chacun peut ainsi se

constituer sa propre collection de fils RSS et la rendre visible sur Internet, via des plateformes telles que *Bloglines* ou via son propre site Internet, moyennant quelques développements. La collection agrégée par le site *Homo Numericus* sur le thème « technologies numériques et société »<sup>1</sup> en constitue un exemple notoire. De même, on retrouvera sur le site du CAREP<sup>2</sup> de Paris les trois fils RSS produits par la veille scientifique et technologique (VST) de l'INRP : actualités des thèses, actualité des ouvrages et sommaires des revues françaises.

Comme nous l'avons évoqué, le RSS ne se limite pas à la diffusion d'informations textuelles : images, photos, documents audios ou vidéos sont (presque) aussi facilement véhiculées que du texte. Le RSS recueille la faveur des internautes avertis pour deux raisons majeures : d'une part, il évite d'accroître – et tend même à diminuer – la quantité des courriels reçus par l'utilisateur, et d'autre part il permet de dépasser l'obstacle des systèmes de filtrage anti-spam et anti-virus associés aux logiciels de messagerie, qui limitent les possibilités de réception de documents non textuels. La popularité actuelle des *podcasts* (flux audios) constitue là encore un indice de cette évolution. Dans le champ de l'éducation, on pourra voir par exemple le site du célèbre *think tank* américain *Educause*, qui offre plus de 500 fils RSS thématiques sur ses propres ressources, et sa plateforme d'actualité qui agrège les blocs-notes de ses membres, avec des flux RSS externes et des entretiens réalisés en *podcast*<sup>3</sup>. Une double tendance se dégage de ces nouveaux modes de diffusion. Du côté des producteurs ou des diffuseurs, on s'inscrit dans un mouvement centrifuge contraire à la logique de « portail » ou bien d'« annuaire » qui a prévalu dans les premières années de l'Internet, alors que, du côté des usagers, domine une dynamique dans laquelle les fonctions de personnalisation sont de plus en plus prégnantes : l'utilisateur crée son propre environnement de travail en agrégeant ses ressources favorites.

Si l'on comprend aisément l'intérêt de cette technologie dite de « *push* » qui, comme son nom l'indique, pousse l'information vers l'utilisateur, son appropriation n'a rien d'intuitif. Il est en effet extrêmement déroutant de cliquer sur un de ces petits logos oranges et d'ouvrir une

1. <http://www.homo-numericus.net/source.php3>

2. CAREP : centre académique de ressources pour l'éducation prioritaire, <http://carep.scola.ac-paris.fr>

3. <http://www.educause.edu/Syndicate/1586> ; <http://connect.educause.edu/>

page de métadonnées composée de balises « *channels* » et « *items* ». Pour être lisibles, ces liens doivent être insérés dans des outils de lecture spécifiques. Certains navigateurs et logiciels de messagerie comportent une telle fonctionnalité, mais il existe aussi des outils dédiés, appelés agrégateurs, qui permettent de se constituer une collection personnelle de fils RSS. Une présentation plus détaillée des fonctionnalités distinctives de ces outils est proposée en annexe de cet article.

## RSS POUR LA VEILLE SUR L'ACTUALITÉ DE L'ÉDUCATION

Les professionnels de la presse ont été parmi les premiers à offrir des services de syndication ; aujourd'hui, les éditeurs membres du GESTE (groupement des éditeurs de services en ligne), par exemple, proposent en téléchargement le lecteur *AlertInfo* qui agrège par défaut les fils d'actualités des principaux quotidiens et magazines de la presse généraliste et professionnelle française et permet à l'utilisateur de paramétrer ses choix (fort heureusement, vu la quantité d'informations disponibles !). L'Administration s'y intéresse également, si l'on en juge par cette fiche d'information publiée par *Service-public.fr* en février 2006 sur les sites publics diffusant leurs actualités en RSS<sup>4</sup>.

S'agissant des actualités de l'éducation, on pourra s'abonner aux flux proposés par *Educnet* ou par le site du Premier ministre (rubrique éducation). Moins formels, le bloc-notes de Jean-Paul Pinte, *Veilles en éducation*, ainsi que les sites du Café pédagogique et de Thot permettent également de suivre, via des fils RSS, l'actualité pédagogique francophone. Du côté de la presse, si la BBC et le *New York Times* possèdent des fils dédiés à l'éducation, rien de tel ne semble disponible dans l'hexagone pour le moment, sans doute parce que ces rubriques n'existent pas toujours en tant que telles dans nos journaux et magazines.

D'autres fils permettent de se tenir informé(e) des mises à jour opérées dans des bases de connaissances ou des banques d'objets pédagogiques : *PhyBase*, *PrimTice* et *Bien Lire* par exemple en France, ou bien *Careo*, *Humbul*,

*Maricopa Learning Exchange* ou encore *Merlot* dans la sphère anglophone. S'agissant des *podcasts*, le site *Education Podcast Network* est un exemple du genre, qui agrège un ensemble d'ores et déjà impressionnant de fichiers audios et vidéos relatifs à l'éducation<sup>5</sup>. D'autres exemples en sciences humaines et sociales peuvent être trouvés sur les sites suivants : *SOSIG* au Royaume-Uni, *LiensSocio*, *Calenda* et *Album des sciences sociales* en France. Ce recensement, non exhaustif, vise simplement à illustrer la prolifération croissante des flux RSS dans les champs qui nous concernent. Si trouver le(s) fil(s) RSS d'un site donné relève parfois d'un parcours du combattant (dans les rubriques « aide » ou « alerte », dans le pied de page, etc.), il suffit parfois d'utiliser un moteur de recherche généraliste avec une requête de type « rss » ou « syndication » et *nom du site* pour gagner un peu de temps. La recherche de flux thématiques peut également être opérée via des collections réalisées par des spécialistes disciplinaires, comme celle d'*Homo Numericus*, déjà évoquée :

- annuaire des flux RSS sur l'éducation, la FOAD et les TICE sur le site *Tribune e-formation* (francophone) ;
- signets pédagogiques agrégés par le Récit (Réseau pour le développement des compétences par l'intégration des technologies) au Québec ;
- *Rss-Xpress* : flux éducation et culture recensés par le JISC (Joint Information Systems Committee) pour le Royaume-Uni<sup>6</sup>.

Alors que ces listes sont présentées plutôt comme des collections à feuilleter (URL du fil, courte description et dernières actualités), certains sites ressemblent davantage à des méta-moteurs ou à des moissonneurs, en fournissant une interface de recherche unique à la fois sur les « *channels* » et sur les « *items* » : une sorte d'accès au texte intégral pour interroger à la fois les titres et les articles. Pour quelques exemples de ces services innovants, on peut consulter les sites suivants :

- la collection *Edu-Rss* établie par Stephen Downes : plus de 160 sources concernant l'éducation ;
- *LisFeeds* : plus de 240 sources internationales dans le domaine des bibliothèques et des sciences de l'information ;

4. <http://www.geste.fr/alertinfo/home.html> ; [http://www.service-public.fr/accueil/decouverte/decouvrir\\_rss.htm](http://www.service-public.fr/accueil/decouverte/decouvrir_rss.htm)

5. <http://epnweb.org/>

6. <http://www.tribune-eformation.com/filinfos.php3> ; <http://recit.org/signets/sedna/> ; <http://rssxpress.ukoln.ac.uk/>

– pour la sphère francophone, *Vcoop*, service coopératif de veille lancé par la FING : plus de 200 sources sur les technologies numériques<sup>7</sup>.

## RSS POUR LA VEILLE BIBLIOGRAPHIQUE

L'activité de veille bibliographique, au cœur de nos préoccupations, bénéficie également de la mise à disposition de la technologie RSS, même si les exemples sont essentiellement issus de la sphère anglo-saxonne pour l'instant. Les applications les plus fréquentes portent sur la surveillance des sommaires de revues, des nouveautés inscrites aux catalogues des libraires et des éditeurs, des nouvelles acquisitions des bibliothèques et des derniers dépôts effectués dans des archives ouvertes.

### *Pour les sommaires de revues*

Les éditeurs et agrégateurs de revues électroniques ont été les premiers à offrir un service de surveillance des sommaires via RSS.

Si Oxford University Press a été pionnier pour les SHS, aujourd'hui l'American Psychological Association, Cambridge University Press et l'University of Chicago Press lui ont emboîté le pas, et d'autres comme Emerald ont déjà annoncé la mise à disposition prochaine d'un tel service d'alerte<sup>8</sup>.

Du côté des agrégateurs, c'est Ingenta qui a ouvert le bal, suivi par Extenza et par Proquest qui, pour sa part, a opté pour un mode de diffusion par bouquets thématiques et non par titre de revue. ScienceDirect et Ebsco devraient rejoindre les rangs très prochainement<sup>9</sup>.

En France, le site *Revues.org* propose également des fils d'actualités pour chacun des périodiques mis en ligne à l'aide du logiciel *Lodel*. Mais tout prestataire de revues de sommaires peut également promouvoir de tels

services. La VST de l'INRP propose ainsi de suivre l'actualité d'une sélection de revues francophones en éducation, alors que les bibliothèques de l'enseignement supérieur commencent à offrir des fils sur les revues correspondant à leurs abonnements. Cette tendance devrait d'ailleurs se renforcer avec le projet *ToCRoSS* (Table of Contents by Really Simple Syndication) développé par le JISC (Joint Information Systems Committee) au Royaume-Uni, en collaboration notamment avec Emerald, pour permettre l'intégration automatique des sommaires des revues souscrites dans les catalogues de bibliothèques : l'extension sera réalisée en *open source* et pourra donc être exploitée à la fois par les éditeurs de logiciels et par les bibliothèques<sup>10</sup>.

### *Pour les catalogues en ligne*

Le principe de l'alerte est identique avec les catalogues d'éditeurs, de libraires et de bibliothèques. L'exemple d'Amazon, qui propose des fils RSS par catégorie et sous-catégorie de produits, est éclairant. D'une manière générale, les précurseurs sont à chercher du côté des éditeurs spécialisés en informatique. Dans la sphère francophone, seules les éditions Eyrolles semblent ainsi offrir un service d'alerte par RSS, à la fois sur les nouveautés d'un rayon et sur les résultats d'une navigation thématique ou d'une recherche par mots-clés. D'autres fils personnalisables sont également mis à disposition par les éditions O'Reilly, dans leur bibliothèque numérique *Safari Bookshelf*. S'agissant des bibliothèques de l'enseignement supérieur, les exemples sont nombreux, surtout aux États-Unis. Nous citerons celui des bibliothèques de l'université d'Illinois Urbana Champaign, qui offrent un service de création de fils RSS à la volée sur les nouvelles acquisitions, par domaine ou par cote Dewey<sup>11</sup>.

7. [http://www.downes.ca/xml/edu\\_rss.htm](http://www.downes.ca/xml/edu_rss.htm) ; <http://www.lisfeeds.com/> ; <http://www.vcoop.net/>

8. <http://www.oxfordjournals.org/help/rss.html> ; <http://www.apa.org/rss/homepage.html> ; <http://journals.cambridge.org/action/byFeeds?journalId=HEP&type=rss> ; <http://www.journals.uchicago.edu/rss.html>

9. [http://www.ingentaconnect.com/about/researchers/subscribing\\_to\\_rss\\_feeds](http://www.ingentaconnect.com/about/researchers/subscribing_to_rss_feeds) ; <http://www.proquest.com/syndication/rss/> ; <http://www.extenza-eps.com/>

10. <http://www.revues.org/> ; [http://www.jisc.ac.uk/index.cfm?name=project\\_tocross](http://www.jisc.ac.uk/index.cfm?name=project_tocross)

11. <http://www.amazon.com/exec/obidos/subst/xs/syndicate.html/002-8478710-8748007> ; [http://www.eyrolles.com/Accueil/Aide/presentation\\_syndication.html](http://www.eyrolles.com/Accueil/Aide/presentation_syndication.html) ; <http://safari.oreilly.com/?mode=file&file=tools> ; <http://www.library.uiuc.edu/newtitles/>

### **Pour la littérature grise et les archives ouvertes**

La littérature grise, encore mal représentée sur Internet, ne bénéficie guère de ces nouvelles technologies facilitant leur dissémination. Concernant les thèses et les mémoires en France, nous signalerons les deux fils proposés par MemSic et TEL ainsi que celui de la VST à l'INRP. Les fils thématiques (9 pour l'éducation) en provenance de la base Proquest des thèses et mémoires, *PDTD* (ProQuest Dissertations & Theses database), donnent pour chaque titre le type de document (*dissertation*, *PhD* ou *doctoral thesis*) et l'université de souenance, mais la consultation en texte intégral est réservée aux abonnés<sup>12</sup>.

Plus généralement, l'ensemble des archives ouvertes gérées par le logiciel *Eprints* dispose de cette fonctionnalité d'alerte<sup>13</sup>. C'est le cas au Canada pour l'archive de l'université de Québec à Montréal, et en France pour le serveur thématique *ArchiveSic* et les serveurs institutionnels de l'ENS LSH et des universités Lyon 2 et Paris 10 Nanterre.

## **2. DE LA GESTION INDIVIDUELLE À LA GESTION COLLECTIVE DE BIBLIOGRAPHIES**

### **LES SERVICES BIBLIOGRAPHIQUES EN LIGNE**

Force est de constater que les annuaires bibliographiques qui occupent les rayonnages des salles chercheurs de nos bibliothèques n'ont pas encore trouvé d'équivalent électronique, si ce n'est dans quelques cas isolés. Ces bibliographies sont en effet le plus souvent statiques, disponibles en HTML ou en PDF, et font rarement l'objet d'une actualisation régulière. À l'heure où la dématérialisation des contenus scientifiques, notamment via l'essor des archives ouvertes, devient un passage obligé pour la promotion des recherches, il est regrettable que ce volet bibliographique, pourtant lié à des données fortement structurées, demeure en quelque sorte le parent pauvre de l'information scientifique numérique. L'utilité de ces corpus bibliographiques, et

surtout de leur mise à jour périodique et de leur normalisation, relève cependant de l'évidence. Et les chercheurs ne sont pas sans savoir combien ce travail peut s'avérer chronophage, tout en demeurant accessoire dans l'activité globale de recherche.

Là encore, la mutualisation peut non seulement faciliter l'édition de bibliographies répondant aux normes de présentation en vigueur, mais aussi permettre des démarches d'appropriation enrichies, articulant gestion individuelle et gestion collective. Sur le principe qu'une référence donnée, sélectionnée pour sa pertinence, intéresse l'ensemble des chercheurs travaillant sur le thème, ne serait-ce qu'à l'échelle d'un laboratoire ou d'une équipe, les possibilités d'enrichissement de cette référence, par exemple à l'aide d'annotations et de citations, donnent tout leur sens à ce travail de saisie ou d'import préalable dans une application dédiée.

Les services bibliographiques électroniques sont globalement de trois ordres :

- des catalogues de références, liés à une base de données bibliographiques, qui disposent de fonctionnalités d'export dans différents formats (*Bibtex* ou *RIS*, le plus souvent) ;
- des applications dites « sociales », dont les références proviennent des usagers possédant un compte sur le site producteur ;
- des logiciels de gestion de bibliographies qui, eux, intègrent à la fois l'import et l'export pour la réalisation automatisée de bibliographies normalisées.

Toutefois, ces fonctionnalités sont rarement cumulatives. Si, chronologiquement, ce sont les bases de données bibliographiques qui ont en premier offert des formats d'export (*Blackwell Synergy* et *HighWire Press* par exemple), suivies progressivement par les archives ouvertes (celles gérées par *Eprints* notamment), cette fonctionnalité ne présente un intérêt que si elle est associée à un logiciel de bibliographie capable de lire ce format. Dans cette même catégorie, on signalera également le service offert par l'université allemande Trier, *DBLP*<sup>14</sup>, qui se présente comme une collection de références bibliographiques en « *computer science* », réalisée par une indexation systématique d'articles de

12. <http://www.proquest.com/proquest/rss/disstheses.shtml>

13. <http://www.eprints.org/software/archives/>

14. <http://dblp.uni-trier.de/>

revues et d'actes de colloques sélectionnés et dépouillés. Une initiative qui, malgré une interface un peu frustrée, pourrait valablement en inspirer d'autres...

Dans la deuxième catégorie, on trouve des sites tels que *Connotea* (*Nature*) et *CiteUlike*<sup>15</sup> qui répondent à une logique de mutualisation par les usagers. Si *Connotea* se présente davantage comme un outil de gestion de signets partagés, qui est à la communauté scientifique ce que des sites tels que *Del.icio.us* et *Furl* sont au grand public, *CiteUlike* gère véritablement des références bibliographiques en offrant des fonctionnalités d'import/export compatibles avec *EndNote* et *Bibtex*. Dans les deux cas, ces services franchissent un pas supplémentaire par rapport à la base *DBLP*, en ouvrant à tous la collecte de sites ressources ou de références, moyennant la création d'un compte sur leur site. Ils s'inscrivent dans la mouvance de ce qu'on appelle désormais les « logiciels sociaux » ou « outils de navigation sociale », qui postulent que la valeur ajoutée de ces sélections repose sur l'intelligence collective produite par l'activité des contributeurs. De tels outils répondent tout d'abord à un besoin, pour soi-même et/ou pour les autres, de stockage et d'annotations des ressources que l'on utilise : l'utilisateur peut en effet à tout moment choisir de rendre public tout ou partie de sa « bibliothèque virtuelle » privée. De plus, ces outils sont accessibles en permanence depuis n'importe quel ordinateur connecté à Internet.

Une des principales plus-values réside toutefois dans la dimension collaborative du service. À un premier niveau de mutualisation, vos collègues peuvent bénéficier de vos découvertes en s'abonnant par courriel ou par RSS aux mises à jour de l'ensemble ou de quelques catégories de votre archive. À un deuxième niveau, l'application permet de repérer les utilisateurs qui indexent les mêmes ressources que vous, de prendre connaissance de leurs commentaires et de parcourir, par rebond, leur propre collection. L'apposition de « tags » personnalisés, sortes d'étiquettes correspondant à des mots-clés libres associés aux ressources, est appelée « folksonomie » : elle permet cette navigation sociale et la création d'alertes transversales sur tel ou tel mot-clé, indépendamment du contributeur.

## LE CAS DE WIKINDX

Dans la troisième catégorie, les logiciels les plus connus (*EndNote*, *Reference Manager* ou *RefWorks*) sont des applications propriétaires qui, bien que bénéficiant d'une interface Internet, sont toujours commercialisées en fonction du nombre d'utilisateurs et donc restreintes, au mieux, à un usage en intranet. Ces logiciels s'avèrent peu performants dans l'articulation entre gestion individuelle et gestion collective de ressources bibliographiques. Le logiciel libre *Wikindx*<sup>16</sup>, sur lequel nous nous attarderons un peu, représente un des rares outils disponibles qui s'inscrit dans une logique multi-utilisateurs et multi-bibliographies. Il peut être installé sur un poste de travail ou sur un serveur Internet, pour un usage personnel ou collectif. Outre la fonction classique de saisie / modification et les possibilités d'import / export aux formats *Bibtex* et *EndNote*, il permet à l'utilisateur d'indexer librement les notices à l'aide de macro-catégories et de mots-clés, de les annoter avec des citations et des paraphrases. En termes d'édition, *Wikindx* propose l'export d'une sélection de références en HTML ou en RTF, selon différents styles bibliographiques prédéfinis (APA, Harvard, MLA, etc.) ou personnalisés via un outil d'édition intégré à l'application. La dernière version du logiciel (3.3) comprend un traitement de texte, qui facilite la gestion dynamique des citations dans un document en cours de rédaction. Dans un usage multi-utilisateurs, chacun dispose d'un compte, peut définir ses options personnelles (normes bibliographiques et styles de citation, notamment) et se constituer ses propres bibliographies de travail à l'aide d'une sélection de références : les sous-ensembles ainsi créés sont étroitement liés à leur utilisateur, qui seul peut ajouter ou enlever des références à sa bibliographie ; mais ils n'affectent pas le contenu de la base bibliographique générale, sorte de « pot commun » dans lequel sont agrégées l'ensemble des notices saisies.

Bien que complexe de prime abord, l'application fournit des services inédits qui prennent toute leur valeur dans la logique « wiki » soutenant le travail collaboratif et qui justifient pleinement l'effort de familiarisation initial. Un outil extrêmement puissant, donc, qui peut être

15. <http://www.connotea.org/> ; <http://www.citeulike.org/>

16. <http://wikindx.sourceforge.net/>

utilisé par les membres d'une même équipe de recherche pour agréger leurs références et partager des commentaires, tout en s'assurant de la « réusabilité » des informations ainsi capitalisées. Chaque utilisateur peut par ailleurs établir ponctuellement une sélection de références en vue de la constitution d'une bibliographie à insérer dans une publication donnée, et ce sans avoir à se préoccuper de saisir à nouveau ces références ni de les mettre en forme. Pour des exemples d'implémentation, on pourra consulter les bibliographies de la Wikimedia allemande et celles de la VST de l'INRP<sup>17</sup>, dont la production est depuis peu articulée avec les thèmes traités dans ses *Lettres d'information* mensuelles.

### 3. RÉSEAUX ET RÉPUTATION : VERS DE NOUVEAUX OUTILS POUR L'ANALYSE SCIENTOMÉTRIQUE

#### FACTEUR D'IMPACT ET CITATIONS CROISÉES

La détermination de « facteurs d'impact » est un service, notamment offert par Thomson Scientific, qui a pour objet d'évaluer la notoriété d'une revue et, par ricochet, celle des chercheurs qui publient dans cette revue. Ce service est basé sur l'analyse des citations dans la littérature scientifique. Dans la catégorie « éducation »<sup>18</sup> le *Social Sciences Citation Index* propose par exemple l'évaluation du facteur d'impact de plus d'une centaine de revues, pour l'essentiel nord-américaines.

Avec l'accroissement des documents en texte intégral, notamment des revues en accès libre sur Internet – on pourra s'en faire une idée en parcourant les répertoires *DOAJ* et *OpenJ Gate* –, les outils d'évaluation utilisés pour déterminer les facteurs d'impact demeurent insuffisants, en particulier pour ce qui concerne l'identification des experts. Avec une couverture linguistique marginale des revues non anglophones et une identification des revues ancrée, de fait, dans une temporalité relativement longue, des dispositifs alternatifs doivent se développer pour élargir la conception de la réputation basée, sur l'évaluation par les pairs<sup>19</sup>.

Cette problématique n'est pas nouvelle et certains sites marchands ont déjà adopté depuis plusieurs années des fonctionnalités permettant de guider l'internaute dans ses choix : la librairie Amazon s'appuie sur ses statistiques de vente (et donc sur la réputation de tel ou tel ouvrage) pour proposer des ouvrages thématiquement similaires à celui consulté ; le site d'enchères Ebay permet de prendre connaissance des notations respectives des acheteurs et des vendeurs afin d'évaluer le degré de confiance qui peut être placé dans telle ou telle transaction.

S'agissant des publications scientifiques, de plus en plus de plateformes bibliographiques permettent désormais d'identifier les références citantes d'une publication donnée, voire d'y accéder par rebond si les abonnements souscrits par l'institution autorisent la consultation du texte intégral : une fonctionnalité utile pour repérer les principaux auteurs actifs sur une thématique de recherche donnée.

C'est le cas bien évidemment de *Google Scholar*, lancé fin 2004, qui indique pour chaque résultat de recherche le nombre de citations correspondantes. Le service *Citeseer*, né d'une collaboration entre la National Science Foundation et Microsoft Research, indexe des publications en informatique et en sciences de l'information et offre des statistiques de citations, des liens sur les articles citants et cités ainsi que des citations en contexte. Compatible avec le protocole OAI (*open archive initiative*), ce service fonctionne comme un métamoteur pour interroger les archives ouvertes, rendant par là-même la majorité des références accessibles en texte intégral. Un outil remarquable, mais dont l'interface demeure là encore assez déroutante<sup>20</sup>.

#### LE CAS DES BLOCS-NOTES : DE LA POPULARITÉ PAR LES PAIRS À L'ANALYSE DES RÉTROLIENS

Un bloc-notes est, a minima, un site Internet dont les contenus sont publiés en ordre chronologique inverse, généralement selon une périodicité régulière. Il se caractérise par une grande simplicité d'usage et permet de publier directement sur Internet, seul ou à plusieurs, des

17. <http://bibliography.wikimedia.de/> ; [http://wikindx.inrp.fr/biblio\\_vst/index.php](http://wikindx.inrp.fr/biblio_vst/index.php) et [http://wikindx.inrp.fr/biblio\\_encyclen/index.php](http://wikindx.inrp.fr/biblio_encyclen/index.php)

18. <http://www.thomsonscientific.com/cgi-bin/jrnlst/jloptions.cgi?PC=J>  
<http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/>

19. <http://www.doaj.org/> et <http://www.openj-gate.com/> ; <http://www.isihighlycited.com/>

20. <http://scholar.google.com/> ; <http://citeseer.ist.psu.edu/>

textes, mais aussi des photos (*photoblog*), des vidéos (*videoblog* ou *vlog*) et des sons (*audioblog*), via un formulaire qui s'apparente à celui des courriels.

Techniquement, un bloc-notes dispose de plusieurs fonctionnalités, activées ou non par l'utilisateur, qui en conditionnent les usages :

- le *blogroll* : la liste des sites « amis » qui traitent de problématiques proches des centres d'intérêts du ou des « blogueurs » ; un territoire qui délimite l'espace informel dans lequel les échanges s'inscrivent, une sorte de « micro blogosphère », donc ;
- les commentaires : le plus souvent non modérés, ils permettent aux lecteurs d'apporter un point de vue ou un éclairage complémentaire sur chaque billet publié ; on assiste parfois à un véritable échange proche de celui des forums de discussion ;
- les rétroliens (*trackbacks*) : ils correspondent en quelque sorte à des commentaires distants signalés sur le bloc source au niveau de l'article, selon un dispositif automatisé de renvois sur les blocs citants ; un moyen donc d'émuler la conversation entre blocs-notes ;
- les mots-clés : ils permettent l'indexation thématique des billets et sont parfois compatibles avec des outils de partage de signets, sur le principe de la folksonomie déjà évoqué. Dans ce cas, l'apposition d'un mot-clé ou d'une étiquette crée automatiquement une association avec l'espace de gestion de signets en ligne, ouvert par le ou les « blogueurs » sur une plateforme dédiée, ce qui a pour effet de démultiplier le référencement d'un même billet et d'accroître sa visibilité sur la toile.

La hiérarchisation anté-chronologique des informations s'accompagne donc le plus souvent de modes de navigation alternatifs, opérant une proximité thématique à la fois interne et externe au bloc et favorisant les résonances entre des personnes dont les préoccupations se font écho, indépendamment de leurs statut, rattachement et zone géographique. Ces caractéristiques fonctionnelles ne conditionnent en rien la nature énonciative des contenus publiés : d'une part tout site « personnel » n'épouse pas forcément la forme d'un bloc-notes et d'autre part l'association courante entre bloc et journal intime réduit considérablement la diversité des usages. Une simple consultation des catégories

« *weblogs* » de l'OpenDirectoryProject ou de Yahoo, par exemple, permet de se représenter cette diversité des pratiques d'écriture<sup>21</sup>. Et le nombre de chercheurs utilisant un tel outil de publication pour donner à penser leurs problématiques de recherche ne cesse de croître.

Cet univers relationnel n'est pas sans intérêt pour la question de l'évaluation de la réputation qui nous intéresse présentement. On distingue deux types d'approches :

- la popularité auprès des pairs, qui permet d'identifier les fils RSS des blocs-notes les plus souscrits et les mieux notés par les utilisateurs enregistrés. Cette fonctionnalité est proche des liens corrélés (*related links*) proposés par certains moteurs de recherche ou sites commerciaux. Elle est basée sur le concept statistique d'affinités et elle est généralement disponible sur des plateformes dédiées à la lecture de fils RSS, telles que *Bloglines* ;
- la popularité par analyse des rétroliens (*tracking*), c'est-à-dire des liens pointant vers tel ou tel bloc-notes ou tel ou tel article publié sur un bloc ; le système s'apparente au calcul de la pertinence popularisé par Google et fournit une sorte de facteur d'impact.

Si ces fonctionnalités de rétroliens ne sont que partiellement exploitées par le moteur *Google Blog Search* (avec un lien « références » qui pointe sur les blocs citant le billet courant), d'autres moteurs de recherche généralistes ou des moteurs expérimentaux tels que celui du MIT Media Laboratory (*Blogdex*) offrent des indicateurs plus ou moins fins pour évaluer l'activité de ces blocs-notes : repérer les plus populaires, identifier des tendances thématiques, tracer les conversations entre blogueurs, etc. Parmi ces moteurs généralistes, nous retiendrons les plus connus : *Technorati* (avec un index de 41 millions de blocs en mai 2006), *Blogpulse* (30 millions) et *PubSub* (24 millions)<sup>22</sup>.

Il est tout à fait envisageable que, dans un avenir proche, les algorithmes utilisés, appliqués à des collections thématiques spécialisées, permettent de dégager des tendances, de repérer des experts, de suivre les échanges entre chercheurs, etc. Pour que de tels outils ouvrent de réelles perspectives qui viendraient compléter et enrichir celles fournies par la métrique traditionnelle, la masse critique est importante et la limite d'ordre culturelle est

21. [http://dmoz.org/Computers/Internet/On\\_the\\_Web/Weblogs/](http://dmoz.org/Computers/Internet/On_the_Web/Weblogs/) ;  
[http://dir.yahoo.com/Computers\\_and\\_Internet/Internet/World\\_Wide\\_Web/Weblogs/](http://dir.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Weblogs/)

22. <http://blogsearch.google.fr/> ; <http://www.technorati.com/> ; <http://www.blogpulse.com/> ; <http://www.pubsub.com/>



forte : une telle diversification repose sur l'hypothèse que les outils de navigation sociale bénéficieront d'une audience de plus en plus large et que le bloc-notes, et donc l'activité réflexive et l'auto-promotion qui sous-tendent son écriture, deviendront un mode de publication à part entière pour le chercheur qui, loin de s'opposer à l'énonciation scientifique traditionnelle, l'enrichit par les mises en perspective et les échanges suscités. Un pas qui ne sera certainement pas franchi sans réticence de la part d'une majorité des chercheurs, en particulier en France...

## CONCLUSION

Cette revue de technologies innovantes pour la veille scientifique explore quelques pistes, entre réalité et prospective, pouvant nourrir notre réflexion à la fois sur les usages émergents des chercheurs, mais aussi sur la pluralité des stratégies à mettre en œuvre à l'égard de l'information scientifique. Trois points retiendront notre attention dans cette conclusion. En premier lieu, l'hybridation et la démultiplication croissantes des accès à l'information rendues possibles par ces technologies assurent une visibilité biaisée de la production scientifique, limitée à quelques « *happy few* » qui ont su adopter très vite les nouveaux outils de promotion disponibles. En second lieu, les traditionnelles polémiques sur la validité de ces contenus librement accessibles doivent, nous semble-t-il, être replacées dans un contexte de transition, alors que de nouveaux modèles éditoriaux, encore immatures, sont en gestation. Enfin, nous soulignerons l'articulation croissante des logiques individuelles et des logiques collectives dans l'appropriation et l'évaluation des ressources scientifiques numériques. Des évolutions qui préfigurent ce que d'aucuns appellent le « *web 2.0* », et pour lesquelles il nous faudra encore patienter avant d'en dessiner plus nettement les contours.

Laure ENDRIZZI

chargée d'études et de recherche,  
cellule de Veille scientifique et technologique, INRP

## RÉFÉRENCES BIBLIOGRAPHIQUES

- ARCHAMBAULT E., VIGNOLA GAGNE E. (2004). *L'utilisation de la bibliométrie dans les sciences sociales et les humanités*. Conseil de recherches en sciences humaines du Canada.
- ARKO R. A., GINGER K. M., KASTENS K. A. *et al.* (2006). Using Annotations to Add Value to a Digital Library for Education. *D-Lib Magazine*, 12(5).
- BILLS D. B., HOLLIMAN S., LOWE L. *et al.* (2006). « The new mobile scholar and the effective use of information and communication technology ». *First Monday*, vol. 11, n° 4.
- DAVID S. & PINCH T. (2006). « Six degrees of reputation: The use and abuse of online review and recommendation systems ». *First Monday*, vol. 11, n° 3.
- DOWNES S. (2004). « The Buntine Oration : Learning networks ». In *Australian College of Educators and Australian Council of Educational Leaders conference*. Perth, Australia.
- HAMMOND T., HANAY T. & LUND B. (2004). « The role of RSS in science publishing: Syndication and annotations on the web ». *D-Lib Magazine*, vol. 10, n° 10.
- HERRING S. C., SCHEIDT L. A., BONUS S. & WRIGHT E. (2005). « Weblogs as a bridging genre ». *Information, Technology & People*, vol. 18, n° 2, p. 142-171.
- TONKIN E. (2006). « Folksonomies: The Fall and Rise of Plain-text Tagging ». *Ariadne*, n° 47.

### **Blocs-notes scientifiques**

- (Coll.). *Urfistinfo : Actualités des sciences de l'information*. <http://www.urfist.info/>
- (Coll.). *Many2Many: a group weblog on social software*. <http://many.corante.com/>
- (Coll.). *Into the blogosphere [rhetoric community and culture of weblogs]*. <http://blog.lib.umn.edu/blogosphere/>
- DOWNES S. *Stephen'web*. <http://www.downes.ca/>
- EFIMOVA L. *Mathemagenic: Giving birth to learning*. <http://blog.mathemagenic.com/>
- FIEDLER S. *Seblogging: Weblogs, CMS, and personal Webpublishing for learning and education* <http://seblogging.cognitivearchitects.com/>
- PAQUET S. *Seb's open research*. <http://radio.weblogs.com/0110772/>
- RICHARDSON W. *Weblogg-ed: The read/write web in the classroom*. <http://www.weblogg-ed.com/>
- SCHEIDT L. A. *Professional Lurker : Comments by an academic in cyberspace*. <http://www.professional-lurker.com/>

### **Bibliographies**

- (Coll.). *Reports and Surveys*. The NewPR/Wiki. <http://www.thenewpr.com/wiki/pmwiki.php?pagename=Resources.ReportsAndSurveys>
- (Coll.). *Theses and dissertations*. The NewPR/Wiki. <http://www.thenewpr.com/wiki/pmwiki.php?pagename=Resources.ThesesAndDissertations>

(Coll.). *The effect open access and downloads ("hits") on citation impact: bibliography of studies*. The Open Citation Project. En ligne : <http://opcit.eprints.org/oacitation-biblio.html>.

## ANNEXE : LES OUTILS DE LECTURE RSS

Nous évoquerons ici quelques outils utiles à la lecture de flux RSS. Pour plus d'informations, nous renvoyons au guide réalisé par Julia Jumeau de l'URFIST de Paris, qui constitue une entrée en matière incontournable<sup>23</sup>.

Pour être lisible, le lien d'un flux RSS doit être inséré dans un outil de lecture dédié, compatible RSS 2.0 et/ou Atom (les deux formats les plus fréquemment utilisés). Précisons toutefois que si la naissance de ces formats prend racine dans des logiques techniques et commerciales concurrentielles, leur différenciation ne présente guère d'intérêt pour l'utilisateur et la procédure d'abonnement est complètement transparente, quel que soit le format choisi.

Ces flux peuvent donc être lus avec plusieurs types d'outils plus ou moins spécialisés. Certains logiciels de messagerie tels que *Thunderbird* (via la création d'un compte « nouvelles RSS et blogs ») et certains navigateurs tels que *Safari*, *Opera* et *Firefox* intègrent ces fonctionnalités de lecture. Si ces deux solutions – messagerie et navigateur – restent viables pour gérer un petit nombre d'abonnements et/ou pour avoir rapidement accès à des sites que l'on surveille quotidiennement, elles demeurent toutefois insuffisantes pour un usage plus intensif, d'autant que l'information véhiculée par un flux RSS n'est pas pérenne.

Pour bénéficier cependant de fonctionnalités de gestion avancées, il convient de se tourner vers des outils de lecture natifs : les agrégateurs. Ces outils permettent par exemple de classer les fils souscrits dans des dossiers ; ils proposent une interface de navigation et des fonctions de recherche, autorisent l'archivage et l'annotation d'items sélectionnés et disposent de fonctionnalités d'import / export (format OPML). On distingue deux types d'agrégateurs : ceux qui doivent être installés sur le poste de travail (*Amphetadesk*, *BlogNavigator*, *RssOwl*, etc.) et ceux qui sont gérés directement en ligne (*Bloglines* et *MetaRss* par exemple). Ces derniers requièrent la création d'un compte sur le site distant : identifiant et mot de passe ouvrent l'accès à un espace personnel, privatif ou consultable par tous, au choix, pour gérer ses propres abonnements.

Des plateformes comme *My Yahoo!* ou *Netvibes* font preuve d'une plus grande polyvalence en associant par exemple la syndication de contenus avec la gestion de signets classiques, la consultation des courriels, l'intégration de services d'informations (météo, comparateur de prix, etc.) et la gestion d'agenda, pour constituer une véritable page d'accueil personnalisée sur Internet.

Bien évidemment, le fait que la lecture de ces flux ne soit encore que partiellement intégrée à nos outils quotidiens (messagerie et navigateur) constitue un frein réel à sa démocratisation : encore une nouvelle interface dont il faut s'approprier le fonctionnement... Mais nul doute que la dissémination extrêmement rapide de la technologie RSS va contribuer à la rendre de plus en plus soluble dans notre environnement électronique...

---

23. <http://www.ext.upmc.fr/urfist/rss/index.htm>