

Plaidoyer pour un réseau d'inventaires des résultats de la recherche

Jacques Ducloy

► **To cite this version:**

Jacques Ducloy. Plaidoyer pour un réseau d'inventaires des résultats de la recherche. Oct 2004, 2004. <sic_00001147>

HAL Id: sic_00001147

https://archivesic.ccsd.cnrs.fr/sic_00001147

Submitted on 1 Dec 2004

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.

Plaidoyer pour un réseau d'inventaires des résultats de la recherche

Jacques DUCLOY (*)
Jacques.Ducloy@inist.fr

(*) INIST-CNRS, 2 allée du Parc de Brabois, 54510 Vandoeuvre lès Nancy Cedex (France).

Mots clefs :

Besoin information, Veille technologique, Recherche scientifique, Coopération entreprise, Marché information, Information scientifique technique, Publication, Libre accès, Base donnée bibliographique, Document électronique.

Keywords:

Information need, Technological watch, Scientific research, Firm cooperation, Information market, Scientific technical information, Publication, Open access, Bibliographic database, Electronic document.

Palabras clave :

Necesidad información, Vigila tecnológica, Investigación científica, Cooperacion empresa, Mercado información, Información científica técnica, Publicacion, Acceso libre, Base datos bibliográfica, Documento electrónico.

Résumé

Dans cet article nous étudions les insuffisances du dispositif d'Information Scientifique et Technique (IST) actuel en regard des besoins de la veille scientifique en milieu industriel ou de ceux du pilotage de la recherche. Nous proposons des réflexions allant dans le sens d'une appropriation de l'IST par le monde scientifique en utilisant les avancées techniques des technologies du document structuré et en profitant des mouvements d'opinion autour des archives ouvertes. Enfin nous donnons des idées directrices pour la construction de réseaux d'inventaires.

1 Introduction

La qualité d'une prestation de veille dépend directement de la validité et de la représentativité des sources d'information. Pour l'information scientifique, la concentration industrielle a conduit à la construction de monopoles éditoriaux dont on peut craindre une vision de moins en moins diversifiée du monde de la recherche et de la technologie. Ces données risquent donc de devenir de moins en moins adaptées à la résolution de problèmes opérationnels, tels que la recherche de partenaires, qui demandent des éléments spécialisés ou régionaux. De leur côté, les universités ou établissements publics à caractère scientifique ou technique voient leur échapper la maîtrise de l'information scientifique dont ils ont besoin pour leur pilotage.

Pour y remédier, nous proposons une réflexion militante sur l'appropriation de l'Information Scientifique et Technique (IST) par le monde de la recherche. Les initiatives autour du libre accès à l'information scientifique (open access [16]) vont naturellement dans ce sens. Mais les réflexions actuelles sont souvent focalisées sur la seule diffusion individualisée des articles, sans prendre en compte la globalité des besoins. Or les nouvelles technologies, du document électronique au web sémantique, permettent justement de compléter ce dispositif pour constituer des réseaux d'inventaires et de services mieux adaptés aux besoins.

Dans une première partie, nous analyserons les limites de l'offre actuelle d'IST par rapport aux besoins de la veille scientifique et du pilotage de la recherche. Nous donnerons ensuite des éléments d'explication sur l'origine de ces dysfonctionnements ou plus précisément sur la façon dont la recherche a perdu la maîtrise de la communication scientifique. Puis nous montrerons comment l'ensemble de la palette des opportunités apportées par les nouvelles technologies peut consolider la base apportée par le libre accès. Enfin, nous évoquerons les mesures d'accompagnement à mettre en œuvre pour favoriser une telle politique.

2 Analyse des besoins d'informations produites par la recherche

2.1 Besoins venant du monde industriel

Les besoins des industriels en information scientifique peuvent être examinés suivant trois points de vue : Quoi de neuf ? Qui fait quoi ? Comment ?

A la question « **quoi de neuf ?** », les grandes bases de données, efficaces pour réaliser un « état de l'art », ne sont pas toujours bien adaptées à la détection de l'innovation. En effet, elles privilégient les revues à haut facteur d'impact dont les articles demandent une solide maturation pour traverser le filtre de l'évaluation par les pairs. Elles ont donc souvent plusieurs années de retard sur la réflexion technologique. Pour prendre un exemple adapté au public de VSST, nous nous sommes intéressés à l'utilisation de la technologie SGML/XML pour l'analyse de l'information dans les années 90 [5][7]. Ce sujet a commencé à émerger dans des cercles assez fermés 4 à 5 ans plus tard. Il est devenu un signal faible « facile à repérer » en 97/98 à la suite de l'émergence du standard XML. Enfin, c'est maintenant un signal fort visible dans toutes les recherches bibliographiques. Un élargissement de la recherche d'information vers des sources spécialisées « moins prestigieuses » ou un meilleur traitement de la littérature grise permet souvent de gagner plusieurs années.

La question « **qui fait quoi ?** » est probablement au cœur de la problématique « recherche industrie » car elle suit immédiatement la détection d'une alerte. Elle se formalise souvent

comme : « où puis-je trouver un expert ? ». Là encore, les grandes bases de données atteignent leurs limites. Des traitements trop simples vont permettre de repérer des grands acteurs, aux agendas saturés, localisés dans un autre continent ! Or dans un premier temps, la demande réelle est souvent : « j'ai besoin de rencontrer quelqu'un d'assez compétent, mais surtout assez neutre et disponible pour avancer dans ma réflexion ». En pratique, en reprenant le thème « veille et xml » une bonne réponse pourrait être : « Dans l'université voisine, il y a un DESS de veille où interviennent des enseignants issus d'un laboratoire d'informatique ».

Enfin la question « **comment ?** » met les grandes bases bibliographiques dans une position de plus en plus difficile. En effet, le format des publications scientifique (limitation à une dizaine de pages) exclut toute description approfondie. Ensuite, plus la problématique s'affine et plus elle fait appel à des données précises et spécifiques du domaine. Voici par exemple un problème récemment posé dans un contexte de fabrication de produit pharmaceutique. Après avoir défini une famille de composants, l'entreprise nous a contactés pour affiner sa recherche de molécules, par exemple en analysant une collection de brevets pour définir un ensemble de formules chimiques non revendiquées, ou encore, en explorant la littérature scientifique pour trouver des résultats publiés¹. La codification des éléments factuels tels que les formules, non prises en compte dans les bases généralistes, devient alors indispensable.

Les grandes bases bibliographiques sont des outils de première approche, pertinents pour repérer une activité scientifique « pas trop silencieuse ». Une stratégie d'investigation informationnelle dans un contexte d'innovation demande des sources multiples, avec des composantes thématiques et régionales. Autrement dit, du côté de la recherche, il faudrait améliorer notablement le signalement des résultats et des compétences. En réalité, les éléments bibliographiques éditoriaux ne sont qu'un élément d'un inventaire plus large (ce qui ne veut pas dire plus volumineux !).

2.2 Besoins venant du Monde de la Recherche

Le monde de la recherche, pour sa veille stratégique propre, son évaluation et son pilotage est un gros consommateur d'IST. Ici encore, les acteurs commerciaux de l'IST ont ciblé des activités « rentables » centrées sur l'accès à la publication de références et l'évaluation individuelle. Ces services ne coïncident pas toujours avec les besoins de la recherche, pris à différents niveaux depuis le poste de travail du chercheur jusqu'au pilotage des institutions.

Au niveau individuel, il convient de distinguer la recherche d'information dans la spécialité du chercheur et des investigations plus larges. Dans le premier cas, les relations personnelles, créées lors des comités de programme, et complétées par des ressources très spécialisées de formes diverses, depuis les mails jusqu'aux wikis, constituent l'ossature des réseaux de veille des laboratoires. Le recours aux bases de données est donc faible.

En revanche, l'universitaire chargé d'un cours qui dépasse sa spécialité de recherche, le chercheur qui veut ouvrir sa problématique dans un contexte multidisciplinaire, ou celui qui cherche des partenaires pour monter un programme européen se retrouve dans une position voisine de l'industriel. Il rencontrera les mêmes besoins (informations de proximité ou à caractère factuel) et les mêmes frustrations.

¹ Ce problème n'a en fait pas été résolu mais nous sert d'exemple significatif des demandes que nous devrions pouvoir satisfaire à moyen terme en travaillant sur la codification des formules chimiques.

L'inadéquation des sources avec les besoins se creuse davantage lorsqu'il faut aborder un niveau collectif (pilotage ou évaluation). Dans un récent rapport [4], la Conférence des Présidents d'Université mettait l'accent sur la difficulté de mesurer le poids de la recherche conduite dans les universités françaises. Une des difficultés majeures est de repérer (et de regrouper) les organismes de recherche nationaux. Cette remarque est partagée par d'autres institutions ayant des missions d'évaluation telles que l'UNIPS du CNRS. Elle est corrélée par la multiplication des observatoires créés à différents niveaux (université, délégation régionale du CNRS, EPST...).

Plus grave encore, les multiples défauts du système d'évaluation basé sur les indicateurs de l'Institute for Scientific Information (ISI) sont régulièrement dénoncés. La bibliographie est abondante sur le détournement du système à travers les auto-citations pour l'évaluation des chercheurs. Dans les interventions orales des réunions spécialisées, les responsables se plaignent de ne pas pouvoir réguler l'activité des chercheurs qui développeraient de plus en plus souvent une stratégie de publication et non une véritable démarche de recherche². Un extrait de l'avis concernant la recherche du projet de loi de finances³ pour 2004 est particulièrement explicite à propos des méthodes d'évaluation des chercheurs. « *Ces dernières restent, en effet, essentiellement axées sur la prise en compte de publications scientifiques au détriment d'autres éléments de la carrière du chercheur, à commencer par les résultats obtenus* ». Le rapporteur, qui s'appuie sur des analyses du Comité National d'Evaluation de la Recherche, considère donc que les publications ne sont pas représentatives des résultats de la recherche (dont la veille scientifique ou le pilotage de la recherche ont précisément besoin).

A un niveau plus global, les indicateurs fondés sur les revues cœur sont excellents pour l'entraînement des chercheurs « nobelisables ». Mais sont-ils suffisants pour structurer la recherche ? Le problème est maintes fois évoqué notamment pour l'évaluation de la recherche des pays en voie de développement. Il est probablement aussi crucial pour l'évaluation des moyens de transfert de technologie ou la diffusion des connaissances.

Les annonces récentes de la société Elsevier sur sa solution Scopus peuvent laisser craindre une évolution vers un système où les articles et mécanismes d'évaluation seraient aux mains d'un unique opérateur commercial. Le site officiel⁴ donne encore peu d'informations précises sur le produit au moment où cet article est rédigé mais ne laisse planer aucun doute sur l'ampleur des ambitions : plus de 13 000 titres avec les citations vérifiées. Le monde de la recherche semble donc perdre la maîtrise de la communication scientifique.

3 Vers un nouveau modèle de la communication scientifique ?

Comment infléchir cette tendance et créer une alternative mieux adaptée aux besoins des industriels et des chercheurs ? Une meilleure analyse des bouleversements actuels subis par le monde de l'édition constitue un élément de réponse potentiel.

² La difficulté de trouver de la littérature scientifique sur ce point illustre parfaitement un effet pervers du mécanisme des publications. En effet, le traitement des auto-citations est un sujet qui se prête au développement d'algorithmes sur lesquels un bon chercheur peut publier plusieurs articles par an. En revanche, on trouve beaucoup moins de documentation sur la difficulté de diriger des chercheurs car ce phénomène est difficile à mesurer et les remèdes ne sont pas évidents. Ce problème est pourtant nettement plus important que le précédent.

³ http://www.senat.fr/rap/a03-075-7/a03-075-7_mono.html

⁴ Scopus : <http://www.scopus.com/scopus/home.url>

3.1 Origine de la perte du pilotage de la communication scientifique par la recherche

Les analyses les plus répandues ont souvent privilégié les paramètres techniques liés à la fabrication et surtout à la diffusion du document. Nous n'irons pas plus loin dans l'analyse de ce phénomène déjà décrit en profondeur par Jean-Claude GUÉDON [11][12][13], et que nous résumons ici en quelques mots⁵ :

Eugene Garfield [10] dans le cadre de l'ISI a proposé des indicateurs pour l'évaluation des travaux scientifiques basés sur l'analyse des citations. Dans le même mouvement, les revues ont elles-mêmes été évaluées et ont bénéficié d'un facteur d'impact. Les grands éditeurs ont alors mobilisé leurs ressources financières pour acquérir les revues ainsi repérées. Ensuite, par des mécanismes de concentration industrielle classiques, un tout petit groupe d'acteurs (dont le nombre converge vers un) s'est approprié un monopole de fait sur l'édition scientifique évaluée par les pairs.

Le système informationnel géré par les grands acteurs de l'IST privilégie une conception très individualiste reposant sur deux axes : l'accès au document et l'évaluation individuelle. Or, les besoins de veille ou d'évaluation des organismes publics ou privés ne coïncident pas toujours avec cette offre. Leur satisfaction demande des informations plus variées, mieux triées, ou plus facilement exploitables par des traitements statistiques.

3.2 Limites des chaînes documentaires sur un modèle industriel, même dans un cadre académique

Pour contrer ce phénomène, une première approche consiste à bâtir une solution alternative en utilisant un modèle de production analogue à celui de l'édition commerciale. Parmi les travaux de synthèse traitant de ce sujet citons la thèse de Souad ODEH [17]. Dans cet article nous donnerons quelques éléments de chiffrage en indiquant les limites de cette approche.

Pour donner des ordres de grandeur, nous pouvons nous appuyer sur la production de la base PASCAL qui est d'environ 450 000 notices par an à partir de 4000 à 5000 titres de périodiques. Pour mettre en oeuvre une solution qui soit une alternative réelle, il faudrait traiter au moins 10 000 titres par an, soit 600 000 à 1 000 000 notices, améliorer la précision de l'indexation et introduire des mécanismes d'évaluation (citations). Toutes ces opérations ne sont pas complètement automatisables et demandent un traitement manuel incluant une lecture sommaire de l'article. Les chiffres généralement avancés pour ce type de traitement sont compris entre 25 et 50 € par article. On peut donc considérer un chiffre compris entre 20 et 100 millions € comme une première estimation de la charge annuelle. On arrive au même ordre de grandeur en considérant la mobilisation d'environ 400 personnes, effectif nécessaire pour couvrir toutes les disciplines scientifiques.

Sur un plan financier, la création d'une telle entité⁶ n'est pas insurmontable à l'échelle européenne. Avec le montage du projet Galileo [3], l'expérience montre que la Commission sait prendre une initiative nettement plus coûteuse avec une motivation suffisante. Dans une problématique plus proche, l'Europe a réussi à mettre en place l'Office Européen des Brevets

⁵ Reprises de [6].

⁶ Les variantes d'implantation sont multiples, depuis un centre à effectif limité (moins de 100 personnes) sous-traitant la production par appel d'offre, jusqu'à une institution de 500 personnes en passant par des solutions distribuées s'appuyant sur des institutions existantes (FIZ, INIST...).

dont les effectifs dépassent les 5 000 employés [18]. Cela dit, de nombreux obstacles seraient à lever d'un point de vue politique car il s'agit en fait de concurrencer des entreprises fortement implantées en Europe (Elsevier par exemple !). Sur un plan technique, la coopération des éditeurs pour obtenir les métadonnées n'est pas du tout évidente car ceux-ci produisent également des bases documentaires multidisciplinaires. En supposant la levée de ces obstacles, quelles seraient les retombées et les limites?

Cette solution vise d'abord à donner à la recherche une indépendance stratégique en IST. Imaginons donc, en nous projetant dans un futur éloigné, 15 ans par exemple, l'hypothèse d'un effondrement des grands acteurs commerciaux ou un blocus politique. Une analyse sommaire montre immédiatement que la vraie denrée stratégique est la publication elle-même et non ses métadonnées. C'est donc sur l'information primaire que doit se situer la priorité.

Au niveau de la qualité de service, l'introduction d'un élément concurrentiel est un facteur positif face à un monopole qui se dessine. De même, le pilotage serait amélioré en contrôlant mieux la sélection des titres. Mais, une analyse détaillée des besoins de veille ou de pilotage montre l'importance de services diversifiés en fonction des disciplines scientifiques ou des activités. Pour des raisons économiques, liées à la rentabilité, les acteurs commerciaux ont tendance à homogénéiser au maximum leurs services. Pour le même type de raisons, liées à la saine gestion des ressources publiques, les résultats risqueraient d'être identiques.

Autrement dit, ce type de solution ne conduit pas à une réelle appropriation, et ne répond pas aux besoins analysés plus haut. Il faut donc travailler sur un autre modèle en approfondissant les spécificités du document numérique.

4 Opportunités offertes par le document numérique

S'agissant de documents numériques, nous allons maintenant aborder un paramètre parfois sous estimé dans l'analyse paradigmatique de la situation de l'IST : la structuration du contenu. En effet, le document numérique n'est pas seulement un nouveau support plus commode à manipuler que le papier. Par son aptitude aux traitements, il ouvre réellement de nouvelles perspectives. Sur un plan général et théorique, nous renvoyons ici le lecteur aux travaux du RTP DOC du département STIC du CNRS qui « *se propose de préciser la notion de document dans son passage au numérique à partir de recherches qui privilégient plutôt la forme (comme un objet matériel ou immatériel), le signe (comme un porteur de sens) ou la relation (comme un vecteur de communication)* » [19]. Pour construire un dispositif mieux adapté à la veille, le document numérique apporte un ensemble d'opportunités que nous allons brièvement passer en revue.

4.1 Libre accès à la communication scientifique

Le document numérique autorise le libre accès à la communication scientifique et les archives ouvertes. Pour faire le tour du sujet, le lecteur pourra consulter le site spécialisé de l'INIST⁷. En pratique, le chercheur produit directement, et depuis longtemps, ses articles en format électronique. Il peut maintenant les rendre accessibles à l'ensemble de la communauté.

On reprochait parfois aux articles en libre accès leur manque de « valeur scientifique ». Cet argument n'est plus d'actualité avec la mise en place de comités de lecture comme dans les revues Public Library of Science (PLOS)⁸ ou BioMed Central⁹ (qui proposent d'ailleurs de

⁷ Libre accès : <http://www.inist.fr/openaccess/>

⁸ PLoS : <http://www.plos.org/>

nouveaux modèles économiques). Avec la déclaration de Berlin [16], signée par des grands organismes de recherche et non plus seulement par des individus, le libre accès conquiert une validité institutionnelle.

Par rapport à la problématique de cet article, le libre accès présente cependant des limites. Sur un plan institutionnel, l'organisation actuelle est issue d'initiatives au départ individuelles, par exemple le Centre pour la Communication Scientifique Directe (CCSD)¹⁰, dans lesquelles l'objectif essentiel est d'augmenter la visibilité des chercheurs (voir par exemple les positions de Stevan HARNARD [14] autour du thème « publier ou périr »). Ces priorités se retrouvent dans les solutions logicielles adoptées (Eprints¹¹, HAL, DSpace¹²) où la recherche d'information, au sens documentaire, n'est pas la priorité actuelle (que le lecteur essaye par exemple de faire le point sur la position de tel ou tel acteur du libre accès en consultant les sites d'archives ouvertes). De même, les consignes de dépôt sont souvent trop laxistes pour permettre une exploitation ultérieure, dans un cadre de veille notamment, qui demande par exemple le respect de formats ouverts et structurés.

Le mécanisme de communication directe associé au libre accès est un outil puissant pour la constitution d'inventaires. Les solutions actuelles privilégient actuellement la lecture humaine d'un document au détriment d'un réel traitement. Pour intégrer complètement les besoins liés à la veille ou à l'évaluation, elles constituent en fait un substrat qui doit être amélioré et complété en utilisant la panoplie des autres opportunités technologiques.

4.2 Opportunités technologiques ouvertes au libre accès pour répondre à la globalité des besoins

Pour permettre une meilleure exploitation de la communication scientifique, nous proposons d'analyser les possibilités complémentaires aux archives ouvertes sous deux aspects :

- Comment répondre à la satisfaction directe du veilleur qui veut faire le point sur un sujet et espère trouver sa réponse en ligne ?
- Comment pouvoir mieux traiter l'information en cas d'échec de la stratégie précédente ou pour des besoins plus synthétiques (évaluation) ?

4.2.1 Revues électroniques

Au moment où se multiplient les portails de revues électroniques, il n'est pas inutile de rappeler qu'une revue n'est pas seulement un simple réservoir ou même un outil d'évaluation d'articles reçus de façon aléatoire... Pour reprendre l'exemple du veilleur qui veut faire le point sur l'OAI, une revue électronique telle que dlib¹³ est performante parce que sa politique éditoriale met naturellement les évolutions et permet donc de les repérer. Une reprise effective d'une activité éditoriale par le monde de la recherche est donc un atout essentiel, notamment pour faire connaître les sujets sur lesquels elle a progressé et où elle cherche à coopérer.

Sur un plan technique, on peut maintenant composer un article dans une DTD normalisée (DocBook, TEI, Erudit...) avec des conditions de confort équivalentes à celles de la bureautique. Pour répondre aux besoins de traitement (pour fabriquer ultérieurement et automatiquement des indicateurs utilisant les citations par exemple), une production structurée « à la source » est également fondamentale.

⁹ <http://www.pubmedcentral.org>

¹⁰ CCSD : <http://www.ccsd.cnrs.fr/>

¹¹ Eprints : <http://www.eprints.org>,

¹² DSpace : <http://www.dspace.org>

¹³ D-Lib Magazine : <http://www.dlib.org>

4.2.2 Métadonnées et portails thématiques

Une revue qui développe une politique éditoriale forte donnera naturellement le point de vue d'une communauté¹⁴, or le veilleur a besoin d'une vision plus critique. Une des vocations des portails est d'apporter une vision plus élargie. Là encore, une valeur ajoutée éditoriale est un apport précieux. Le lecteur intéressé par l'OAI peut consulter par exemple la bibliographie de Charles W. Bailey [1] pour voir l'efficacité des sélections et l'éclairage d'un weblog.

Pour faciliter la fabrication de tels portails ou pour permettre la réalisation d'études, la disponibilité de réservoirs de métadonnées est bien entendu essentielle. On dispose maintenant d'une large panoplie d'outils pour les réaliser. Au niveau générique on trouve des protocoles tels que OAI-PMH¹⁵ et des formats de métadonnées (RDF, RSS, Dublin Core plus ou moins Qualifié). Pour répondre à des besoins spécifiques (formules chimiques, mathématiques, localisations géographiques, etc) on dispose également de batteries de solutions qui peuvent cohabiter par l'utilisation des espaces de nom (namespace). Autrement dit, parallèlement à une interopérabilité informatique apportée par xml et son ingénierie, on assiste à la construction d'une interopérabilité à usage sémantique.

Sur un plan strictement technique, la communauté de la recherche dispose des solutions pour produire une information répondant aux besoins de la veille, le problème vient plutôt de leur mise en œuvre.

4.2.3 Ontologies et indexation automatique

Ce passage en revue des opportunités technologiques serait incomplet sans référence aux attentes des programmes tournant autour du web sémantique [2]. Ils sont censés améliorer considérablement la puissance des outils de recherche d'information, aussi bien en mode assisté (réponse immédiate à des besoins de veille) que « fortement expert ».

La mise en œuvre de ces techniques demande cependant un changement culturel majeur : il faudra passer du concept de « prêt à tirer » au « prêt à traiter ». Ce traitement du contenu va amener le chercheur à une appropriation de plus en plus forte des techniques éditoriales ou documentaires. Pour faire fonctionner les nouveaux moteurs de recherche, il faudra construire une solide modélisation du domaine scientifique auquel se réfère un document. S'agissant d'un champ de recherche en émergence, qui sera capable de modéliser les relations d'une discipline nouvelle sinon le chercheur lui-même ? Produire une ontologie efficace demande de réaliser des compromis permanents, qui ne peuvent être formulés que par le spécialiste.

Le même type de remarque s'applique également aux applications de l'indexation automatique.

5 Vers un réseau d'inventaires

Nous sommes donc confrontés à un paradoxe : face à un ensemble très diversifié de besoins, la situation actuelle privilégie une offre monolithique, au moment même où la technologie permet des réponses très spécialisées dans un contexte d'interopérabilité. Comment utiliser ces opportunités pour construire des réseaux, en utilisant la dynamique des inventaires, fondation pour de la construction de bases bibliographiques ou de portails.

¹⁴ La revue dlib citée plus haut, financée par la NSF, la DARPA et le CNRI est naturellement influencée par la stratégie américaine, même si elle se veut ouverte au résultats internationaux.

¹⁵ Protocol for Metadata Harvesting

5.1 Idées directrices

5.1.1 De la hiérarchie au réseau

Intuitivement, et par tradition, la consolidation de sites institutionnels pour constituer des grandes bases de données se traduit par des organisations hiérarchisées (Figure 1).

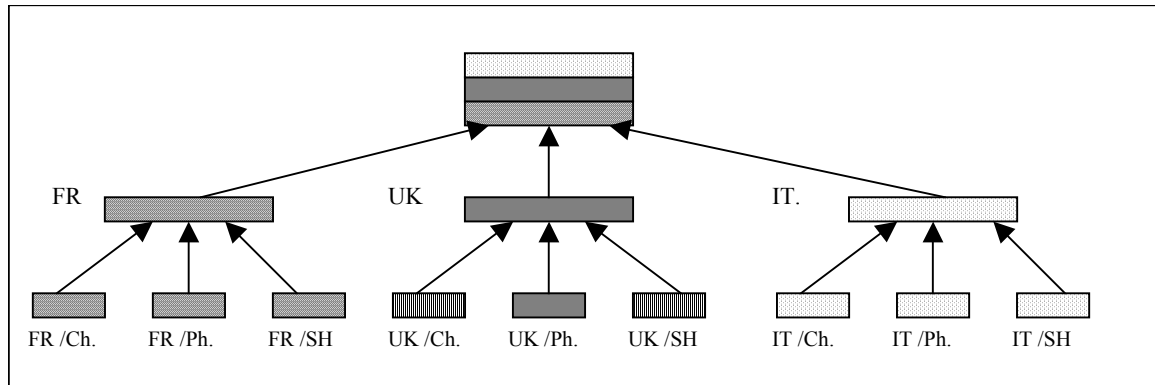


Figure 1 Architecture intuitive d'un réseau de moissonnage.

Dans cet exemple, des serveurs avec des composantes thématiques (Chimie, Physique, Sciences Humaines) sont consolidés au niveau national (France, Royaume-Uni, Italie), puis au niveau international. Cette architecture, utilisée avec rigueur, conduit aux mêmes types d'inconvénients que ceux évoqués plus haut (paragraphe 3.2). En réalité les protocoles tels que OAI-PMH autorisent des architectures plus souples (Figure 2).

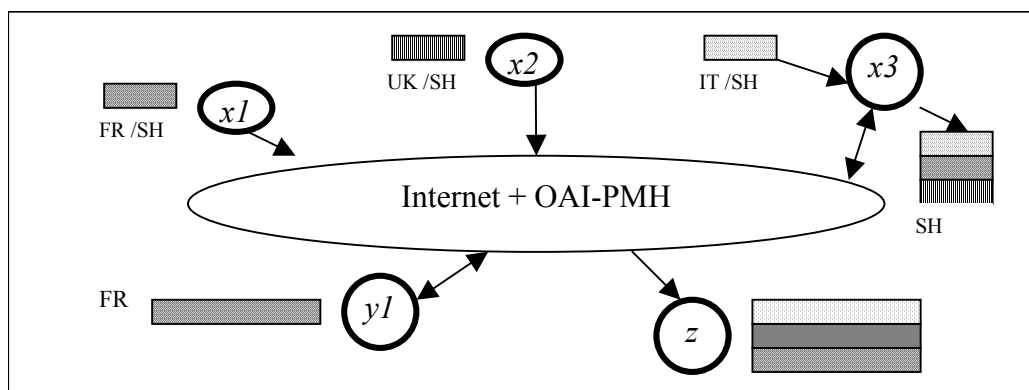


Figure 2 Architecture réseau.

Dans cet exemple (Figure 2), les rôles sont plus diversifiés. Les sites $x1$ et $x2$ déposent leurs documents sur des archives ouvertes ; $x3$ fait de même et construit un portail thématique en Sciences Humaines ; $y1$ entretient par moissonnage un serveur national et z construit un portail généraliste (à partir de sites y). À côté de vastes réservoirs d'informations, généralistes et donc fatalement limités, une telle architecture permet la multiplication d'initiatives spécialisées plus orientées vers des besoins précis et adaptées à chaque thématique ou problématique.

5.1.2 Asynchronisme, appropriation et maîtrise technologique

Dans leur mise en œuvre, les deux démarches précédentes sont radicalement différentes. Dans le premier cas, une équipe technique peut développer des solutions informatiques et définir des consignes qui seront respectées à tous les étages. Dans l'approche réseau, une réflexion

spécifique doit être menée à chaque nœud, ce qui induit des changements profonds dans les méthodes de travail ou de conception.

Une première difficulté est liée à l'interopérabilité des solutions envisagées. Il est parfaitement légitime qu'une institution fasse cohabiter deux objectifs :

1. mettre son information à la disposition de tiers sans savoir exactement ce que ceux-ci vont en faire - ce qui est en contradiction avec de nombreuses méthodologies de conception informatique.
2. réaliser un portail mettant en avant sa visibilité et offrant des services à haut degré d'intégration - ce qui ressemble à une approche plus classique.

Les méthodes informatiques couramment enseignées en système d'information privilégient des solutions synchrones, s'implémentant facilement avec un SGBD mais supposant une analyse exhaustive de tous les traitements envisageables. Dans une approche réseau, on est obligé d'adopter une démarche radicalement différente où l'on doit désynchroniser la constitution des données mises en libre accès de la construction d'un portail. De plus, chaque élément actif du réseau doit définir sa propre stratégie en respectant des contraintes d'interopérabilité.

Cela signifie naturellement une appropriation de la problématique de l'IST dans une dimension complète : « que dois-je mettre en ligne pour atteindre un veilleur avec efficacité, et comment ? ». Cela suppose également une maîtrise technologique permettant de mettre en œuvre ou de paramétrer des applications personnalisées à partir de solutions génériques.

5.2 Nouveau rôle des acteurs de la communication scientifique

Comment se déclinent le développement des initiatives et les besoins d'appropriation au niveau des acteurs de la recherche et plus précisément de la communication scientifique ?

5.2.1 Chercheur : auteur, linguiste, éditeur !

Le point le plus crucial est celui du chercheur qui est aussi un auteur. Nous l'avons vu, tout serait pratiquement résolu si tous les auteurs mettaient toute leur production en ligne en respectant la normalisation. La consultation du CCSD¹⁰ ou d'autres sites d'archivage montre qu'un outil, même s'il devient de plus en plus convivial, ne suffit pas pour enclencher un mécanisme de collecte exhaustive. Un niveau avancé de culture informationnelle et des mécanismes incitatifs sont donc indispensables, nous en parlerons dans les sections suivantes.

Mais cela ne suffit pas. Nous l'avons évoqué, les nouveaux outils liés par exemple au web sémantique demandent une implication de plus en plus forte des chercheurs, les seuls vrais experts d'un domaine scientifique en mutation. Par exemple, le choix d'un système de codification chimique ou mathématique¹⁶ demande un haut niveau d'expertise du domaine et une excellente maîtrise algorithmique. Les besoins de formation sont donc conséquents.

Il est ainsi essentiel que des chercheurs puissent s'investir des activités d'ingénierie linguistique ou éditoriale relevant de leur spécialité. En pratique, la situation française présente un handicap majeur car une telle activité n'est pas valorisée au niveau de l'évaluation. Dans ce contexte, la dichotomie administrative entre les ingénieurs et les

¹⁶ La NSF vient par exemple de lancer un programme de R&D sur les moteurs de recherche utilisables sur des formules mathématiques.

chercheurs pose un nouvel obstacle. Pour faire face au transfert de connaissances entre la recherche et la formation, on a su créer des statuts d'enseignants chercheurs. Peut être faudrait-il donner un cadre aux ingénieurs chercheurs ?

5.2.2 Environnement local, laboratoire, bibliothèque

L'environnement scientifique du chercheur (équipe, laboratoire) joue bien entendu un rôle primordial sur les motivations de chercheurs et la mise en place de procédures de diffusion (par un serveur local ou dépôt sur un ou plusieurs centres tels que le CCSD bien identifiés en fonction de la thématique). Mais c'est également à ce niveau que peuvent être initialisées des opérations thématiques bien ciblées (portail thématique, revue électronique) intégrant les besoins potentiels de différentes communautés.

Sur un plan logistique, les bibliothèques et centres de documentation des laboratoires académiques peuvent jouer un rôle capital en adaptant leur fonctionnement à cette nouvelle situation. Du temps du support papier, leur fonction était centrée sur l'importation d'informations utiles et leur gestion. Aujourd'hui l'information source est directement utilisable par les utilisateurs. Le rôle des infomédiaires de la recherche publique devrait maintenant rendre prioritaire l'exportation des documents et métadonnées produites par leur communauté. Le cadre technique (normes, outils...) existe pour construire par exemple un portail « moissonnable » dans un contexte OAI, mais, là encore, les besoins de formation ne doivent pas être sous-estimés¹⁷.

5.2.3 Environnement national

Au niveau national, on constate une mobilisation de plus en plus forte des institutions. La plupart des organismes de recherche ont été alertés par les coûts de plus en plus prohibitifs des abonnements. Ils ont mis en œuvre de nouvelles initiatives (par exemple le consortium COUPERIN¹⁸). Ils reconnaissent le caractère hautement stratégique de l'IST, ceci est particulièrement clair dans un rapport¹⁹ réalisé par une mission IST au sein du CNRS. De plus, l'amélioration du dispositif national d'IST demande une meilleure utilisation des moyens existants, une recherche de synergie au niveau international et pas forcément une abondance de ressources complémentaires. On peut donc espérer un fort soutien institutionnel.

Dans l'immédiat, l'obstacle le plus difficile vient probablement de la résistance au changement de la part des acteurs de l'IST eux-mêmes. Nous l'avons évoqué à propos du passage du mode hiérarchisé au mode réseau, les changements de paradigmes en cours dans le monde de la communication scientifique demandent de nouvelles façons d'agir qui sont parfois en opposition radicale avec les anciennes. Il ne s'agit plus seulement de **faire** mais aussi d'apporter un soutien logistique pour **que d'autres fassent**. Les opérateurs nationaux qui géraient des chaînes de production doivent se reconvertir sur des fonctions de moissonnage, normalisation, formation et animation des réseaux de terrain.

¹⁷ La production des documents normalisés dans une DTD telle que DocBook ou TEI est plus complexe que la création d'une base Texto avec des règles locales...Voici d'autres exemples montrant le type de problèmes qui seront de plus en plus souvent traités au niveau du laboratoire. Pour faire cohabiter plusieurs systèmes de codification tels que MeSH ou Rameau, la normalisation des métadonnées demande une bonne maîtrise de la technologie XML. Une codification arborescente des affiliations, (où par exemple l'Unité de Recherche et d'Innovation de l'INIST est codée par /FR/CNRS/INIST/URI au lieu de XX999) apporte infiniment plus de souplesse que les tables traditionnelles mais demande une excellente maîtrise des arborescences, au sens mathématique du terme.

¹⁸ <http://couperin.cines.fr/>

¹⁹ non rendu public au moment de la rédaction de cet article.

Dans ce contexte, tous les acteurs participants à l'évaluation jouent un rôle important, et notamment les sociétés savantes et comités de programme. En effet, il s'agit d'inciter les chercheurs à produire une information structurée avec des éléments d'indexation normalisés. Dans un passé récent, des communautés de mathématiciens ou de physiciens ont réussi dès 1980 à faire produire des articles en LaTeX dans un environnement peu convivial (emacs) grâce à l'implication de comités de programme qui rendaient son usage obligatoire.

6 Conclusion

L'émergence de l'imprimerie est un paradigme bien connu des historiens et l'idée de comparer l'apparition des NTIC à l'invention de Gutenberg n'est plus nouvelle. L'histoire nous montre l'ampleur des mutations attendues comme l'atteste la littérature abondante qui lie à l'imprimerie le développement de la Réforme [8][9] ou l'apparition de la législation sur la propriété intellectuelle [20]. Les bouleversements des mécanismes de l'innovation sont donc prévisibles et correspondent à un phénomène quasiment naturel.

Les acteurs publics de la recherche sont temporairement en situation de faiblesse pour répondre convenablement aux besoins d'information scientifique de l'industrie ou même pour leurs propres besoins, car ils doivent traiter une information dont les droits d'exploitation appartiennent aux éditeurs. En réalité, maintenant que les verrous techniques se lèvent progressivement, la recherche dispose d'un avantage très significatif dans une stratégie d'appropriation de la production de son information. La communauté scientifique, qui produit l'information concernée, dispose de fait d'un avantage colossal sur le monde commercial pour maîtriser l'ensemble du dispositif.

Le monde de l'information scientifique se trouve dans une situation paradoxale où la concentration industrielle n'a jamais été aussi forte alors que la technologie permet une approche totalement inverse. Les effets pervers et des insuffisances de cette situation sont dénoncés lorsqu'on raisonne au niveau collectif. En revanche, la recherche d'un confort individuel par rapport à l'évaluation ou à l'acquisition encourage ce modèle monopolistique.

Le terme « plaidoyer » utilisé dans le titre de cet article s'adresse donc essentiellement à l'ensemble des acteurs de la production scientifique. Des progrès significatifs ont été accomplis dans les mouvements liés au libre accès à l'information scientifique où les revendications individuelles ont fait place à des prises de positions institutionnelles. Celles-ci débouchent déjà sur des initiatives opérationnelles sous forme de mise en place de nouveaux services d'archivage ou d'édition électronique. Il faut aller encore plus loin.

Et les veilleurs, doivent-ils simplement attendre que le monde de la recherche se transforme pour espérer des jours meilleurs ?

Le plaidoyer les concernent aussi. Ils vont opérer dans un nouvel environnement où ils devront acquérir ou perfectionner un double savoir-faire. De nombreux organismes de formation, académiques ou industriels ont tendance à vouloir former de « très bons utilisateurs de systèmes documentaires ». Cela ne suffit pas. Certes, les veilleurs doivent devenir d'excellents utilisateurs de la navigation dans l'Internet pour repérer des sources signalant déjà les signaux faibles qu'ils recherchent ou recelant les synthèses dont ils ont besoin. Mais, en cas d'échec de cette stratégie, ils doivent devenir des spécialistes de l'interopérabilité pour tirer le maximum de profit de l'information disponible, autrement dit : des experts capables d'intervenir là où les « très bons utilisateurs » échouent

7 Remerciements

Je tiens à remercier mes collègues de l'INIST, et notamment Catherine AUBIER, Anne-Marie BADOLATO, Valérie DEVAUX et Francis ANDRE, pour leurs précieux conseils dans la rédaction de cet article.

8 Bibliographie

- [1] BAILEY C., Scholarly Electronic Publishing Bibliography, Version 54, 13/7,2004 <http://info.lib.uh.edu/sepb/sepb.html> (consulté le 16 Août 2004)
- [2] BERNERS-LEE T., HENDLER J. et LASSILA O., *The Semantic Web*, Scientific American, 2001 284(5), pp.34-43
- [3] Commission Européenne, EUROPA - Energie et Transports - GALILEO, 2003 http://europa.eu.int/comm/dgs/energy_transport/galileo/index_fr.htm (consulté le 18 Avril 2004)
- [4] Commission des Présidents d'Université, Quelques éléments sur la place de la recherche conduite dans les universités au sein de la recherche publique en France. Rapport de la CPU http://www.cpu.fr/Telecharger/2004-02Colloque/Place_rechercheUniversites.pdf (consulté le 18 Avril 2004)
- [5] DUCLOY J., GRIVEL L., LAMIREL J.-C., POLANCO X. et SCHMITT L, *INIST's experience in Hyper-Document building from bibliographic Databases*, Proceedings of RIAO'91 - Barcelone, Avril 1991
- [6] DUCLOY J., *Le français dans les publication scientifiques, comment contrer le déclin du français*, colloque RIFRAM Paris 2003
- [7] DUSOULIER N. et DUCLOY J., *Processing of data and exchange of records in a scientific and technical information center. Forlats : what for ?* UNIMARC/CCF Workshop Florence (IFLA/UNESCO), 05-07 Juin 1991.
- [8] EISENSTEIN, E., *The Printing Press as an Agent of Change*, Cambridge University Press, 1979
- [9] EISENSTEIN E., *The Printing Revolution in Early Modern Europe* (Cambridge, U.K.: Cambridge University Press, 1983)
- [10] GARFIELD E., *Citation analysis as a tool in journal evaluation*, Science 178:471-79, 1972.
- [11] GUEDON J.-C., *Les journaux électroniques - L'importance de la mobilisation contre les pratiques de certains éditeurs commerciaux*, Rencontres de Biblio-fr, 1998
- [12] GUEDON J.-C., *Numériser les revues savantes : d'un commerce à l'autre*. La Recherche, 335 (Oct 2000), pp 78-85.
- [13] GUEDON J.-C., *Open Access Archives: from scientific plutocracy to the republic of science*, IFLA journal. [IFLA j..] 2003 , vol. 29 , no 2
- [14] HARNAD S., *The Self-Archiving Initiative*. Nature 410: 1024-1025 (2001) <http://www.nature.com/nature/debates/e-access/Articles/harnad.html> (vérifié le 09/08/04)
- [15] HARNAD, S. (2003) *Self-Archive Unto Others as Ye Would Have Them Self-Archive Unto You*. The Australian Higher Education Supplement. <http://www.ecs.soton.ac.uk/~harnad/Temp/unto-others.html>
- [16] Max Planck Gesellschaft, *Berlin Declaration* Berlin 2003 <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html> (vérifiée le 08/09/2004)
- [17] ODEH S., *Le positionnement des intermédiaires en économie de l'information numérique : le cas de l'industrie bibliographie*, Thèse Lyon 2004.
- [18] OEB (Organisation Européenne des Brevets), *Quelques faits et chiffres* http://www.european-patent-office.org/epo/facts_figures/facts2002/pdf/facts_figures_02.pdf
- [19] PEDAUQUE R., *Document : forme, signe et relation, les re-formulations du numérique..* Article. 28 avril 2003. Working paper. http://archivesic.ccsd.cnrs.fr/sic_00000413.html
- [20] WOODLEY B., *The Impact of Transformative Technologies on Governance: Some Lessons from History*, October 2001 url (vérifiée 03/2004) : http://www.iog.ca/publications/transformative_tech.pdf