

La diffusion de l'information documentaire et des actualités en format RSS : un exemple de mise en place au Centre de Documentation en Santé Publique de Lausanne
Pablo Iriarte

► **To cite this version:**

Pablo Iriarte. La diffusion de l'information documentaire et des actualités en format RSS : un exemple de mise en place au Centre de Documentation en Santé Publique de Lausanne. Document numérique et société, 1ère édition, sous la dir. de Ghislaine Chartron et Evelyne Broudoux. Actes de la conférence organisée dans le cadre de la Semaine du document numérique à Fribourg (Suisse) les 20 et 21 septembre 2006. ADBS Éditions, 2006. 344 p. Collection Sciences et techniques de l'information, 2006. <sic_00079211v3>

HAL Id: sic_00079211

https://archivesic.ccsd.cnrs.fr/sic_00079211v3

Submitted on 25 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La diffusion de l'information documentaire et des actualités en format RSS

Pablo Iriarte

pablo.iriarte@chuv.ch

*Centre de Documentation en Santé Publique (CDSP)
Bibliothèque Universitaire de Médecine (BiUM)
Centre Hospitalier Universitaire Vaudois (CHUV) – Lausanne*

RÉSUMÉ. Dans une société de l'information en constante mutation, des pratiques innovantes émergent et une nouvelle « économie de l'attention » se dessine peu à peu. Au centre des nouveaux usages et de la nouvelle vague du « web 2.0 », la diffusion de l'information en format RSS est en train de modifier radicalement la façon dont humains et machines s'approprient l'information en provenance d'Internet, ainsi que les méthodes de veille et de diffusion sélective, employées par les professionnels de l'information. Ce texte est destiné à faire partager l'expérience acquise au Centre de Documentation en Santé Publique de Lausanne dans la mise en place et le développement des flux RSS, tant sur le site web (flux thématiques des actualités et nouvelles acquisitions) que dans l'OPAC (flux contextuels selon les critères de recherche)

MOTS-CLÉS : flux RSS, veille, DSI, XML, Web 2.0

1. Introduction

Dans une société de l'information en constante mutation, des pratiques innovantes émergent et une nouvelle « économie de l'attention » (LE CROSNIER, 2004, p. 12) se dessine peu à peu. Au centre de ces nouveaux usages, la diffusion de l'information en format *RSS*¹ est en train de modifier radicalement la façon dont humains et machines s'approprient l'information en provenance d'Internet, ainsi que les méthodes employées par les professionnels de l'information pour tenter de maîtriser son flux exponentiel.

Après les blogs² et les médias en ligne, premiers créateurs du contenu à avoir utilisé les possibilités du langage *XML*³ pour partager massivement l'information publiée dans leur site web, d'autres acteurs ont aussi commencé à offrir des *flux RSS*⁴ (pour avoir un répertoire plus détaillé, voir l'annexe « l'offre et la demande d'information en format RSS ») :

- Moteurs de recherche (*Google news*⁵, *Yahoo! news*⁶, *Wikio*⁷, *ASK*⁸, etc.)
- Bases de données bibliographiques (*PubMed*⁹, *SCOPUS*¹⁰, etc.)
- Editeurs scientifiques (*Nature Publishing Group*¹¹, *Oxford University Press*¹², *Blackwell*¹³, *Sage*¹⁴, etc.)¹⁵
- Archives institutionnelles (*ArXiv*¹⁶, *HAL*¹⁷, etc.)

¹ RSS est utilisé comme acronyme de « *Really Simple Syndication* », « *Rich Site Summary* », « *RDF Site Summary* » ou une autre variante de ces termes. Pour plus de détails, voir la page explicative faite par l'ADBS : <http://www.adbs.fr/site/repertoires/outils/rss.php>

² Outil de publication web personnel appelé aussi *weblog*, *carnet web*, *joueb*...

³ Extensible Markup Language (<http://www.w3.org/XML/>)

⁴ Appelés aussi parfois « *files RSS* », « *files d'info* », « *web feeds* », « *RSS feeds* », « *Atom feeds* »...

⁵ <http://news.google.com>

⁶ <http://news.yahoo.com>

⁷ <http://www.wikio.com>

⁸ Pour la partie « Blogs & Feeds » (<http://www.ask.com/?tool=bls>)

⁹ <http://www.pubmed.org>

¹⁰ <http://www.scopus.com>

¹¹ <http://npg.nature.com>

¹² <http://www.oxfordjournals.org>

¹³ <http://www.blackwell-synergy.com>

¹⁴ <http://www.sagepub.com>

¹⁵ L'ensemble de l'offre RSS actuelle des éditeurs scientifiques (environ 2000 titres) représente le **14% des titres électroniques** accessibles à l'Université de Lausanne et au CHUV

¹⁶ <http://arxiv.org>

¹⁷ <http://hal.ccsd.cnrs.fr>

En effet, cette forme de diffusion est parfaitement adaptée à l'activité de veille documentaire, tout en étant moins intrusive et plus facile à gérer que l'envoi régulier d'actualités et de références bibliographiques par courrier électronique (technique connue aussi sous le terme de « *push* »).

Cette généralisation de RSS comme format privilégié pour la diffusion des nouvelles informations et pour la *syndication de contenu*¹⁸ a contribué à la création de logiciels de lecture des flux RSS¹⁹ (également appelés *agrégateurs*) de tous genres et pour tous les usages possibles. A son tour, l'utilisation de plus en plus étendue de ces outils par les internautes a poussé à la mise en place des flux RSS dans les systèmes de gestion de contenu (CMS²⁰) utilisés dans les nouveaux sites web, dans un véritable processus de « *RSSification* » du *World Wide Web* (WWW).

Grâce à ce « cercle vertueux » (FIEVET, 2004, p. 192), le format RSS a acquis en quelques années la masse critique nécessaire à tout format d'échange pour s'imposer dans l'univers technologique de l'information numérique. De la même façon que les ondes radio et télévision coexistent, le WWW en format *HTML*²¹ et celui en format XML/RSS sont devenus deux véritables canaux parallèles mais complémentaires pour transmettre les informations sur Internet.

Travaillant dans un domaine scientifique, les collaborateurs de la Bibliothèque Universitaire de Médecine (BiUM) à Lausanne et, tout particulièrement les documentalistes du Centre de Documentation en Santé Publique (CDSP)²², ont très tôt été confrontés à l'émergence de ces nouvelles applications pour la production et la diffusion de l'information. La base de données PubMed, produite par la *National Library of Medicine*²³, est depuis plusieurs années à la pointe des développements technologiques et sert de référence en matière de service d'information pour les chercheurs et praticiens du domaine biomédical.

¹⁸ Anglicisme qui est utilisé sur Internet pour parler de la re-publication automatique sur un site B des dernières informations publiées par un site A, en se servant du flux RSS du site A par exemple

¹⁹ Outils apparentés aux logiciels de gestion du courrier électronique, chargés de gérer, d'actualiser et d'afficher les informations provenant des flux RSS. Il existe actuellement un bon nombre de ces lecteurs, pour toutes les plateformes ou hébergés sur Internet, gratuits ou payants. Voir par exemple la liste donnée par l'URFIST de Paris : <http://www.ext.upmc.fr/urfist/rss/agregateur.html>

²⁰ Content Management System

²¹ HyperText Markup Language (page officielle : <http://www.w3.org/MarkUp/>)

²² Le CDSP est une section de la BiUM produisant un site web et une base de données bibliographique orientée vers la recherche thématique. Il a rejoint les locaux de la BiUM fin 2004

²³ <http://www.nlm.nih.gov>

Cependant, les services d'information documentaire dans leur ensemble, quel que soit leur domaine, ne sont pas restés indifférents à cette évolution. En effet, un tiers des professionnels utilisent maintenant les flux RSS²⁴ dans l'activité de veille ou pour leurs besoins courants d'information (informations générales, autoformation, loisirs...) (BROCHARD, 2005). Certains d'entre eux sont aussi devenus acteurs dans la *blogosphère*²⁵ (GARREAU, 2005) ou participent à des *wikis*²⁶ collaboratifs.

D'autre part, de plus en plus de bibliothèques et centres de documentation maintiennent des blogs (VOGEL, 2005) ou produisent des flux RSS pour des besoins très divers, qui vont de la communication d'informations pratiques concernant le service (horaires, manifestations, etc.), à la publication en flux des produits documentaires plus ou moins élaborés (listes thématiques des dernières acquisitions, dossiers documentaires, etc.) (ÇELIKBAŞ, 2004). Des centres de documentation spécialisés diffusent aussi par ce biais un condensé d'actualités touchant leur domaine d'activité (JUMEAU, 2005).

Si la veille est l'un des domaines les plus bouleversés par l'arrivée du format RSS (DESCHAMPS, 2003 et 2005), la *diffusion sélective de l'information* (DSI) est aussi en passe de l'être. En effet, des outils permettant de générer automatiquement des courriels à partir des flux RSS et de gérer les listes de diffusion qui y sont associées²⁷ commencent à être disponibles sur le web. Cette nouvelle application du format RSS permet d'amener l'information disponible dans les flux aux utilisateurs qui n'emploient pas d'agrégateur ou qui préfèrent la messagerie électronique comme outil de travail. En outre, des outils de mixage²⁸

²⁴ 60% des professionnels de l'information connaissent les flux RSS, et 33% les utilisent (GARREAU, 2005)

²⁵ Il existe plusieurs répertoires de blogs faits par des professionnels de l'information ou par des institutions, comme par exemple celui du projet « Open directory » (http://pscontent.com/od2/opendirectory.php?browse=/Reference/Libraries/Library_and_Information_Science/Weblogs/) ou celui du « Libdex » (<http://www.libdex.com/weblogs.html>)

²⁶ Outil de publication web instantané et ouvert aux modifications des utilisateurs. Il est utilisé par exemple pour le projet Wikipédia (<http://fr.wikipedia.org/wiki/Wiki>). Dans le monde des bibliothèques, il existe plusieurs wikis comme Bibliopedia (<http://biblio.wikia.com/>) ou LISWiki (<http://www.liswiki.com>)

²⁷ Par exemple FeedBlitz (<http://www.feedblitz.com>), FeedBurner (<http://www.feedburner.com/fb/a/publishers/emailsub>), Squeet (<http://www.squeet.com>) ou Zookoda (<http://www.zookoda.com>)

²⁸ Par exemple RSS Mix (<http://www.rssmix.com/>), Feed Digest (<http://www.feeddigest.com/>), FeedRinse (<http://www.feedrinse.com/>) ou xFruits (<http://www.xfruits.com/>)

et de filtrage²⁹ des flux RSS permettent de créer des flux personnalisés combinant plusieurs sources d'information et adaptés à chaque profil de veille.

Malgré cette évolution et l'univers des possibilités qui s'ouvre grâce à l'utilisation de RSS (JDEY, 2006), la plus grande partie de l'information produite et gérée par les bibliothèques et centres de documentation est encore loin d'être disponible selon ce nouveau moyen de diffusion. Plusieurs causes expliquent probablement ce retard (BROCHARD, 2005) : la méconnaissance de RSS par une bonne partie encore des professionnels et des utilisateurs, le « déficit de compétences techniques dans les bibliothèques » et le « désintérêt ou méconnaissance » de la plupart des éditeurs commerciaux de systèmes intégrés de gestion de bibliothèques (SIGB) qui n'ont pas la même vitesse de réaction face aux changements que certains logiciels libres comme *Koha*³⁰ et *PMB*³¹, qui ont intégré dernièrement la production des flux RSS pour les nouvelles acquisitions ou l'affichage des flux externes dans le catalogue. De nombreux CMS open source intègrent aussi cette technique depuis quelque temps³², tout comme le navigateur *Mozilla Firefox*³³ qui l'a fait en 2005, bien avant son grand concurrent propriétaire *Explorer*³⁴ qui prévoit de le faire seulement dans sa version 7 annoncée pour cette année.

La production native des flux RSS, thématique ou selon les critères de recherche, est cependant annoncée pour les futures versions d'une partie des logiciels propriétaires de gestion documentaire³⁵, mais le sera-t-elle sans coûts supplémentaires ?

²⁹ Par exemple Feed Findings (<http://www.feedfindings.com/>) et Feed Digest

³⁰ <http://www.koha.org>

³¹ <http://www.sigb.net/>

³² Par exemple Drupal (<http://drupal.org>), SPIP (<http://www.spip.net>), Lodel (<http://www.lodel.org>), Mambo (<http://www.mamboserver.com>) ou Zope (<http://www.zope.org>)

³³ <http://www.mozilla.com/firefox/>

³⁴ <http://www.microsoft.com/windows/ie/>

³⁵ SIGB Horizon de Dynix (<http://www.sirsidynix.com/Resources/Pdfs/Solutions/Products/Unicorn.pdf>), Aleph 500 de Ex-Libris (<http://www.exlibrisgroup.com/newsdetails.htm?nid=456>), Vubis Smart de Geac (http://www.library.geac.com/object/Rel2.4.1_LIB.html), Cadic (<http://www.cadic.fr>) et Alexandrie de GB-Concept (<http://www.gbconcept.com/media/rss/filrss.htm>)

2. Un pas vers le web 2.0

La diffusion et le partage de l'information en format RSS fait partie, avec les blogs, les wikis et les podcasts³⁶, de la dernière grande révolution de la sphère Internet, qui a fortement modifié les rôles d'auteur, éditeur et lecteur (BROUDOUX, 2005), mettant l'utilisateur au centre de l'univers informatif. RSS est ainsi devenu le dénominateur commun entre les différentes pièces du nouveau système, il permet le développement d'une multitude d'applications qui peuvent alors agir sur l'ensemble des éléments simultanément.

En effet, l'apparition de ces outils interactifs de publication web instantanée, ainsi que les éditeurs en ligne de type WYSIWYG³⁷, les outils sociaux de catégorisation³⁸ et de partage de signets, références bibliographiques ou images (*Del.icio.us*³⁹, *Connotea*⁴⁰ et *Flickr*⁴¹, par exemple), développés dans la plupart des cas grâce à l'existence et à la maturité d'une plateforme complète, basée uniquement sur des logiciels libres et orientée web (plateforme « LAMP » : système d'exploitation *Linux*⁴², serveur *Apache*⁴³, logiciel de gestion des bases de données *MySQL*⁴⁴ et langage de script *PHP*⁴⁵), sont autant de phénomènes qui ont contribué à un développement très rapide d'un nouveau web, dont certains ont voulu marquer le passage ou le saut technologique en le nommant « *web 2.0* »⁴⁶. Plus adaptées à l'échange dynamique d'informations entre les machines grâce à

³⁶ Contraction de « iPod » et de « broadcasting ». Forme de flux RSS auquel on ajoute le lien vers des fichiers multimédia qui sont alors disponibles directement à partir du lecteur RSS ou téléchargeables automatiquement dans un baladeur numérique.

³⁷ WYSIWYG est l'acronyme de la locution anglaise « What You See Is What You Get ». Les interfaces de ce type sont utilisées dans les logiciels de mise en page et surtout dans les plateformes de blogging pour écrire facilement pour le web sans connaître le langage HTML

³⁸ Aussi appelée *tagging* ou *folksonomie*

³⁹ <http://del.icio.us/>

⁴⁰ <http://www.connotea.org/>

⁴¹ <http://flickr.com/>

⁴² <http://www.linux.org>

⁴³ <http://www.apache.org>

⁴⁴ MySQL est un logiciel libre de gestion de bases de données de type SQL (Structured Query Language). Site officiel : <http://www.mysql.com/>

⁴⁵ PHP est l'acronyme récursif de « PHP Hypertext Preprocessor ». PHP est un langage de script qui est très utilisé pour créer des sites web dynamiques. Site officiel : <http://www.php.net/>

⁴⁶ Voir l'article fondateur de Tim O'Reilly « What Is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software » <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>. Version française : <http://web2rules.blogspot.com/2006/01/what-is-web-20-par-tim-oreilly-version.html>

l'utilisation de XML, les applications de type 2.0 devraient permettre de mieux répondre à la diversité des besoins et pratiques des utilisateurs.

D'un web centré sur les serveurs et les grands centres créateurs de contenus, nous sommes passés à un web centré sur les utilisateurs/clients. L'essor des applications basées sur *AJAX*⁴⁷ (*Live.com*, *del.icio.us*, *Flickr*, etc.) et l'utilisation des services web ou des *APIs*⁴⁸ (mises à disposition par exemple par *PubMed*⁴⁹, *CrossRef*⁵⁰, *Google*⁵¹, *Yahoo!*⁵², *Amazon*⁵³, *Del.icio.us*⁵⁴, ou *Bloglines*⁵⁵) est un bon exemple de la façon dont l'information est actuellement traitée par les applications « *behind the screen* ».

L'exemple en date le plus frappant de cette révolution « *anti-copernicienne* », qui remet l'utilisateur au centre du système, pourrait être Wikio, un outil – en phase de test – dans la lignée de digg.com, qui a pour slogan « *Le média c'est vous* ». Il s'agit d'une base de données d'actualités francophones, constituée par l'agrégation de milliers de flux RSS (dont celui du CDSP) en provenance des médias, de sites web et des blogs⁵⁶. Chaque utilisateur peut y ajouter de l'information, la commenter ou la plébisciter. Toute nouvelle information indexée est catégorisée automatiquement en fonction de son contenu et classée selon une arborescence thématique maintenue par des documentalistes.

Comme c'est le cas dans Wikio, d'une manière générale, l'une des caractéristiques communes à toute application de type social ou « 2.0 » reste l'abondante production des flux RSS qui sont offerts aux utilisateurs/acteurs à des niveaux différents de granularité de l'information :

⁴⁷ « Asynchronous JavaScript And XML ». C'est un ensemble de techniques qui permet à une page web d'échanger des informations externes sans devoir être actualisée. Voir l'article fondateur de Jesse James Garrett « Ajax: A New Approach to Web Applications » (<http://www.adaptivepath.com/publications/essays/archives/000385.php>)

⁴⁸ « Application Programming Interface ». Il définit la manière dont les applications peuvent communiquer entre elles

⁴⁹ http://eutils.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html

⁵⁰ http://www.crossref.org/02publishers/openurl_info.html

⁵¹ <http://www.google.com/apis/>

⁵² <http://developer.yahoo.com/>

⁵³ <http://www.amazon.com/gp/browse.html/103-5407007-6535845?node=3435361>

⁵⁴ <http://del.icio.us/help/api/>

⁵⁵ <http://www.bloglines.com/services/api/>

⁵⁶ Voir l'interview de son directeur Pierre Chappaz : « Traiter de l'info en ne référant que des médias traditionnels serait réducteur » <http://www.liberation.fr/page.php?Article=369355>

- Flux généraux ou chronologiques (verticaux)
- Par thème, par « tag » ou par auteur (horizontaux)
- Spécifiques selon les critères de recherche (ponctuels)
- Pour les commentaires ou les citations (parallèles)
- Pour chaque élément isolé, pour chaque billet d'un blog et ses commentaires, chaque page d'un wiki et ses modifications (individuels)

Après avoir suivi avec une attention particulière toutes ces innovations depuis 2005, nous avons essayé d'adapter notre activité de veille pour y intégrer les flux RSS publiés par des blogs spécialisés, par des bases de données bibliographiques, des éditeurs scientifiques et des sites web institutionnels, etc. En outre, dans le but d'introduire l'ensemble des collaborateurs du service à l'utilisation de ces nouveaux outils, nous avons essayé de gérer en interne la publication d'un blog⁵⁷ et d'un wiki⁵⁸ tout en organisant plusieurs formations pour les collaborateurs sur les blogs et le format RSS, ainsi que sur la façon de les exploiter dans l'activité quotidienne de veille à l'aide d'un agrégateur adapté⁵⁹.

3. Implémentation des flux RSS : site web et OPAC

Une fois familiarisés avec ces outils, l'étape suivante a été logiquement l'implémentation de la diffusion des informations produites par le CDSP en format RSS. En effet, lors de la restructuration du site web réalisée entre octobre et novembre 2005, nous avons opté pour la réalisation en interne du développement web nécessaire permettant d'offrir différents types de flux RSS, tant sur le site web⁶⁰ que dans l'OPAC⁶¹ de la base de données du réseau documentaire en santé publique *SAPHIR*⁶² dont le CDSP est le responsable.

Parmi les différents types de flux RSS, nous avons choisi le format 2.0⁶³ pour les raisons suivantes :

⁵⁷ <http://blog.bium.ch>

⁵⁸ <http://www.bium.ch/wiki/doku.php>

⁵⁹ C'est finalement l'agrégateur web « social » Bloglines (<http://www.bloglines.com>) qui a été choisi après une petite phase de test de quelques agrégateurs en ligne. Nous avons dû renoncer d'emblée aux lecteurs RSS installés sur les postes, car notre service informatique craignait une surcharge de la bande passante générée par une multiplicité des lecteurs individuels.

⁶⁰ <http://www.saphirdoc.ch/cdsp.htm>

⁶¹ Online Public Access Catalog

⁶² *Swiss Automated Public Health Information Resources* (<http://www.saphirdoc.ch>)

⁶³ <http://www.rssboard.org/rss-specification>

- C'est le format le plus simple à implémenter
- C'est le plus utilisé d'après les chiffres trouvés sur le web⁶⁴
- Il a été choisi par PubMed, ce qui représente une garantie de compatibilité
- Il est accepté par Google comme « *sitemap* »⁶⁵

Outre un gain de temps espéré dans l'activité de veille faite directement par les utilisateurs ou par l'intermédiaire des documentalistes, ainsi que les aspects positifs purement formateurs liés à l'appropriation d'une nouvelle technologie et à la maîtrise d'une norme de type XML, nous avons aussi voulu apporter, avec ces développements, des améliorations dans cinq points sensibles qui nous touchent particulièrement :

1. Améliorer la visibilité de l'activité de notre centre
2. Faciliter l'activité de veille sur l'OPAC
3. Mettre en place une DSI qui puisse couvrir une plus grande partie de l'ensemble des ressources électroniques ou papier disponibles
4. Faciliter l'indexation du contenu de l'OPAC par les moteurs de recherche comme Google et réduire la charge du serveur web
5. Permettre la syndication du contenu produit par le CDSP dans les sites web partenaires

Pour le site web du CDSP, les flux sont générés dynamiquement à partir d'une recherche implicite selon la thématique associée aux documents catalogués et aux nouvelles entrées du site (actualités, agenda et adresse pour la formation continue). L'icône {xml.jpg} placée à côté de chaque thème permet un repérage facile.

Pour la base de données du réseau SAPHIR, le développement réalisé permet la création des flux RSS générés dynamiquement selon les critères de recherche. Ainsi, quand une recherche est effectuée dans l'OPAC, outre la liste de résultats habituels, ou le message signalant son absence, l'utilisateur obtient l'*URL*⁶⁶ du flux RSS correspondant aux critères de recherche utilisés, signalé par ce même icône, comme cela se fait habituellement :

{opac.jpg}

⁶⁴ 68% des flux selon syndic8 (<http://www.syndic8.com/stats.php?Section=rss#tabtable>)

⁶⁵ <https://www.google.com/webmasters/sitemaps/docs/en/other.html#feed>

⁶⁶ Uniform Resource Locator (<http://www.w3.org/Addressing/URL/Overview.html>)

Figure 1 : page des résultats dans l'OPAC avec l'icône XML signalant la présence du flux RSS contextuel

La souplesse du module de publication web de notre SIGB⁶⁷ nous a permis de créer facilement les pages XML dynamiques en appliquant le même schéma de fonctionnement que celui qui intervient pour la publication des résultats de recherche de l'OPAC avec les adaptations nécessaires au format XML :

{html.jpg}

Figure 2 : parallélisme entre les deux formes de publication web : HTML et RSS

Pour les flux thématiques en provenance du site web l'adaptation est similaire, seul l'élément <category> est ajouté pour décrire la thématique transversale.

Bien que la plupart des producteurs de flux RSS préfère la publication des extraits sommaires⁶⁸, par souci de protection de ses données ou envie d'attirer le lecteur vers la page web originale, nous avons opté pour la syndication des informations presque aussi complètes que celles données dans l'OPAC. Seul le résumé documentaire est omis dans le flux car il n'est actuellement disponible que pour les abonnés identifiés sur le site.

Le but du développement étant centré sur la diffusion, nous avons introduit les métadonnées complètes des documents, ajoutant l'indexation et la date de création de la notice par rapport aux métadonnées sommaires affichées dans les listes de l'OPAC, mais seulement au niveau de l'élément <description>, sans les utiliser de manière plus structurée à l'aide des éléments *Dublin Core*⁶⁹. Ceci serait souhaitable dans une deuxième phase de développement.

Pour introduire les métadonnées dans cet élément, nous avons utilisé la technique la plus courante et la plus pratique, aussi utilisée par PubMed, qui consiste à introduire directement le code HTML à l'intérieur d'une section *CDATA*⁷⁰ pour indiquer au *parseur XML*⁷¹ d'ignorer son contenu (autrement les balises HTML

⁶⁷ Alexandrie, édité par la société française GB-Concept (<http://www.gbconcept.com>)

⁶⁸ 84% des flux proposent uniquement des informations sommaires selon une étude menée par Pheedo en 2006 : http://www.pheedo.info/pheedread/Pheedo_Pheed_Read_3_Spring.2006.pdf

⁶⁹ <http://web.resource.org/rss/1.0/modules/dc/>

⁷⁰ Le texte à exclure du parsing doit être précédé par <! [CDATA[et terminée par]]> <http://en.wikipedia.org/wiki/Cdata>

⁷¹ Logiciel chargé d'analyser le fichier XML et d'en extraire les informations

qui ne respectent pas la norme stricte *X-HTML*⁷², ou chaque symbole « & », provoquerait une erreur fatale⁷³). Cette technique a quelques avantages importants : elle permet d'une part de réutiliser, presque sans faire de changements, le code de programmation interne⁷⁴ chargé de générer la page HTML de la notice complète dans l'OPAC. D'autre part, l'affichage des informations dans les agrégateurs est plus convivial car il garde une certaine mise en page (retours de ligne, listes à puces, tableaux, gras, italique,...). En outre, cela donne la possibilité de naviguer à partir des liens hypertextes situés à l'intérieur de cet élément : lien vers le texte intégral, par auteur, revue, descripteur et autres liens utiles.

{bloglines.jpg}

Figure 3 : affichage dans l'agrégateur Bloglines des éléments syndiqués par le site web

4. Résultats

Actuellement, l'application offre les flux RSS suivants, qui ont généré au total 39263 requêtes entre janvier et avril 2006, soit **le 23% des requêtes du serveur web dans cette période** :

Flux RSS verticaux :

- Pour l'ensemble des documents catalogués dans la base de données du réseau SAPHIR⁷⁵ : nouvelles acquisitions, articles dépouillés, sites web, etc.
- Pour l'ensemble de l'information produite par le CDSP⁷⁶, que ce soit au niveau du site web (actualités, agenda et adresses pour la formation continue) ou de la base de données SAPHIR (documents catalogués par le CDSP)

Flux RSS horizontaux :

⁷² <http://www.w3.org/TR/xhtml1/>

⁷³ <http://www.w3.org/TR/2004/REC-xml-20040204/#dt-fatal>

⁷⁴ Langage 4D (<http://www.4d.com/>)

⁷⁵ <http://www.saphirdoc.ch/rss.xml>

⁷⁶ <http://www.saphirdoc.ch/rss.xml?q=cdsp>

- Par thèmes du CDSB : chaque information introduite au niveau du site web et chaque document catalogué par le CSDP dans la base de données SAPHIR se voit attribuer un ou plusieurs thèmes à partir d'une liste préétablie de 47 catégories⁷⁷

Flux RSS ponctuels :

- Pour toute recherche effectuée dans l'OPAC de la base de données SAPHIR
- Pour chaque sous-ensemble de documents établi à l'aide des recherches sous-jacentes (« dossiers » documentaires ou bibliographies dynamiques proposées dans l'OPAC)

Après quelques mois d'expérience et une utilisation en hausse non négligeable de la part des utilisateurs (malgré le manque de publicité à ce stade de développement), nous commençons à explorer également toute une série de nouvelles applications qui deviennent possibles grâce à la création de ces flux RSS :

- Utilisation du flux RSS en tant que « *sitemap* » dans *Google*. Ceci permet une meilleure indexation par le moteur de recherche du nouveau contenu introduit dans la base de données
- Création des listes de diffusion par mail selon des critères de recherche (*DSI*) grâce à des outils comme *FeedBlitz* ou *Zookoda*, qui diffusent le contenu du flux RSS en forme de courriels
- Création d'un tableau dynamique avec les dernières nouvelles (concernant un thème en particulier par exemple) facilement re-publiable par un site partenaire en ajoutant simplement une ligne javascript dans le code *HTML* de la page⁷⁸

Nous avons également exploré la possibilité d'ajouter dans l'OPAC l'information syndiquée par les éditeurs de journaux scientifiques auxquels le service a accès. Dans ce sens, nous avons ajouté un champ à notre base de données pour répertorier les flux RSS des périodiques gérés qui en possèdent. Ainsi, grâce à l'utilisation du service web *Feed2JS*⁷⁹ qui transforme le flux RSS en code javascript, quand la notice détaillée d'un périodique possédant un flux RSS (édité

⁷⁷ http://www.saphirdoc.ch/cdsp_themes.htm

⁷⁸ en utilisant par exemple l'outil « JavaScript RSS Box Viewer » (<http://p3k.org/rss/>)

⁷⁹ <http://jade.mcli.dist.maricopa.edu/feed/>

par *Nature* ou *Oxford University Press* par exemple) est affichée dans l'OPAC, le contenu du dernier numéro de la revue est chargé automatiquement dans la même fenêtre.

5. Développements futurs

Quelques fonctionnalités manquent encore aux flux proposés et pourraient être développées pour compléter notre offre :

- Créer des flux disponibles dans la navigation à travers les index et les liens internes (par auteur, revue, collection ou descripteur). Pour cela il faudrait développer aussi un « permalien »⁸⁰ pour les notices d'autorité.
- Ajouter des métadonnées, avec les éléments *Dublin Core*, aux items qui concernent les documents.
- Récupérer les critères de recherche pour les réutiliser dans le titre des flux dynamiques
- Produire des flux RSS dans les autres formats utilisés actuellement (*1.0*⁸¹ et *Atom*⁸²)
- Créer une feuille de style *XSL*⁸³ pour afficher les pages XML des flux d'une façon plus lisible pour les personnes
- Pousser le développement des flux RSS 2.0 pour arriver au format « *open search* »⁸⁴, véritable service web orienté vers la recherche fédérée

⁸⁰ Contraction de « Lien permanent » venue du monde des blogs. Afin d'exploiter l'élément <guid> (globally unique identifier) dans les flux, nous avons développé ce type de lien à partir d'une recherche implicite par l'identifiant unique que chaque notice possède dans la base de données. Vu son intérêt, il est aussi utilisé maintenant dans l'OPAC, et affiché d'une façon explicite dans chaque notice complète

⁸¹ <http://web.resource.org/rss/1.0/>

⁸² <http://www.atomenabled.org/>

⁸³ Extensible Stylesheet Language (<http://www.w3.org/Style/XSL/>)

⁸⁴ <http://opensearch.a9.com/>

6. Conclusion

Ce développement, qui a fait émerger les subtilités et les contraintes liées à XML dans ce mécanisme de diffusion en format RSS, s'est révélé une excellente opportunité pour appréhender une nouvelle technique et gagner en savoir-faire concernant la production des flux les plus personnalisables et utiles pour la veille, les flux générés en fonction des critères de recherche⁸⁵. Il a également apporté des améliorations à l'OPAC à travers l'introduction du contenu dynamique en provenance des revues proposant des flux RSS.

Même si la langue française est l'une des plus représentées dans la « blogosphère »⁸⁶, les éditeurs des revues électroniques et des bases de données bibliographiques ainsi que les bibliothèques et centres de documentation francophones ont pris un certain retard dans l'introduction des flux RSS pour diffuser l'information qu'ils produisent ou qu'ils gèrent. Cependant, nous sommes dans une étape critique de transition, très proche probablement du moment de « percolation » (LE CROSNIER, 2004, p. 26), et les bibliothèques et centres de documentation ne vont certainement pas tarder à rejoindre le mouvement de la diffusion en format RSS.

Pour la pratique de veille, chaque site web, chaque catalogue ou base de données bibliographiques qui propose des flux, engendre un soulagement et un gain de temps très appréciable, d'autant plus si ces flux peuvent être personnalisés selon des critères de recherche ou des thématiques suffisamment fines.

Une fois que nous franchissons la porte de cette nouvelle façon d'intégrer l'information dans un bon agrégateur, nous ne pouvons pas imaginer faire marche arrière, sauf peut-être si nous arrivons au stade de « *Feed Overload Syndrome* »⁸⁷. Par contre, nous attendons presque avec impatience le moment où d'autres

⁸⁵ Des moteurs de recherche commencent aussi à proposer un choix de flux RSS générés à partir des termes de recherche et exportables parfois en format OPML (Outline Processor Markup Language, <http://www.opml.org/>) comme par exemple *Kebberfegg* (<http://www.researchbuzz.org/tools/kebbefegg.pl>), *Google Blog Search* (<http://www.google.com/blogsearch>), *RSS Micro* (<http://www.rssmicro.com/>) *Keyword Search Feed Directory* (<http://www.keyword-search-feeds.com/>)

⁸⁶ La deuxième après l'anglais si l'on croit aux chiffres donnés par la revue *Wired* dans son article « *Vive les Blogs !* » (<http://www.wired.com/news/culture/0,1284,67273,00.html>)

⁸⁷ Voir à ce sujet le billet du blog de Bill Burnham « *Saving RSS: Why Meta-feeds will triumph over Tags* » (http://billburnham.blogs.com/burnhamsbeat/2005/01/saving_rss_why_.html)

ressources électroniques (*Web of Science*⁸⁸, bases de données de la plateforme *OVID*⁸⁹, périodiques édités par *Elsevier*⁹⁰, *Springer*⁹¹, *Taylor & Francis*⁹², *Lippincott*⁹³, *Karger*⁹⁴ ou *Masson*⁹⁵) commenceront à proposer des flux RSS.

Si le projet du web sémantique⁹⁶ est encore loin devant nous, nous pouvons affirmer aujourd'hui que les avantages des échanges en XML sont chaque jour plus évidents. Il est donc essentiel de maîtriser cette technique actuellement car, si elle continue à se développer sur le web comme il serait souhaitable, elle deviendra probablement de plus en plus sous-jacente et « cachée » par les applications, comme c'est le cas actuellement dans les portails personnalisés de type *My Yahoo!*⁹⁷ ou dans la page d'accueil personnalisée de Google⁹⁸

Les blogs nous montrent une voie à suivre, celle de l'innovation générée à partir de l'appropriation, par l'intelligence collective d'une communauté, des techniques de l'information et de la communication. C'est bien cette voie que les bibliothèques et centres de documentation devraient emprunter pour collaborer activement à la nouvelle étape 2.0⁹⁹.

⁸⁸ <http://scientific.thomson.com/products/wos/>

⁸⁹ <http://www.ovid.com>

⁹⁰ <http://www.elsevier.com>

⁹¹ <http://www.springer.com>

⁹² <http://taylorandfrancis.metapress.com>

⁹³ <http://www.lww.com/>

⁹⁴ <http://www.karger.com/>

⁹⁵ <http://www.masson.fr>

⁹⁶ <http://www.w3.org/2001/sw/>

⁹⁷ <http://my.yahoo.com/s/rss-faq.html>

⁹⁸ Voir l'article consacré au sujet par *Search engine Watch* en mai 2005: "*Google Launches Personalized Home Page*" <http://searchenginewatch.com/searchday/article.php/3506541>

⁹⁹ Voir aussi les articles consacrés à la "bibliothèque 2.0" :

« Web 2.0: Building the New Library » <http://www.ariadne.ac.uk/issue45/miller/>

« Library 2.0 and "Library 2.0" » <http://cites.boisestate.edu/v6i2a.htm>

« The L20 Manifesto » <http://eltuo.pbwiki.com/>

7. Bibliographie

BARTHE, E. (2005). RSS ou Le futur de la veille. *Precisement.org* [en ligne]. [Consulté le 28 mai 2006]. Disponible en ligne :

<http://www.precisement.org/blog/article.php3?id_article=92>

BROCHARD, J-C., DE DARAN, H., COUDRIN, D., HOUPIER, J-C., SIMON, C. (2005). *Utilisation des fils RSS en bibliothèque* [en ligne]. Mastère professionnel, Lyon, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB). [Consulté le 28 mai 2006]. Disponible en ligne :

<http://memsic.ccsd.cnrs.fr/mem_00000249.html>

BROUDOUX, E., GRESILLAUD, S., LE CROSNIER, H., LUX-POGODALLA, V. (2005) *Construction de l'auteur autour de ses modes d'écriture et de publication* [en ligne]. H2PTM'05. [Consulté le 28 mai 2006]. Disponible en ligne: <http://archivesic.ccsd.cnrs.fr/sic_00001552.html>

ÇELIKBAŞ, Z. (2004). *What is RSS and how it can serve libraries*. [en ligne]. E-prints in Library and Information Science [en ligne]. [Consulté le 29 mai 2005]. Disponible en ligne : <<http://eprints.rclis.org/archive/00002531/>>

DESCHAMPS, C. (2003). Le Rss pour la veille. *Veille Magazine*. Disponible en ligne sur le blog de l'auteur : < <http://www.outilsfroids.net/news/839.shtml>> [Consulté le 29 mai 2005]

DESCHAMPS, C. (2005). Veille et Rss, comment passer à la vitesse supérieure. *Outils Froids* [en ligne]. [Consulté le 28 mai 2006]. Disponible en ligne :

<<http://www.outilsfroids.net/news/1121.shtml>>

GARREAU, A. (2005). *Les blogs entre outil de publication et espace de communication : un nouvel outil pour les professionnels de la documentation* [en ligne]. Maîtrise des sciences de l'information et de la documentation, CAO, Université. [Consulté le 28 mai 2006]. Disponible en ligne :

<http://memsic.ccsd.cnrs.fr/mem_00000273.html>

FIEVET, C., TURRETTINI, E. (2004). *Blog story* [imprimé]. Paris : Eyrolles. VIII-306 p. ISBN 2-7081-3158-3

- FITCHER, D. (2003). Why and How to Use Blogs to Promote Your Library's Services? Marketing library services. *Information Today, Inc* [en ligne]. Vol. 17, n°6. [Consulté le 28 mai 2006]. Disponible en ligne : <<http://www.infotoday.com/mls/nov03/fichter.shtml>>. ISSN 08055-8750
- HAMMOND, T., HANNAY, T., LUND, B. (2004). The Role of RSS in Science Publishing : Syndication and Annotation on the Web. *D-Lib Magazine* [en ligne]. Vol 10, n°12. [Consulté le 28 mai 2006]. Disponible en ligne : <<http://www.dlib.org/dlib/december04/hammond/12hammond.html>>. ISSN 1082-9873
- LE CROSNIER, H. (2004). L'évolution des modèles éditoriaux confrontés aux documents numériques. In : *Publier sur Internet : séminaire INRIA, 27 septembre - 1er octobre 2004, Aix-les-Bains* [imprimé]. Paris : ADBS. 248 p. ISBN 2-84365-072-0
- JDEY, A. (2006). 100 idées pour utiliser RSS. *Vtech* [en ligne]. [Consulté le 28 mai 2006]. Disponible en ligne : <http://vtech.canalblog.com/docs/100id_esRSS.pdf>
- JUMEAU, J. (2005). *Utiliser des fils RSS pour effectuer une veille informationnelle active* [en ligne]. URFIST de Paris/ Ecole des Chartes. [Mis à jour le 17 mai 2006] [Consulté le 28 mai 2005]. Disponible en ligne : <<http://www.ext.upmc.fr/urfist/rss/RSS.htm>>
- VOGEL, T., GOANS, D. (2005). Delivering the News with Blogs: The Georgia State University Library Experience. *Internet Reference Services Quarterly*. Vol. 10, n°1. ISSN 1087-5301