



**HAL**  
open science

**Weblogs : un nouveau paradigme pour les systèmes  
d'information et la diffusion de connaissances ?  
Applications et cas d'usage en contexte de veille et  
d'intelligence économique.**

Olivier Ertzscheid

► **To cite this version:**

Olivier Ertzscheid. Weblogs : un nouveau paradigme pour les systèmes d'information et la diffusion de connaissances ? Applications et cas d'usage en contexte de veille et d'intelligence économique.. Apr 2005. sic\_00001433

**HAL Id: sic\_00001433**

**[https://archivesic.ccsd.cnrs.fr/sic\\_00001433v1](https://archivesic.ccsd.cnrs.fr/sic_00001433v1)**

Submitted on 2 May 2005

**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.

**Titre :**

Weblogs : un nouveau paradigme pour les systèmes d'information et la diffusion de connaissances ?

**Sous-titre :**

Applications et cas d'usage en contexte de veille et d'intelligence économique.

**Auteur :** Olivier Ertzscheid

**Mél :** [olivier.ertzscheid@gmail.com](mailto:olivier.ertzscheid@gmail.com)

**Thème(s) de la conférence :**

- La représentation orientée utilisation des informations
- Veille et Intelligence économique
- Usages

## **Résumé :**

L'explosion des weblogs est aujourd'hui avérée sur le net. Ils sont aujourd'hui en France et depuis déjà quelques années outre-atlantique un vrai phénomène de masse. Un certain nombre d'indicateurs permettent de comparer l'engouement pour leur utilisation à celui rencontré pour l'usage du courrier électronique il y a de cela quelques années, et leur taux de croissance à celui du web à ses débuts.

Après avoir brièvement présenté le **contexte d'émergence de ces outils et leur principales fonctionnalités** nous montrerons comment leur **couplage avec les technologies RSS** (Really Simple Syndication) permet, lorsqu'ils sont déployés dans le cadre d'un **système d'information (SI) finalisé**, de se recentrer sur le **contenu** et sur les **usages** dudit SI.

Nous prendrons l'exemple du processus de veille dans un contexte d'intelligence économique pour démontrer comment les **usages en entreprise** de ces weblogs permettent non seulement d'**optimiser le taux de couverture d'un processus de veille** au travers de la **syndication de contenu** – un individu pouvant surveiller quotidiennement (monitoring) plusieurs centaines de sites – mais également et surtout d'agir au niveau de la **qualification** (sourcing) **des informations** sous surveillance en autorisant pour le veilleur des niveaux de granularité inédits au niveau du fond comme de la forme des contenus surveillés, nous permettant de parler de **granularité adaptative et contextuelle**.

Nous centrerons notre démonstration sur les **étapes d'identification et de collecte** des sources d'information dans le cadre d'un processus de veille et montrerons le **gain qualitatif et quantitatif** que représentent ces outils pour une démarche **d'intelligence économique** au niveau des actions de veille que celle-ci peut englober

Nous montrerons enfin comment le **positionnement technique** de ces outils (RSS, RDF, XML<sup>1</sup>) permet de dépasser leurs actuelles limitations et en fait la première application « résistante aux usages » du **web sémantique**, permettant également d'envisager des schémas d'**interopérabilité** valides avec des **progiciels** nécessitant une forte **structuration de l'information**.

## **Mots-clés :**

Weblogs – Intelligence économique – (processus de) Veille – Granularité adaptative – Granularité contextuelle – Usages – Syndication de contenu – Systèmes d'information – Web sémantique – RSS – XML – Monitoring.

---

<sup>1</sup> RDF : Ressource Description Framework  
XML : eXtensible Markup Language

## Introduction

Après l'avènement de l'hypertexte et la naissance conjointe du « web », après celui de l'e-mail et des listes et groupes de discussion, après l'essor et l'engouement relatif aux technologies de la messagerie instantanée, et conjointement aux développements des technologies WiFi et de l'internet sans fil, les « weblogs » ou « blogs » ou « carnets web » sont la dernière évolution marquante liées aux technologies de l'information et de la communication.

Un Weblog est une page personnelle, généralement alimentée par un seul individu et constituée des courts messages affichés et archivés automatiquement selon un ordre anté-chronologique. Mais plusieurs déclinaisons de cette forme originelle sont possibles<sup>2</sup> : ainsi il existe des weblogs collaboratifs entre plusieurs auteurs, des weblogs d'entreprise qui suppléent à toute la chaîne de gestion de la relation client (CRM), font office d'Intranets, absorbent les tâches liées à la gestion de projet, au management des connaissances, etc.

Cet article tentera de mesurer l'impact à court terme que ces nouveaux outils de publication et de gestion de l'information peuvent avoir sur des actions de veille en Intelligence économique.

## 1. Contexte d'émergence de ces outils et première typologie pour la veille.

### 1.1. Les raisons du succès : fonctionnalités principales.

L'explosion des weblogs est aujourd'hui avérée sur le net. Ils sont en France et depuis déjà quelques années outre-atlantique, un vrai phénomène de masse. L'une des premières pages web consacrées au phénomène en recensait 23 en Janvier 1999. Ils sont 4.1 millions en Septembre 2003<sup>3</sup>, 6.5 Millions en Décembre 2004<sup>4</sup>. Ces chiffres qui continuent de suivre une courbe exponentielle sont cependant à nuancer pour les raisons suivantes :

- sur les 4.1 millions de l'étude Perseus, seuls 1.4 millions étaient actifs, c'est à dire mis à jour au moins une fois tous les trois mois ;
- quels que soient les chiffres annoncés, aucune étude ne recense la totalité des weblogs existants, le nombre des plateformes et solutions d'hébergement étant lui-même délicat à inventorier de manière fiable.<sup>5</sup>

Néanmoins, un certain nombre d'indicateurs et des retours d'expérience de plus en plus nombreux permettent de comparer l'engouement pour leur utilisation à celui rencontré pour l'usage du courrier électronique il y a de cela quelques années, et leur taux de croissance à celui du web à ses débuts. « *Potential for growth within the entreprise similar to the growth of instant messaging* » [MAY 04]

Les raisons de ce succès sont liées à plusieurs facteurs convergents :

- Gratuité : l'apparition, dès Juillet 1999, sur le marché d'outils grands publics gratuits et faciles d'emploi dans lesquels il n'est plus nécessaire d'avoir une connaissance même minimale d'HTML pour publier des pages sur le net. Le plus utilisé de ces outils est Blogger ([www.blogger.com](http://www.blogger.com)).

---

<sup>2</sup> <http://www.google.com/search?hl=en&lr=&q=define%3Ablog&btnG=Search>

<sup>3</sup> [http://www.perseus.com/corporate/news\\_shell.php?record=51](http://www.perseus.com/corporate/news_shell.php?record=51)

<sup>4</sup> [http://biz.yahoo.com/prnews/041118/nyth043\\_1.html](http://biz.yahoo.com/prnews/041118/nyth043_1.html)

<sup>5</sup> ainsi certaines communautés telle « skyblogs » (<http://www.skyblog.com/>) service de création et d'hébergement gratuit offert aux auditeurs de la radio Skyrock, ont rapidement atteint le chiffre du million de blogs créés.

- Facilité : le mode de publication simple et intuitif : quel que soit l'outil, tout se déroule à partir d'une interface semblable à celle de l'envoi d'un courrier électronique.
- Gestion automatique de l'information :
  - o les informations « postées » sont automatiquement archivées, l'utilisateur pouvant choisir la fréquence de cet archivage (mensuel, hebdomadaire, quotidien).
  - o chaque information se voit automatiquement attribuée un URL permanent.
  - o chaque information est « indexée » à l'aide de balises « méta » minimales : auteur, date de publication, titre<sup>6</sup>.
- Services collaboratifs :
  - o chaque information postée peut faire l'objet de commentaires, liés au message et également archivés.
  - o gestion des droits : il est possible d'ouvrir des droits différents en lecture et en écriture.

Comme le résume rapidement cet inventaire non exhaustif, les weblogs offrent ainsi au plus grand nombre, les attendus techniques minimaux de tout système d'information « élaboré » : indexation, gestions des droits, archivage, recherche. C'est d'ailleurs à ce titre qu'ils remplacent d'ores et déjà dans certains grands groupes les architectures lourdes et trop souvent sous-utilisées de type Intranet.

*"Les blogs et leur écosystème, immiscés entre le courrier électronique et le web, sont le maillon manquant dans la chaîne de communication... Ils se rapprochent de la promesse originelle du web, dans la mesure où ils sont le premier outil qui a rendu facile -ou au moins simplifié- l'action de publier sur le web". [GIL 04]*

## 1.2. Typologie pour la veille.

Parmi l'ensemble des weblogs se développant sur le net, il est possible d'isoler trois macro-catégories [BLO 02] :

1. des journaux personnels, égotistes, simples « consignes » d'états d'âme quotidiens,
2. des carnets de note (« *notebooks* ») qui proposent un contenu sous forme d'approches et de synthèses (« *focused essays* »)
3. et enfin des weblogs-filtres (« *filters* ») annonçant colloques, séminaires, parutions, manifestation scientifiques, etc ...

Même si de nombreux weblogs sont une hybridation des catégories listées ci-dessus, elles n'en demeurent pas moins un filtre pertinent pour l'analyse. A ce titre, les deux dernières catégories constituent des sources intéressantes pour les actions de veille, sources qui pourraient, à terme, et sous réserve que l'explosion quantitative du phénomène se confirme, devenir incontournables voire remplacer le totalité des sources web habituellement consultées (sites web, forums, etc ...) avec, comme nous le verrons dans la suite de cet article un gain de précision et d'efficacité considérable pour le veilleur.

A titre d'exemple voici quelques weblogs francophones et anglo-saxons labellisés « Intelligence économique » ou « Business Intelligence » et appartenant aux deux dernières catégories<sup>7</sup> :

---

<sup>6</sup> Sur certaines plateformes de création et d'hébergement de weblogs il est également possible d'ajouter un niveau d'indexation thématique par la création de rubriques et/ou de mots-clés. Chaque message posté est alors également indexé et archivé selon la catégorie à laquelle il appartient.

- Carnets :
  - o <http://totem-consult.blogspot.com>
  - o <http://infoguerre.blogspot.com/>
  - o <http://www.aurorawdc.com/ci/>
- filtres :
  - o <http://www.itfacts.biz/> (ce site ne délivre qu'une information strictement factuelle)
- hybrides (à la fois filtre et carnet) :
  - o <http://www.deanie.ultimteam.com/>
  - o <http://www.outilsfroids.net>
  - o <http://www.urfist.info>

## 2. RSS : Une technologie au service de la veille.

### 2.1. RSS : définition et caractéristiques.

Les weblogs se distinguent des forums et des communautés en ligne de type Usenet en cela qu'il sont initialement des espaces de publication et non de discussion. L'information publiée, même si son statut varie considérablement d'un site à un autre est ainsi en général immédiatement capitalisable. Cela est notamment rendu possible par l'utilisation du format RSS qui, s'il n'est pas l'apanage exclusif des weblogs, y a trouvé l'un de ses usages les plus pertinents en terme de diffusion et de traitement d'information via le processus de syndication.

*« RSS is an extensible metadata description and syndication format, currently used for a number of applications, including news and other headline syndication, weblog syndication, etc. »*<sup>8</sup> [NAN 04]

Très concrètement, la syndication de contenu rend possible sur des logiciels clients de type « agrégateurs », l'affichage structuré en xml des informations publiées sur un site. Il est ainsi possible de recevoir uniquement les derniers contenus d'un site et d'en garder une trace, un archivage chronologique.

La différence avec les outils classiques de monitoring est de taille puisque ce type de diffusion permet d'éviter nombre des biais traditionnellement associés au processus de veille (fausses mises à jour, localisation de la partie effectivement mise à jour, ...) tout en conservant la plupart des fonctionnalités classiques (cycles programmables, gestion d'arborescence, etc.)

### 2.2. Enjeux et usages pour la veille et l'IE

Le processus de veille peut être découpé de la manière suivante:

- « *Le **ciblage** est la détermination de l'environnement à scruter. Il vise à spécifier les acteurs et les thèmes à surveiller, ainsi que des sources d'information à mettre en oeuvre;*
- *La **traque** consiste à désigner les personnes susceptibles de collecter des informations sur l'environnement ciblé, leur attribuer leurs missions de collecte d'informations et leur donner les méthodes et outils utiles à cette activité.*

---

<sup>7</sup> Liens actifs au 15/01/2005

<sup>8</sup> « RSS est un format extensible de description et de syndication de métadonnées, couramment utilisé par nombre d'applications parmi lesquelles la syndication d'actualités, de titres de presse, de weblogs, etc. »

- La **circulation** des informations est ici interne à l'entreprise et consiste à mettre en commun les informations collectées dans l'environnement et les rendre accessibles aux décideurs
- Le **traitement** des informations vise leur transformation en véritables forces motrices pour l'action et les prises de décisions. » [BLA 97]

Blanco et Lesca exposent ce cycle dans un contexte de veille stratégique. Nous l'envisagerons dans la suite de cet article comme la méthodologie *a minima* permettant de rendre compte de tout type d'action de veille, indépendamment de sa nature (économique, technologique, documentaire) et de sa finalité (stratégique, anticipative, commerciale, etc.)

Nous nous centrerons principalement sur l'étape de ciblage pour argumenter et exemplifier le gain qualitatif et quantitatif des weblogs mais il y aurait également beaucoup à dire sur les deux dernières étapes : les weblogs sont en effet de plus en plus couramment utilisés comme modèle d'architecture pour le SI de l'entreprise (Intranet<sup>9</sup>), exploitant à plein les fonctionnalités de diffusion et de traitement dont ils disposent et qu'ils offrent sous une forme intuitive et automatisée aux utilisateurs dudit système (auto-archivage, création de catégories, indexation croisée – catégories et héritage multiples - , recherche filtrées, etc.)

### 2.3. Gains liés au ciblage.

Les opérations liées au ciblage pourraient apparaître délicates pour le choix des sources de type « weblogs » à mettre sous surveillance, au nom des critères évoqués précédemment<sup>10</sup> ; critères auxquels s'ajoute le manque d'outil de recherche global inventoriant et donnant accès à un index suffisamment complet de weblogs. En effet les outils (moteurs et annuaires) de recherche de weblogs sont pléthore et si certains<sup>11</sup> offrent un réel avantage quantitatif, il n'existe pas encore de « google des blogs ».

Ces limites étant posées, elles n'en constituent pas pour autant une réelle difficulté. Il existe en effet dans la « blogosphère », sur une thématique, un secteur, une industrie ou un produit donné ce que l'on appelle des « *pundit* » [GLA 04], experts auto-proclamés et dont les sites bénéficient d'un nombre de liens entrants (« backlinks ») qui tendent à asseoir cette expertise lorsqu'elle semble justifiée.

De plus, des analyses reposant sur des méthodes statistiques de mesure appliquées à l'analyse des réseaux sociaux [HER 04] permettent d'établir très rapidement dans la topologie de la blogosphère la présence de weblogs incontournables, constituant sur des thèmes et/ou des terrains parfois très pointus, une « A-list », liste de référence.

Les weblogs y figurant présentent la caractéristique d'être à la fois les plus lus, les plus représentatifs et les plus cités dans les médias et reçoivent de ce fait le plus grand nombre de liens entrants. Sur certains domaines, le repérage de quelques-uns de ces weblogs appartenant à la « A-list » est suffisant pour un premier ciblage de l'opération de veille. Reste ensuite à appliquer une méthodologie classique de type « fil d'Ariane » visant à recenser les liens présents sur les weblogs choisis de la « A-list » pour que le ciblage devienne exhaustif, permettant d'inclure dans le monitoring des sources, d'autres weblogs relevant d'un signal beaucoup plus « faible ».

Nous aurons plus tard l'occasion de mettre en évidence les gains qualitatifs et quantitatifs mais restons-en pour l'instant au nouveau potentiel qu'offrent les weblogs en termes de sources entrant dans une étape de ciblage : du fait de leur facilité de mise en place,

<sup>9</sup> A titre d'exemple on pourra se référer à ce témoignage estimant le gain sur un cahier des charges de refonte d'un Intranet de 4 mois avec des solutions classiques à 10 jours avec des plateformes blogs.

<http://blog.ronez.net/index.php?p=37>

<sup>10</sup> Voir le point 1.1 relativement au recensement délicat des blogs existants.

<sup>11</sup> [www.technorati.com](http://www.technorati.com), [www.daypop.com](http://www.daypop.com), [www.feedster.com](http://www.feedster.com)

d'usage et d'alimentation, les weblogs sont un facteur d'incitation qui met à disposition des veilleurs de nouvelles sources pour la seule raison que de nouveaux individus se mettent à publier des informations qui réclamaient auparavant de passer des filtres techniques démotivants (connaissance du code HTML, gestion d'un espace d'hébergement, gabarit du site et de la page, etc...). Quelle que soit leur provenance, et à tous les échelons hiérarchiques de l'entreprise, ces « nouveaux » acteurs de la diffusion d'information publient et communiquent, avec, la plupart du temps, une liberté de ton que rend possible l'absence de processus de validation ou de chaîne éditoriale<sup>12</sup>. Ainsi PDGs, CEOs et autres experts sectoriels deviennent de nouvelles sources disponibles pour les traqueurs.

### **3. Apports qualitatifs et quantitatifs des weblogs dans un processus de veille.**

Pour le veilleur, les gains liés à l'utilisation des weblogs sont autant qualitatifs que quantitatifs, ce double avantage étant rendu possible par la possibilité de syndication de contenu qu'ils proposent. L'ensemble du processus de veille dispose alors, de manière simple, gratuite (ou avec un coût dérisoire) et sans nécessité d'une grande maîtrise technologique, d'outils lui permettant d'appliquer à tout type de contenu informationnel une granularité adaptative et contextuelle jusque-là impossible.

#### **3.1. Gains quantitatifs.**

Les weblogs ont pour premier avantage de permettre d'étendre considérablement le spectre d'une veille : un individu peut en effet « manuellement », chaque jour depuis son poste de travail, vérifier et éditer le contenu mis à jour d'un nombre de sites considérable<sup>13</sup>. Sachant que seuls 20 % des weblogs existants publient une nouvelle information dans un délai inférieur à 24 heures, le suivi quotidien ou bi-quotidien d'un immense corpus devient possible et il est protégé des biais classiques (spam, fausses mises à jour ...).

#### **3.2. Gains qualitatifs.**

La granularité de l'information sur laquelle les weblogs permettent de travailler dans une optique de réingénierie<sup>14</sup> se décline selon deux axes :

- en termes de contenu tout d'abord : il est possible en sus des fils d'information standards fournis par les sites web, de mettre en place des filtres internes au contenu même du site. Le magazine américain « wired » propose ainsi de syndiquer uniquement les articles de l'éditorialiste Meg Ryan (spécialiste de la question des weblogs). Et cela est bien entendu possible pour d'autres éditorialistes et journalistes de Wired
- en terme de format ensuite, la plupart des sites proposant des contenus à syndiquer le font sous différentes options : syndication longue (contenu des billets postés), syndication courte (titre du billet seulement), syndication longue avec commentaires (tous les commentaires postés sont à leur tour syndiqués)

---

<sup>12</sup> dans des limites qui sont évidemment celles fixées par l'entreprise pour le cas de weblogs hébergés sur le site web de l'entreprise. Cette « liberté de ton » et la diffusion d'informations normalement restreintes ou pouvant nuire à l'image de l'entreprise (y compris cette fois dans le cas de weblogs hébergés sur des sites indépendant de l'entreprise) à déjà valu à certains des procédures de licenciement : voir par exemple

<http://deanie.ultimteam.com/index.php?2005/01/13/118-licencie-pour-avoir-blogue>

<sup>13</sup> ne disposant pas à ce jour de données fiables issues de cellules de veille en entreprise je ne peux que faire part de mon expérience personnelle qui, en une heure, me permet de monitorer plus de 700 fils RSS issus aussi bien de sites webs d'entreprise, de sites académiques, de sources presse et de bases de données.

<sup>14</sup> Dans l'optique de l'étape de « traitement » chez Blanco & Lesca (cf supra)



Ajoutons à cela qu'il est également possible sur un nombre croissant de sites d'établir une syndication personnalisée sur mot-clé (tous les articles de « wired » contenant le mot « internet » par exemple<sup>15</sup>) et nous aurons une idée exacte de cette granularité adaptative et contextuelle que proposent les weblogs à leurs utilisateurs.

L'information ainsi récupérée n'arrivant pas « brute » mais déjà élaborée selon une DTD XML il est alors possible de la chaîner à l'intérieur du SI de l'entreprise avec une capitalisation immédiate désormais possible et dispensant des longues et coûteuses étapes de pré-processing et/ou de validation par des experts (documentalistes par exemple).

### 3.3. Une topologie propice à la veille.

La topologie des weblogs semble propice à la mise en place d'actions de veille et d'intelligence économique dans la mesure où les weblogs isolés ou très faiblement connectés représentent la majorité sur l'ensemble des blogs disponibles. :

« *The blogosphere appears to be selectively interconnected, with dense clusters in parts, and blogs minimally connected in local neighborhoods, or free-floating individually, constituting the majority.* »<sup>16</sup> [HER 04]

La topologie de cet espace favorise donc l'émergence de signaux faibles qu'un travail de sourcing méticuleux conjugué à certains des outils présentés plus haut peut rendre facilement repérables et capitalisables.

Même si cette topologie est encore naissante et susceptible de grands bouleversements, (le taux de croissance des weblogs étant, rappelons-le, comparable à celui des pages web à leurs débuts) la structure de la blogosphère en fait en outre une structure idéale pour l'étude des phénomènes de propagation de l'information. [ADA 04] parle à ce propos d'« *information epidemics* » et proposent un outil de visualisation et un algorithme « *I-Rank* »<sup>17</sup> permettant « *to visualize the paths specific infections take through networks.* »<sup>18</sup>

Les weblogs présentent ainsi la double caractéristique d'être un terrain d'expérimentation directement exploitable mais également un terrain de recherche pour l'analyse et la validation des phénomènes de flux informationnels, phénomènes auxquels les disciplines (scientométrie, bibliométrie) permettant de théoriser les processus et actions de veille doivent leur naissance.

### 3.4. Veille mais aussi guerre économique

Du fait de cette topologie particulière, de la visibilité presque immédiate de l'information ainsi éditée dans les moteurs de recherche, et des facilités offertes au référencement, les weblogs constituent également de redoutables outils au service de pratiques relevant de la désinformation ou de la guerre économique. Ainsi la veille du lancement international de l'Airbus A 380, Randy Baseler, Vice-Président du Marketing de Boeing, a choisi d'ouvrir son blog personnel<sup>19</sup> stigmatisant (selon lui) les « *erreurs de*

---

<sup>15</sup> <http://search.wired.com/wnews/default.asp?format=rss&query=blogs>. Il suffit ici de changer le dernier mot de l'URL pour créer autant de fils de syndication que souhaités.

<sup>16</sup> « *La blogosphère apparaît interconnectée mais de manière selective, avec en partie des clusters assez denses, des weblogs faiblement connectés dans leur proche voisinage, et enfin d'autres flottant librement et qui constituent la majorité.* ».

<sup>17</sup> I-Rank se différencie du PageRank de Google car il travaille sur les liens non plus explicites mais sur « the inferred implicit structure linking of blogs. »

<sup>18</sup> « *visualiser les chemins infectieux spécifiques au travers des réseaux.* »

<sup>19</sup> <http://www.boeing.com/randy/>

*jugement* » associées à ce gros porteur et espérant probablement contrebalancer une couverture médiatique essentiellement élogieuse.

#### **4. Convergence des outils et des sources : vers de nouveaux enjeux technologiques ?**

##### **4.1. Outils et sources : un couplage réussi.**

Parmi les sources Internet, que celles-ci offrent une information structurée (Bases de donnée en ligne<sup>20</sup>) ou non (forum, mailing list, ...), nombreuses sont celles qui ont d'ores et déjà opéré un basculement vers de modèles de diffusion d'information orientés RSS. Il s'agit là de signaux forts stigmatisant le changement de paradigme occasionné par le format RSS.

Ainsi, la totalité des messages postés sur des forums de discussion du service « Yahoogroups<sup>21</sup> » sont désormais accessibles au format RSS. Idem pour la partie « actualités » du moteur MSN<sup>22</sup>. Il est également possible de syndiquer les résultats d'une requête sur l'Open Directory et ce soit sur un mot-clé, soit sur une des rubriques ou sous-rubriques de l'annuaire<sup>23</sup>.

En parallèle de plus en plus d'outils de type « clients » combinent les sources RSS émergentes et les sources plus classiques. Les outils suivants permettent aussi bien de monitorer des pages web et des fils RSS que des newsgroups ou des comptes POP3 (messagerie électronique) : Active Refresh<sup>24</sup>, Newzcrawler<sup>25</sup>.

Les services de « veille » comme GoogleAlert<sup>26</sup> proposent désormais de choisir la réception de ses alertes par email ou sous forme de fil RSS<sup>27</sup>. D'autres services comme Botablog<sup>28</sup> permettent de créer des agents qui monitorent autant de fils rss que souhaités en envoient un mail d'alerte. Chaque « bot » (robot) ainsi créé peut monitorer jusqu'à 25 fils d'information.

Devant la richesse pour l'analyse et la sériation des tâches liées à une veille qu'offre RSS il est également possible de créer un fil de syndication pour des pages n'en disposant pas, y compris pour des sites web dits « classiques ». Il s'agit encore ici de fonctionnalités en phase « Béta » et donnant en l'occurrence des résultats assez aléatoires mais nonobstant très pertinents sur certains sites<sup>29</sup>.

##### **4.2. Enjeux technologiques.**

En faisant évoluer le web pour le rendre plus conforme à ses ambitions initiales, à savoir, être un outil de diffusion de contenu, les weblogs changent la nature du web. Par le couplage technique mais pouvant demeurer transparent pour l'utilisateur final avec RSS et XML qu'ils permettent d'atteindre ils peuvent être considérés comme l'une des premières pierres « résistante aux usages » d'un web sémantique<sup>30</sup> à portée de tous. Ils ont en effet en

---

<sup>20</sup> <http://www.biomedcentral.com/info/about/rss>

<sup>21</sup> <http://groups.yahoo.com>

<sup>22</sup> <http://fr.newsbot.msn.com/>

<sup>23</sup> [www.xmlhub.com/odp\\_feed.php](http://www.xmlhub.com/odp_feed.php)

<sup>24</sup> <http://www.activerefresh.com/>

<sup>25</sup> <http://www.newzcrawler.com/>

<sup>26</sup> <http://www.googlealert.com/>

<sup>27</sup> pour une liste plus complète consulter notamment : <http://www.outilsfroids.net/news/839.shtml>

<sup>28</sup> <http://www.botablog.com/>

<sup>29</sup> [www.feedfire.com](http://www.feedfire.com)

<sup>30</sup> <http://www.semanticweb.org/>

commun avec ce dernier un certain nombre de composants annoncés comme déterminants<sup>31</sup> : outre XML et RDF, figurent dans ces composants l'ajout automatique de métadonnées certes minimales (auteur, date, titre) mais standardisées et pouvant être réinjectées ou adaptées à des SI existants, voire alimenter certaines ontologies.

## 5. Conclusion.

La perception que nous avons d'Internet est intimement liée à la nature des espaces de diffusion de contenu qu'il englobe. On trouve ainsi l'espace « web », souvent assimilé par une métonymie trompeuse à Internet, espace « web » qui est lui-même composé des pages personnelles et de sites d'entreprises ou d'institutions, ces derniers pouvant bénéficier d'un accès réservé dans le cadre d'un intranet par exemple. On trouve ensuite un deuxième espace qui est celui de l'échange de courriers électroniques groupant Usenet, différents forums et newsgroups et bien entendu l'échange de mails de personne à personne. Internet existe essentiellement au travers de ces deux espaces.

Or les weblogs impactent déjà considérablement et (probablement) durablement chacun d'eux :

- ils remplacent progressivement les pages personnelles (nécessitant une maîtrise d'HTML) en tant qu'outils de publication techniquement transparents et intuitifs ;
- ils se substituent de plus en plus aux intranets, faisant là encore tomber une barrière technique et emportant de fait l'adhésion immédiate des utilisateurs en encourageant au lieu de décourager la mutualisation de contenus, sans laquelle aucune capitalisation n'est possible, et qui est à ce jour la principale pierre d'achoppement des « anciens » intranets ;
- concernant l'échange de courrier électronique là encore les weblogs apparaissent comme un outil de substitution pérenne. La plupart des sites jouissant avant l'arrivée des weblogs d'une audience et d'une renommée avérées ont tous migré vers des modes de diffusion de type fils rss<sup>32</sup>. Idem pour la totalité des forums Yahoogroups.

Seul semble résister encore (et probablement pour longtemps) l'usage du mail dans un contexte de personne à personne, le rendant ainsi à sa vocation initiale et lui redonnant une pertinence en tant qu'outil que de nombreux travers lui avaient ôté (spam, pourriels, surabondance, etc.).

De notre point de vue, les weblogs et leur utilisation dans un contexte de veille est encore largement sous-évaluée, aussi bien dans la sphère publique que dans la sphère privée même si des percées significatives voient régulièrement le jour<sup>33</sup>

Cette sous-évaluation est dépendante de deux variables :

- **le changement de modèle et de pratiques** dans la mise en œuvre de systèmes d'information dits « élaborés »<sup>34</sup>, changement qui se heurte à nombre de **résistances** essentiellement « **culturelles** » et liées, pour certaines entreprises, aux **modes de management** en vigueur,
- les quelques limites techniques dont souffrent encore les weblogs et qui sont liées à leur référencement et à leur indexation par les outils de recherche : une offre pléthorique d'outils présentant un faible taux de couverture et des différences fondamentales dans les techniques d'indexation utilisées (full-text et/ou métadonnées

---

<sup>31</sup> <http://www.semanticweb.org/about.html>

<sup>32</sup> InternetActu, Automatesintelligents, Abondance ...)

<sup>33</sup> ainsi le progiciel Evolution de la société Digimind ([www.digimind.fr](http://www.digimind.fr)) permet désormais de monitorer des fils RSS au travers de requêtes élaborées.

<sup>34</sup> On utilisera alors des plateformes de type CMS (Content Management Systems)

– titre – seulement et/ou mesure des backlinks) rendent le sourcing et le ciblage encore délicats<sup>35</sup>.

A l'inverse, un risque existe de voir se reproduire sur une échelle là encore inédite une surcharge cognitive liée au phénomène de surabondance informationnelle. De plus en plus de sites (d'actualités notamment) faisant le choix de la syndication de contenus, le veilleur s'expose alors au « Feed Overload Syndrome »<sup>36</sup>. Cependant, s'agissant rappelons le de blocs d'informations balisés (selon une DTD XML) il est possible d'envisager de mettre en œuvre des méta-flux d'informations utilisant mots-clés et marqueurs (balises) pour redistribuer automatiquement les contenus dans des arborescences et des dossiers préstructurés, en attendant de pouvoir les traiter ultérieurement.

Enfin remarquons davantage pour ouvrir des perspectives que pour conclure que dans les domaines de la communication de crise<sup>37</sup>, de la gestion de la relation client, du marketing, de la veille évidemment, mais également de la gestion des connaissances<sup>38</sup> (KM), les weblogs donnent lieu chaque jour à de nouvelles expérimentations<sup>39</sup> et innovations, marquant là encore un changement paradigmatique d'importance en apportant une valeur ajoutée souvent immédiatement mesurable.

## Bibliographie

- [ADA 04] Adar E. (et al.) « *Implicit Structure and the Dynamics of Blogspace* », Workshop on the Weblogging Ecosystem, 13th International World Wide Web Conference, May 18th, 2004. Pre-print available on line : <http://www.hpl.hp.com/research/idl/papers/blogs/blogspace-draft.pdf>. Consulté le 31/01/2005.
- [BLA 97] Blanco S., Caron M-L., Lesca H., « *Implantation d'une veille stratégique pour le management stratégique, proposition d'un modèle conceptuel et premières validations.* » 5ième Conférence Internationale de Management Stratégique (AIMS), Montréal, Canada, 25-27 juin 1997. [ en ligne ] <http://www.strategie-aims.com/montreal/lescaeta.pdf>, consulté le 30/01/2005.
- [BLO 02] Blood R., **The Weblog Handbook : Practical Advice on Creating and Maintaining your Blog**, Cambridge, MA: Perseus Publishing. 2002.
- [GIL 04] Gillmor D., **We the Media (hardback) : Grassroots Journalism by the People, for the People**, July 2004, O'Reilly Media, Inc.,CA.
- [GLA 04] Glance N.-S., Hurst M., Tomokiyo T., “*BlogPulse: Automated Trend Discovery for Weblogs*”, [en ligne] <http://www.blogpulse.com/papers/www2004glance.pdf> , consulté le 31/01/2005.

---

<sup>35</sup> Nul doute que si le phénomène et l'intérêt des blogs se confirme nous verrons bientôt apparaître un onglet « Blog » dans l'interface de Google. Son positionnement et son rachat du premier service d'hébergement et de création de weblogs Blogger ([www.blogger.com](http://www.blogger.com)) est un indice fort allant dans ce sens.

<sup>36</sup> « Feed » est la désignation d'un « fil d'information », donc d'une adresse de syndication. Sur ce sujet voir notamment : [http://billburnham.blogs.com/burnhamsbeat/2005/01/saving\\_rss\\_why\\_.html](http://billburnham.blogs.com/burnhamsbeat/2005/01/saving_rss_why_.html)

<sup>37</sup> “Crisis Communications Blogs Use Them Cautiously” <http://tinyurl.com/43ozk>

<sup>38</sup> <http://www.lgi.ecp.fr/~dudezert/images/CER0501.PDF>

<sup>39</sup> nombre d'entreprises utilisent ainsi des K-Blogs (Knowledge blogs) pour animer des communautés de pratique ou faire de la localisation d'expertise. D'autres (les mêmes le plus souvent) se servent également de P-Blogs (Project Blogs) qui prennent en charge toutes les étapes d'une gestion de projet.

- [HER 04] Herring S. et al., "*Conversations in the Blogosphere: An Analysis "From the Bottom Up"*", Proceedings of the Thirty-Eight Hawaiï International Conference on System Sciences (HICS-38). Los Alamitos:IEEE Press. [en ligne] <http://www.blogninja.com/hicss05.blogconv.pdf>, consulté le 31/01/2005.
- [MAY 04] Mayfield R., "*Blogs bubble into business*", [en ligne] <http://www.socialtext.com/weblog/040128enterprise.html> , consulté le 30/01/2005.
- [NAN 04] Nanno T. (et al.), "*Automatically collecting, monitoring, and mining japanese weblogs*" [en ligne] <http://lr-www.pi.titech.ac.jp/blogwatcher/paper/www2004-838.pdf>, consulté le 30/01/2005.