

Appropriation des réseaux d'inventaires scientifiques par les entités de recherche en émergence

Jacques DUCLOY (*), Lucile GRASSET(**)

Jacques.Ducloy@inist.fr, Lucile.Grasset@cirad.fr

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

(**) Cirad, Avenue Agropolis, 34398 Montpellier Cedex 5 (France)

Mots clefs :

Besoin information, Veille technologique, Recherche scientifique, Marché information, Renforcement des capacités, Information scientifique technique, Publication, Libre accès, Base donnée bibliographique, Document électronique.

Keywords:

Information needs, Technological watch, Scientific research, Information market, Capacity Building, Scientific technical information, Publication, Open access, Bibliographic database, Electronic document.

Résumé

Dans cet article nous étudions les insuffisances du dispositif d'Information Scientifique et Technique (IST) actuel en regard des besoins des structures de recherche en émergence (recherches pionnières et/ou dispositifs en structuration). Nous proposons des réflexions allant dans le sens d'une appropriation de l'IST par le monde scientifique en utilisant les avancées des technologies du document structuré et en profitant des mouvements d'opinion autour des archives ouvertes. Nous donnons des idées directrices pour la construction de réseaux d'inventaires et présentons une initiative éditoriale et collective nommée ARTIST pour appuyer cette construction.

1 Introduction

Cette communication constitue la suite d'un article[7] publié pour le congrès VSST'04¹. Nous avons alors montré les insuffisances du dispositif d'Information Scientifique et Technique par rapport aux besoins du transfert de technologie et donné des éléments de réponse par la création de réseaux d'inventaires. Nous voudrions ici donner un éclairage complémentaire, plus proche du pilotage du monde de la recherche.

Plus précisément nous nous focalisons sur les besoins des entités en émergence dans le monde de la recherche. Le terme « émergence » est naturellement inspiré de l'expression « pays en émergence », pour lesquels le montage de réseau des opérateurs, laboratoires ou centre de ressources, et le pilotage de l'ensemble sont complexes. Les grandes bases bibliographiques et d'évaluation qui sont adaptés aux pays fortement industrialisés s'avèrent très difficiles à

¹ VSST : Veille Stratégique Scientifique et Technologique, Toulouse 25-29 Octobre 2004 <http://atlas.irit.fr/>

Pour une meilleure lisibilité, nous avons repris une très grande partie du texte de cette communication antérieure dans cet article.

utiliser dans ce contexte. En effet, l'absence d'un ensemble de références de taille significative rend inopérante toute confrontation à un ensemble mondialisé. Ce problème nous semble en réalité très proche des recherches pionnières, qui même dans les pays les plus actifs de la recherche, ont beaucoup de difficulté à émerger.

Pour compléter cette introduction, dans une première partie nous approfondirons la question : les grandes revues scientifiques sont-elles adaptées à l'émergence de nouveaux thèmes de recherche ou de nouvelles structures d'aide à l'innovation ? Sont-elles adaptées à la structuration des activités de Recherche dans les pays en émergence ?

Dans une deuxième partie, nous donnerons un état de l'art sur les nouveaux modèles de la communication scientifique. Avec le développement du libre accès à l'information scientifique [18] celle-ci est une activité en pleine mutation. Une grande partie de la littérature sur le sujet privilégie dans son analyse la réponse à deux catégories de besoins : l'accès à l'information par le chercheur d'une part et son évaluation individuelle² de l'autre, et non la globalité des besoins tels que le pilotage de la recherche. Or les nouvelles technologies, du document électronique au web sémantique, permettent justement de compléter ce dispositif pour constituer des réseaux d'inventaires de résultats (et pas seulement de publications) mieux adaptés à ces besoins.

Dans une troisième partie nous donnerons des éléments techniques et organisationnels pour la fédération de réseau d'inventaires. Nous présenterons également une initiative nommée ARTIST que nous mettons en place pour favoriser les coopérations associées.

2 Analyse des besoins d'information produites par la recherche

Le lectorat potentiel de la recherche est vaste et ses besoins sont multiples et variés. Nous qualifierons de « bibliothéconomique universel » la simple acquisition de connaissance pour approfondir des besoins plus précis liés à un passage à l'action comme une recherche de partenaire ou une procédure d'évaluation.

2.1 Besoins bibliothéconomiques « universels »

Par un moyen quelconque, un utilisateur acquiert une information plus ou moins complète ou structurée sur une publication, dans une bibliographie par exemple, et il cherche à accéder au contenu. Nous ne citons ici ce type de besoin que pour mémoire car la littérature sur ce sujet est particulièrement abondante. En effet, la plupart des travaux menés sur les bibliothèques électroniques, et plus particulièrement les archives ouvertes, héritent des traditions bibliothéconomiques et sont focalisées sur ce thème.

Nous analyserons maintenant les besoins qui sortent de ce cadre.

2.2 Quoi de neuf ?

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.

² On pourrait presque parler de notation...

Sur les recherches pionnières, voici un exemple adapté à l'ingénierie de l'information : nous nous sommes intéressés à l'utilisation de la technologie SGML/XML pour l'analyse de l'information dans les années 90 [5][8]. 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 avec l'émergence du standard XML. C'est maintenant un signal fort devenu une évidence.

Dans les domaines de recherche directement adaptés aux pays en émergence, comme la littérature sur les ressources naturelles, les grandes bases de données comme AGRIS sont construites à partir des apports de bases régionales ou spécialisées. Les pays qui ne disposent pas d'une telle infrastructure de constitution de bases de données rencontrent dans les faits des difficultés sérieuses pour accéder aux données et une quasi-impossibilité pour y figurer.

La question « qui fait quoi » devient : comment diffuser vers la communauté internationale un sujet nouveau, surtout lorsqu'il vient d'un système de recherche en émergence.

2.3 Qui fait quoi ?

Les chercheurs se considèrent souvent comme très autonomes dans la recherche de compétences concernant leur spécialité. 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 vu comme une deuxième approche.

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 international se retrouve dans une position voisine de celle d'un acteur extérieur à la recherche (industriel...). Ici encore, pour répondre à la question « qui fait quoi ? », 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 enseignement sur l'intelligence économique où interviennent des enseignants issus d'un laboratoire d'informatique ».

La question « **qui fait quoi ?** » est probablement au cœur du mécanisme de transfert de technologie, et particulièrement dans les systèmes de recherche en émergence. Elle se formalise souvent comme : « où puis-je trouver un expert disponible ? ». Les annuaires de compétence sont par exemple des compléments fondamentaux aux archives bibliographiques³.

³ A titre d'exemple, un des premiers services ouverts par l'IMIST (Institut Marocain d'Information Scientifique et Technique) dans sa phase de préfiguration est précisément un annuaire des compétences. Voir : <http://www.imist.ma/competences/>.

2.4 Comment ?

Enfin la question « **comment ?** » met les grandes bases bibliographiques dans une position de plus en plus difficile. En effet, le format des publications scientifiques (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.

Les grandes bases bibliographiques sont des outils de première approche, pertinents pour repérer une activité scientifique « peu silencieuse ». Or une stratégie d'investigation informationnelle dans un contexte d'innovation demande des sources multiples, avec des composantes thématiques et régionales. Les bases de données institutionnelles proposent quant à elles un panorama plus large des activités de recherche et de développement des entités de recherche. Elles offrent notamment, l'accès à des références de littérature grise et la localisation des documents. Trop peu encore permettent l'accès au texte intégral⁴.

En résumé, le monde de la recherche devrait améliorer notablement le signalement des résultats et des compétences. Les éléments bibliographiques éditoriaux ne sont qu'un élément d'un inventaire plus large (ce qui ne veut pas dire plus volumineux !).

2.5 Le pilotage 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 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. Dans un récent rapport [4], la Conférence des Présidents d'Université mettait précisément l'accent sur la difficulté de mesurer le poids de la recherche conduite dans les universités françaises.

En pratique, pour traiter des données au niveau des institutions de recherche ou des collectivités territoriales, une des difficultés majeures consiste à repérer (et regrouper) les organismes de recherche nationaux. Cette remarque est partagée par d'autres institutions ayant des missions d'évaluation telles que l'OST⁵, l'UNIPS⁶ du CNRS. Elle est corrélée par les multiples tentations de création d'observatoires à différents niveaux (université, délégation régionale des établissements de recherche...).

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 recherche⁷. Un

⁴ IRD : http://www.bondy.ird.fr/pleins_textes/

⁵ OST : Observatoire des Sciences et des Techniques <http://www.obs-ost.fr/fr/>

⁶ UNIPS : l'Unité d'Indicateurs de Politique Scientifique , <http://www.cnrs.fr/DEP/entite/unips/unips.htm>

⁷ 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.

extrait de l'avis concernant la recherche du projet de loi de finances⁸ 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'Évaluation 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 officiel⁹ ne laisse planer aucun doute sur l'ampleur des ambitions : plus de 14 000 titres de périodiques avec les citations vérifiées. De même, une lecture des accords¹⁰ entre CrossRef et Google Scholar est particulièrement explicite : *“if there are multiple versions of an article shown in the Google Scholar search results, the first link will be to the publisher's authoritative copy”*. Le monde de la recherche semble donc perdre la maîtrise de la communication scientifique.

La part des coopérations internationales dans la production scientifique des pays émergents a une importance non négligeable. Mis à part, L'Afrique du Sud, l'Égypte et le Nigéria¹¹, les autres pays publient en coopération avec des équipes internationales (50% pour les pays anglophones, 60 à 80% pour les pays francophones)[23]. La visibilité des activités scientifiques et technologiques des pays africains passe essentiellement par des collaborations avec l'Europe et l'Amérique du Nord. En effet, seules 5% des publications lient entre eux plusieurs pays d'Afrique, les réseaux Sud-Sud sont donc rares. L'alliance avec les équipes occidentales permet de publier dans les titres de revues à facteur d'impact, en revanche, les réseaux régionaux ne permettent pas d'aboutir à un tel résultat. Cependant, il serait pertinent de renforcer ces collaborations entre pays émergents, les « problématiques » ne s'arrêtant pas aux frontières. Publier, sur quels supports ? L'avènement de nouveaux modèles d'édition¹² pourrait offrir une opportunité de rendre visible la production des réseaux entre pays émergents.

⁸ http://www.senat.fr/rap/a03-075-7/a03-075-7_mono.html

⁹ Scopus : <http://www.info.scopus.com/>

¹⁰ Voir CrossRef Newsletter de février 2005 : <http://www.crossref.org/01company/10newsletter.html>

¹¹ La majorité de la production scientifique de ces trois pays est prioritairement nationale, seule, 40% de leur production est cosignée avec des équipes internationales.

¹² Citons, le programme PERI de l'INASP est un des nombreux programmes de soutien à l'information de recherche <http://www.inasp.info/peri/indexfr.html>

De même, le programme Technologies de l'information et de la communication et appropriation des savoirs de l'AUF constitue également une dynamique dans ce domaine : <http://www.auf.org/programmes/programme4/>

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.

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 [12][13][14], et que nous résumons ici en quelques mots¹³ :

Eugene Garfield [11] 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.

Signalons également le rapport du Wellcome Trust [21] qui donne une analyse très détaillée de la situation économique de la communication scientifique. 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.

Pour contrer ce phénomène, une première approche consisterait à bâtir une solution alternative en utilisant un modèle de production analogue à celui de l'édition commerciale, autrement dit en consolidant les bases de données bibliographiques d'origine publique. Parmi les travaux de synthèse traitant de ce sujet citons le site de la « National Federation of Abstracting and Information Services (NFAIS¹⁴) ». Dans les travaux en langue française citons la thèse de Souad ODEH [19], ou, sur le modèle économique de l'édition scientifique, l'ouvrage de Thierry CHANIER[3]. Dans l'article de VSST[7] nous avons montré les limites de ce type de solution. Il faut donc travailler sur un autre modèle en approfondissant les spécificités du document numérique.

3.2 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

¹³ Reprises de [6].

¹⁴ NFAIS : <http://www.nfais.org/>

travaux du programme RTP DOC¹⁵ 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)* » [20]. Pour construire un dispositif mieux adapté à la veille, le document numérique apporte un ensemble d'opportunités que nous allons brièvement passer en revue.

3.3 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'INIST¹⁶. 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)¹⁷ ou BioMed Central¹⁸ (qui proposent d'ailleurs de nouveaux modèles économiques). Avec la déclaration de Berlin [18], signée par des grands organismes de recherche et non plus seulement par des individus, le libre accès conquiert une légitimité 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)¹⁹, dans lesquelles l'objectif essentiel est d'augmenter la visibilité des chercheurs (voir par exemple les positions de Stevan HARNARD[16][16] autour du thème « publier ou périr »). Ces priorités se retrouvent dans les solutions logicielles adoptées (Eprints²⁰, HAL, DSpace²¹) 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 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.

3.4 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 :

¹⁵ Le RTP DOC est une action du Département Sciences et Technique de l'Information et de la Communication (STIC) du CNRS.

¹⁶ Libre accès : <http://www.inist.fr/openaccess/>

¹⁷ PLoS : <http://www.plos.org/>

¹⁸ <http://www.pubmedcentral.org>

¹⁹ CCSD : <http://www.ccsd.cnrs.fr/>

²⁰ Eprints : <http://www.eprints.org>,

²¹ DSpace : <http://www.dspace.org>

- 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) ?

3.4.1 Revues électroniques

Au moment où se multiplient les portails de revues électroniques²², 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... Reprenons l'exemple du veilleur qui fait le point sur l'OAI, une revue électronique telle que D-lib Magazine²³ est performante parce que sa politique éditoriale met en évidence 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 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.

3.4.2 Métadonnées, portails thématiques et syndication

Une revue qui développe une politique éditoriale forte donnera naturellement le point de vue d'une communauté²⁴, 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-PMH²⁵ et des formats de métadonnées (RDF, RSS, Dublin Core plus ou moins Qualifié). Ces approches technologiques, en particulier la syndication, ouvrent de nouvelles perspectives dans le domaine de la publication scientifique [15]. Cette innovation technologique pourrait permettre en particulier aux scientifiques des pays émergents de gérer leurs réseaux et de publier les savoirs développés au Sud, et plus généralement de se réappropriier le pilotage de la communication scientifique.

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.

²² Pour mémoire, rappelons que le rapport de forces entre les publications évaluées par les pairs dans le circuit commercial est de 23 000 titres à comparer aux 1200 revues en libre accès.

²³ D-Lib Magazine : <http://www.dlib.org>

²⁴ 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 aux résultats internationaux.

²⁵ Protocol for Metadata Harvesting

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.

3.4.3 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.

4 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.

4.1 Idées directrices

4.1.1 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).

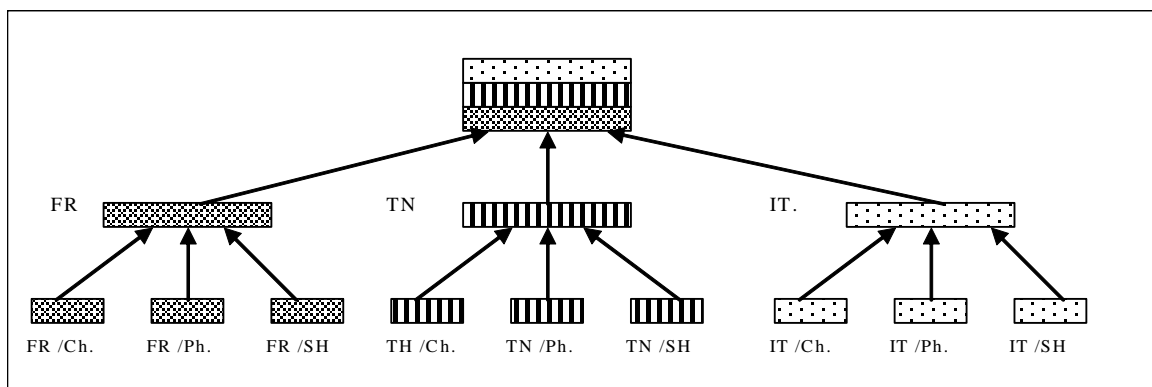


Figure 1 Architecture intuitive d'un réseau de moissonnage.

Dans cet exemple, des serveurs avec des composantes thématiques (Chimie, Physique, Sciences Humaines) sont consolidés au niveau national (France, Tunisie, Italie), puis au

niveau international. Cette architecture, utilisée avec rigueur, conduit aux mêmes types d'inconvénients que ceux reprochés au modèle commercial. En effet elle conduit de fait à une agrégation du dispositif de production et aboutit à des phénomènes de monopole. En réalité les protocoles tels que OAI-PMH autorisent des architectures plus souples (Figure 2).

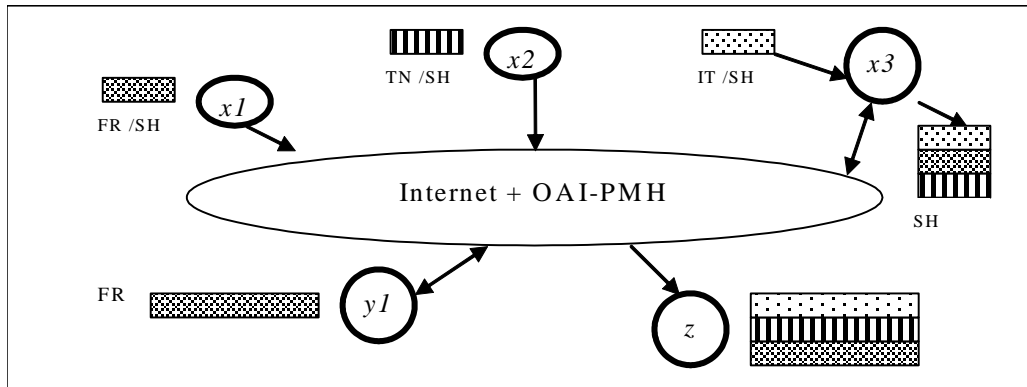


Figure 2 Architecture réseau.

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). A 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.

4.1.2 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.

4.2 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 ?

4.2.1 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 CCSD¹⁹ 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 mathématique²⁶ 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 dans 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 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 ?

4.2.2 Environnement local, laboratoire, bibliothèque

L'environnement scientifique du chercheur (équipe, laboratoire) joue bien entendu un rôle primordial dans les motivations des chercheurs pour la mise en place de procédures de diffusion vers des centres de dépôt au niveau local, institutionnel ou thématique. Mais c'est également dans les laboratoires que peuvent être initialisées des opérations thématiques bien ciblées (portail thématique, revue électronique) intégrant les besoins potentiels de différentes communautés scientifiques.

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

²⁶ La NSF vient par exemple de lancer un programme de R&D sur les moteurs de recherche utilisables sur des formules mathématiques.

portail « moissonnable » dans un contexte OAI, mais, là encore, les besoins de formation ne doivent pas être sous-estimés.

4.2.3 Environnement national

Les organismes de recherche ont été alertés par les coûts de plus en plus prohibitifs des abonnements. Ils sont donc de plus en plus mobilisés. En France, les universitaires ont mis en œuvre de nouvelles initiatives comme le consortium COUPERIN. De son côté, le CNRS a annoncé la création d'une direction de l'IST au cœur de sa stratégie scientifique. Les grands établissements multiplient les échanges et mutualisent leurs ressources²⁷.

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 plus haut, les changements de paradigme 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 avant tout d'apporter un soutien logistique pour **que d'autres fassent**. Les opérateurs nationaux qui géraient des chaînes de production doivent se reconverter sur de nouvelles fonctions telles que : organisation des réseaux de moissonnage de métadonnées, normalisation, formation et animation des réseaux de terrain.

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.

4.2.4 Environnement international

L'objectif ultime de cette stratégie est bien entendu de renforcer les collaborations internationales et régionales pour créer des réservoirs de compétences et de données dans lesquelles les communautés initiales développent leur visibilité. Avec une masse critique de données représentatives des activités de la recherche dans le monde, il sera possible de construire des observatoires ou pôles de références au travers les communautés obtiendront leur reconnaissance.

Par exemple, dans ce contexte, le projet financé par le Ministère des Affaires étrangères français, intitulé « Aide à la structuration et au développement des systèmes d'information et de communication (SIST) dans les établissements d'enseignement supérieur et de recherche en Afrique et à Madagascar »²⁸, propose de réaliser un portail d'accès aux ressources internationales et nationales d'une douzaine de pays, d'offrir un espace éditorial et pédagogique où les experts pourront échanger et co-construire des projets, co-publier et accroître le dialogue des équipes du Sud entre elles ou avec les équipes du Nord. Cette initiative permettra, en complément des grandes bases de données bibliographiques de renforcer la visibilité de la littérature grise produite au Nord, et surtout d'offrir une « vitrine » des connaissances produites par le Sud.

²⁷ Par exemple, la création d'une maison d'édition commune à l'INRA au Cirad, au CEMAGREF et à l'IFREMER.

²⁸ <http://www.sist-sciencesdev.net/>

4.3 L'Initiative ARTIST

Pour faire progresser les acteurs dans cette voie, l'INIST, en coopération avec diverses organisations de la Recherche, comme le RTP-DOC a lancé une initiative éditoriale et coopérative nommée ARTIST (Appropriation par la Recherche des Technologies de l'IST²⁹). ARTIST rassemble d'une part des spécialistes de l'IST ou de son ingénierie (chercheurs ou praticiens) et, d'autre part, des acteurs de la recherche. Ils ont en commun la volonté d'œuvrer pour une appropriation par la recherche de son dispositif de communication scientifique et d'évaluation.

4.3.1 Soutien éditorial à la recherche sur les nouvelles technologies de l'IST

Cette activité vise à faire lever les verrous technologiques ou sociologiques qui bloquent une stratégie d'appropriation. Cette problématique est multidisciplinaire et couvre un vaste panorama d'activités : édition électronique, constitution de bases de données, activités terminologiques, veille etc. On adoptera un point de vue spécialisé en traitant des questions innovantes dans lesquelles un mécanisme d'appropriation par la communauté scientifique est clairement identifié.

En pratique, nous visons deux types d'activités : ouverture d'un espace de discussion par des forums spécialisés, et édition d'une revue électronique. Voici quelques exemples de thèmes actuellement retenus : Architectures fédératives et Métadonnées, Infométrie et Recherche, La Langue des Sciences.

4.3.2 Observatoire de l'appropriation

Le travail théorique exposé plus haut demande une bonne analyse des besoins de la communauté scientifique et un bon repérage de l'état de l'art. C'est pourquoi nous proposons de monter en parallèle un observatoire des nouvelles pratiques de la recherche vis à vis des technologies de l'IST, en focalisant la vision « appropriation de la technologie ».

En pratique, ce dispositif doit repérer les expériences d'appropriation, en identifiant les problèmes d'appropriation et en analysant les résultats obtenus. Il est alimenté par des pratiques de veille avec une forte accentuation thématique, et en privilégiant les aspects coopératif (et non un recensement exhaustif). Pour illustrer ce propos, sur un thème comme la chimie, il s'agit d'identifier les acteurs du changement dans l'IST, ceux qui travaillent sur la codification des formules chimiques dans une orientation système ouvert par exemple, et non de réaliser un inventaire de producteurs...

4.3.3 Le portail coopératif ARTIST

La matérialisation de ces activités est réalisée dans le cadre d'un portail d'édition coopérative³⁰. Nous avons ouvert un prototype de service, qui utilise à ce jour la solution

²⁹ Ce sigle peut également se lire comme Appropriation de l'IST à travers les Technologies.

³⁰ Portail ARTIST : <http://artist.inist.fr/>

SPIP comme base de développement. Ce premier dispositif apporte une infrastructure de dialogue et de diffusion d'information.

Il fonctionne également comme une expérience pilote. En effet, son utilisation met en évidence des besoins de natures variées qui font appel à des logiques différentes. Par exemple, dans les activités à connotation recherche, comment passer d'une logique « forum » à celle d'une évaluation en bonne et due forme ? De même dans une activité de type « observatoire et veille », comment marier une organisation éditoriale hiérarchique pour une partie des informations avec une approche encyclopédique pour une autre ?

En fait, nous rencontrons des difficultés tout à fait classiques de celles qui émergent sur le terrain. Nous avons également le projet de travailler dans un cadre fédératif, et donc constituer un réseau d'inventaires de contributions sur le thème de l'appropriation.

5 Conclusion

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.

Les acteurs publics de la recherche sont donc 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.

Des progrès significatifs ont été accomplis dans les mouvements du libre accès à l'information scientifique et les revendications individuelles ont fait place à des engagements institutionnels. Les établissements scientifiques créent de nouveaux services d'archivage ou d'édition électronique. Sur le terrain, il faut aller beaucoup plus loin pour amener les acteurs de la recherche à entreprendre une véritable révolution technologique. L'initiative ARTIST répond à ce type d'objectifs en matière de sensibilisation ou de formation.

Les acteurs de l'IST, chercheurs ou praticiens, doivent-ils craindre cette nouvelle complexité ? Les historiens ont travaillé sur les changements de paradigme issus de l'émergence de l'imprimerie. Pour entrevoir les bouleversements prévisibles dus au document numérique ou à l'Internet, il suffit de parcourir la littérature abondante qui lie à l'invention de Gutenberg le développement de la Réforme [9][10] ou l'apparition de la législation sur la propriété intellectuelle [22]. Du côté de la science, les chercheurs, pour bénéficier des avantages de l'imprimerie, ont été obligés d'apprendre à lire et à écrire, ils ont dû passer outre le scepticisme de Socrate qui reprochait à l'écriture de véhiculer une illusion³¹. Peut-être doivent-ils maintenant apprendre à lire et à écrire en numérique...

³¹ On trouvera un extrait du PHEDRE de Platon avec cette citation sur le site Theuth : <http://theuth.univ-remes.fr/theuth-liste.html> (consulté le 25/2/2005).

6 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] CHANIER T., *Archives ouvertes et édition scientifique*, L'Harmattan, Paris 2005.
- [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 publications scientifiques, comment contrer le déclin du français*, colloque RIFRAM, Paris 2003
- [7] DUCLOY J., *Plaidoyer pour un réseau d'inventaires des résultats de la recherche*, VSST, Toulouse 2004
- [8] DUSOULIER N. et DUCLOY J., *Processing of data and exchange of records in a scientific and technical information center .Formats : what for ?* Unimarc/CCF Workshop Florence (IFLA/UNESCO), 05-07 Juin 1991.
- [9] EISENSTEIN, E., *The Printing Press as an Agent of Change*, Cambridge University Press, 1979
- [10] EISENSTEIN E., *The Printing Revolution in Early Modern Europe* (Cambridge, U.K.: Cambridge University Press, 1983)
- [11] GARFIELD E., *Citation analysis as a tool in journal evaluation*, Science 178:471-79, 1972.
- [12] GUEDON J.-C., *Les journaux électroniques - L'importance de la mobilisation contre les pratiques de certains éditeurs commerciaux*, Rencontres de Biblio-fr, 1998
- [13] GUEDON J.-C., *Numériser les revues savantes : d'un commerce à l'autre*. La Recherche, 335 (Oct 2000), pp 78-85.
- [14] GUEDON J.-C., *Open Access Archives: from scientific plutocracy to the republic of science*, IFLA journal. [IFLA j..] 2003 , vol. 29 , no 2
- [15] HAMMOND T., HANNAY T., LUND B, *The role of RSS in Science Publishing : Syndication and Annotation on the Web*. D-Lib Journal Magazine December 2004
- [16] 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)
- [17] 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>
- [18] Max Planck Gesellschaft, *Berlin Declaration* Berlin 2003 <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html> (vérifiée le 08/09/2004)
- [19] ODEH S., *Le positionnement des intermédiaires en économie de l'information numérique : le cas de l'industrie bibliographie*, Thèse Lyon 2004.
- [20] 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
- [21] Wellcome Trust, *An Economic Analysis of Scientific Research Publishing*, Londres http://www.wellcome.ac.uk/doc_WTD003181.html (vérifié le 22/09/04). Traduit en français sur <http://www.inist.fr/wellcome/>
- [22] 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
- [23] WAAST R., *La science en Afrique à l'aube du XXIème siècle : Les coopérations scientifiques en Afrique*, 2001, 27 p.