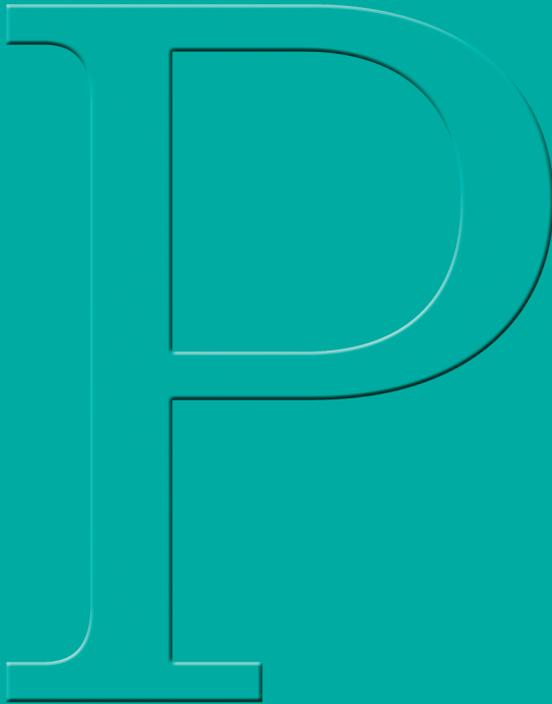


colloque

PRATIQUES DOCUMENTAIRES
NUMÉRIQUES À L'UNIVERSITÉ

sous la direction de Ghislaine Chartron, Benoît Epron
et Annaïg Mahé



PRESSES DE L'ENSIB

PAPIERS

+++++

colloque

PRATIQUES DOCUMENTAIRES
NUMÉRIQUES À L'UNIVERSITÉ

+++++

PAPIERS

**SOUS LA DIRECTION DE
THIERRY ERMAKOFF**

++++
La collection Papiers a pour ambition d'explorer de nouveaux champs de recherche autour des sciences de l'information et des bibliothèques. Elle donne aux auteurs l'occasion de produire une réflexion nouvelle, originale, et propose de nouvelles lectures des domaines d'expertise de l'enssib.

++++

PRESSES DE L'enssib

École nationale supérieure des sciences
de l'information et des bibliothèques
17-21 boulevard du 11 novembre 1918
69623 Villeurbanne Cedex
Tél. 04 72 44 43 43 – Fax 04 72 44 43 44

Pour suivre l'actualité de
nos publications :

< <http://www.enssib.fr/presses> >.

Pour obtenir la version imprimée (34 €) : commande auprès des libraires ou sur la librairie en ligne *Le comptoir des presses d'universités (LCDPU)*

< <http://www.lcdpu.fr/> >.

Supports : broché et PDF
ISBN papier : 978-2-910227-88-3
ISBN numérique : 979-10-91281-07-2
Nombre de pages : 226
Format : 15 x 23 cm
Année de parution : 2012
Prix de la version imprimée : 34 €

colloque

PRATIQUES DOCUMENTAIRES
NUMÉRIQUES À L'UNIVERSITÉ

Anita Beldiman-Moore, Emma Bester, Caroline Bruley,
Marin Dacos, Ilham Derfoufi, Benoît Epron, Jenny Fry,
Isabel Galina, Anne Gentil-Beccot, Paul Huntington,
Annaïg Mahé, Delphine Merrien, Nikoleta Pappa,
Melissa Terras, Claire Warwick et Anna Wojciechowska

Ont coordonné cet ouvrage :

Ghislaine Chartron

Benoît Epron

Annaïg Mahé

Cet ouvrage est la restitution de la journée d'étude Intd - enssib - Urfist, « Diversité des pratiques documentaires numériques dans les champs scientifiques », organisée à l'enssib le 2 juillet 2009.

Pratiques documentaires numériques à l'université [Ressource électronique] / sous la direction de Ghislaine Chartron, Benoît Epron et Annaïg Mahé. - Villeurbanne : Presses de l'enssib, cop. 2012. - Données textuelles (1 fichier PDF). - L'impression du document génère 227 pages. - (Collection Papiers). - Disponible gratuitement sur : <http://www.enssib.fr/presses/catalogue/> ISBN 979-10-91281-07-2 (édition électronique)

Rameau :

Culture de l'information

Recherche de l'information

Chercheurs

Bibliothèques universitaires - - Ressources Internet

Information électronique - - Utilisation

Édition électronique

Dewey : 027.7

© enssib, 2012.

+++++

SOMMAIRE

+++++

Introduction

par Annaïg Mahé et Benoît Epron 7

Les pratiques informationnelles des chercheurs dans l'enseignement supérieur et la recherche : regards sur la décennie 2000-2010

par Annaïg Mahé 11

Lumière sur le Web universitaire : l'influence de la culture disciplinaire sur les représentations en ligne

par Jenny Fry 43

Pratiques informationnelles en physique des particules

par Anne Gentil-Beccot 77

Pratiques documentaires et pratiques d'auto- archivage des mathématiciens et informaticiens en France

par Anna Wojciechowska 89

Pratiques numériques de chercheurs : reflet de la discipline, l'exemple des sciences de l'éducation

par Ilham Derfoufi 121

Bibliothèques, ressources d'information et
utilisateurs de ressources électroniques dans
les sciences humaines

*par Claire Warwick, Melissa Terras, Isabel Galina,
Paul Huntington et Nikoleta Pappa* 139

Que savons-nous de l'identité, des comportements
et des attentes des lecteurs de Revues.org en 2008
et 2009 ?

par Marin Dacos et Emma Bester 175

Les ressources numériques entre notoriété et
usages : enquêtes sur les pratiques et attentes
des lecteurs de la bibliothèque de Sciences Po
(2005-2008)

par Anita Beldiman-Moore 193

Un panorama des pratiques et usages des e-books
au sein des établissements Couperin

par Caroline Bruley et Delphine Merrien 209

Liste des auteurs

..... 225

par Annaïg Mahé et Benoît Epron

++++
INTRODUCTION
++++

+++++

INTRODUCTION

+++++

La part de la production scientifique accessible en ligne ne cesse de croître et, si elle ne constitue pas (encore) la totalité de cette production, elle n'en est pas loin. Parallèlement, les pratiques de diffusion, d'échanges et de recherche d'information en ligne deviennent une part intégrante de l'activité de recherche. Du moins c'est le portrait général que l'on peut faire actuellement, car la réalité du terrain est loin d'être homogène. De nombreuses études ont fait état de l'importance du critère disciplinaire notamment, dans les variations de pratiques et d'usages de l'information en ligne. D'autres facteurs institutionnels ou culturels sont également à prendre en compte. À l'heure où se développent de nombreux contenus et services d'accès, il devient utile d'appréhender cette diversité afin de mieux comprendre les modes d'adaptation et d'appropriation de ces contenus en ligne selon les communautés d'usage, et de pouvoir au moins les accompagner, au mieux les anticiper, voire les déployer, grâce à l'identification *a minima* des « bonnes pratiques ».

En effet, à l'heure de la maturité croissante de l'offre documentaire numérique et de ses potentiels dans tous les champs scientifiques, que connaissons-nous vraiment de la réalité des pratiques documentaires à l'université et dans les organismes de recherche ? Cette manifestation avait pour objectif de contribuer à répondre à cette question.

UNE OFFRE DOCUMENTAIRE EN ÉVOLUTION

+++++

Au cours de la dernière décennie, les pratiques documentaires des chercheurs ont connu des évolutions significatives, corollaires au développement d'une offre de ressources numériques toujours plus importante. Au début des années 2000, le bouleversement de l'offre éditoriale des revues scientifiques, notamment dans le domaine sciences, techniques et médecine (STM), conduit à réinterroger les pratiques documentaires des chercheurs.

En 2002, alors que les catalogues de revues des grands éditeurs des domaines scientifiques, techniques et médicaux se mettent en ligne et que les *Big Deals* deviennent une réalité de l'économie de l'édition scientifique,

un ouvrage vient poser quelques jalons sur ces évolutions récentes de la documentation numérique scientifique¹. De façon complémentaire à l'analyse des évolutions du marché de l'information scientifique et technique, plusieurs terrains issus de différentes disciplines (physique, astronomie, biologie moléculaire, économie-gestion, études littéraires, etc.) et de différents contextes (au sein de différents organismes de recherche, d'un laboratoire, d'une bibliothèque, etc.) sont ici investis au travers d'observations empiriques de différentes natures.

Si le support numérique devient partie intégrante du paysage informationnel des chercheurs, les spécificités des modes de communication selon les champs scientifiques ne disparaissent pas avec l'avènement de ce nouveau support, bien au contraire, et il paraît dès lors primordial de comprendre ces caractérisations particulières de chaque domaine afin de comprendre les enjeux et les limites de la transition numérique².

Aujourd'hui le contexte a encore évolué : l'offre est plus importante et s'étend progressivement à l'ensemble des disciplines, proposant des modalités d'accès et des modèles économiques divers. L'usage des revues numériques dans les champs scientifiques est installé et il accompagne le déploiement de l'usage de ressources numériques dans de nouvelles directions.

Le paysage éditorial numérique à l'université est d'autant plus bouleversé qu'il voit apparaître aux côtés des revues numériques l'affirmation de formes éditoriales inédites ainsi qu'une offre de livres numériques en croissance qu'il convient d'intégrer dans les collections constituées par les acteurs universitaires de la documentation.

Il semblerait judicieux d'observer dans quelle mesure l'expérience acquise dans le développement d'une offre de revues numériques pour l'université joue ici un rôle tant pour les producteurs de contenus et les éditeurs que pour les bibliothèques universitaires.

1. Chartron Ghislaine, *Les chercheurs et la documentation numérique : nouveaux services et usages*, Paris, Éditions du Cercle de la Librairie, 2002.

2. Chartron Ghislaine, « Éléments pour une approche comparée de la publication scientifique », in *Forum universitaire. La communication scientifique en quatre dimensions*, Colloque annuel du centre de coopération interuniversitaire franco-québécoise, ministère des Finances, de l'Économie et de la Recherche, Montréal, Archives nationales du Québec, 2003.

LA NÉCESSAIRE COMPRÉHENSION DES USAGES NUMÉRIQUES

+++++

Aujourd'hui, face à cette nouvelle offre, il devient crucial de prendre du recul sur les usages réels, et non plus seulement souhaités, déclarés ou rapportés.

Le questionnement s'écarte alors d'une vision prospective ou prescriptive pour s'attacher à une observation directe des pratiques et usages documentaires à l'université.

L'ensemble de travaux et d'analyses récents couvre une palette d'observations des pratiques illustrant cette diversité selon différents points de vue : méthodologies convoquées (quantitatives et qualitatives), différents supports (notamment revues et ouvrages), différents champs disciplinaires (physique des hautes énergies, mathématiques et informatique, sciences de l'éducation, sciences politiques). Cette journée visait également à favoriser les rencontres et futures collaborations afin d'accroître nos capacités de compréhension des pratiques informationnelles et documentaires dans les mondes académiques et universitaires.

Les textes ici rassemblés sont issus des interventions de la journée d'étude du 2 juillet 2009, en dehors de la synthèse sur les pratiques informationnelles spécifiquement rédigée à l'occasion de cet ouvrage.

par Annaïg Mahé

+++++

LES PRATIQUES
INFORMATIONNELLES
DES CHERCHEURS
DANS L'ENSEIGNEMENT
SUPÉRIEUR ET
LA RECHERCHE :
REGARDS SUR LA
DÉCENNIE 2000-2010

+++++

+++++

LES PRATIQUES INFORMATIONNELLES DES CHERCHEURS DANS L'ENSEIGNEMENT SUPÉRIEUR ET LA RECHERCHE : REGARDS SUR LA DÉCENNIE 2000-2010

+++++

Les recherches sur les pratiques informationnelles ont une longue tradition, se développant en parallèle à l'évolution même de la science et de la croissance de la production de connaissances depuis la seconde moitié du xx^e siècle. L'avènement du numérique n'a fait que relancer les interrogations et l'intérêt de multiplier les études pour suivre au plus près l'évolution permanente des pratiques. En conséquence, il existe une multitude de travaux, la plupart très localisés et de ce fait difficiles à généraliser et à conceptualiser [Rowlands, 2007 ; Tenopir *et al.*, 2009]. Dans un tel contexte, les travaux de synthèse ou de méta-analyse de ces résultats morcelés sont particulièrement bienvenus pour apporter une vision plus globale [Tenopir, 2003 ; Rowlands, 2007 ; Silipigni Connaway et Dickey, 2010].

PRÉCISIONS SUR LES SOURCES

+++++

Avant d'entrer dans le vif du sujet, il nous paraît nécessaire d'apporter quelques précisions sur les documents qui nous ont permis de réaliser cette synthèse. Parmi la foison d'analyses et d'enquêtes publiées sur le sujet, nous avons retenu, sur la décennie passée, celles qui nous paraissent les plus larges, en termes de populations ou de problématiques, ainsi que quelques travaux récents plus spécifiques à une population ou un type d'outils. Si une grande part de ces études sont majoritairement anglo-saxonnes, cela est dû à des contextes historiques et institutionnels particuliers : aux États-Unis, des études sur ces problématiques sont

régulièrement financées par des organismes à but non lucratif (OCLC¹, CLIR, Andrew Mellon Foundation) ; au Royaume-Uni, elles sont commanditées par des organismes gouvernementaux ayant spécifiquement en charge d'analyser et de promouvoir les évolutions des nouvelles technologies dans l'enseignement supérieur (AHRC, JISC, RIN¹), et elles sont réalisées par des experts ou des équipes universitaires spécialisées dans ces problématiques². À cela s'ajoutent des études financées par des associations d'éditeurs (ALPSP¹) ou des consortiums plus larges comme le récent programme européen PEER¹. Ces études citent elles-mêmes d'autres travaux internationaux (notamment en Finlande ou en Australie). Un certain nombre d'études sont également issues du monde des professionnels qui s'interrogent sur l'évolution des services qu'ils apportent à leurs communautés d'utilisateurs.

PRATIQUES INFORMATIONNELLES ET NOTION DE DISCIPLINE

+++++

La notion de discipline est un invariant des études de la communication scientifique et des pratiques informationnelles. Dans les années 1970 ont été posées les fondations d'une heuristique des pratiques informationnelles des chercheurs où la variable disciplinaire apparaît comme une des variables majeures de la diversité des pratiques de communication et des comportements informationnels des communautés scientifiques [Garvey, 1979 ; Merton, 1973].

Aujourd'hui, cette heuristique n'est toujours pas démentie, loin s'en faut : en effet, le constat général d'une intégration irréversible du support numérique dans les pratiques documentaires des chercheurs s'accompagne de cet autre constat, sans cesse confirmé : celui du maintien dans le monde numérique de l'existence des particularités disciplinaires [Kling et McKim, 2000 ; Tenopir, 2003 ; Rowlands, 2007 ; CIBER: University College London, 2008 ; Silipigni Connaway et Dickey, 2010 ; Harley

-
1. OCLC: Online Computer Library Center, < <http://www.oclc.org/uk/en/global/default.htm> >. CLIR: Council on Library and Information Resources, < <http://www.clir.org/> >. AHRC: Arts and Humanities Research Council, < <http://www.ahrc.ac.uk/Pages/default.aspx> >. JISC: Joint Information Systems Committee, < <http://www.jisc.ac.uk/> >. RIN: Research Information Network, < <http://www.rin.ac.uk/> >. ALPSP: Association of Learned and Professional Society Publishers, < http://alpsp.org/ngen_public/ >. PEER: Publishing and the Ecology of European Research, < <http://www.peerproject.eu/> > (consultés le 10 novembre 2011).
 2. Parmi ces équipes universitaires : les départements de sciences de l'information à l'université de Loughborough ou à l'University College London qui abrite le Centre for Digital Humanities ou le très prolifique CIBER (Center for Information Behaviour and the Evaluation of Research: < <http://ciber-research.eu/index.html> >, consulté le 10 novembre 2011).

et al., 2010]. Ces différences de comportements peuvent s'expliquer assez naturellement par le fait que, selon les disciplines et les communautés, les chercheurs utilisent différents types de ressources informationnelles et des canaux spécifiques de communication et de diffusion des résultats de recherche [Sparks, 2005 ; Meadows, 2006 ; Heimeriks *et al.*, 2008]. Au-delà de ce constat dans les pratiques des chercheurs, il apparaît aussi que la discipline est actuellement la forme dominante et la plus évidente de représentation des communautés scientifiques, qui se définissent à travers des associations, des revues et autres formes institutionnelles particulièrement visibles et « officielles », ne rendant pas toujours compte de la diversité des formes que peuvent prendre les communautés scientifiques.

Malgré cela, ou peut-être à cause de cela, bien souvent, la notion de communauté scientifique n'est pas interrogée dans le cadre des analyses des comportements informationnels, et de ce fait, elle reste généralement une unité d'analyse définie d'emblée, voire un implicite méthodologique. Il s'avère alors que les constats restent souvent assez généraux et que peu d'études de terrain ont exploré de manière systématique les différences disciplinaires et organisationnelles dans les comportements informationnels des chercheurs [Sparks, 2005 ; Rowlands, 2007]. Depuis quelques années, certains travaux tentent de proposer des cadres théoriques explicatifs et soulèvent dans le même temps la difficulté de définir l'unité d'analyse pertinente [Unsworth, 2000 ; Fry et Talja, 2005 ; CIBER: University College London, 2008 ; Velden et Lagoze, 2008]. Cela est d'autant plus problématique que l'évolution de la science a des conséquences directes sur les modes de fonctionnement des disciplines elles-mêmes [Rowlands, 2007 ; Brown et Swan, 2007] : augmentation des collaborations et des projets pluridisciplinaires³, naissance de nouvelles disciplines encore peu reconnues, et cela aussi bien dans les domaines des sciences physiques, sciences de la vie, médecine, sciences de l'ingénierie (STM), que dans les sciences humaines et sociales.

Des études spécifiquement centrées sur les frontières plus ou moins bien définies des disciplines risquent, par conséquent, de ne pas favoriser la prise en compte de ces évolutions. Cela vaut aussi pour les supports et outils étudiés. Aujourd'hui, une majorité d'études porte sur le développement des revues scientifiques en ligne : la revue est le support traditionnel

3. Une étude du Research Information Network (RIN) indique que 38 % des chercheurs interrogés déclarent participer à des projets interdisciplinaires [Research Information Network et RightsCom, 2006].

majeur de la communication scientifique et c'est celui qui a connu la plus forte expansion numérique ces dernières années. Or, cela ne permet de prendre en compte ni les variations d'usage entre les différents types de supports, ni la variété des ressources et services en développement, et au fur et à mesure que l'offre se développe, l'explication des différences disciplinaires par les variations de disponibilité des ressources n'est plus satisfaisante [Talja *et al.*, 2007]. Certains notent également l'intérêt de la variable institutionnelle [Nicholas *et al.*, 2009 ; Niu *et al.*, 2010]. L'idéal réside alors dans la possibilité de faire ressortir les spécificités de communautés particulières mais aussi les points communs et les pratiques que l'on peut « généraliser » jusqu'à un certain point, en attendant de pouvoir bénéficier de cadres théoriques suffisamment opérationnels [Sparks, 2005]. Pour aller dans ce sens, l'idéal, d'un point de vue méthodologique cette fois-ci, est de multiplier les approches, si possible dans un même cadre d'analyse : observations, entretiens, questionnaires, analyses de logs⁴ [Fry, 2006 ; Brown et Swan, 2007 ; Warwick, Terras, Huntington *et al.*, 2008 ; Tenopir *et al.*, 2009].

Une grande partie des études des comportements informationnels dans les domaines académiques et universitaires est logiquement ciblée sur les premiers domaines scientifiques directement concernés par les évolutions de l'offre numérique, et plus spécifiquement la physique, l'astrophysique, les mathématiques, les sciences biomédicales, ou les sciences de l'ingénieur. De plus en plus de travaux sont menés sur de nouveaux corpus : les sciences humaines et sociales (les sciences économiques, également pionnières du numérique, ayant déjà fait l'objet de nombreuses études), ainsi que les étudiants [Niu *et al.*, 2010], en parallèle avec la massification des cohortes dans l'enseignement supérieur. Les études sur les domaines interdisciplinaires se multiplient également, suivant en cela les développements récents de la science. Des études récentes proposent aussi quelques premiers retours sur de nouveaux supports/services en développement, tels que les e-books ou les outils du Web 2.0. Enfin, comme nous l'avons déjà précisé, une partie des études est menée par ou pour des professionnels des bibliothèques académiques et universitaires afin de mieux connaître les populations desservies et d'évaluer l'adéquation des services rendus à l'évolution de leurs pratiques et besoins. La synthèse

4. Selon la plupart des études basées sur cette méthode, l'approche de l'analyse des logs est encore récente et insuffisamment éprouvée pour permettre de tirer le meilleur parti des masses d'informations récoltées.

proposée suivra donc le cheminement suivant : après les constats généraux, s'appliquant très largement aux domaines scientifiques, techniques et médicaux, nous ferons un point sur le libre accès, sur les sciences humaines et sociales, sur les nouveaux supports (ouvrages numériques) et outils (Web 2.0), sur les limites et perspectives, et, enfin, sur la place et les rôles des bibliothèques et des professionnels de l'information.

DES CONSTATS GÉNÉRAUX

+++++

Une partie des études sur les pratiques informationnelles dans le monde académique et universitaire porte sur des populations de chercheurs ou d'usagers⁵ assez largement indifférenciées, au mieux décrites en grandes catégories, et souvent avec un biais vers les domaines STM. Ces études apportent un certain nombre de constats généraux qu'il est intéressant de rappeler, d'autant plus que cela donne aussi une vision des pratiques plus axée sur la prise en compte des caractéristiques communes et des fonctions fondamentales de l'activité scientifique que sur ses spécificités disciplinaires ou organisationnelles [Unsworth, 2000 ; Sparks, 2005].

Selon les domaines scientifiques, le support majeur de dissémination de l'information varie (de même que les volumes de production ou les taux de coécriture) : cela peut être l'article de revue pour les sciences biomédicales, le pre- ou postprint dans les sciences physiques et de l'ingénieur, et un ensemble plus large en sciences sociales [Swan et Brown, 2005]. Parmi cette diversité, la revue scientifique reste un support particulier très présent, à des degrés divers, dans toutes les disciplines. C'est elle, en effet, qui porte une des fonctions de base parmi les plus essentielles de la communication scientifique : la validation par les pairs des résultats de la recherche et de l'information diffusée. Pour de nombreux domaines, les indicateurs de production liés aux articles et à la revue sont des jalons essentiels de la recherche, et elle est un support de diffusion privilégié,

5. Certaines études concernent des chercheurs et des étudiants. Dans cette synthèse, le terme générique de « chercheurs » englobe aussi les étudiants et leurs pratiques de recherche d'information dans les cadres académiques et universitaires. Des différences liées à l'âge ou à l'expertise sont mentionnées, mais le contexte scientifique reste largement partagé dans les pratiques informationnelles.

mais aussi de reconnaissance sociale, mesurée au prestige de la revue⁶, même lorsque ce n'est pas le support majeur de la communauté⁷.

Pour ces raisons, c'est un des premiers supports à avoir été le plus massivement transposé dans le domaine numérique, d'où le nombre d'études le concernant. Selon de récentes estimations, 96 % des titres en sciences, techniques et médecine, et plus de 86 % dans les domaines des sciences humaines et sociales sont actuellement disponibles en ligne [Cox et Cox, 2008]. En France, plus de 600 000 titres étaient disponibles dans les bibliothèques universitaires en 2008 [Colcanap, 2009]. Cela a facilité une adoption massive et large, et les premiers résultats de grande ampleur d'analyse des *logs* de consultation de plateformes de revues en ligne font état d'une consommation massive sur 90 % des collections étudiées [Nicholas *et al.*, 2009 ; Ware et Mabe, 2009 ; Silipigni Connaway et Dickey, 2010].

Si certaines études rapportent clairement la satisfaction des chercheurs quant à l'offre disponible, ainsi que le fait qu'ils ne semblent pas avoir de difficulté majeure à accéder aux ressources en ligne [Swan et Brown, 2005 ; Ware et Mabe, 2009 ; McClanahan *et al.*, 2010], d'autres travaux indiquent que l'accès aux articles peut être problématique (pas d'abonnement ou d'accès à la base de données ou aux articles de conférences notamment) pour plus d'un tiers des chercheurs interrogés : le « dernier kilomètre » jusqu'au texte intégral se révélant bien souvent poser problème [Research Information Network et RightsCom, 2006 ; JISC, 2010]. Par ailleurs, les chercheurs semblent reproduire des schémas traditionnels dans l'utilisation d'une plateforme de revues et se concentrer sur un ensemble de titres cœur⁸.

Les conséquences de l'augmentation du volume d'informations disponibles sur les comportements de lecture des chercheurs sont directes : si l'on note une croissance importante du nombre d'articles lus dans la

6. Une enquête a noté une corrélation statistique entre le nombre de téléchargements d'articles sur une plateforme et le rang des institutions concernées [Nicholas *et al.*, 2009].

7. Une étude britannique rapporte que la mise en place de la procédure d'évaluation nationale de la recherche (Research Assessment Exercise) pourrait amener l'article de revue à devenir la forme dominante de dissémination des résultats de la recherche, entraînant des convergences entre les disciplines, notamment dans l'usage des ressources et des modes de publication [Swan et Brown, 2005].

8. 5 % des revues comptant pour un tiers, voire la moitié, des consultations totales [Nicholas *et al.*, 2009].

dernière décennie⁹, cela va de pair avec une réduction de la quantité d'attention par unité d'information et la lecture d'une portion toujours plus restreinte de la production annuelle, amenant les chercheurs à développer une « lecture stratégique » [King *et al.*, 2003 ; Tenopir, 2003 ; Rowlands, 2007 ; CIBER: University College London, 2008 ; Ware et Mabe, 2009 ; Niu *et al.*, 2010].

À différents stades de leur activité, les chercheurs sont amenés à explorer des domaines encore inconnus et commencent leurs recherches d'informations sans en connaître précisément l'objet [Research Information Network et RightsCom, 2006]. Il devient alors nécessaire d'élargir au maximum les possibilités de recherche afin de favoriser le processus de sérendipité, le plus favorable à une recherche d'information plus ou moins bien définie.

C'est sans doute la raison pour laquelle des études ont pu faire ce constat apparemment contradictoire que l'éventail des outils de recherche académiques utilisés peut être aussi bien large et généraliste que limité et ciblé sur des ressources spécifiques [Silipigni Connaway et Dickey, 2010 ; Schonfeld et Housewright, 2010]. Cela aboutit sans doute aussi à ce paradoxe relevé par des professionnels de l'information qui notent que les chercheurs demandent l'accès au texte intégral mais ne vont souvent pas plus loin que le résumé et se contentent bien souvent d'une seule base de données (en l'occurrence le Web of Science) malgré les efforts réalisés pour étendre la couverture des ressources mises à disposition [Brown et Swan, 2007]. Du point de vue du chercheur d'information, ce comportement paraît assez logique : la consultation est centrée sur une seule base de données, à la fois suffisamment large dans ses thématiques et suffisamment définie dans ses frontières en termes de types de supports, pour augmenter ses chances de trouver des informations pertinentes en évitant le bruit généré par des outils plus généralistes comme les moteurs de recherche, et sur des stratégies de sélection rapide de l'information par la consultation des résumés ou autres éléments de repérage du contenu, le téléchargement étant réservé aux articles jugés suffisamment pertinents pour une lecture plus attentive. Ce type de consultation est décrit comme du « super-feuilleteur » (*power browse*), et de cette manière, le chercheur

9. Selon les études de Carol Tenopir, les chercheurs lisent 280 articles par an (contre 216 en 2000-2002, 150 en 1977) [Tenopir *et al.*, 2009]. Une autre étude indique des chiffres moindres : plus de 90 % des répondants disent lire plus de 10 articles validés par les pairs par an en moyenne. Un peu plus d'un tiers disent lire entre 51 et 100 articles par an en moyenne et moins d'un quart lisent entre 101 et 250 articles par an [Fry *et al.*, 2009].

satisfait à la fois les critères de quantité et de qualité, mais aussi de commodité et de rapidité en privilégiant les ressources avec accès au texte intégral immédiat [OCLC, 2005 ; Silipigni Connaway et Dickey, 2010].

Lorsque l'information est recherchée dans le contexte d'une activité en cours (telle que la mise en place d'un projet de recherche ou l'écriture d'un article), il est en effet crucial de pouvoir accéder au texte intégral du document, et lorsque ce n'est pas le cas, bien souvent les chercheurs abandonnent s'il n'est pas disponible en ligne immédiatement : le temps de faire une demande « officielle » et d'obtenir l'article, il sera sans doute déjà trop tard. Bien souvent, et à des degrés différents selon les domaines ou les activités concernés, les chercheurs fonctionnent en *just in time* et l'accès aux informations se fait dans ce cadre, générant de la frustration et un déficit d'information potentiel à la moindre barrière technique ou économique [Brown et Swan, 2007 ; Niu *et al.*, 2010].

Un avantage non négligeable des collections en ligne est qu'elles sont accessibles « sur place », et donc encore plus dans le flux même de l'activité que lorsque le chercheur devait se déplacer physiquement jusqu'à la bibliothèque. Qu'il soit choisi ou subi, ce déplacement est généralement vécu comme une rupture dans les activités [Mahé, 2002]. Il est donc très compréhensible que la grande majorité des chercheurs aient une préférence pour les ressources directement accessibles en ligne, même lorsque la bibliothèque est physiquement proche, et qu'ils accèdent généralement aux ressources de leur bureau : selon une étude, c'est en effet le cas pour 90 % des chercheurs en sciences physiques et de la vie, 76 % en sciences sociales et 58 % en sciences humaines (22 % des accès dans ces domaines grands « consommateurs » de bibliothèques se font justement dans ce cadre, qu'une part des chercheurs considère d'ailleurs comme leur « laboratoire ») [Brown et Swan, 2007 ; Tenopir *et al.*, 2009 ; Niu *et al.*, 2010]. Le domicile est le deuxième lieu d'accès en importance, et il est notable de constater que cette étude relève également que 15 % des accès se font en déplacement, c'est-à-dire dans la même proportion que les accès effectués à partir de la bibliothèque. Dans le même ordre d'idée, il est intéressant de noter que près d'un quart des consultations ont lieu en dehors des heures de bureau et 15 % le week-end [Nicholas *et al.*, 2009]. Par ailleurs, les jeunes chercheurs seraient aussi plus nombreux que leurs anciens à obtenir leurs documents du Web plutôt que de la bibliothèque ou des abonnements personnels, ainsi que de plus en plus enclins à effectuer leur lecture directement à l'écran plutôt que sur un support imprimé [Tenopir *et al.*, 2009].

La pratique numérique de la revue est donc de plus en plus ancrée dans les habitudes, et, de fait, l'accès direct à tous types de ressources est devenu une nécessité, ainsi que l'accès au texte intégral, avec des délais d'attente se comptant parfois plutôt en termes de secondes que de jours dans certains domaines [Brown et Swan, 2007 ; McClanahan *et al.*, 2010 ; Silipigni Connaway et Dickey, 2010]. De ce fait, les chercheurs sont particulièrement frustrés devant la dispersion des sources, qui les oblige à multiplier les procédures d'identification et d'interrogation, et sont plutôt favorables aux possibilités de méta-recherche d'information [Niu *et al.*, 2010]. Il est alors tout à fait logique que l'utilisation de Google se soit si facilement et massivement répandue comme point de départ pour une grande majorité des recherches d'informations : une des priorités pour les chercheurs étant de ne pas manquer d'information importante, l'intérêt se porte en premier lieu sur les moteurs de recherche généralistes et les larges corpus de titres et d'articles, d'autant plus que, comme déjà mentionné, si l'article de revue est une ressource clé pour la grande majorité des chercheurs, ils sont aussi nombreux à déclarer utiliser d'autres supports et ressources d'information, tels que les ouvrages ou des sources plus informelles *via* les réseaux personnels [Research Information Network et RightsCom, 2006 ; Silipigni Connaway et Dickey, 2010].

Un certain nombre de travaux se sont penchés plus spécifiquement sur les modes de recherche d'information utilisés par les chercheurs, tant sur le support papier que numérique, indiquant parfois une simple transposition, mais aussi parfois des avantages ou, au contraire, des limites. Concernant la consultation des articles numériques, trois comportements types ont été notés : la recherche de références ou le suivi de citations¹⁰, le feuilletage de revues cœur, et la recherche thématique [Inger et Gardner, 2008]. Pour la recherche de références, les bases de données bibliographiques spécialisées et les moteurs de recherche sont les points d'entrée les plus importants, et sans doute du fait de l'importance des moteurs et des interfaces de recherche dans l'accès aux informations du Web et aux collections d'articles, l'accès à ces dernières se fait de manière croissante par le biais de la recherche et l'utilisation de mots clés, plutôt que par le feuilletage [Ware et Mabe, 2009 ; Silipigni Connaway et Dickey, 2010]. De fait, une grande part des visites aux plateformes de revues mises à disposition proviennent ainsi directement des moteurs de

10. Le suivi de citations désigne le repérage des références pertinentes dans la bibliographie du document initial, aussi défini comme « recherche par chaînage ».

recherche généralistes, et en particulier de Google, et ne durent généralement que quelques minutes, sur un petit nombre de pages, ignorant les fonctionnalités de personnalisation ou de recherche avancée [Inger et Gardner, 2008 ; Nicholas *et al.*, 2009 ; Silipigni Connaway et Dickey, 2010]. Un certain nombre de visites se font aussi directement par la page de la revue, souvent à la suite d'une alerte [Inger et Gardner, 2008].

Qu'en est-il vraiment de la place des moteurs de recherche en général, et de Google en particulier, « cet entonnoir devenu fenêtre sur le monde numérique »¹¹, dans les pratiques de recherche d'information ? Plusieurs études relèvent une évolution notable : si, en 2005, on ne pouvait pas encore parler de « googlisation totale », les chercheurs se basant sur une large gamme d'outils, il semblerait que cela soit différent en 2007, et Google serait devenu la porte universelle aux informations du Web, amenant les chercheurs à se concentrer sur un nombre d'outils limités familiers et jugés satisfaisants [Van Orsdel et Born, 2006 ; Brown et Swan, 2007 ; CIBER: University College London, 2008]. Si la prédominance de Google comme porte d'entrée est avérée et chiffrée, il convient peut-être de relativiser son importance dans les pratiques informationnelles des chercheurs.

Comme nous l'avons mentionné précédemment, les chercheurs ont une préférence pour l'accès le plus direct au contenu le plus large possible. Dans un premier temps, que la recherche d'information soit ciblée ou indéfinie, Google et les moteurs de recherche généralistes donnent suffisamment de réponses aux questions qui leur sont posées, et une étude récente indique clairement la part de plus en plus importante que prend l'utilisation de Google Scholar dans la recherche d'information scientifique, détrônant presque celle des bases de données spécialisées [Herrera, 2011]. En renforçant la satisfaction produite par les résultats de la « recherche floue » [Gallezot *et al.*, 2008], l'utilisation de Google renforce naturellement le besoin d'autonomie des chercheurs et leur « proximité relationnelle » avec les sources d'information [Simonnot et Gallezot, 2009]. En déduire que Google devient le principal outil des chercheurs serait cependant excessif car il semblerait plutôt que la logique de la longue traîne puisse s'appliquer aux ressources utilisées par les chercheurs, un petit nombre de ressources étant connues de beaucoup, et un grand nombre de ressources mentionnées par quelques individus seulement [Research Information Network et RightsCom, 2006]. En fonction des objectifs poursuivis, les

11. Le Crosnier Hervé, préface, in Simonnot et Gallezot, 2009.

chercheurs utilisent une grande variété d'outils, et pas uniquement les moteurs généralistes : bases de données bibliographiques et moteurs de recherche spécialisés, portails et catalogues de bibliothèques, plateformes thématiques, réseaux disciplinaires et personnels [Swan et Brown, 2005 ; Research Information Network et RightsCom, 2006 ; CIBER: University College London, 2008]. Selon les domaines, les préférences ne sont pas les mêmes : portails thématiques pour les sciences médicales, bases de données bibliographiques en sciences de la vie, ressources des bibliothèques et moteurs de recherche pour les sciences humaines et sociales [Fry *et al.*, 2009]. Et il apparaît clairement que les chercheurs en sciences de la vie sont comme des poissons dans l'eau numérique des revues en ligne : la littérature est particulièrement abondante dans ces disciplines et ils comptent pour la moitié de la consommation des plateformes de revues analysées ; et il en va de même pour les bases de données spécialisées [Inger et Gardner, 2008 ; Nicholas *et al.*, 2009].

Dans un deuxième temps, la pratique de Google et du monde documentaire numérique évolue. Au fur et à mesure que la masse d'informations disponibles en ligne se développe, et que les chercheurs gagnent en expertise des outils, grandit le besoin d'outils de recherche et d'accès à la littérature scientifique plus sophistiqués. Une étude récente indique que les deux interfaces de recherche préférées mentionnées par les chercheurs de manière équivalente sont celle de la bibliothèque et de Google, reflétant ainsi la conscience que les chercheurs ont des limites de Google mais peut-être aussi l'évolution des sites de bibliothèques vers une ergonomie « *google-like* » [Niu *et al.*, 2010]. Les usagers sont conscients de la différence entre la littérature de recherche formelle et le contenu Internet de base [Silipigni Connaway et Dickey, 2010], et au-delà des outils permettant l'accessibilité aux ressources, ceux qui assurent la garantie de la validité scientifique du contenu correspondent à une nécessité fondamentale. En ce sens, la validation effectuée par les revues reste un outil majeur qui n'a pas seulement un intérêt pour le scientifique en tant qu'auteur, mais aussi en tant que lecteur, d'autant plus lorsque l'expertise du domaine est moindre (lors du démarrage d'un nouveau projet ou dans le cadre d'une recherche pluridisciplinaire, par exemple), d'où le maintien de l'importance de ce support dans les pratiques [Tenopir *et al.*, 2009 ; Fry *et al.*, 2009].

CONSTATS SPÉCIFIQUES AU LIBRE ACCÈS

+++++
Certaines des études mentionnées dans cette synthèse portent spécifiquement sur le développement du mouvement du libre accès à l'information scientifique, les autres englobant éventuellement cette thématique dans l'analyse des pratiques informationnelles. De manière générale, il est clair que, même si ces problématiques sont plus largement maîtrisées par les professionnels de l'information, les chercheurs commencent à en avoir une bonne connaissance, notamment grâce aux nombreux débats en cours, mais de manière moindre dans les sciences humaines et sociales [Brown *et al.*, 2006 ; Brown et Swan, 2007 ; Fry *et al.*, 2009]. Selon les domaines, et sans doute aussi selon l'évolution de la prise en main de ces questions par les institutions, la connaissance respective des archives thématiques et institutionnelles évolue. Les premiers résultats du projet PEER indiquent que s'il existe une préférence générale pour les archives thématiques, les chercheurs manifestent aussi un intérêt pour les archives institutionnelles, les deux étant généralement pensées en lien, et ceci malgré les tensions existantes entre les politiques institutionnelles et les pratiques disciplinaires [Fry *et al.*, 2009].

On y retrouve logiquement de grandes variations disciplinaires, les domaines les plus propices étant la physique, les mathématiques et l'informatique¹², et les avantages du libre accès sont clairement différents selon les disciplines : accès libre aux articles en sciences de la vie, le facteur de rapidité de diffusion comptant moins pour les sciences médicales, tandis que les sciences humaines et sociales y gagneraient en augmentation des citations [Brown et Swan, 2007 ; Fry *et al.*, 2009]. Les taux de dépôts peuvent aussi varier considérablement, une étude rapportant que la moitié des physiciens déposent de manière régulière contre 18 % des chercheurs en biologie et sciences médicales, le taux le plus bas étant dans les sciences humaines, et les chercheurs sont nombreux à ne pas connaître les éléments juridiques concernant leurs publications [Swan et Brown, 2005]. L'étude de PEER rapporte que les sites de dépôts les plus cités sont arXiv, PubMed Central, Citeseer, Hal et RePEc [Fry *et al.*, 2009]. Par ailleurs, une très forte majorité de chercheurs considèrent qu'il est important de savoir quelle est la version de l'article qu'ils ont trouvé, cette notion du statut du document et de la validation par les pairs étant parti-

12. Voir l'article d'Anna Wojciechowska : « Pratiques documentaires et pratiques d'auto-archivage des mathématiciens et informaticiens en France », p. 89 dans cet ouvrage.

culièrement forte pour les sciences humaines et sociales. Globalement, si la connaissance des problématiques du libre accès se répand, son implémentation dans les pratiques des chercheurs reste encore insuffisamment développée dans de nombreuses disciplines, les notions mêmes de libre accès n'étant pas toujours clairement définies, souvent encore confondues avec les revues en accès « libre » (c'est-à-dire payé en amont) via les portails des bibliothèques ou des éditeurs [Martineau et Clénet, 2006 ; Fry *et al.*, 2009].

CONSTATS SPÉCIFIQUES AUX SCIENCES HUMAINES ET SOCIALES

Parallèlement aux enquêtes et analyses générales qui illustrent plus largement les domaines scientifiques, techniques et médicaux, les travaux spécifiquement dédiés aux sciences humaines et sociales se sont multipliés depuis quelques années suivant la maturation plus récente de l'offre et des usages. Les études britanniques font généralement la différence entre les sciences sociales d'un côté et les arts et sciences humaines de l'autre [Warwick, Terras, Galina *et al.*, 2008] et ont plus souvent porté sur cette deuxième catégorie. Une synthèse détaillée en français en a récemment été faite [Derfoufi, 2009]. Les analyses se font aussi souvent sur des catégories larges, et il existe encore peu d'études complètes spécifiques à des disciplines (en dehors des sciences économiques déjà mentionnées, de la linguistique ou des sciences de l'éducation, par exemple).

Dès le début des années 2000, des constats ont été faits sur l'adoption large des technologies numériques par les sciences humaines et sociales et les potentialités pour la recherche apparaissent clairement et de manière bien plus forte qu'en sciences, techniques et médecine où la pratique dominante est particulièrement concentrée sur un accès plus traditionnel au contenu de l'article scientifique [Brockman *et al.*, 2001 ; Unsworth, 2003]. Les sciences humaines et sociales ont, en effet, des pratiques liées à la documentation scientifique très spécifiques et, de ce fait, des attentes particulières du support numérique dont la disponibilité croissante est encore récente [Brown *et al.*, 2006].

Une étude a comparé les grandes lignes des spécificités qui différencient les sciences sociales des sciences humaines [Bass *et al.*, 2005]. De manière générale, les chercheurs en sciences sociales développent plus de collaborations et utilisent les revues comme sources principales. Ayant besoin d'informations au-delà des frontières de leur discipline, les sources validées sont en effet primordiales. En sciences humaines, les chercheurs

travaillent de manière plus individuelle et utilisent une grande gamme de documents, parmi lesquels les ouvrages en particulier, l'objectif étant d'apporter une interprétation personnelle de leur objet d'étude. Une autre enquête indique que les chercheurs de sciences humaines travaillent plus souvent seuls (à 62 % contre 51 % pour les économistes et 48 % en sciences sociales) [Talja *et al.*, 2007]. Cependant, si la production scientifique est plus individualisée en sciences humaines et sociales que dans les autres disciplines scientifiques, l'activité scientifique globale n'en est pas pour autant une activité complètement isolée. Les chercheurs ont une activité d'autant plus collective qu'elle est informelle, et travaillent avec des collègues indépendamment de l'institution, du pays ou de la discipline, même si la recherche hors des frontières disciplinaires établies reste difficile [Unsworth, 2003 ; Lougee, 2006]. Pour certains, cependant, cette forte proportion de chercheurs travaillant seuls et sans financement de projets peut expliquer le peu de connaissance qu'ils ont des ressources disponibles, les informations circulant en premier lieu par les collègues [Huxley *et al.*, 2006].

À la différence des sciences sociales, pour lesquelles le terrain est leur laboratoire, en sciences humaines, c'est souvent la bibliothèque qui est l'équivalent du laboratoire puisque c'est là que les chercheurs trouvent le matériel nécessaire à la recherche [Lougee, 2006 ; Brown et Swan, 2007]. L'importance de la bibliothèque et des ressources primaires est en effet une constante de ces domaines [Buchanan *et al.*, 2005 ; Huxley *et al.*, 2006 ; Brown et Swan, 2007 ; Warwick, Terras, Galina *et al.*, 2008]. Les activités fondamentales sont la lecture, l'écriture et la relecture, et les chercheurs ont besoin d'outils de manipulation des documents (pour l'analyse, les annotations) [Derfoufi, 2009]. Si les chercheurs en sciences humaines lisent en moyenne moins d'articles de revues et des articles plus anciens, ils lisent plus d'ouvrages, de matériaux primaires et de manuscrits [Tenopir *et al.*, 2009]. Pour trouver les documents nécessaires, les chercheurs consacrent une grande part de leur temps aux recherches bibliographiques (dans sa thèse, Ilham Derfoufi rapporte les résultats d'une enquête du CNRS qui indiquent que les chercheurs consacrent plus de 65 % de leur temps à ce type d'activité).

Non seulement les chercheurs se déplacent fréquemment pour leurs recherches documentaires – véritable « travail d'enquête et d'archéologie » – et fréquentent plusieurs bibliothèques dans lesquelles ils apprécient le libre accès aux ouvrages, mais ils maintiennent aussi souvent des collections personnelles de documents, généralement à leur domicile et,

de ce fait, non accessibles aux autres chercheurs [Charra, 2006 ; Lougee, 2006]. Cependant, ces spécialistes de la recherche d'information sont aussi bien souvent des spécialistes de la gestion des « piles » de documents : ces collections personnelles sont, en effet, plus ou moins bien organisées, souvent sans véritable réflexion sur la pérennité, la conservation et l'accessibilité, généralement par manque de temps, d'argent, d'espace et d'expertise [Lougee, 2006]. Certaines collections peuvent ainsi être « perdues » pour la collectivité lorsqu'un chercheur arrive en fin de carrière [Charra, 2006].

Parallèlement à cette forte consommation des bibliothèques physiques, celle des bibliothèques numériques est aussi en augmentation [Charra, 2006] et, avec le développement croissant de l'offre numérique, les chercheurs prennent l'habitude d'une économie mixte des ressources [Brown *et al.*, 2006 ; Lougee, 2006 ; Warwick, Terras, Galina *et al.*, 2008]. La croissance de cette offre n'est pas uniquement due à la mise à disposition par les éditeurs de collections de revues ou d'ouvrages ; elle est aussi nourrie par les nombreuses initiatives générées par les chercheurs eux-mêmes, et bien plus structurées que la gestion de leurs propres collections, telles que la numérisation de données d'archives ou la collaboration avec des experts pour la mise en ligne et la conservation de ces collections, mais aussi la mise en ligne de données secondaires et le développement de bases de données permettant des fonctionnalités telles que l'annotation ou le commentaire [Unsworth, 2003 ; Lougee, 2006].

De fait, si l'émergence du numérique a été plus tardive dans ces domaines, les chercheurs n'en sont pas moins conscients que cela constitue un changement important dans leur recherche, ne serait-ce que par la vitesse et la facilité d'accès à une plus grande variété de ressources, la majorité d'entre eux considérant avoir un niveau intermédiaire de compétences dans l'utilisation de ces nouvelles ressources [Huxley *et al.*, 2006]. Comme dans les autres domaines, les chercheurs ont une préférence pour de larges collections de références plutôt que des ressources trop spécialisées. Par contre, la recherche par chaînage (suivi de citations) et le feuilletage sont les méthodes les plus largement utilisées¹³, et comme nous l'avons déjà mentionné précédemment, cela peut s'avérer d'autant plus problématique que l'accès aux documents dans les collections numériques se fait de plus en plus par le biais de mots clés, souvent mal adaptés aux

13. Une enquête sur des ressources finlandaises indique que près de 30 % des accès aux articles proviennent du suivi de citations [Talja *et al.*, 2007].

termes disciplinaires et à la recherche par concepts [Warwick, Terras, Galina *et al.*, 2008 ; Derfoufi, 2009 ; Niu *et al.*, 2010].

Les chercheurs en sciences humaines et sociales ont des attentes très fortes et très spécifiques face au support numérique, et l'inadéquation des technologies à ces spécificités de la recherche a amené très tôt des critiques, notamment sur l'absence d'uniformité dans les systèmes d'information (protocoles de recherche, formats de citations) ou de fonctionnalités telles que l'annotation [Brockman *et al.*, 2001]. Cette inadéquation, couplée à un volume de matériel numérique encore insuffisant (ouvrages, archives), explique en grande partie que les pratiques restent encore largement traditionnelles et limitées à des outils connus et familiers, contrairement aux autres domaines scientifiques où la forte utilisation des plateformes de revues reste largement traditionnelle plutôt par adéquation aux habitudes de consultation de la littérature scientifique, les fonctionnalités avancées proposées étant toujours très peu utilisées et plébiscitées [Warwick, Terras, Galina *et al.*, 2008 ; Derfoufi, 2009].

CONSTATS SPÉCIFIQUES AUX OUVRAGES NUMÉRIQUES ET OUTILS WEB 2.0 ÉMERGENTS

+++++
Depuis quelques années, la mise en ligne des collections s'étend aux ouvrages tandis qu'une nouvelle génération d'outils et de technologies émerge au service du Web et commence à se répandre dans l'univers de la recherche scientifique [Gallezot et Le Deuff, 2009], et l'analyse de l'intégration de ces nouveaux supports et formes de communication dans les communautés scientifiques en est à ses débuts.

Concernant les ouvrages numériques, les premiers résultats sont clairs et plutôt cohérents : les usagers, de plus en plus familiers des collections numériques, adoptent très rapidement et massivement les nouveaux supports et sont demandeurs d'une offre étendue à la littérature grise et aux collections spéciales [Brown et Swan, 2007 ; JISC, 2009]. La maturation des outils, et notamment des tablettes de lecture, permet aussi de développer le confort de la lecture à l'écran, pratique qui commence à se répandre [JISC, 2010]. Cette lecture, cependant, n'est pas une lecture linéaire, mais plutôt de référence et d'extraction d'informations ponctuelles [Springer, 2008 ; JISC, 2009]. La majorité des éditeurs scientifiques, dans tous les domaines, ont commencé à proposer des collections d'ouvrages en ligne, et différents modèles économiques cohabitent, du *print-on-demand*

à l'accès en ligne gratuit¹⁴. L'offre et les technologies associées doivent encore devenir matures et moins confuses à l'usage, mais les communautés de chercheurs et d'étudiants sont clairement très demandeuses de ce nouveau support, et les études à venir permettront de mieux comprendre les pratiques et besoins des usagers [JISC, 2009].

Concernant les outils émergents du Web 2.0, les résultats sont plus mitigés et, selon les études, des conclusions différentes sont tirées sur leur apport aux communautés scientifiques [Silipigni Connaway et Dickey, 2010]. De récentes études font état du fait que si une majorité de chercheurs déclare utiliser au moins occasionnellement un ou plusieurs outil(s) ou service(s) du Web 2.0, l'usage fréquent ou intensif demeure encore l'exception plutôt que la règle, et il est relevé qu'ils ne seraient que moins de 5 % à utiliser ces nouvelles formes de communication que sont les blogs, wikis et autres médias sociaux, certains allant jusqu'à juger que ces outils n'occasionnent qu'une perte de temps ou même qu'ils sont dangereux et inappropriés à l'activité scientifique [Niu *et al.*, 2010 ; Research Information Network, 2010]. Les différentes études s'accordent à dire que ce sont les canaux formels traditionnels qui dominent largement, tandis que ces nouvelles propositions ne font qu'émerger. De nombreux chercheurs, en effet, considèrent ces initiatives avec intérêt, mais, n'y voyant pas suffisamment de bénéfices, ne les intègrent pas pour autant dans leur quotidien, et s'il existe là encore de fortes variations disciplinaires, il ne semble pas pour autant que des changements importants soient envisagés à court ou moyen terme [Harley *et al.*, 2010 ; Niu *et al.*, 2010 ; Research Information Network, 2010]. En effet, le facteur générationnel semble peu pertinent et ce ne sont pas toujours les plus jeunes qui se révèlent les usagers les plus intensifs. Comme il est noté dans une étude, le partage des résultats de la recherche en amont de leur publication validée peut être problématique : la recherche en cours circule au sein des réseaux personnels et des collèges invisibles, mais la transposition numérique de ces réseaux relativement informels est encore à réaliser [Harley *et al.*, 2010]. Dans ce sens, les initiatives institutionnelles, garantes d'une certaine qualité scientifique des contenus diffusés *via* ces nouveaux canaux,

14. OAPEN Library est un projet européen qui propose, depuis septembre 2010, une collection d'ouvrages en sciences humaines en libre accès : < <http://www.oapen.org> >.

peuvent certainement favoriser leur appropriation par une majorité de chercheurs¹⁵.

LIMITES ET PERSPECTIVES

+++++

Un des constats qui se dégagent de ce panorama des pratiques informationnelles et documentaires des chercheurs est la prégnance dans l'activité de recherche de normes de communication formelles relativement conservatrices, centrées autour du processus de la validation par les pairs de la production scientifique, et dont la forme dominante actuelle est l'article publié dans une revue [Ware et Mabe, 2009 ; Harley *et al.*, 2010]. Concernant les outils émergents notamment, ce ne sont pas toujours les plus jeunes chercheurs les plus actifs, leur intérêt principal étant de suivre et de reproduire les normes en vigueur dans leur domaine, et non pas d'innover à ce niveau s'ils veulent s'intégrer et obtenir la reconnaissance de leur communauté [Research Information Network, 2010 ; Harley *et al.*, 2010]. Les chercheurs seniors, déjà bien reconnus, peuvent prendre plus de liberté ; de même il peut être parfois plus facile de prendre des risques dans les domaines récents ou nouveaux car leur développement peut être plus soutenu par l'institution, notamment en sciences humaines et sociales [Harley *et al.*, 2010].

Un deuxième constat peut être celui de techniques encore insuffisamment matures : les barrières à l'usage recensées sont encore souvent une faible usabilité, une complexité élevée des outils mis à disposition, ou au contraire une insuffisance de sophistication selon les attentes, et une absence d'interopérabilité des différents systèmes de recherche, empêchant ainsi les usagers de se concentrer sur le contenu [Brown *et al.*, 2006 ; Martineau et Clénet, 2006 ; Brown et Swan, 2007 ; Ware et Mabe, 2009]. Ces derniers semblent aussi de plus en plus conscients des limites de Google : pratique pour démarrer une recherche d'information ou pour accéder à du contenu déjà connu, les chercheurs reconnaissent que son

15. Des universités britanniques mettent à disposition de leurs usagers, étudiants et chercheurs, des plateformes de blogs institutionnels. En France, les propositions de plateformes scientifiques collectives se multiplient, particulièrement dans les sciences humaines et sociales : *hypotheses.org*, jeune et dynamique « plateforme de carnets de recherche en sciences humaines et sociales » du Centre pour l'édition électronique ouverte (Cléo), < <http://hypotheses.org/> >, ou le tout récent « réseau social des sciences de l'homme » *Risoma* de la Maison des sciences de l'homme, < <http://www.risoma.org/> >. Les chercheurs sont aussi de plus en plus nombreux à intégrer des sites de réseaux professionnels ou scientifiques tels que LinkedIn, Viadeo ou ResearchGate.

interface simpliste n'est cependant pas adaptée aux recherches bibliographiques, et la quantité de résultats obtenue ne facilite pas la sélection des informations pertinentes et de qualité [Niu *et al.*, 2010]. D'un autre côté, la multiplicité des interfaces des bases de données spécialisées ne facilite pas la tâche lorsque chacune nécessite un apprentissage particulier à une structure à chaque fois différente [JISC, 2010].

À côté de cette insuffisance des techniques, un troisième constat peut être fait : celui d'une certaine insuffisance dans les compétences de maîtrise de l'information, et il semblerait qu'une partie des usagers, apparemment confiants dans leurs capacités à utiliser les outils de recherche d'information, en vient à confondre leurs capacités « techniques » avec leurs capacités informationnelles, notamment dans l'évaluation de la qualité des informations [JISC, 2010]. Une étude a ainsi constaté une forte corrélation entre les difficultés d'accès et une adoption limitée des ressources en ligne et au contraire une correspondance entre une forte utilisation, de fortes compétences et une satisfaction de l'usage [Huxley *et al.*, 2006]. Cela est d'autant plus vrai que ces usagers sont jeunes, tandis que la maîtrise des outils augmente avec la maîtrise du domaine [CIBER: University College London, 2008 ; JISC, 2010]. Comme nous l'avons déjà mentionné, cela est d'autant plus vrai dans les sciences humaines et sociales où l'on constate que les stratégies de recherche ne sont pas toujours efficaces, du fait d'une mauvaise connaissance des techniques de recherche, mais aussi du fait de termes disciplinaires mal adaptés. Non seulement les chercheurs utilisent peu de mots clés et rarement les critères de recherche avancée, mais les mots clés utilisés sont souvent généraux et imprécis ou mal adaptés à la recherche d'information conceptuelle [Buchanan *et al.*, 2005 ; Huxley *et al.*, 2006].

Toutes les études s'accordent d'ailleurs sur la nécessité de développer les outils de manière à les adapter aux pratiques et besoins des utilisateurs, et cela semble d'autant plus crucial en sciences humaines et sociales où l'inadéquation est grande entre les attentes et besoins des chercheurs et les possibilités actuellement permises par les outils disponibles. Il semblerait en effet plus simple d'adapter les outils aux manières de faire plutôt que l'inverse, voire d'aider les chercheurs à développer les outils qui leur correspondent et qui favorisent non seulement l'autonomie de la recherche d'information mais aussi l'appropriation collective et collaborative des fonds et corpus existants et en développement [Brown *et al.*, 2006 ; Lougee, 2006 ; Sukovic, 2008 ; Bradley, 2008 ; Pouyllau, 2010].

PLACE ET RÔLES DES BIBLIOTHÈQUES ET DES PROFESSIONNELS DE L'INFORMATION

+++++

Le numérique éloigne les chercheurs et les étudiants de la bibliothèque et il devient de plus en plus difficile pour les professionnels de l'information de suivre l'évolution de leurs pratiques et de leurs besoins et donc de s'y adapter et de proposer les réponses les plus adéquates en termes de mise à disposition de ressources ou de formations. Avec l'augmentation de la mise à disposition des ressources en ligne, la baisse de fréquentation des bibliothèques est visible partout, notamment dans les domaines des sciences, techniques et médecine, entraînant une désintermédiation croissante de la bibliothèque et un risque de perte de pertinence de son rôle cœur [Brown et Swan, 2007 ; Schonfeld et Housewright, 2010]. L'utilisation massive des moteurs de recherche, mais aussi des sites scientifiques spécifiques, favorise une méconnaissance des ressources mises à disposition par les bibliothèques, d'autant plus que l'accès à ces ressources se fait de plus en plus directement par les moteurs, avec la conséquence d'une confusion fréquente entre les ressources en « libre accès » et « l'accès libre » aux ressources de l'institution [OCLC, 2005 ; Brown et Swan, 2007]. Habités à l'ergonomie des moteurs de recherche, les chercheurs attendent le même type d'interface des sites de bibliothèques, permettant un accès simplifié et direct à l'information à partir d'une seule requête sur une interface unique interrogeant des sources diversifiées [Verry-Jolivet, 2001 ; Niu *et al.*, 2010].

Ainsi, le fort besoin d'autonomie des chercheurs les amène parfois à ressentir une dichotomie entre « le temps court de la recherche et celui plus long de la bibliothèque » : la bibliothèque est alors vécue comme une contrainte et des stratégies de contournement et d'évitement sont mises en place [Verry-Jolivet, 2001 ; Bernaudin *et al.*, 2005]. Cette divergence de vue est d'ailleurs largement réciproque, la majorité des chercheurs ne connaissant pas le rôle et les actions menées par leur bibliothèque, et les bibliothécaires ayant surtout le sentiment que les chercheurs ont besoin d'être formés à la bonne utilisation des ressources, tandis que ces derniers n'en voient pas la nécessité [Research Information Network, 2010 ; Silipigni Connaway et Dickey, 2010]. Des influences réciproques et des pressions se mettent en place entre les comportements des chercheurs et les développements des bibliothèques, les uns étant parfois en avance sur les autres [Brown et Swan, 2007].

De fait, les attentes des chercheurs envers les bibliothèques sont multiformes, et leurs perceptions évoluent lentement [Silipigni Connaway et Dickey, 2010]. D'un côté, les bibliothèques sont encore largement perçues comme des lieux de travail et d'accès aux ressources physiques, et plus particulièrement aux ouvrages. D'un autre côté, ils savent que les ressources sélectionnées par la bibliothèque sont des ressources de qualité, alors que cette qualité n'est pas toujours clairement identifiable pour les informations obtenues *via* les moteurs de recherche [OCLC, 2005]. Les chercheurs considèrent avant tout les professionnels de la documentation comme des fournisseurs, des gestionnaires et des gardiens des ressources numériques [Bernaudin *et al.*, 2005 ; Brown et Swan, 2007]. Et en effet, les professionnels sont bien conscients des changements drastiques concernant l'environnement des bibliothèques et du fait que l'on est passé d'une logique de stock à une logique de flux [Muet, 2009]. Le temps et les budgets consacrés aux ressources numériques ont considérablement augmenté et, paradoxalement, le fait d'amener l'information aux usagers les rend de moins en moins visibles [Inger et Gardner, 2008]. Ces changements sont en effet certainement trop récents pour être perçus par la majorité des chercheurs, encore largement peu au fait des actions menées par les bibliothèques [Brown et Swan, 2007]. Celles-ci doivent avant tout renforcer la visibilité des ressources numériques et papier en les intégrant systématiquement dans leurs catalogues, en développant des métadonnées de qualité et des outils d'intégration des différents systèmes d'information [Brown et Swan, 2007 ; Inger et Gardner, 2008 ; MacColl, 2008]. Elles doivent aussi renforcer l'usabilité de leurs services et continuer à former leurs usagers à l'utilisation des outils et à la maîtrise de l'information [Verry-Jolivet, 2001 ; JISC, 2010]. Face aux évolutions de la science, notamment l'augmentation de la recherche interdisciplinaire et du volume des données de la science, et aux demandes de publics de plus en plus diversifiés, elles doivent surtout défendre et revendiquer leur place au sein de leur institution, voire développer des stratégies de marketing de manière à acquérir une identité forte [Brown et Swan, 2007 ; Silipigni Connaway et Dickey, 2010]. Enjeux passionnants mais difficiles dans des contextes économiques et institutionnels qui ne favorisent pas toujours ces entreprises : les budgets ne suivant pas toujours les besoins¹⁶

16. Une étude britannique note ainsi qu'un rapport des années 1960 (the Parry Report) recommandait un taux de financement des bibliothèques de 6 % du budget de son institution : de 4-5 % à l'époque, il a baissé depuis à 3 % [Brown et Swan, 2007].

et les instances ne mettant pas toujours suffisamment en avant le lien nécessaire entre le développement d'une recherche et d'un enseignement de qualité avec les ressources documentaires nécessaires [Tesnière, 2008 ; Utrecht University Library, 2009].

RÉFÉRENCES BIBLIOGRAPHIQUES

+++++

Bass Abigail, Fairley Jill, Fox Kara et Sullivan Jennifer, *The Information Behavior of Scholars in the Humanities and Social Sciences*, Washington, University of Washington, 2005.

Bernaudin Anne-Claire, Charra Gaëlle, Charrier-Arrighi Nathalie, Desalme Aubierge et Pinet Nicolas, *La bibliothèque du scientifique*, Mémoire de recherche pour le diplôme de conservateur de bibliothèque, 2005.

[En ligne] < <http://www.enssib.fr/bibliotheque-numerique/document-627> > (consulté le 26 septembre 2010).

Bradley John, "Pliny: a model for digital support of scholarship", *Journal of Digital Information*, vol. 9, n° 1, 2008.

[En ligne] < <https://journals.tdl.org/jodi/rt/printerFriendly/209/198> > (consulté le 26 septembre 2010).

Brockman William S., Neumann Laura, Palmer Carole L. et Tidline Tonya J., *Scholarly Work in the Humanities and the Evolving Information Environment*, Digital Library Federation, Council on Library and Information Resources, Washington, D.C., 2001.

[En ligne] < <http://www.clir.org/pubs/reports/pub104/pub104.pdf> > (consulté le 26 septembre 2010).

Brown Sheridan et Swan Alma, *Researchers' Use of Academic Libraries and their Services, A Report Commissioned by the Research Information Network and the Consortium of Research Libraries*, 2007.

[En ligne] < <http://www.rin.ac.uk/files/libraries-report-2007.pdf> > (consulté le 26 septembre 2010).

Brown Stephen, Ross Robb, Greengrass Mark, Bryson Jared et Gerrard David, *RePAH: A User Requirements Analysis for Portals in the Arts and Humanities. Final Report*, Arts and Humanities Research Council, ICT in Arts and Humanities Programme, 2006.

[En ligne] < <http://repah.dmu.ac.uk/report/> > (consulté le 26 septembre 2010).

Buchanan George, Cunningham Sally Jo, Blandford Ann, Rimmer Jon et Warwick Claire, "Information seeking by humanities scholars (pre-print)", in *Lecture Notes in Computer Sciences*, vol. 3652, *Research and Advanced Technology for Digital Libraries*, 9th European Conference (ECDL 2005), Vienna, Austria, September 18-23, 2005, pp. 218-229. [En ligne] < <http://www.ucl.ac.uk/annb/docs/gbsjcabjrcwECDLpreprint.pdf> >, et chez Springer < <http://www.springerlink.com/content/enf96ymmn8b3/#section=582881&page=1&locus=0> > (consultés le 26 septembre 2010).

Charra Gaëlle, *Pratiques de recherche documentaire et attentes des publics de chercheurs en lettres et sciences humaines. Étude à partir du cas de la bibliothèque Denis Diderot*, Mémoire d'étude pour le diplôme de conservateur, 2006.

[En ligne] < <http://enssibal.enssib.fr/bibliotheque/documents/dcb/charra.pdf> > (consulté le 26 septembre 2010).

CIBER: University College London, *Information Behaviour of the Researcher of the Future*, British Library/JISC, 2008.

[En ligne] < http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf > (consulté le 26 septembre 2010).

Colcanap Grégory, « Acquérir la documentation électronique pour l'enseignement supérieur et la recherche », *Bulletin des bibliothèques de France*, 2009, t. 54, n° 6, pp. 18-21.

[En ligne] < <http://bbf.enssib.fr/consulter/bbf-2009-06-0018-003> > (consulté le 26 septembre 2010).

Cox John et Cox Laura, *Scholarly Publishing Practice, Third Survey 2008: Academic Journal Publishers' Policies and Practices in Online Publishing*, ALPSP, 2008.

[En ligne] < <http://www.ingentaconnect.com/content/alpsp/spp3> > (consulté le 26 septembre 2010).

Derfoufi Ilham, *Information scientifique pour la recherche en éducation : analyse comparée de l'offre et des usages entre la France et le Royaume-Uni*, Thèse de doctorat, Université Claude Bernard, Lyon 1, 2009.

Fry Jenny, "Studying the scholarly Web: how disciplinary culture shapes online representations", *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, vol. 10, n° 1, 2006.

[En ligne] < <http://cybermetrics.cindoc.csic.es/articles/v10i1p2.pdf> > (consulté le 26 septembre 2010).

Fry Jenny, Oppenheim Charles, Spezi Valérie et White Sonya, *PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories, Baseline Report*, 2009.

[En ligne] < http://www.peerproject.eu/fileadmin/media/reports/Final_revision_-_behavioural_baseline_report_-_20_01_10.pdf > (consulté le 26 septembre 2010).

Fry Jenny et Talja Sanna, "The cultural shaping of scholarly communication: explaining e-journal use within and across academic fields", in *Proceedings of the American Society for Information Science and Technology*, vol. 41, n° 1, 2005, pp. 20-30.

Gallezot Gabriel et Le Deuff Olivier, « Chercheurs 2.0 », *Les Cahiers du numérique*, 2009, n° 2, pp. 15-31.

Gallezot Gabriel, Roland Michel et Araszkievics Jacques, « La recherche floue », In *Actes de la deuxième conférence Document numérique et société : traitements et pratiques documentaires : vers un changement de paradigme ?*, Paris, CNAM, 17-18 novembre 2008.

[En ligne] < http://archivesic.ccsd.cnrs.fr/sic_00340835/fr/ > (consulté le 26 septembre 2010).

Garvey William D., *Communication: the Essence of Science. Facilitating Information Exchange among Librarians, Scientists, Engineers and Students*, Oxford, Pergamon Press, 1979.

Harley Diane, Acord Sophia Krzys, Earl-Novell Sarah, Lawrence Shannon et King C. Judson, *Assessing the Future Landscape of Scholarly Communication: an Exploration of Faculty Values and Needs in Seven Disciplines*, Center for Studies in Higher Education, UC Berkeley, 2010.

[En ligne] < http://escholarship.org/uc/cshe_fsc > (consulté le 26 septembre 2010).

Heimeriks Gaston, van den Besselaar Peter et Frenken Koen, "Digital disciplinary differences: an analysis of computer-mediated science and

'Mode 2' knowledge production", *Research Policy*, vol. 37, n° 9, 2008, pp. 1602-1615.

Herrera Gail, "Google scholar users & user behaviors: an exploratory study", *College and Research Libraries*, vol. 72, n° 4, 316-330, 2011.

[En ligne] < <http://crl.acrl.org/content/72/4/316> >.

Huxley Leslie, Mullings Christine, Hodos Tamar et Jones Diane, *Gathering Evidence: Current ICT Use and Future Needs for Arts and Humanities Researchers*, University of Bristol, 2006.

[En ligne] < http://www.ahrcict.rdg.ac.uk/activities/strategy_projects/reports/bristol/gathering_evidence_final.pdf > (consulté le 26 septembre 2010).

Inger Simon et Gardner Tracy, *How Readers Navigate to Scholarly Content, Comparing the Changing User Behaviour between 2005 and 2008 and its Impact on Publisher Web Site Design and Function*, 2008. [En ligne] < <http://www.sic.ox14.com/howreadersnavigatetoscholarlycontent.pdf> > (consulté le 26 septembre 2010).

JISC, *JISC National E-books Observatory Project: Key Findings and Recommendations*, 2009.

[En ligne] < <http://www.jiscebooksproject.org/reports/finalreport> > (consulté le 26 septembre 2010).

JISC, *User Behaviour Observational Study: User Behaviour in Resource Discovery*, 2010.

[En ligne] < <http://www.jisc.ac.uk/publications/programmerelated/2010/ubirdfinalreport.aspx> > (consulté le 26 septembre 2010).

King Donald W., Tenopir Carol, Montgomery C. H. et Aerni S. E., "Patterns of journal use by faculty at three diverse universities", *D-Lib Magazine*, n° 9, 2003.

[En ligne] < <http://www.dlib.org/dlib/october03/king/10king.html> > (consulté le 26 septembre 2010).

Kling Rob et McKim Geoffrey, "Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication", *Journal of the American Society for Information Science*, vol. 51, n° 14, 2000, pp. 1251-1338.

Lougee Wendy Pradt, *A Multi-Dimensional Framework for Academic Support. A Final Report*, Submitted to the Andrew W. Mellon Foundation from the University of Minnesota Libraries, 2006.

[En ligne] < <http://www1.lib.umn.edu/about/mellon/docs.phtml> > (consulté le 26 septembre 2010).

MacColl John, *The Prolific RIM: Mapping Research Support Activity in Libraries*, 2008.

[En ligne] < <http://www.oclc.org/research/events/2008-11-05k.pdf> > (consulté le 26 septembre 2010).

Mahé Annaïg, *La communication scientifique en (r)évolution : l'intégration des revues électroniques dans les pratiques informationnelles de chercheurs en sciences de la nature comme révélateur des mutations du modèle traditionnel de la communication scientifique*, Thèse de doctorat, Université Claude Bernard, Lyon 1, 2002.

[En ligne] < <http://www.enssib.fr/bibliotheque/documents/theses/mahe/mahe.pdf> > (consulté le 26 septembre 2010).

Martineau Catherine et Clénet Geneviève, *Pratiques documentaires et nouveaux modes de publication. Résultats d'une enquête comparative*, 2006.

[En ligne] < http://archivesic.ccsd.cnrs.fr/docs/00/09/70/43/PDF/synthese_commune_2.pdf > (consulté le 26 septembre 2010).

McClanahan Kitty, Wu Lei, Tenopir Carol et King Donald W., "Embracing change: perceptions of e-journals by faculty members", *Learned Publishing*, vol. 23, n° 3, 2010, pp. 209-223.

Meadows Jack, "The users of e-publishing and their communication behavior", in *Proceedings ELPUB 2006 Conference on Electronic Publishing*. Banskó, 2006.

[En ligne] < http://elpub.scix.net/data/works/att/sup1_elpub2006.content.pdf > (consulté le 26 septembre 2010).

Merton Robert K., *The Sociology of Science. Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, 1973.

Muet Florence, « Mutations de l'enseignement supérieur et perspectives stratégiques pour les bibliothèques universitaires », *Documentaliste – Sciences de l'information*, 2009, vol. 46, n° 4, pp. 4-12.

Nicholas David, Rowlands Ian, Huntington Paul, Clark David et Jamali Hamid R., *E-Journals: their Use, Value and Impact*, Research Information Network, 2009.

[En ligne] < <http://www.rin.ac.uk/system/files/attachments/E-journals-report.pdf> > (consulté le 26 septembre 2010).

Niu Xi, Hemminger Bradley M., Lown Cory, Adams Stephanie, Brown Cecilia, Level Allison, McLure Merinda, Powers Audrey, Tennant Michele R. et Cataldo Tara., "National study of information seeking behavior of academic researchers in the United States", *Journal of the American Society for Information Science and Technology*, vol. 61, n° 5, 2010, pp. 869-890.

OCLC, *Perceptions of Libraries and Information Resources*, 2005.

[En ligne] < <http://www.oclc.org/reports/2005perceptions.htm> > (consulté le 26 septembre 2010).

Pouyllau Stéphane, *Construire le Web de données pour les sciences humaines et sociales. Les digital humanities au service de la recherche en SHS. Note informationnelle et propositions du Centre national pour la numérisation de sources visuelles*, 2010.

[En ligne] < http://archivesic.ccsd.cnrs.fr/sic_00494227/fr/ > (consulté le 26 septembre 2010).

Research Information Network, *If you Build it, will they Come? How Researchers Perceive and Use Web 2.0*, 2010.

[En ligne] < http://www.rin.ac.uk/system/files/attachments/web_2.0_screen.pdf > (consulté le 26 septembre 2010).

Research Information Network et RightsCom, *Researchers and Discovery Services: Behaviour, Perceptions and Needs*, 2006.

[En ligne] < <http://www.rin.ac.uk/system/files/attachments/Researchers-discovery-services-report.pdf> > et < <http://www.rin.ac.uk/system/files/attachments/Researchers-discovery-services-report.pdf> > (consulté le 26 septembre 2010).

Rowlands Ian, "Electronic journals and user behavior: a review of recent research", *Library and Information Science Research*, vol. 29, n° 3, 2007, pp. 369-396.

Schonfeld Roger C. et Housewright Ross, *Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies*, Ithaka S+R, 2010.

[En ligne] < <http://www.ithaka.org/ithaka-s-r/research/faculty-surveys-2000-2009/Faculty%20Study%202009.pdf> > (consulté le 26 septembre 2010).

Silipigni Connaway Lynn et Dickey Timothy J., *The Digital Information Seeker. Report of Findings from Selected OCLC, RIN and JISC User Behaviour Projects*, 2010.

[En ligne] < <http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf> > (consulté le 26 septembre 2010).

Simonnot Brigitte et Gallezot Gabriel, *L'entonnoir. Google sous la loupe des sciences de l'information et de la communication*, Caen, C&F, 2009.

Sparks Sue, *Disciplinary Differences and Needs*, JISC – Rightscom Ltd, 2005.

[En ligne] < http://www.jisc.ac.uk/uploaded_documents/Disciplinary%20Differences%20and%20Needs.doc > (consulté le 26 septembre 2010).

Springer, *Ebooks. The End User Experience*, 2008.

[En ligne] < http://www.springer.com/cda/content/document/cda_downloaddocument/eBooks+-+the+End+User+Experience?SGWID=0-0-45-608298-0 > (consulté le 26 septembre 2010).

Sukovic Suzana, “Convergent flows: humanities scholars and their interactions with electronic texts”, *The Library Quarterly*, vol. 78, n° 3, 2008, pp. 263-284.

Swan Alma et Brown Sheridan, *Open Access Self-Archiving: an Introduction, JISC, HEFCE Report*, 2005.

[En ligne] < <http://eprints.ecs.soton.ac.uk/11006/> > (consulté le 26 septembre 2010).

Talja Sanna, Vakkari Pertti, Fry Jenny et Wouters Paul, “Impact of research cultures on the use of digital library resources”, *Journal of the American Society for Information Science*, 58 (11), 2007, pp. 1674-1685.

Tenopir Carol, *Use and Users of Electronic Library Resources: an Overview and Analysis of Recent Research Studies*, Washington, Council on Library and Information Resources, ERIC, 2003.

[En ligne] < <http://www.clir.org/pubs/abstract/pub120abst.html> > (consulté le 26 septembre 2010).

Tenopir Carol, King Donald W., Spencer Jesse et Wu Lei, "Variations in article seeking and reading patterns of academics: what makes a difference?", *Library & Information Science Research*, vol. 31, n° 3, 2009, pp. 139-148.

Tesnière Valérie, *Politique scientifique et politique documentaire des universités : quelles articulations ?*, Inspection générale des bibliothèques, 2008.

[En ligne] < http://media.enseignementsup-recherche.gouv.fr/file/Concours_2008/32/8/rapportdefinitif_poldoc-polsc_VT_19-08-08_33328.pdf > (consulté le 26 septembre 2010).

Unsworth John, "Scholarly primitives: what methods do humanities researchers have in common, and how might our tools reflect this?", *Symposium on "Humanities Computing: Formal Methods, Experimental Practice" Sponsored by King's College London*, 2000.

[En ligne] < <http://people.lis.illinois.edu/~unsworth/Kings.5-00/primitives.html> > (consulté le 26 septembre 2010).

Unsworth John, "The Humanist: 'Dances with wolves' or 'Bowls alone'?", in *Scholarly Tribes and Tribulations: how Tradition and Technology are Driving Disciplinary Change*. Washington, Association of Research Libraries, 2003.

[En ligne] < <http://www.arl.org/bm~doc/unsworth.pdf> > (consulté le 26 septembre 2010).

Utrecht University Library, *Partner in Science: 2009-2012 Policy Plan*, 2009.

[En ligne] < <http://www.uu.nl/university/library/EN/about/Documents/Strategic%20Plan%20University%20Library%20Utrecht%202009%-2012.pdf> > (consulté le 26 septembre 2010).

Van Orsdel Lee C. et Born Kathleen, "Journals in the time of Google", *LibraryJournal.com*, 2006.

[En ligne] < <http://www.libraryjournal.com/article/CA6321722.html> > (consulté le 26 septembre 2010).

Velden Theresa et Lagoze Carl, "The transformation of scientific communication systems in the digital age. Towards a methodology for comparing scientific communication cultures", in *Proceedings of Workshop Oxford e-Research*. 2008.

[En ligne] < <http://www.cs.cornell.edu/%7Etvelden/pubs/2008-eresearch08.pdf> > (consulté le 26 septembre 2010).

Verry-Jolivet Corinne, "Pratiques et attentes des chercheurs. La médiathèque scientifique de l'Institut Pasteur", *Bulletin des bibliothèques de France*, 2001, t. 46, n° 4, pp. 26-30.

[En ligne] < <http://bbf.enssib.fr/consulter/bbf-2001-04-0026-002> > (consulté le 26 septembre 2010).

Ware Mark et Mabe Michael, *The STM Report: an Overview of Scientific and Scholarly Journal Publishing*, 2009.

[En ligne] < http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf > (consulté le 26 septembre 2010).

Warwick Claire, Terras Melissa, Galina Isabel, Huntington Paul et Pappa Nikoleta, "Library and information resources and users of digital resources in the humanities", *Program: Electronic Library and Information Systems*, vol. 42, n° 1, 2008, pp. 5-27.

Warwick Claire, Terras Melissa, Huntington Paul et Pappa Nikoleta, "If you build it will they come? The LAIRAH study: quantifying the use of online resources in the arts and humanities through statistical analysis of user log data", *Literary and Linguistic Computing*, vol. 23, n° 1, 2008, pp. 85-102.

par Jenny Fry

traduit par Laure Manceau

+++++

LUMIÈRE SUR LE WEB
UNIVERSITAIRE :
L'INFLUENCE
DE LA CULTURE
DISCIPLINAIRE SUR
LES REPRÉSENTATIONS
EN LIGNE

+++++

+++++

LUMIÈRE SUR LE WEB UNIVERSITAIRE : L'INFLUENCE DE LA CULTURE DISCIPLINAIRE SUR LES REPRÉSENTATIONS EN LIGNE¹

+++++

INTRODUCTION

+++++

Cet article a pour but d'établir un pont entre les caractéristiques culturelles fondamentales de divers domaines d'études et la façon dont les activités de recherche sont représentées sur le Web universitaire, en mettant l'accent sur l'analyse qualitative des liens hypertextes². Pour ce faire, nous nous demanderons si la théorie de Whitley [1984] concernant l'organisation intellectuelle et sociale des sciences, bien connue dans le champ des études des sciences et des techniques, peut s'appliquer au contexte du Web. La théorie de Whitley s'appuie sur deux concepts capitaux : le degré de *mutual dependence* qui existe entre les chercheurs ou les disciplines en matière de contribution au corps des connaissances, et le degré de *task uncertainty*, qui touche à la coordination et à l'interprétation des méthodes de recherche et leurs résultats. L'application de ces deux concepts, intimement liés et relatifs, constitue un véritable défi car on ne peut les mesurer en termes absolus. En examinant le comportement des communautés de chercheurs en mode « connexion » ou « déconnexion », cet article fournit un cadre explicatif à la compréhension des phénomènes intellectuels et sociaux qui sous-tendent les relations hypertextuelles dans le Web universitaire.

Comme le soulignent Harries *et al.* [2004] ainsi que Thelwall [2006], malgré l'intérêt actuel pour les relations hypertextuelles et les nombreuses statistiques indiquant que l'on peut associer les liens Internet à l'importance et à la productivité d'une discipline, aucun lien de cause

1. N.D.T. : cet article a été publié sous le titre "Studying the scholarly Web: how disciplinary culture shapes online representations", dans *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, 10 (1), 2006. [En ligne] < <http://cybermetrics.cindoc.csic.es/articles/v10i1p2.pdf> >.

2. Pour une typologie des analyses de liens, voir Thelwall [2006].

à effet n'a été démontré, et les efforts pour fournir un cadre théorique permettant de donner un sens probant à ces résultats en sont restés au stade de tentatives. Pour Harries et ses collaborateurs [2004], il est urgent de remédier à cette situation car la compréhension lacunaire de la création des liens Internet constitue un obstacle au développement du champ émergent qu'est la webométrie.

Dans ses recherches sur les motivations sociales qui sous-tendent la création des liens Internet, Thelwall [2006] signale les limites d'une analyse de liens à grande échelle et prône la triangulation des méthodes qualitative et quantitative. Cet article tente de pallier les lacunes concernant la théorie sociale de la webométrie en puisant uniquement dans les techniques de l'ethnographie virtuelle [Hine, 2000 ; Marcus, 1995], en vue de contribuer au concept de sphère Internet³ développé par Schneider et Foot [2002]. Schneider et Foot ont développé ce concept dans un certain nombre de contextes non universitaires, en particulier dans le cadre d'une étude sur les campagnes électorales en ligne. Ils définissent une sphère Internet comme étant un « ensemble de ressources numériques définies recouvrant plusieurs sites Internet jugés pertinents ou en rapport avec un thème ou un objet central ». Dans cette étude, nous avons pris pour « thème central » les frontières épistémologiques de deux champs d'études précis.

Ces domaines sont : la théorie de l'argumentation d'une part, champ interdisciplinaire au confluent des études en philosophie, linguistique et communication ; et la linguistique pragmatique d'autre part, champ également interdisciplinaire regroupant des aspects de la linguistique computationnelle, tels que le traitement automatique du langage naturel, et des domaines de la linguistique plus traditionnelle, tels que la sociolinguistique, la linguistique contrastive et la morphologie.

La linguistique pragmatique se caractérise par un très haut degré de travail collaboratif, lequel se réalise en termes de projets à l'échelle nationale et dont les résultats aboutissent, entre autres, à l'élaboration de corpus [Becher, 1989]. L'étude de cas de linguistique pragmatique s'appuie sur un important projet de corpus national aux Pays-Bas : le *Corpus Gesproken Nederlands* (CGN)⁴, le corpus de néerlandais parlé.

La théorie de l'argumentation se caractérise, quant à elle, par des écoles de pensée plutôt que par des projets [Becher, 1989]. Notre étude

3. N.D.T. : nous choisissons de traduire *websphere* par « sphère Internet » dans cet article.

4. CGN : < <http://lands.let.kun.nl/cgn/> >.

de cas reflète cette donnée et se concentre sur ce que les chercheurs de ce domaine nomment « l'École d'Amsterdam », groupe de recherche basé dans le département de communication du langage, de théorie de l'argumentation et de rhétorique de l'université d'Amsterdam.

Si ces deux études de cas tombent sous la coupe de la même discipline mère – la linguistique –, il n'en demeure pas moins que les objectifs, l'identité culturelle et la visibilité sur Internet de leur communauté universitaire respective, diffèrent sur de nombreux points. Les sphères Internet imbriquées ont été construites en se basant sur des entretiens avec les deux chefs de projet du CGN et un membre de l'École d'Amsterdam ainsi que sur des données collectées par le biais de Microsoft Site Analyst et de moteurs de recherche⁵.

Les résultats pointent vers des indicateurs qualitatifs potentiellement étalonnables qui rendent ce que Cronin et ses collaborateurs [1998] décrivent comme les traces liminales de l'estime de ses pairs, d'influence sur la réputation et les ressources, et la légitimation des problèmes, approches et techniques de recherche.

L'ÉTUDE DES PRATIQUES DES CHERCHEURS UNIVERSITAIRES SUR INTERNET

+++++
Les études sur la communication universitaire ont utilisé divers niveaux de granularité comme unités d'analyse. Cela reflète la production du savoir dans une multitude de configurations sociales et institutionnelles se chevauchant parfois, comme les groupes de chercheurs, les projets, les collègues invisibles, les domaines spécialisés, les laboratoires, les départements et les disciplines universitaires, dont les frontières manquent parfois de précision.

Price [1963] et Crane [1972] ont utilisé la notion de communautés scientifiques orientées vers la recherche se regroupant autour de problèmes ou de méthodes de recherche comparables. Le concept de communauté scientifique est souvent décrit comme idyllique et consensuel et en tant que tel a été critiqué [Swales, 1998]. Price a inventé le terme de collègue invisible en 1963 : il s'agit pour lui d'un réseau social d'une capacité d'environ 200 membres regroupés autour d'un noyau d'une vingtaine de chercheurs. Price a expliqué qu'au-delà de cette capacité, le domaine de

5. Voir les annexes : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html > (consulté le 3 mai 2004).

recherche arrivait à saturation et que les groupes se divisaient afin de former de nouvelles disciplines spécialisées. Des chercheurs ont développé ce concept dans un certain nombre de disciplines. En 1983, Kochen a, par exemple, eu recours à l'estimation de la taille d'un groupe de recherche de pointe en modélisation mathématique afin de simuler l'évolution des disciplines spécialisées. En sociologie, Crane [1972] s'est quant à lui appuyé sur la notion de collège invisible de Price pour exemplifier la diffusion des connaissances *via* les réseaux des communautés scientifiques. Dans des travaux plus récents, Bruckner, Ebeling et Scharnhorst [1989] ont utilisé la théorie de Price pour simuler des instabilités dans les systèmes scientifiques évolutionnistes et ont poursuivi le modèle de Kochen.

S'orientant vers une vision moins idyllique des communautés scientifiques où la compétition et la répartition du travail sont capitales, Bourdieu a introduit la notion de champ intellectuel en 1988. Selon la définition de Bourdieu, les champs sont des espaces fluides basés sur le principe de la compétition – plutôt que sur le consensus, comme il en est question dans le concept des « communautés de pratique » [Wenger, 1998] –, principe selon lequel des communautés organisées en réseau sont en compétition pour les mêmes problèmes et les mêmes ressources.

La notion de discipline constitue, quant à elle, une façon plus institutionnelle d'envisager les communautés scientifiques. Les disciplines scientifiques sont indissociables des institutions pédagogiques et de formation, et représentent des monopoles de marché de certains types de connaissances [Whitley, 1984]. Salter et Hearn [1996, p. 23] ont réuni les thèmes récurrents qui tentent de définir ce qui caractérise une discipline : constellation de sujets, perspectives et méthodes ; approche dominante, avec des perspectives critiques ; reconnaissance institutionnelle sous la forme de départements, journaux et conférences ; communauté autoproclamée d'universitaires ; et méthodes pour inculquer ou imposer l'adhésion à la culture de ladite discipline. Mais le concept de discipline académique est loin d'être simple. Le rapport de Heilbron [2004] montre que l'histoire des disciplines est différente selon les pays. Lenoir [1997] souligne que la majeure partie de la recherche littéraire ne se limite pas à une seule discipline, mais fait appel aux travaux de plusieurs disciplines. L'existence de départements universitaires pertinents ne définit qu'en partie une discipline, et il ne s'ensuit pas automatiquement que chaque département représente une discipline [Becher, 1989]. Dans l'étude de la visibilité sur Internet, l'analyse hypertextuelle a tendance à associer les départements universitaires aux domaines de recherche apparentés. Par exemple, une

étude de Heimeriks, Hörlesberger et Van den Besselaar [2003] portant sur la communication et la collaboration dans les relations entre université, industrie et gouvernement, a utilisé les pages d'accueil de départements universitaires comme représentations de champs de connaissances de mode 2 [Gibbons *et al.*, 1994] ; pourtant, un seul département de cette étude de cas représentait en réalité les activités de recherche apparentées. De telles associations peuvent mener à des résultats trompeurs et à des erreurs d'interprétation dans les interrelations en jeu dans l'étude des liens hypertextes. Afin de saisir au plus juste la représentation numérique des sociétés de la production du savoir, cet article se concentre sur un niveau d'agrégation moindre que celui des départements universitaires, optant plutôt pour les projets et les groupes de recherche en tant qu'études de cas imbriquées dans un champ intellectuel.

Bien que Whitley [1984] ait basé sa théorie sur la comparaison des disciplines, Fry [2006] a démontré qu'elle pouvait être utilisée pour étudier les pratiques de chercheurs dans un domaine spécialisé, dans la mesure où les concepts de *mutual dependence* et de *task uncertainty* sont valables aussi bien pour l'organisation interne (répartition du travail et compétition) d'une communauté de spécialistes, l'environnement scientifique plus large de la discipline dont ils dépendent, que pour le monde de la science. Par exemple, la position stratégique qu'occupe une spécialité au sein de sa discipline mère peut constituer un facteur majeur dans la détermination de son autonomie, de sa cohérence et de la direction de ses recherches. Suivant le degré de granularité – discipline ou spécialité –, les communautés scientifiques peuvent différer dans leur degré relatif de *mutual dependence* et de *task uncertainty*, tout comme des spécialités peuvent s'avérer plus ou moins coordonnées, intégrées et standardisées que leur discipline mère.

Les études sur la communication universitaire à l'intérieur des réseaux numériques tendent à privilégier l'analyse du social sur le cognitif ou l'inverse. Par exemple, Kling et McKim [2000] se sont concentrés sur l'influence qu'ont des structures sociales telles que la reconnaissance ou la récompense sur les publications universitaires sur le Web, sans prendre en compte les caractéristiques de la structure des connaissances en question. En raison de leur caractère interdépendant, les deux dimensions de Whitley permettent à ces deux éléments du travail universitaires d'être envisagés comme valides du point de vue de l'analyse. Ils se subdivisent en quatre éléments analytiques qui se recoupent parfois et sont liés soit au contrôle de la réputation et à la coordination des stratégies de

recherche et des priorités intellectuelles, à la « dépendance stratégique » et à l'« incertitude stratégique », soit à la coordination des standards de compétence, des techniques de recherche et des résultats, à la « dépendance fonctionnelle » et à l'« incertitude technique ».

Bien que cette étude ait porté sur un phénomène Internet, il était important de convoquer l'environnement plus général des universitaires de chaque étude de cas afin d'interpréter les traces révélées par la collecte des données en ligne et leur analyse. Comme le signalent Wouters et de Vries [2004], les études qui se focalisent sur les caractéristiques numériques de la communication universitaire ont tendance à négliger les aspects intellectuel, social et institutionnel. Les recherches ayant le Net pour objet ont montré que les frontières entre mondes réel et virtuel sont souvent indistinctes [Hine, 2000 ; Haythornthwaite et Wellman, 1998], ce qui implique qu'une distinction entre mondes « connecté » et « déconnecté » soit problématique. Le concept de monde « déconnecté » est utilisé ici avec prudence. Sa fonction principale est d'indiquer que les chercheurs ont passé du temps à examiner les espaces traditionnels alloués à la communication [Marcus, 1995], tels que littérature grise, conférences et articles de revue.

CRÉATION D'UNE SPHÈRE INTERNET

+++++
En raison de l'amélioration qualitative du concept de sphère Internet, notre méthodologie n'a pas visé à saisir dans sa globalité une sphère Internet pour la linguistique pragmatique et la théorie de l'argumentation au niveau de la discipline. Nous avons privilégié une approche intégrée de l'étude de cas, en nous appuyant sur les répartitions du travail typiques de chaque domaine, ce qui a créé la base d'une sphère Internet égocentrée.

Le choix de ces deux études de cas est fonction de plusieurs critères, dont l'accessibilité et la renommée hors ligne de leurs membres. L'URL qui représentait collectivement chaque étude de cas au niveau institutionnel, comme le serveur de l'université, a été utilisée en tant qu'URL centrale de chaque sphère Internet⁶.

Le concept de sphère Internet se base sur la production du Web considérée du point de vue du producteur. Étant donné qu'il s'agissait d'une étude sur l'intégration et la coordination de la recherche et la reconnaissance de

6. Pour la première URL de chaque sphère Internet, voir les annexes : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html > (consulté le 10 novembre 2001).

ses pairs, il était nécessaire d'étendre la notion de production au contexte des sphères Internet universitaires afin de prendre en compte l'interconnectivité entre plusieurs sphères. Les liens dirigés vers les URL centrales ont donc également été inclus dans la collecte et l'analyse des données.

Étant donné la différence d'approche dans la production numérique de ces deux études de cas, il fallait aussi une approche différente dans la construction de leur sphère Internet respective. Par exemple, une différence majeure réside, dans le cas du CGN, dans la production « directe » d'une sphère Internet par le biais d'une URL centrale < <http://lands.let.ru.nl/cgn> > hébergée par l'université catholique de Nimègue, dont le contenu, produit par le chef de projet, est la représentation numérique « officielle » du projet. Le moteur de recherche Google.nl a ensuite été utilisé pour identifier les liens entrants accumulés au fil du temps. Ainsi, l'URL « officielle » du projet CGN était le centre de la sphère Internet.

Le contraste avec l'École d'Amsterdam, dont la production en matière de sphère Internet est principalement indirecte, est important. Ses membres ont d'ailleurs, dans un premier temps, décliné l'invitation à participer à cette étude au motif qu'ils ne se considéraient pas comme producteurs ni même utilisateurs de contenu numérique. Les pages d'accueil institutionnelles des membres du groupe hébergées par l'université d'Amsterdam < <http://cf.hum.uva.nl/data/afd/neerlandistiek/tar-english/> >⁷ ne contenaient que très peu d'informations en rapport avec les activités de recherche de l'École d'Amsterdam. Cette URL a initialement été choisie comme nœud central de la sphère Internet, mais elle ne contenait qu'un lien externe et ne possédait de lien entrant sur aucune autre page, et n'était donc pas reliée à un réseau d'URL externes. Il a donc fallu chercher un remplaçant à l'URL centrale de la sphère Internet de l'École d'Amsterdam plus à même de refléter la présence du groupe de chercheurs sur Internet. En 1986, suite à la première conférence internationale sur la théorie de l'argumentation qui s'était tenue à l'université d'Amsterdam, le directeur de l'École d'Amsterdam et d'autres membres ont fondé l'International Society for the Study of Argumentation (ISSA). C'est suite à l'entretien au cours duquel cette information a été révélée que l'URL du site Internet de l'ISSA < <http://cf.hum.uva.nl/issa/> > a été utilisée comme URL centrale de la sphère Internet. En sus de cette conférence

7. N.D.T. : cette URL qui a servi de base pour le corpus ne fonctionne plus. Celle du Department of Speech Communication, Argumentation Theory and Rhetoric de l'université d'Amsterdam est < <http://www.hum.uva.nl/argumentation> > (consulté le 10 novembre 2011).

annuelle, l'École d'Amsterdam a créé une revue internationale, *Argumentation: an International Journal on Reasoning* < <http://www.kluweronline.com/issn/0920-427X> >⁸, et cela joue également un rôle important dans la présence de l'École d'Amsterdam sur Internet. Les sites Web de l'ISSA et de la revue internationale dirigée par le directeur de l'École d'Amsterdam diffèrent de celui du projet « officiel » du CGN dans la mesure où ils n'ont pas été produits par des membres de l'École d'Amsterdam en tant que représentation directe des activités de recherche du groupe, mais plutôt comme représentation d'activités professionnelles et de leurs publications. En construisant une sphère Internet ethnologiquement renseignée, qui n'en demeure pas moins fonctionnelle (dépendant des liens sortants de l'URL centrale), il a été possible de distinguer une certaine représentativité de l'École d'Amsterdam, bien que ses membres aient perçu leur groupe comme non utilisateur et non producteur de contenu numérique⁹.

Pour les besoins de ce projet, j'ai défini une sphère Internet et sa présence sur le Web comme incluant les types d'inscriptions et d'actions numériques suivants :

URL centrale d'une page ou d'un site produit(e) par les membres d'un groupe / projet représentant la recherche collective ou les activités professionnelles du groupe / projet (telles qu'identifiées en entretien)	nœud central	sphère Internet produite directement
URL sortantes du nœud central (identifiées par Microsoft Site Analyst)		
URL entrantes vers le nœud central (identifiées via l'outil de recherche de liens entrants de Google.nl)		sphère Internet étendue produite indirectement
URL non reliées au nœud central, mais jugées, en entretiens et lors d'échanges d'e-mails avec des participants à l'étude, pertinentes pour la recherche et les activités professionnelles du groupe/projet		
URL au concept proche contenant des évocations du contenu de chaque étude de cas, mais non reliées par un lien hypertexte (identifiées en recherchant des extraits exacts via Google.nl)		présence sur le Web – sphère Internet potentielle

8. Le décompte des liens entrants figurant sur les pages en ligne de revues papier est compliqué par le fait que certains éditeurs, tels que Kluwer, fournissent de nombreuses options en matière de liens. N.D.T. : depuis 2004, la revue est publiée chez Springer : < <http://www.springerlink.com/openurl.asp?genre=journal&issn=0920-427X> > (consulté le 10 novembre 2011).

9. Ma requête initiale pour leur participation ayant été déclinée au motif de cette perception.

Ainsi définie, chaque sphère Internet a été enrichie de données ethnologiques, construite longitudinalement et validée en envoyant la liste des URL constituant chaque sphère Internet initiale aux membres sélectionnés pour l'étude de cas. Les URL centrales initiales ont été confirmées lors d'entretiens avec les membres clés de chaque étude de cas. Puis l'on a eu recours à Microsoft Site Analyst pour identifier les pages vers lesquelles renvoyait chaque URL centrale. Les liens entrants¹⁰ ont, quant à eux, été identifiés par des recherches régulières via l'outil de Google.nl prévu à cet effet, environ six fois en douze mois, et les résultats ont été révisés afin de supprimer les liens autoréférentiels et les URL lourdes¹¹. Ces opérations ont mis au jour de multiples sites pour le projet du CGN hébergé par diverses universités partenaires aux Pays-Bas et en Flandre, mais tous étaient reliés au site Internet « officiel » hébergé aux Pays-Bas, ainsi qu'un intranet protégé par mot de passe pour le projet hébergé par l'université d'Utrecht. Elles ont également identifié deux liens réciproques entre le projet du CGN et la Fondation néerlandaise pour la science (NWO), cofondatrice du projet, et l'Union de la langue néerlandaise (Nederlandse Taalunie).

La représentativité de chaque cas sur Google.nl a également fait partie de l'analyse. Pour ce faire, des expressions précises ont été recherchées afin de trouver des pages en rapport avec le sujet mais non reliées via un lien hypertexte. Dans un contexte de recherches liées à la discipline, cette étude de la présence sur le Web a abouti à la représentation d'une sphère Internet potentielle pour chaque étude de cas, qui était bien plus importante dans les deux cas que la sphère Internet existante. Par exemple, la recherche des termes *Corpus Gesproken Nederlands* (CGN), corpus de néerlandais parlé, aboutit à environ 900 URL, alors que le nombre d'URL dans la sphère Internet du CGN est de 63. De même, la recherche du mot « pragma-dialectique », nom familier et champ de recherche de l'École d'Amsterdam, sur Google.nl a abouti à 860 URL¹² (pour une liste des 30 premiers résultats de chaque recherche), comparé à 28 URL sur la

10. Voir Björneborn et Ingwersen [2004] pour une typologie des liens.

11. Il est possible d'exclure les liens autoréférentiels sur Alta Vista en ayant recours aux opérateurs booléens (voir Rousseau [1997] pour une utilisation de l'outil de recherche avancée d'Alta Vista concernant les liens entrants) et, au cours de cette étude, il semblerait que Google ait reconfiguré son algorithme et exclut dorénavant automatiquement les liens autoréférentiels des résultats de recherche.

12. Voir les annexes V et VI : [en ligne] : < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

sphère Internet. Ces recherches ont également été menées *via* le moteur de recherche Alta Vista en cas de couverture limitée.

Pour comparer : le cas de la linguistique pragmatique concerne une sphère Internet de projet, qui reflète la structure des sites Web construits autour de projets dans le domaine de la linguistique comparée, tandis que le cas de la théorie de l'argumentation s'appuie sur une sphère Internet d'école de pensée, qui reflète aussi sa présence sur le Net au sens large.

MESURE DE LA MUTUAL DEPENDENCE ET DU TASK UNCERTAINTY

+++++
Les degrés relatifs de *mutual dependence* et de *task uncertainty*, caractéristiques de l'identité culturelle de chaque discipline, ont été estimés grâce à des entretiens approfondis avec deux participants du projet CGN et un participant de l'École d'Amsterdam¹³ ainsi qu'à une analyse de littérature spécialisée par des participants à l'étude de cas. Ces dimensions de *mutual dependence* et *task uncertainty* au moyen desquelles notre étude a jugé les disciplines concernées incluent la taille de la communauté nationale et internationale, le degré du contrôle sur les divers canaux de diffusion, les genres prédominants de communication, la position stratégique au sein de la discipline mère, la présence d'un auditoire profane, des questions d'ordre pédagogique (telles que la discipline fait-elle partie d'un programme d'études ?), et le rôle des gardiens de l'information tels que les éditeurs.

Le tableau 1, inspiré de Whitley [1984], illustre les degrés relatifs de *mutual dependence* de la linguistique pragmatique et de la théorie de l'argumentation en s'appuyant sur ce qui a été découvert à propos de l'organisation intellectuelle et sociale de chaque discipline.

13. Voir les annexes I et II sur le détail des questions : [en ligne] : < http://cybermetrics.cindoc.csic.es/articulos/v10i1p2_Appendices.html >.

Tableau 1 : différences de degré de dépendance stratégique et fonctionnelle entre la linguistique pragmatique et la théorie de l'argumentation

		Dépendance fonctionnelle	
		Basse	Haute
Dépendance stratégique	Basse		<i>Linguistique pragmatique</i> : « Groupes de chercheurs spécialisés poursuivant des objectifs distincts au moyen de procédures standardisées. Importante coordination des résultats et haut degré de spécialisation des sujets, mais peu de considération pour la hiérarchie des objectifs. » [Whitley, 1984]
	Haute	<i>Théorie de l'argumentation</i> : « Écoles de recherche entretenant des liens étroits et poursuivant des objectifs distincts au moyen de procédures différentes. Haut degré de coordination au sein des écoles, mais faible entre elles. Importante compétition pour dominer la discipline. » [Whitley, 1984]	

Selon Whitley, un degré de dépendance fonctionnelle croissant débouche sur un haut degré de spécialisation et de différenciation des problèmes et des objectifs, et donc à la création de procédures spécifiques pour les gérer. Un haut degré d'interdépendance au sein d'une discipline spécialisée mène, quant à lui, à une importante coordination des résultats et des domaines spécialisés. À mesure que le degré de dépendance stratégique augmente, la coordination des problèmes, objectifs et procédures s'intensifie elle aussi. La compétition augmente également lorsque des scientifiques issus de différentes sous-disciplines tentent de convaincre leurs collègues de l'aspect crucial de leurs problèmes et de leurs objectifs pour leur domaine.

L'environnement hors ligne de la linguistique comparée indique un degré de dépendance fonctionnelle plus important que dans la théorie de l'argumentation. Ceci est sûrement dû en grande partie au fait que ses objectifs s'orientent vers la production, ce qui requiert l'intégration des techniques. Et cela va à l'encontre du pluralisme intellectuel des disciplines axées sur la théorie telle que la théorie de l'argumentation.

Généralement parlant, le chevauchement disciplinaire entre différents domaines d'effort intellectuel convergeant au sein du champ, tels que la linguistique computationnelle, la sociolinguistique, la phonologie et la lexicographie, est désormais considéré comme admis, non sans controverse préalable [Sampson, 2003 ; Lawler et Dry, 1998 ; Borsley et Ingham, 2002]. En conséquence, tandis qu'en théorie de l'argumentation, la rhétorique tend à avoir recours à des expressions conciliantes comme « se conformer » ou « envers », les textes de linguistique pragmatique tendent à employer des métaphores guerrières – « gagné », « bataille » – lorsqu'il s'agit de refléter la position du domaine au sein de la discipline mère [Sampson, 2003]. Cette donnée suggère que si le degré de dépendance fonctionnelle est plus important en linguistique pragmatique, celui de dépendance stratégique est plus bas en théorie de l'argumentation.

Certaines frontières intellectuelles et sociales de la linguistique pragmatique sont plus contestées que d'autres. L'interface avec la linguistique théorique est la frontière la plus controversée et la plus discutée dans la littérature de linguistique générale¹⁴. Dans le contexte plus général de la discipline mère, le degré de dépendance stratégique n'a pas encore été déterminé avec certitude, mais peut être considéré comme beaucoup plus faible que dans un champ bien établi tel que la physique des particules.

Depuis la fin des années 1990, la littérature produite par les membres de l'École d'Amsterdam contient la rhétorique d'intégration des deux principaux paradigmes contradictoires [Van Eemeren et Houtlosser, 2001]. Afin de déterminer le degré de *mutual dependence* dans ce domaine, la volonté de consensus décelée dans la littérature a été interprétée comme une indication que le degré de dépendance stratégique est relativement bas, mais que l'École d'Amsterdam et ses alliés travaillent à l'augmentation du degré de dépendance stratégique, en particulier entre les groupes, organisations et institutions qui représentent les diverses écoles de pensée au sein de la théorie de l'argumentation. Cela est nécessaire afin d'exploiter les ressources telles que la réputation, le public et le personnel, et de faire progresser le programme de recherche encore naissant vers l'institutionnalisation et la survie en tant que discipline.

Les remarques de Whitley sur les conséquences culturelles d'un degré de dépendance stratégique croissant trouvent un écho prolongé dans l'étude de cas de la théorie de l'argumentation. Au début des années 1990,

14. Voir Borsley et Ingham [2002] pour une critique typique dont la linguistique pragmatique fait l'objet.

les membres de l'École d'Amsterdam ont produit un certain nombre de monographies [Van Eemeren et Grootendorst, 1994] proposant un programme de recherche interdisciplinaire au sein de la théorie de l'argumentation, en se basant sur l'approche pragma-dialectique. La création de l'International Society for the Study of Argumentation (ISSA) en 1986 peut également être considérée comme un moyen de placer l'École d'Amsterdam, et donc l'étude de la théorie de l'argumentation telle qu'elle se pratique aux Pays-Bas, dans une position essentielle au sein de la communauté internationale de la théorie de l'argumentation. L'ISSA a également mis sur pied un programme de bourses, qui renforce le positionnement stratégique de l'École d'Amsterdam au sein de la discipline mère en termes de contrôle de réputation. Le financement par le Netherlands Institute for Advanced Study in the Humanities and Social Sciences d'un petit groupe de recherche axé sur « Des sophismes comme violations des règles du discours argumentatif » confirme d'autant plus l'impression d'une École d'Amsterdam entourée d'un réseau international aux mailles serrées. Parmi les experts de l'argumentation qui faisaient partie de ce programme figurent le président du département de la communication orale, de la théorie de l'argumentation et de la rhétorique ainsi que d'autres partisans importants de l'approche pragma-dialectique qui, eux-mêmes, étaient, pour beaucoup, membres de ce département à l'époque.

Tableau 2 : différences de degré d'« incertitude technique » et d'« incertitude stratégique » entre les deux études de cas

Technical task uncertainty		
	Bas	Haut
Strategic task uncertainty	Bas	
	Haut	<i>Linguistique pragmatique</i> : « Prévisibilité, stabilité et visibilité des résultats importantes. Consensus sur la façon d'interpréter et de coordonner ces résultats. Problèmes et objectifs assez divers, instables et non clairement organisés. »

Selon Whitley, un degré croissant d'« incertitude technique » débouche sur une organisation de travail des communautés intellectuelles caractérisée par une forte dépendance au contrôle direct et personnel du travail

effectué, ce qui mène à de trop grandes fluctuations dans les méthodes de travail et à un manque de fédération des objectifs de recherche au niveau international. L'« incertitude stratégique » inclut l'incertitude quant aux priorités intellectuelles, l'importance des sujets de recherche et les méthodes privilégiées de traitement de ces sujets, les récompenses potentielles de différentes stratégies de recherche en matière de renommée, et la pertinence des résultats pour les objectifs intellectuels collectifs [Whitley, 1984]. Un degré d'« incertitude stratégique » croissant débouche donc sur une plus grande dépendance à un petit groupe de collègues pour l'accès au matériel et à la renommée, d'où la nécessité d'aboutir à un consensus sur les priorités intellectuelles et une hiérarchie des sujets de recherche.

Le degré de « dépendance stratégique » de la linguistique pragmatique est mitigé par un degré d'« incertitude technique » relativement élevé. Les tâches qu'implique typiquement la construction d'un corpus incluent le plan, l'enregistrement, la transcription, l'indexation, l'annotation et la mise au point d'un logiciel d'exploitation. Il y a beaucoup de décisions à prendre sur la façon d'aborder ces tâches, et la littérature foisonne de débats sur la meilleure approche à adopter. Il existe, par exemple, plusieurs niveaux d'annotation de données empiriques : syntaxique, prosodique, ou encore phonétique, et chacun d'eux offre différentes possibilités analytiques aux futurs utilisateurs inconnus. Beaucoup de tâches nécessitent la création de critères techniques. En 2000, Atwell et ses collaborateurs ont mené une étude sur l'usage d'outils normés d'analyse grammaticale en linguistique et ont découvert une grande fluctuance dans l'emploi des schémas d'analyse.

La tendance vers un degré plus élevé de « dépendance stratégique » au sein de la communauté de la théorie de l'argumentation signifie que dans un mouvement simultané le degré d'« incertitude stratégique » diminuera.

PRINCIPAUX CONTRASTES DANS LES CARACTÉRISTIQUES DU CONTENU DES DEUX SPHÈRES INTERNET

+++++

Une différence capitale entre les sphères Internet du projet CGN et de l'École d'Amsterdam réside dans la taille en termes de : nombre d'URL, identification du contenu des URL, destination des liens sortants de

chaque URL centrale en termes de caractéristiques nationale (identifiés par le code pays et le nom du domaine de premier niveau – *country code Top Level Domain* ou ccTLD¹⁵), organisationnelle (reconnue, entre autres, grâce au nom du domaine de premier niveau – TLD¹⁶) et de contenu ; et l'origine des URL entrantes vers l'URL centrale en termes de caractéristiques nationale, organisationnelle et de contenu (la liste des URL qui constituent chacune des deux sphères Internet¹⁷).

Caractéristiques de la sphère Internet du CGN produite directement et indirectement

+++++
Au sein du site Web « officiel » du CGN, représenté par l'URL < <http://lands.let.ru.nl/cgn> >¹⁸, figuraient 55 pages contenant une description détaillée du projet, la justification de ses objectifs, des échantillons de données, des références bibliographiques, des rapports techniques et d'autres sur l'état d'avancement des recherches.

Des documents étaient également disponibles au format PDF, PPT et postscript. Le tableau 3 résume les principales dimensions de la sphère Internet du CGN.

Tableau 3 : résumé de la sphère Internet du CGN

Nombre de pages de l'URL centrale	55	
Nombre de liens sortant de l'URL centrale	8	
URL sortantes identifiées par les noms ccTLD et TLD	.nl	5
	.fr	1
	.de	1
	.it	1
	.at	1
Nombre de liens entrant vers l'URL centrale	39	

15. La liste des noms de domaine de premier niveau code pays est disponible sur < <http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/index.html> >.

16. De plus amples informations sur les noms de domaine de premier niveau sont disponibles sur < <http://www.icann.org/en/tlds/> > (consulté le 10 novembre 2011).

17. Voir les annexes III et IV pour la liste des URL : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

18. Le 3 mai 2004.

URL entrantes identifiées par les noms ccTLD et TLD	.nl	20
	.be	9
	.org	5
	.de	2
	.dk	1
	.ie	1
	.at	1
Nombre d'URL ajoutées par les participants		14
Nombre d'URL de la plateforme		62

Il y avait 8 liens sortants du site et 39 liens entrants (à l'exclusion des liens internes). Sur les 8 liens sortants, 4 appartenaient au ccTLD **.nl** :

- la Fondation néerlandaise pour la science (NWO), qui a financé le projet conjointement avec le gouvernement flamand ;
- l'Union de la langue néerlandaise (Nederlandse Taalunie), à qui revient la propriété du corpus ;
- l'intranet du CGN hébergé par l'université d'Utrecht ;
- PRAAT, outil phonétique utilisé pour le projet et mis au point par le département de phonétique de l'université d'Amsterdam.

Les 4 liens sortants restants étaient des noms ccTLD européens :

- ELDA (Evaluations and Language Resources Distribution Agency), basée en France et en charge de la diffusion du corpus du CGN ;
- EAGLES, norme internationale du génie de la langue basée en Italie ;
- Negr@corpus, corpus allemand de textes de presse ;
- le dernier, enregistré **.at (Autriche)**, n'est plus en service et redirige l'internaute vers Nederlandse Taalunie ; le lexique de l'URL suggère également que la destination du lien était à l'origine néerlandaise.

Sur une période d'un an, d'août 2003 à juillet 2004, 39 URL entrantes vers l'URL du projet CGN ont été identifiées grâce à l'outil de recherche

de Google.nl. Vingt de ces URL appartenait au ccTLD.**nl**¹⁹. Ces URL incluent des sous-pages de la page « officielle » du CGN produites par des institutions partenaires du projet, des pages créées par des utilisateurs du corpus, des assistants-chercheurs d'anciens projets, le département du langage de l'université catholique de Nimègue où travaille le chef de projet de la section néerlandaise du corpus, des listes de liens qui dirigent les gens vers des ressources sur la technologie du langage, des ateliers de recherche, l'Union de la langue néerlandaise (Nederlandse Taalunie) et la Fondation néerlandaise pour la science (NWO).

Sur les 19 URL entrantes en dehors du nom ccTLD.**nl**, 9 appartenait au ccTLD.**be**. Les URL du domaine **.be** représentent en partie la participation des partenaires du projet basés en Belgique en charge de la section flamande du corpus. Toutefois, les URL des principaux acteurs belges impliqués dans la mise au point et l'utilisation du corpus ne sont pas représentées dans la sphère Internet produite de façon directe. Ces URL ont été ajoutées à la sphère Internet étendue produite de façon indirecte par le chef de la section flamande du projet au cours de l'étape de validation de la construction de la sphère Internet.

Cinq des URL entrantes appartenait au nom de domaine **.org** et étaient hébergées par deux organisations néerlandaises : DBNL (Digitale Bibliotheek voor de Nederlandse) (1 lien) et l'Union de la langue néerlandaise (Nederlandse Taalunie) (3 liens). Le dernier lien provenait de l'HLT (Human League Technologies), basé en Europe.

Les 5 URL entrantes restantes étaient aussi hébergées en Europe. Deux sous le nom de domaine **.de** : une page sur la philologie aux Pays-Bas hébergée par l'université libre de Berlin et comportant du contenu coécrit par le chef de projet en charge de la section néerlandaise ; une autre hébergée par un centre d'étude de la langue néerlandaise basé à Leipzig. Le lien entrant appartenant au domaine **.dk** représentait le site Web Danish Dependency Treebank, le *treebank* étant un arbre syntaxique, une technique d'analyse utilisée par le CGN.

19. Voir l'annexe III pour une liste complète des URL entrantes et des intitulés de pages : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

Caractéristiques de la sphère Internet de l'École d'Amsterdam produite directement et indirectement

+++++

- Au cours de la période d'étude, l'URL centrale de l'École d'Amsterdam ne contenait que 3 pages (une page d'accueil décrivant les objectifs de l'ISSA, une page contenant des détails sur la conférence annuelle, et une page recensant les récipiendaires de la bourse de l'ISSA) et 2 liens sortants (un pointait vers un associé de l'École d'Amsterdam basé dans le département de communication de l'université de Louisville [.edu], et l'autre vers la revue *Argumentation: an International Journal on Reasoning*, dirigée par le président de l'École d'Amsterdam et publiée par Kluwer, maison basée aux Pays-Bas [.nl]²⁰). Malgré le manque de contenu en rapport avec la recherche, d'informations concernant les cours, d'articles et de dossiers téléchargeables, l'URL de l'ISSA profitait de 17 liens entrants. En général, les pages Web universitaires bénéficient de peu de liens entrants [Thelwall et Harries, 2004] ; c'est pourquoi il est surprenant que l'URL de l'ISSA compte plus de liens entrants que de pages de contenu. Le tableau 4 résume les principales dimensions de la sphère Internet de l'École d'Amsterdam.

Tableau 4 : résumé de la sphère Internet de l'École d'Amsterdam

Nombre de pages de l'URL centrale	3	
Nombre de liens sortant de l'URL centrale	2 (en rapport avec la recherche) 36 (informations touristiques)	
URL sortantes identifiées par les noms ccTLD et TLD	.edu	1
	.nl	1
Nombre de liens entrant vers l'URL centrale	17	

20. Trente-six liens sortants figurant sur les pages de l'ISSA étaient sans rapport avec la recherche, il s'agissait de renseignements touristiques destinés aux participants à la conférence annuelle de l'ISSA.

URL entrantes identifiées par les noms ccTLD et TLD	.edu	5
	.com	4
	.ca	2
	.ru	1
	.org	1
	.nl	1
	.uk	1
	.ch	1
	.de	1
Nombre d'URL ajoutées par les participants	6	
Nombre d'URL de la plateforme	28	

Par opposition à l'orientation nationale de la sphère Internet du CGN, 16 des liens entrants de la sphère Internet de l'École d'Amsterdam émanaient de l'extérieur du domaine **.nl** et étaient de nature internationale. L'unique lien entrant émanant du domaine **.nl** appartenait au site Web d'un membre actif de la liste de discussion ARGTHRY (principale liste de discussion de la communauté de la théorie de l'argumentation) ayant créé un site Internet personnel conséquent consacré à l'étude de la théorie de l'argumentation. Le contenu de la page d'accueil du lien entrant a trait au 9^e Atelier international sur le raisonnement non monotone et le lien référençant la conférence de l'ISSA de 2002 est intégré dans une liste de liens en rapport avec le sujet.

D'autres pages extérieures au domaine **.nl** qui pointaient vers le site de l'ISSA incluent : "Philosophy resources on the Internet: EpistemeLinks" (**.com**), site commercial de vente de produits en rapport avec la philosophie ; "Rhetoric resources on the Web" (**.edu**), qui fait partie des pages d'accueil d'un membre du département d'anglais de l'université de Wisconsin-Madison ; "Tim Gelder's critical thinking on the Web: Institutes, Centers and Societies" (**.org**) ; "The Reasoning Page" (**.edu**), liste annotée de ressources en ligne dans l'étude de l'argumentation.

Le site Web de Vale Press (**.com**) pointait 3 liens sortants vers l'URL de l'ISSA. Vale Press est un éditeur spécialisé dans le champ de la pensée critique. Ils publient notamment la collection « SIC SAT » en théorie de l'argumentation, support de publication des monographies écrites par les membres de l'École d'Amsterdam. Les liens sortants qui pointaient vers le site de l'ISSA apparaissaient à deux reprises sur la page d'accueil de Vale

Press : dans le calendrier, avec la référence à la conférence annuelle de l'ISSA, et dans la liste des liens les plus populaires. Ces liens étaient tous les deux intégrés à proximité de 2 liens dirigeant vers l'Association pour la logique informelle et la pensée critique (AILACT), qui possède sur son site un lien pointant vers la revue de l'ISSA. Le troisième lien sortant figurant sur le site Web de Vale Press pointant vers l'URL de l'ISSA était intégré à une liste de titres de monographies.

Il y avait 5 liens entrants émanant du domaine **.edu** (dont deux ont déjà été décrits plus haut). L'un d'eux émanait du site Web de la Stanford Encyclopedia of Philosophy sous l'entrée *Informal Logic* (logique informelle). Les 2 URL entrantes restantes du domaine **.edu** sont des pages personnelles d'universitaires des départements de rhétorique ou de communication et apparaissent en tant que ressources utiles, d'une part, et associations et conférences en rapport avec la communication, d'autre part.

Il y avait 2 URL entrantes appartenant au domaine **.ca** (Canada). L'une dirigeait vers le site Web du *Journal of Informal Logic: Reasoning and Argumentation in Theory and Practice*, dans une liste de liens pointant vers d'autres revues et ressources, qui contient également un lien sortant vers l'URL de la revue de l'ISSA. Le second lien entrant du domaine **.ca** pointait vers la Wetaskiwin Telephone Company, qui contient un lien vers un article publié dans les actes de la cinquième conférence de l'ISSA, à Amsterdam, du 25 au 28 juin 2002.

Un membre de l'École d'Amsterdam publie en ligne une revue intitulée *Argumentation, Interpretation and Rhetoric*, en russe et en anglais. L'Institut néerlandais de Saint-Petersbourg est le lien qui relie l'École d'Amsterdam à la Russie et l'URL entrante du domaine **.ru** (Russie) appartenait à une page d'archives référençant la revue *Argumentation, Interpretation and Rhetoric*.

Un ancien étudiant de troisième cycle du département de la communication orale, de la théorie de l'argumentation et de la rhétorique avait créé un lien entrant dans ses pages personnelles, appartenant au domaine **.de** (Allemagne). Le domaine **.uk** (Royaume-Uni) comptait un lien entrant, figurant dans les pages Web d'un atelier de pré-conférence de l'ECIA 2002 (Conférence européenne sur l'intelligence artificielle) sur les modèles computationnels de l'argumentation naturelle. Le dernier lien entrant, du domaine **.ch** (non listé dans la liste des codes pays ISO 3166), n'a en réalité qu'un lien trivial avec les activités de l'École d'Amsterdam, étant donné qu'il figure sur les pages d'un centre international de la santé et de la coopération, listé sous diverses associations industrielles et universitaires.

Les URL ajoutées par le participant à l'étude de cas sont toutes en rapport avec les ateliers et conférences majeurs sur la théorie de l'argumentation auxquels les membres de l'École d'Amsterdam ont participé. Étant donné que les pages Web de Kluwer, ancien éditeur de la revue de l'ISSA, étaient intimement liées à celles du site de l'ISSA dans la mesure où il existait un lien réciproque entre elles et où les éditeurs jouent un rôle primordial dans la sphère Internet de l'École d'Amsterdam, l'on a également procédé à une recherche de liens entrants à partir de l'URL de la revue < <http://www.kluweronline.com/issn/0920-427X> >²¹ en utilisant Google.nl²². Les 23 liens entrants pointant vers la revue de l'ISSA, tous extérieurs au domaine .nl, ne seront pas évoqués en détail mais fournissent un contexte plus large et illustrent les différences qui existent lorsqu'une sphère Internet est construite avec pour unité d'analyse différentes représentations d'une communauté universitaire. La figure 1 résume les principaux aspects des sphères Internet de chaque étude de cas :

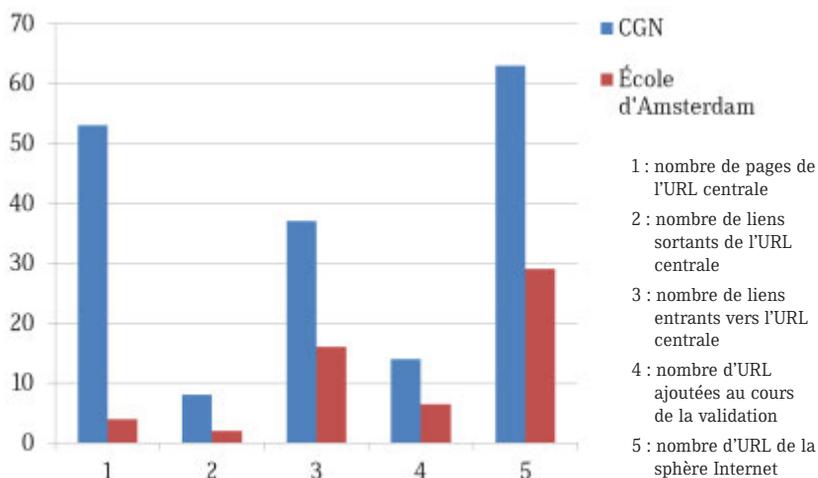


Figure 1 : comparaison des principaux aspects des sphères Internet du CGN et de l'École d'Amsterdam

21. N.D.T. : voir note 8.

22. Voir l'annexe IV pour une liste des URL entrantes : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10ilp2_Appendices.html >.

Présence sur le Web : une sphère Internet potentielle

+++++

La présence sur le Web de chaque étude de cas telle que représentée sur Google.nl a fourni un contexte lié à la discipline. En tant que telle, elle donne une image de ce que pourrait être la sphère Internet de chaque étude de cas, en termes de nombre d'URL et de genres prédominants, bien plus importante dans les deux cas que la sphère Internet existante. Par exemple, la recherche de l'expression exacte « corpus gesproken nederlands » aboutit à environ 900 URL, alors que le nombre d'URL dans la sphère Internet du CGN n'atteint que 63. De même, la recherche du mot « pragma-dialectique », nom familier et champ de recherche de l'École d'Amsterdam, sur Google.nl a abouti à 860 occurrences, comparé à 28 URL sur la sphère Internet. Cet important contraste soulève la question de la pertinence du concept de sphère Internet comme outil d'étude de la production du Web universitaire. Les recherches les plus récentes ont été menées le 1^{er} mai 2005 et les 30 premières occurrences sont recensées dans les annexes V et VI²³.

Il existe des problèmes de comparabilité entre les représentations numériques des deux disciplines obtenues par le biais de moteurs de recherche, pour des raisons linguistiques [Vaughan et Thelwall, 2004 ; Bar-Ilan, 2005], et les problèmes associés aux requêtes multilingues [Moukdad, 2002]. Pour le projet du CGN, les résultats sont différents lorsque l'on entre la traduction anglaise de « corpus de néerlandais parlé » dans la barre de recherche, mais étant donné la faible représentation de l'anglais dans la sphère Internet du CGN, il a été estimé plus pertinent d'avoir recours au néerlandais. Cependant, malgré l'aspect rudimentaire de la méthode, nous avons pu observer que le nombre d'URL faisant référence dans le contenu de leurs pages au projet du CGN ou aux membres de l'École d'Amsterdam était plus important que celles qui contenaient véritablement un lien sortant pointant vers l'URL centrale de chaque étude de cas. Cette donnée souligne les limites de la prise en compte exclusive des liens hypertextes comme critère de l'interconnectivité entre les acteurs en jeu.

Un échantillonnage manuel des 30 premiers et 10 derniers résultats de chaque série de recherches a été réalisé afin de vérifier la pertinence des occurrences. On a jugé de l'identification de chaque URL en s'appuyant sur le résumé informationnel contenu dans la structure de l'URL, comme

23. Voir les annexes V et VI : [en ligne] < http://cybermetrics.cindoc.csic.es/articles/v10i1p2_Appendices.html >.

les caractéristiques de l'institution, le titre de la page, et le contenu de pages lui-même lorsque les métadonnées ne fournissaient pas suffisamment de renseignements pour pouvoir estimer l'identification.

Dans le cas du *Corpus Gesproken Nederlands* (CGN), ce sont des liens pointant vers des ressources telles que des corpus, des outils, des chercheurs, des organisations, des mémoires, des rapports et du matériel pédagogique qui arrivaient en tête des résultats de recherche. Les librairies et bibliothèques en ligne représentant les publications de la linguistique pragmatique n'apparaissaient qu'après les 60 premiers résultats. Ce qui contraste avec les recherches de la « pragma-dialectique », où les équivalents numériques des monographies papier arrivaient en tête de liste.

L'identification des URL obtenues pour les requêtes de la théorie de l'argumentation s'orientait vers les versions numériques de monographies trouvées sur les pages Web d'éditeurs, de librairies électroniques et de bibliothèques en ligne. Elles comprenaient celles écrites ou dirigées par des membres de l'École d'Amsterdam.

TRACES DE MUTUAL DEPENDENCE ET DE TASK UNCERTAINTY DANS LA REPRÉSENTATION NUMÉRIQUE DES DISCIPLINES INTELLECTUELLES

++++
Il est donc clair que la production numérique au sein de la linguistique pragmatique et de la théorie de l'argumentation possède des caractéristiques bien distinctes. Dans l'étude de cas du CGN, nous observons une sphère Internet locale et décentralisée (locale dans le sens où pages et sites sont organisés au niveau du projet et non de la discipline ; décentralisée dans la mesure où l'impératif organisationnel de la sphère Internet obéit davantage à la notion de dépendance fonctionnelle qu'à celle de dépendance stratégique), qui s'adresse en conséquence à un public national de chercheurs, du fait que la discipline ne soit pas dominée par une approche unique et ne soit pas régie par une hiérarchie de problèmes. Cela signifie qu'il est moins nécessaire de démontrer au domaine scientifique l'importance des problèmes, techniques et aboutissements en matière de recherche. Dans le cas de la théorie de l'argumentation, en revanche, la production numérique est centralisée au niveau de la discipline, avec des pages et des sites organisés autour de l'activité professionnelle des écoles de recherche et des communautés de chercheurs représentant un paradigme dominant. L'impératif organisationnel est centralisé et s'appuie ici sur l'idée de « dépendance stratégique », où ce sont la hiérarchisation des

questions à traiter et l'édification d'une réputation qui priment. La sphère Internet du CGN a, par exemple, une orientation nationale très marquée, avec un nombre négligeable d'URL extérieures au domaine **.nl**. La sphère Internet de l'École d'Amsterdam a, au contraire, une orientation internationale, avec une majorité d'URL extérieures au domaine **.nl**.

En termes de public scientifique hors ligne pour le CGN, les résultats de recherche ont été disséminés au fil de conférences internationales et de revues spécialisées, et des articles publiés par des membres du projet ont été cités dans des revues répertoriées par l'ISI Web of Knowledge. Toutefois, la linguistique pragmatique est dominée par la recherche en langue anglaise, et l'analyse hypertextuelle de la sphère Internet du CGN semble refléter cette donnée. Des corpus de langue très influents, tels que le corpus national britannique ou le corpus international d'anglais, n'ont pas sur leur site Internet de lien pointant vers l'URL du projet CGN. Et bien que le contenu du projet CGN évoque le corpus national britannique sur la question des critères internationaux à atteindre dans la création des corpus, aucun lien ne pointe vers ce dernier, ni vers aucun autre corpus national.

Cela peut constituer un autre signe de haut degré de *strategic uncertainty* pour le projet CGN, dans le sens où il n'est pas nécessaire que les autres acteurs de la linguistique pragmatique s'alignent sur ce projet pour gagner en renommée ou en légitimité dans la discipline. L'inverse se vérifie dans le cas de la sphère Internet de l'École d'Amsterdam, où la grande proportion de liens entrants comparée à la quantité relativement limitée de pages et de contenu (en termes de diversité des types de dossiers, outil utilisé par la mesure de la représentativité sur Internet) indique la position politique dominante du groupe : les autres chercheurs doivent reconnaître le groupe afin d'être crédibles. Thelwall [2006] a découvert que plus une page Internet est la cible de liens, plus il sera probable qu'elle soit ciblée par des liens nouvellement créés. Donnée qui fait écho au schéma de la naissance de l'estime que Merton [1968] a décrit sous la dénomination « effet saint Matthieu ».

Bénéficier d'une forte représentativité numérique peut s'avérer important pour les groupes appartenant à une discipline émergente, où ils ont besoin d'accroître leur degré de *strategic dependence*, comme c'est le cas pour l'École d'Amsterdam. Si, en revanche, leurs objectifs intellectuels sont mieux servis par une communication classique, comme c'est souvent le cas dans les domaines théoriques, qui nécessitent de longues explications de concepts et d'arguments, alors la communication de style

monographique [Swales, 1998] tend à être le mode de communication dominant. Cela signifie que la représentation numérique d'un courant ou d'une discipline peut être dominée par les gardiens de la communication publiée de façon classique, tels que les bibliothèques et les éditeurs, plutôt que par la communication informelle orientée vers la recherche – comme on le voit dans la prédominance des éditeurs et bibliothèques en ligne dans la sphère Internet de l'École d'Amsterdam. Il est intéressant de constater que bien que l'École d'Amsterdam n'ait pas produit directement une sphère Internet hypertextuelle, sa position stratégique dans le monde hors ligne est reflétée par une sphère Internet étendue produite indirectement comme l'indique le nombre important de liens entrants vers les pages Web de l'ISSA. Le fait que ce site ne contienne que trois pages de contenu en rapport avec la recherche (le reste du site étant consacré au tourisme local et aux informations pratiques concernant le logement en vue de la conférence de l'ISSA), sans lien sortant vers d'autres groupes de recherche ou organisations, et dont les liens entrants dépassent le nombre total de pages et de liens sortants, démontre la position politique dominante du groupe : les autres chercheurs doivent s'aligner sur le groupe afin d'être crédibles, et non créer seulement un lien informationnel [Rousseau, 1999].

Tout comme la sphère Internet de l'École d'Amsterdam, le projet CGN est présent dans des sources secondaires d'information telles que portails thématiques et bulletins d'information. Les deux disciplines ont un public profane, mais en raison de la spécificité de la langue dans le projet CGN et le jargon technique abscons inhérent au domaine lui-même, il est probable que le public profane soit restreint. Lorsqu'on lui a demandé à qui se destinaient les résultats des recherches effectuées par son groupe, un membre de l'École d'Amsterdam a répondu qu'en plus des universitaires des domaines de la logique, de la rationalité, de la rhétorique, de la persuasion et de la dialectique, les étudiants de deuxième et troisième cycles étaient également importants, tout comme les non-universitaires intéressés par les rapports de science vulgarisés.

La communication informelle est un mode de communication dominant dans le milieu de la linguistique pragmatique, où, dans les domaines techniques de la discipline tels que la technologie du discours, le modèle de communication s'apparente davantage à la science computationnelle. Beaucoup d'informations techniques sont communiquées sous la forme de rapports sur l'avancée des recherches, de manuels et de comptes rendus de conférences, et les traces de ces documents non publiés hors ligne

sont très visibles sur Internet. Ceci est également dû au fait que le degré de « dépendance fonctionnelle » est plus élevé qu'en théorie de l'argumentation, ce qui signifie qu'il y a un plus grand besoin de coordination des techniques et des résultats à l'échelle des groupes, de la discipline et du pays. La prédominance du nom de domaine .nl est surprenante au regard de cette interdépendance, surtout dans la mesure où des membres du groupe avaient publié des articles dans des revues internationales et donné des conférences internationales sur la création du corpus. Inversement, bien que le corpus national britannique soit référencé dans des publications et articles produits dans le contexte du projet, il n'existait aucun lien pointant vers ses pages Web. Une attention particulière a en revanche été accordée à la visibilité de références internationales telles qu'EAGLES et PRAAT *via* des liens sortants, afin de s'attribuer une certaine autorité sur le sujet. Un corpus peut s'envisager à la fois comme outil et comme sujet de recherche. Pour les développeurs de corpus tels que le CGN, le projet constitue un sujet de recherche, qui soulèvera des questions de recherche, mais pour ceux qui utiliseront ce corpus, il deviendra un outil pour répondre à des questions. Peut-être qu'à mesure que le CGN deviendra un outil, les réseaux qui se sont formés autour de lui le placeront plus au centre de la scène internationale de la linguistique pragmatique. Il est également possible que sa focalisation sur la langue néerlandaise l'enferme dans un réseau national, malgré sa nouveauté et son intérêt technique potentiel en tant que rare exemple de corpus oral.

Dans la sphère Internet du CGN, la recherche menée par les membres du projet à titre individuel est noyée par la recherche collective, à laquelle on donne l'avantage. Par exemple, le chef du projet et développeur du site officiel du CGN n'a intégré aucun lien sortant vers les pages personnelles des membres du projet, ni même vers la sienne. Bien qu'il existe plusieurs liens pointant vers l'URL centrale du projet sur les sites Web d'anciens étudiants et d'assistants-chercheurs, aucun lien n'émane des pages Web des universitaires reconnus qui ont participé au projet.

Les contributions personnelles des membres de l'École d'Amsterdam sont, elles aussi, submergées par le flot d'informations concernant la recherche collective. Les activités des membres de l'École d'Amsterdam en tant qu'enseignants sont plus visibles que leurs activités de recherche. Il existe diverses raisons capables d'expliquer pourquoi si peu de pages Web sont consacrées aux ressources de la recherche dans le domaine de la théorie de l'argumentation. Il se peut que le degré élevé de « dépendance stratégique », entraînant une compétitivité accrue, dissuade les

chercheurs de partager leurs informations et leurs données. À l'inverse, le degré plus bas de « dépendance fonctionnelle » peut signifier qu'il est moins impératif de partager les ressources.

Des degrés relatifs de *mutual dependance* et de *task uncertainty* peuvent avoir une influence importante sur la taille des sphères Internet (en termes de nombre d'URL), car un degré décroissant de *mutual dependence* (tant fonctionnelle que stratégique) indique que le besoin de coordination et de partage des ressources est moindre. Cela peut se combiner à un degré croissant de *task uncertainty*, ce qui rend la coordination des problèmes, techniques et résultats de recherche plus difficile. Ces différences culturelles rendent problématique la création d'indicateurs numériques standardisés visant à mesurer la productivité des disciplines.

Ces résultats indiquent aussi que lorsque l'on a recours à différentes unités d'analyse, différents niveaux de prééminence et de connectivité peuvent être observés. Dans le cas de l'École d'Amsterdam, s'appuyer sur la communauté professionnelle pour créer une sphère Internet révèle des indications d'évaluation par ses pairs, d'estime et de pouvoir, alors que ce ne serait pas le cas pour une sphère Internet construite autour des pages de leur département.

CONCLUSION

+++++

Comme l'a démontré cet article, les différences entre ce que les disciplines rendent visible sur Internet et ce qui demeure invisible sont très révélatrices. La théorie de Whitley [1984] s'est avérée très utile pour la compréhension de ces différences. Par exemple, dans le cas de l'École d'Amsterdam, le taux important de liens pointant vers le site de l'ISSA peut être interprété comme un degré croissant de « dépendance stratégique » (avec ses exigences inhérentes en matière d'alliances et de construction d'une communauté), alors que le CGN ne partage pas cette position stratégique centrale au sein de sa discipline, ce qui pourrait expliquer le taux relativement bas de liens pointant vers le site Web du projet. En termes de contenu des URL de chaque sphère Internet, le plus haut degré de *strategic uncertainty* au sein du projet CGN, qui suit une tendance générale de la discipline, est reflété par l'emploi d'un langage justificatif pour le plan du corpus et par la façon ostentatoire dont les liens sortants pointant vers des références techniques internationales sont intégrés dans cette justification.

Voici quelques aspects particuliers de l'identité culturelle d'une discipline intellectuelle qui pourraient apporter une contribution importante à la création d'un cadre théorique pour l'interprétation des relations hypertextuelles ainsi qu'un moyen d'identifier les unités d'analyse adéquates :

- objectifs de la communauté de chercheurs, en termes de théorie, production, recueil et partage des données, etc. ;
- méthodes, techniques et outils ;
- public : étudiants de deuxième cycle, public profane, universitaires issus d'autres disciplines, etc. ;
- maturité de la discipline ;
- orientation nationale ou internationale de la discipline ;
- envergure de la discipline.

Capables de conférer un pouvoir interprétatif aux phénomènes sociaux, intellectuels et institutionnels (voir également Beaulieu et Simakova [2006] sur le sujet), les études hypertextuelles ethnographiques sont d'autant plus précieuses qu'elles permettent d'identifier une unité d'analyse plus subtile pour l'étude des différences qui opposent les disciplines en matière de production et de pratiques numériques. Il est toutefois nécessaire d'augmenter proportionnellement ces études afin que le cœur pédagogique ou les activités de recherche des disciplines intellectuelles soient immédiatement identifiables et que des schémas comportementaux plus généraux se fassent jour. Un faible taux de liens universitaires dans la littérature, en particulier en sciences humaines, peut être le résultat d'un problème de méthodologie plutôt que le signe d'une déficience au niveau de la superstructure des universités [Harries *et al.*, 2004] ou de la recherche institutionnelle. Afin d'éviter tout contresens dans l'interprétation des résultats, il est indispensable, au niveau méthodologique, d'être clair sur l'unité d'analyse employée – bien faire la distinction notamment entre département universitaire et discipline afin d'éviter tout amalgame. Si, par exemple, une étude exclusivement quantitative des relations hypertextuelles aux Pays-Bas prenait les départements universitaires pour nœud central, alors les pages Web de l'École d'Amsterdam < <http://cf.hum.uva.nl/data/afd/neerlandistiek/tar-english> >²⁴ feraient apparaître l'École comme un groupe isolé (en raison de l'absence d'URL entrantes ou sortantes). Dans la théorie des réseaux, l'étendue de l'interconnectivité entre les nœuds (par exemple une URL comptant beaucoup de liens entrants) est utilisée pour interpréter des facteurs sociaux tels que la dose de critique

24. N.D.T. : voir note 7.

que s'autorise un chercheur ou un groupe de chercheurs, faisant autorité en tant que producteurs de connaissances, à l'égard d'une communauté scientifique. Utilisé comme indicateur de performance quantitative, le modèle hypertextuel qui se dégagerait de l'École d'Amsterdam en tant que département serait très loin de refléter l'envergure internationale du groupe et le rôle capital qu'il joue au sein de la théorie de l'argumentation.

Les résultats exposés dans cet article démontrent que des approches méthodologiques de l'étude des liens hypertextes et des modèles hypertextuels différentes aboutissent à des représentations des domaines de recherche et des communautés universitaires qui les peuplent différentes. Dans notre cas, c'est le recours aux techniques ethnographiques (entretiens et analyse du discours) allié à l'utilisation d'outils tels que le moteur de recherche de Google et Microsoft Site Analyst (qui s'appuient tous les deux sur l'exploration automatique du Web), qui ont fourni un apport particulier au concept de sphère Internet : celui de rendre les traces de l'identité intellectuelle et sociale de disciplines universitaires spécialisées plus visibles, sur l'Internet s'entend. Dans une étude à plus grande échelle que celle rapportée ici, il aurait peut-être été impossible de trianguler autant de techniques différentes, d'où l'importance de rester vigilant et de bien comprendre les limites que comporte toute approche lorsqu'il s'agit d'interpréter les résultats.

REMERCIEMENTS

++++++
Cet article s'appuie sur des recherches que j'ai menées au sein du groupe Networked Research and Digital Information (NERDI) à l'Institut néerlandais des sciences de l'information et a été financé par l'Académie royale des arts et des sciences néerlandaise. J'aimerais remercier mes collègues du NERDI (devenu le Virtual Knowledge Studio des sciences humaines et sociales) pour leur soutien inestimable au cours de ces recherches et leurs commentaires avisés sur les versions préliminaires de cet article. Je voudrais aussi remercier mes deux correcteurs, le professeur Peter van den Besselaar et le D^r Viv Cothey, dont les commentaires détaillés m'ont permis d'aboutir à une version définitive améliorée.

RÉFÉRENCES BIBLIOGRAPHIQUES

++++++
Atwell Eric, Demetriou George, Hughes John, Schiffirin Amanda, Souther Clive et Wilcock Sean, "A comparative evaluation of modern English

corpus grammatical annotation schemes”, *ICAME Journal*, n° 24, 2000, pp. 7-23.

Bar-Ilan Judit, “How do search engines respond to some non-English queries?”, *Journal of Information Science*, 31 (1), 2005, pp. 13-28.

Beaulieu Anne et Simakova Elena, “Textured connectivity: an ethnographic approach to understanding the timescape of hyperlinks”, *Cybermetrics*, 10 (1), 2006.

[En ligne] < <http://cybermetrics.cindoc.csic.es/articles/v10i1p5.html> > (consulté le 3 juin 2004).

Becher Tony, *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*, Buckinghamshire, Open University Press, 1989.

Björneborn Lennart et Ingwersen Peter, “Toward a basic framework for webometrics”, *Journal of the American Society for Information Science and Technology*, 55 (14), 2004, pp. 1216-1227.

Borsley Robert D. et Ingham Richard, “Grow your own linguistics? On some applied linguists’ views of the subject”, *Lingua*, 112 (1), 2002, pp. 1-6.

Bourdieu Pierre, *Homo Academicus*, Traduit par Peter Collier, Stanford (Ca), Stanford University Press, 1988.

Bruckner Eberhard, Ebeling Werner et Scharnhorst Andrea, “Stochastic dynamics of instabilities in evolutionary systems”, *Systems Dynamics Review*, 2, 1989, pp. 176-191.

Crane Diana, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, London, Chicago, The University of Chicago Press, 1972.

Cronin Blaise, Snyder Herbert W., Rosenbaum Howard, Martinson Anna et Callahan Ewa, “Invoked on the Web”, *Journal of the American Society for Information Science*, 49 (14), 1998, pp. 1319-1328.

Eemeren Frans H. van et Grootendorst Rob, *Argumentation, Communication, and Fallacies: A Pragma-Dialectical Perspective*, Mahwah, New Jersey, Lawrence Erlbaum Associates, 1992.

Eemeren Frans H. van et Grootendorst Rob (dir.), *Studies in Pragma-Dialectics*, Amsterdam, Sic Sat, International Centre for the Study of Argumentation, 1994.

Eemeren Frans H. van et Houtlosser Peter, "Rhetoric in pragma-dialectics", *Argumentation, Interpretation and Rhetoric* [Electronic journal University of St. Petersburg & University of Amsterdam], 1 (1), 2001.

Fry Jenny, "Scholarly research and information practices: a domain analytic approach", *Information Processing and Management*, 42 (1), 2006, pp. 299-316.

Gibbons Michael, Limoges Camille, Nowotny Helga, Schwartzman Simon, Scott Peter et Trow Martin, *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London, Sage, 1994.

Harries Gareth, Wilkinson David, Price Liz, Fairclough Ruth et Thelwall Mike, "Hyperlinks as a data source for science mapping", *Journal of Information Science*, 30 (5), 2004, pp. 436-447.

Haythornthwaite Caroline et Wellman Barry, "Work, friendship and media use for information exchange in a networked organization", *Journal of the American Society for Information Science*, 49 (12), 1998, pp. 1101-1114.

Heilbron Johan, "A regime of disciplines: toward a historical sociology of disciplinary knowledge", in Camic Charles et Joas Hans (dir.), *The Dialogical Turn: New Roles for Sociology in the Post-Disciplinary Age*, Lanham, Rowman & Littlefield Pub Inc, 2004, pp. 23-42.

Heimeriks Gaston, Hörlesberger Marianne et Besselaar Peter van den, "Mapping communication and collaboration in heterogeneous research networks", *Scientometrics*, 58 (2), 2003, pp. 391-413.

Hine Christine, *Virtual Ethnography*, London, Sage, 2000.

Kling Rob et McKim Geoffrey, "Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication", *Journal of the American Society for Information Science*, 51 (14), 2000, pp. 1306-1320.

Kochen Manfred, "Mathematical model for the growth of two specialties", *Science of Science*, 3 (11), 1983, pp. 199-217.

Lawler John M. et Dry Helen Aristar (dir.), *Using Computers in Linguistics: a Practical Guide*, London, Routledge, 1998.

Lenoir Timothy, *Instituting Science: The Cultural Production of Scientific Disciplines*, Stanford (Ca), Stanford University Press, 1997.

Marcus George E., "Ethnography in/of the world system: the emergence of multi-sited ethnography", *Annual Review of Anthropology*, 24, 1995, pp. 95-117.

Merton Robert K., "The Matthew effect in science", *Science*, 159 (3810), 1968, pp. 56-63.

Moukdad Haidar, "Language-based retrieval of Web documents: an analysis of Arabic-recognition capabilities of two major search engines", *Proceedings of the American Society for Information Science and Technology*, 39 (1), 2002, p. 551.

Price Derek J., *Little Science, Big Science*. New York, Columbia University Press, 1963.

Rousseau Ronald, "Sitations: an exploratory study", *Cybermetrics*, 1 (1), 1997.

[En ligne] < <http://cybermetrics.cindoc.csic.es/articles/v1i1p1.html> > (consulté le 21 juin 2004).

Rousseau Ronald, "Daily time series of common single word searches in Alta Vista and NorthernLight", *Cybermetrics*, 2/3 (1), 1999.

[En ligne] < <http://www.cindoc.csic.es/cybermetrics/vol2iss1.html> > (consulté le 10 mai 2005).

Salter Liora et Hearn Alison, *Outside the Lines: Issues in Interdisciplinary Research*, Montréal, McGill-Queen's University Press, 1996.

Sampson Geoffrey, *Are we nearly there yet, Mum?*. Corpus Linguistics 2003 conference, Lancaster, mars 2003.

[En ligne] < <http://www.grsampson.net/index.html> > (consulté le 3 janvier 2004).

Schneider Steven M. et Foot Kirsten A., "Online structure for political action: exploring presidential Web sites from the 2000 American election", *Javnost (The Public)*, 9 (2), 2002, pp. 43-60.

Swales John M., *Other Floors, other Voices: A Textography of a Small University Building*, London, Lawrence Erlbaum Associates, 1998.

Thelwall Mike, "Interpreting social science link analysis research: a theoretical framework", *Journal of the American Society for Information Science and Technology*, 57 (1), 2006, pp. 60-68.

Thelwall Mike et Harries Gareth, "Can personal Web pages that link to universities yield information about the wider dissemination of research?", *Journal of Information Science*, 30 (3), 2004, pp. 243-256.

Vaughan Liwen et Thelwall Mike, "Search engine coverage bias: evidence and possible causes", *Information Processing and Management*, 40 (4), 2004, pp. 693-707.

Wenger Étienne, *Communities of Practice: Learning, Meaning and Identity*, Cambridge, Cambridge University Press, 1998.

Whitley Richard, *The Intellectual and Social Organization of the Sciences*, Oxford, Clarendon Press, 1984.

Wouters Paul et Vries Repke de, "Formally citing the Web", *Journal of the American Society for Information Science and Technology*, 55 (14), 2004, pp. 1250-1260.

par Anne Gentil-Beccot

++++
PRATIQUES
INFORMATIONNELLES
EN PHYSIQUE
DES PARTICULES
++++

+++++

PRATIQUES INFORMATIONNELLES EN PHYSIQUE DES PARTICULES

+++++

La physique des hautes énergies est une discipline qui vise à découvrir les constituants de la matière et comprendre leurs interactions. Elle compte environ 30 000 chercheurs dans différents instituts de recherche et universités répartis dans le monde et constitue une petite communauté scientifique qui possède une culture internationale et collaborative très marquée. D'autres particularités propres à cette communauté l'ont conduite à développer des pratiques informationnelles spécifiques. Elle a en effet adopté, dès les premières heures de la discipline, un système de communication scientifique unique, basé en grande partie sur la prépublication. Ces pratiques ont évolué avec le temps mais, aujourd'hui encore, les prépublications et la littérature grise en général restent une indispensable ressource pour cette discipline, elle est même devenue un moteur qui entraîne de nombreuses évolutions dans le paysage informationnel.

Dans le même temps, la publication dans les revues scientifiques demeure essentielle pour la communauté scientifique, la plupart des prépublications produites étant finalement publiées. En outre, les technologies de l'information deviennent de plus en plus complexes. En plus des outils développés au sein de la communauté, les scientifiques ont la possibilité d'utiliser de nombreux produits comme les bases de données commerciales ou des moteurs de recherche tels que Google ou Google Scholar ; les utilisateurs peuvent désormais accéder, facilement ou non, à d'énormes quantités d'informations. Par ailleurs, les technologies sont plus innovantes chaque jour et de nouvelles interactions avec les utilisateurs se créent. Dans un tel contexte, il est intéressant d'analyser les usages actuels en physique des particules en termes d'information scientifique et de comprendre les évolutions en cours.

Nous commencerons par une description de la culture de la prépublication en physique des particules et du développement des outils adaptés à ces ressources spécifiques, puis nous montrerons quels sont les bénéfices de ce schéma de communication pour la communauté. Enfin, nous

montrons comment elle adapte en permanence ses systèmes d'information en fonction des besoins des utilisateurs et des technologies nouvelles.

LA DIFFUSION DES PRÉPUBLICATIONS, UNE HABITUDE ANCRÉE DEPUIS DES DÉCENNIES

+++++

Dans les années 1950, le délai entre la soumission d'un article scientifique et sa mise à disposition pour le lecteur était inacceptable pour les physiciens des particules. La communauté était déjà composée de deux groupes pour qui la communication – au sein du groupe et avec l'extérieur – était essentielle : durant les premières heures de la discipline, les physiciens expérimentalistes travaillant sur les accélérateurs de particules d'une énergie sans cesse accrue observaient régulièrement de nouvelles découvertes et les théoriciens interprétant ces résultats amélioraient leurs théories et suggéraient de nouvelles expériences. Il était tout simplement hors de question d'accepter des mois de délai dans la diffusion des résultats de recherche – le temps moyen entre la naissance d'une idée et les premiers résultats expérimentaux était de quelques mois seulement –, les prépublications étaient donc la solution idéale.

Pendant des décennies, des physiciens théoriciens et expérimentalistes, désireux de diffuser rapidement leurs résultats de recherche, ont envoyé par courrier postal à tous les grands instituts de recherche en physique des particules des copies de leurs manuscrits tout en les soumettant en parallèle à des revues [Goldschmidt-Clermont, 2002 ; Heuer *et al.*, 2003 ; Aymar, 2009]. Les instituts soutenaient alors financièrement la diffusion des résultats scientifiques de leurs chercheurs, ce qui impliquait des coûts relativement élevés. Les bibliothèques de ces mêmes instituts consacraient également des ressources à l'indexation de toutes ces prépublications, les rendant accessibles aux chercheurs de l'organisation. Cette forme de libre accès assurait une diffusion massive et rapide des résultats scientifiques. Il est à noter que les chercheurs des institutions riches étaient forcément favorisés par ce système. Celles-ci pouvaient en effet payer pour l'envoi massif et étaient, par ailleurs, beaucoup plus susceptibles de recevoir des copies de prépublications d'autres scientifiques en quête de reconnaissance et de visibilité. Les institutions plus petites et plus pauvres avaient donc moins de chance de diffuser leurs résultats et d'être informées des recherches d'autres scientifiques.

Avec l'utilisation croissante d'Internet, le processus s'est poursuivi par voie électronique, ce qui a permis de rendre le système plus équitable.

Avec le développement des technologies, la communauté a créé ses propres outils pour gérer la littérature grise : en 1991, avant même que le Web ait été inventé, Paul Ginsparg, à Los Alamos, a lancé arXiv [arXiv, 2010], le premier serveur de prépublications en physique. Ce nouvel outil a assuré la transition entre une culture du papier obsolète et un nouveau processus électronique, en offrant à tous les scientifiques un moyen facile, gratuit et libre de diffuser leurs prépublications et en supprimant la barrière due au coût de diffusion par la poste des prépublications.

Avec plus de 500 000 articles, arXiv s'est aujourd'hui développé au-delà de la physique des hautes énergies, devenant ainsi le serveur majeur pour de nombreuses autres disciplines [Ginsparg, 1994].

SPIRES [SPIRES, 2010], le premier catalogue électronique de littérature grise [Addis, 1994 ; Kreitz et Brooks, 2003], est né à Stanford Linear Accelerator Center (SLAC), laboratoire à Stanford en Californie, en 1974, et a été développé en collaboration avec DESY, à Hambourg en Allemagne, et Fermilab, à Chicago. Ce catalogue contient à l'heure actuelle des métadonnées d'environ 760 000 articles dans le domaine.

Avec la création du Web, SPIRES est devenu le premier serveur Web [Kunz *et al.*] aux États-Unis en 1992, et des liens ont été créés vers arXiv pour le texte intégral, les prépublications sont donc enfin disponibles sur la toile, avec une accessibilité accrue grâce à l'indexation détaillée incluant les références aux versions publiées.

Aujourd'hui, la communauté produit environ 5 000 articles publiés par an. La grande majorité de ces articles sont publiés dans seulement six revues à comité de lecture [Mele *et al.*, 2006]. Dans la figure 1, on voit que 90 à 100 % des articles publiés dans ces six revues sont également soumis sur arXiv : nous voyons également que cette situation est stable et dure depuis dix ans déjà. En outre, de nombreux scientifiques soumettent de nouveau sur arXiv une version révisée de leur prépublication correspondant à la version finale publiée et incluant les corrections apportées au cours du processus de publication.

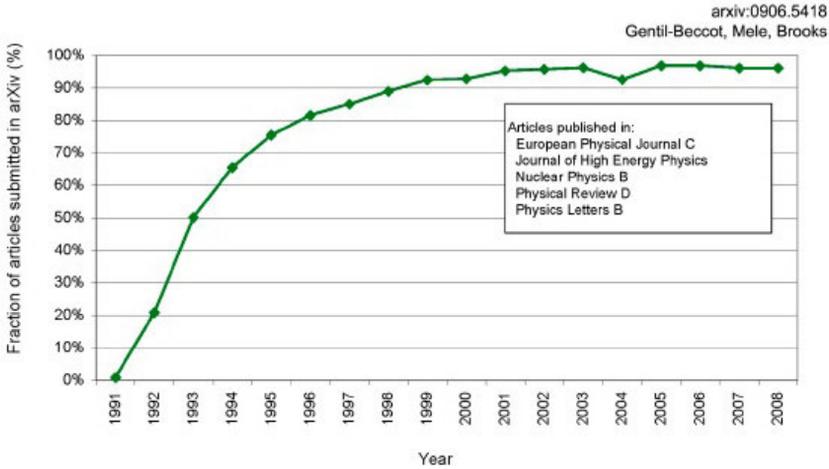


Figure 1 : fraction des articles publiés dans les principales revues de physique des particules et qui ont été aussi soumis sur arXiv.org

Même à l'ère des revues électroniques, par conséquent, les prépublications conservent pleinement leur importance, aujourd'hui arXiv contient la vaste majorité des prépublications (dans la plupart des cas incluant les modifications engendrées par le processus de publication) dans le domaine, cela signifie que presque toute la littérature de la discipline est librement accessible sur le Web. Il est essentiel de préciser que cette situation s'est créée sans aucun mandat : les scientifiques ont développé ce schéma parce qu'il répondait à leur propre besoin, comme nous allons le voir dans la prochaine section.

NÉCESSITÉ D'UNE COMMUNICATION RAPIDE ET IMMÉDIATE

Presque toute la littérature produite en physique des particules est disponible sur arXiv. Mais pourquoi est-ce une réalité pour la physique des hautes énergies, alors que dans certaines disciplines, les dépôts disciplinaires ou institutionnels sont à peine remplis ?

Une étude [Gentil-Beccot *et al.*, *in press*] a été réalisée en 2009 sur l'utilisation réelle d'information dans la communauté en utilisant les citations et les *logs* provenant de SPIRES. Nous voyons dans la figure 2 l'une des conclusions de cette étude. Les articles publiés dans cinq revues pendant l'année 2005 ont été divisés en deux échantillons, ceux qui ont été soumis sur arXiv et ceux qui n'ont pas été soumis sur arXiv, nous voyons

clairement que les articles soumis sur arXiv commencent à collecter les citations bien avant leur publication ; à la date de publication, ils ont en effet déjà acquis 20% du total des citations collectées 24 mois après la publication. Cela s'explique par le fait que les auteurs lisent la prépublication dès qu'elle est disponible sur arXiv, plusieurs mois avant sa publication. Il est à noter que dans la figure 2, les citations issues des prépublications ont été prises en compte au même titre que les citations issues des versions publiées. La rapidité de croissance de la courbe haute s'explique donc, puisqu'il n'y a aucun délai de publication (ni pour le document cité, ou pour celui qui cite). Il y a donc un immense avantage pour les scientifiques à utiliser arXiv : la vitesse de diffusion de l'information. Et, de la même manière que dans les années 1960, les scientifiques envoyaient par la poste leurs prépublications pour diffuser leurs résultats de recherche aussi vite que possible, aujourd'hui, ils utilisent arXiv pour les mêmes raisons, avec, évidemment, une efficacité accrue.

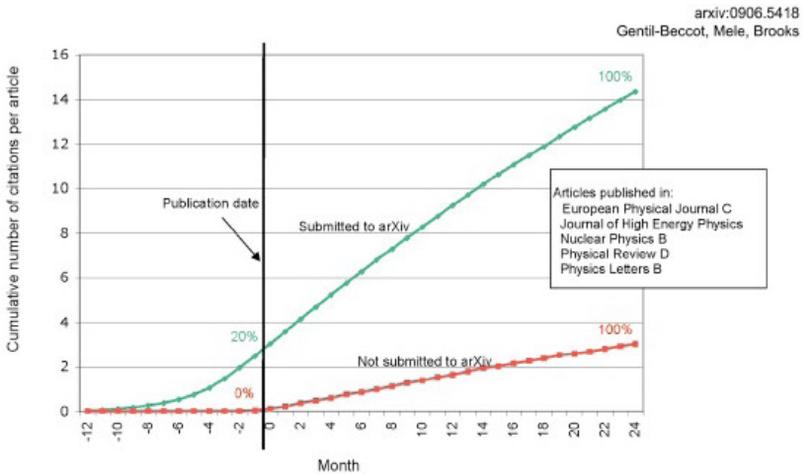


Figure 2 : cumul des citations par article et par mois en fonction de la date de citation par rapport à la date de publication. Les données concernent 4 839 articles de 5 revues majeures en physique des particules : Journal of High-Energy Physics (SISSA/Springer), European Physical Journal C (Springer), Physics Letters B (Elsevier), Nuclear Physics B (Elsevier) et Physical Review D (APS), au cours de l'année 2005.

L'analyse des citations présentée ci-dessus permet de visualiser la rapidité de la communication scientifique dans la physique des hautes énergies, cependant, elle ne peut être lue que comme une « interprétation » des habitudes de lecture réelles. En revanche, l'analyse des logs de SPIRES donne des informations encore plus concrètes sur les usages de lecture des physiciens des particules.

Une enquête [Gentil-Beccot *et al.*, 2009] a montré qu'environ 50% des scientifiques utilisent SPIRES pour une recherche bibliographique (figure 3).

Par conséquent, l'analyse des logs issus de SPIRES devrait donner une représentation claire des habitudes de lecture de la communauté. Les logs ont été récoltés pendant un mois, en octobre 2008 (30 000 clics) [Gentil-Beccot *et al.*, *in press*]. L'étude s'est limitée aux clics provenant d'articles offrant à la fois un lien vers le texte intégral sur arXiv et vers la version publiée sur le site de l'éditeur. Nous avons découvert que dans 82% des cas, arXiv était préféré. L'enquête de 2007 avait également montré que 40% des scientifiques allaient directement sur arXiv sans passer par SPIRES, ils ne sont donc pas comptés dans les données mentionnées ici. La prédominance d'arXiv sur les versions publiées est donc très vraisemblablement plus élevée que ce qui est montré ici.

Plusieurs caractéristiques typiques de la communauté contribuent à expliquer ce résultat. Une des raisons principales est que, dans la plupart des cas, l'auteur resoumet sur arXiv une version révisée de sa prépublication, incluant les corrections apportées par le processus de révision lors de la publication. Par ailleurs, arXiv offre un accès gratuit à son contenu, alors que la version publiée sur le site de la revue est souvent payante. Les chercheurs privilégient donc un accès direct et libre par définition à un accès qui peut leur être refusé car la bibliothèque n'a pas d'abonnement au journal ou parce qu'ils se trouvent en dehors de l'institut. Ces données confirment un résultat de l'enquête faite en 2007 : les répondants devaient indiquer quel système ils utilisaient le plus pour la recherche de prépublications et pour la recherche d'articles publiés. Les résultats ont montré que les habitudes ne changeaient pas dans un cas ou dans l'autre. Lorsque les chercheurs ont besoin d'accéder à la littérature « classique », ils utilisent donc les mêmes systèmes qui ont été initialement conçus pour recenser la littérature grise dans le domaine.

Les revues scientifiques ont donc perdu leur rôle de fournisseurs d'information et de mode de communication scientifique. Toutefois, ces revues continuent à jouer un rôle indispensable, en fournissant un contrôle

de qualité indépendant qui est essentiel dans ce domaine comme dans de nombreuses autres disciplines scientifiques. En effet, l'évaluation des chercheurs (principalement les jeunes) repose encore principalement sur leurs publications dans les revues prestigieuses à comité de lecture. Le rôle que les revues scientifiques de physique des particules conservent donc aujourd'hui est l'enregistrement et surtout la garantie d'une révision scientifique de qualité. Le projet SCOAP³ [Bianco *et al.* ; SCOAP³, 2010] vise à convertir toutes les revues en physique des particules en libre accès, selon un modèle dans lequel la communauté financerait le peer-review des revues plutôt que l'accès proprement dit aux articles.

Si le schéma de communication en physique des particules s'est développé ainsi, c'est donc qu'il a suivi les besoins premiers des utilisateurs, mais il a aussi su tirer parti des évolutions technologiques, et plutôt que de résister, il s'est adapté en fonction.

DES OUTILS EN CONSTANTE ÉVOLUTION

+++++
Grâce à l'association d'arXiv et de SPIRES, les scientifiques ont accès à l'ensemble de la littérature en physique des particules, soit par la pré-publication soit par la version publiée, SPIRES fournit des métadonnées détaillées et les références de publication, et arXiv donne accès au texte intégral de presque tous les articles qui sont ensuite publiés dans les revues. Comme nous l'avons mentionné dans la section précédente (figure 3), l'enquête a démontré qu'arXiv et SPIRES sont les deux principaux outils utilisés par la communauté. 91,4% des personnes interrogées préfèrent ces services à tout autre outil. Ce n'est pas un résultat surprenant, car ces systèmes ont été créés par des acteurs de la communauté pour répondre à ses propres besoins, et ces outils se sont adaptés au fil du temps.

D'autre part, 9% des personnes interrogées déclarent utiliser Google ou Google Scholar. L'enquête a également montré que l'utilisation de Google tend à augmenter chez les plus jeunes. Ceci est un reflet de l'augmentation progressive dans la communauté de chercheurs appartenant à une tranche d'âge qui a été formée à la recherche sur Internet avant leurs études universitaires et leur carrière académique, par opposition à des chercheurs qui ont commencé à utiliser des moteurs de recherche au cours de leur activité professionnelle. Il faut ici noter que Google n'est souvent en fait qu'une passerelle vers d'autres sources. En effet, les utilisateurs de Google trouvent le document qu'ils cherchent parce que des ressources comme

arXiv et SPIRES sont indexées par le moteur de recherche. Ce résultat ne peut donc être considéré comme une preuve que Google devient plus important que les outils développés par la communauté, en revanche, il montre clairement le besoin croissant des utilisateurs d'avoir un accès unique à leurs ressources d'information.

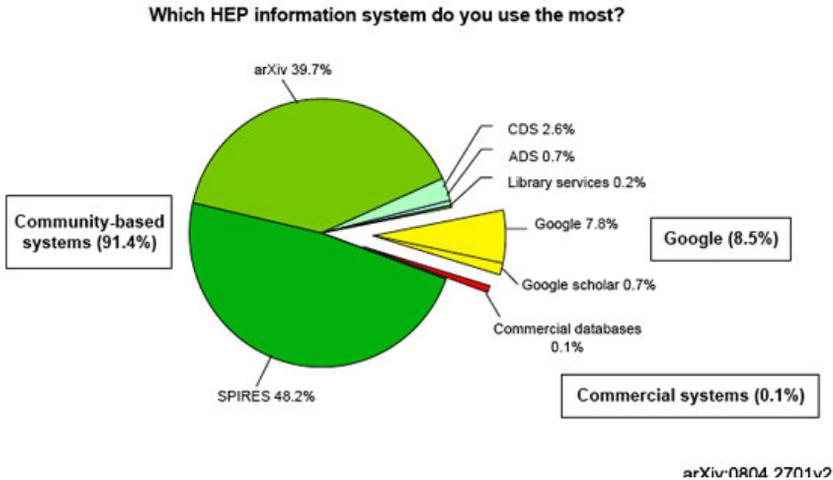


Figure 3 : les ressources d'information préférées des chercheurs en physique des particules

L'enquête a montré par ailleurs que les utilisateurs ont de plus en plus besoin d'avoir un accès plus facile et plus large au contenu, tel que les présentations données aux conférences, la littérature « encore plus grise » (les thèses par exemple) ainsi que les données qui se trouvent derrière les tableaux et les graphiques. L'accès au texte intégral ne suffit plus, les chercheurs ont besoin d'accéder à une information plus complète, plus profonde et plus interconnectée.

Dans cette même étude, on a pu observer que les technologies Web 2.0 prenaient une place non négligeable parmi les exigences des utilisateurs, presque 50 % des personnes qui ont répondu pensent que la « recommandation d'articles », c'est un outil important. Une autre question a tenté d'évaluer le potentiel de la mise en œuvre de fonctionnalités Web 2.0 qui permettraient d'interagir mieux avec l'utilisateur comme le temps que serait prêt à passer un chercheur à commenter/tagger du contenu. 63 % des répondants ont affirmé qu'ils seraient prêts à y passer entre cinq minutes par jour et une heure par semaine, ce qui montre qu'il y a un

immense potentiel d'enrichissement de l'information provenant directement de l'utilisateur. Cette question est essentielle, cette indexation de masse pourrait aider les futurs systèmes d'information à être toujours plus exacts et performants. Nous ne savons pas encore comment cela va évoluer, comment la communauté va utiliser ces nouveaux outils. Mais, ce nouveau type d'information et d'interaction avec l'utilisateur constituera un nouveau défi dans le futur car cela sera sans aucun doute un contenu scientifique qui aura une valeur en tant que telle.

L'enquête réalisée et les discussions entre les quatre plus grands laboratoires de recherche en physique des hautes énergies (CERN, DESY, Fermilab et SLAC), en synergie avec d'autres partenaires (notamment arXiv) et dans un dialogue continu avec les grands éditeurs dans le domaine, ont engendré l'idée d'un nouveau système d'information en HEP. Ce nouveau système est à la croisée des chemins entre l'actuelle base de données SPIRES et une plateforme moderne, Invenio, logiciel open source de bibliothèque numérique [CDS Invenio, 2010]. Ce nouveau système d'information, INSPIRE [INSPIRE, 2010], est développé par une collaboration entre les quatre principaux laboratoires de recherche en physique des particules mentionnés ci-dessus. Il intègre le contenu des bases de données présentes pour accueillir l'ensemble des métadonnées et le texte intégral de toutes les publications en libre accès, passées et futures, comprenant de nouveaux contenus comme les présentations de conférences. Ce nouveau système est appelé à devenir le point d'entrée unique pour l'ensemble de la communauté à tous les documents en physique des particules.

CONCLUSION

+++++

La littérature grise est devenue le canal de communication le plus important parce que les physiciens des particules avaient besoin d'un accès immédiat à l'information. C'est encore le cas, et c'est pourquoi la communauté développe de nouveaux systèmes qui rendent l'accès à l'information encore plus rapide et exhaustif. Mais les besoins évoluent et la communauté doit maintenant aller plus loin : l'accès libre au texte intégral ne suffit plus, les scientifiques veulent une couverture encore plus large et profonde du contenu. Ils ont besoin d'interagir plus directement avec toutes ces ressources. Ce sont tous les défis auxquels le nouveau système d'information HEP INSPIRE doit faire face. Une autre question parallèle est bien sûr l'évolution du rôle de la littérature publiée, qui ne sert plus

directement à la communication scientifique, même si son rôle d'enregistrement et de validation de la qualité ne peut être remis en cause dans le panorama scientifique actuel.

Mais une constante reste dans tous ces mouvements, la communauté elle-même prend la tête de son évolution, guidée par ses propres usages. Par ailleurs, les besoins ne cessent pas d'évoluer avec l'accomplissement des projets, et il sera nécessaire dans le futur de questionner, de mesurer et d'évaluer – encore et toujours – pour adapter au mieux les systèmes qui sont dessinés aujourd'hui.

RÉFÉRENCES BIBLIOGRAPHIQUES

+++++

Addis Louise, *Brief and Biased History of Preprint and Database Activities at the SLAC Library, 1962-1994*, 1994.

[En ligne] < <http://www.slac.stanford.edu/spires/papers/history.html> > (consulté le 24 février 2010).

ArXiv, site Web, [En ligne] < <http://www.arXiv.org> > (consulté le 24 février 2010).

Aymar Robert, "Scholarly communication in high-energy physics: past, present and future innovations", *European Review*, vol. 17, 2009, pp. 33-51, CERN-OPEN-2008-015.

Bianco Stefano *et al*, *Towards Open Access Publishing in High Energy Physics: Report of the SCOAP³ Working Party*, 2007.

[En ligne] < <http://scoap3.org/files/Scoap3WPReport.pdf> > (consulté le 24 février 2010).

CDS Invenio, site Web. [En ligne] < <http://cdsware.cern.ch/invenio/index.html> > (consulté le 24 février 2010).

Gentil-Beccot Anne, Mele Salvatore, Holtkamp Annette, O'Connell Heath et Brooks Travis C., "Information resources in high-energy physics: surveying the present landscape and charting the future course", *Journal of the American Society for Information Science and Technology*, vol. 60, 2009, pp. 150-160, arXiv : 0804.2701v2.

Gentil-Beccot Anne, Mele Salvatore et Brooks Travis C., "Citing and reading behaviours in high-energy physics. How a community stopped worrying

about journals and learned to love repositories”, *Scientometrics*, 84, 2010, pp. 345-355, arXiv : 0906.5418.

Ginsparg Paul, “First steps towards electronic research communication”, *Computers in Physics*, vol. 8, 1994, pp. 390-396.

Goldschmidt-Clermont Luisa, “Communication patterns in high-energy-physics”, *High-Energy Physics Libraries Webzine*, n° 6, 2002.

[En ligne] < <http://library.web.cern.ch/library/Webzine/6/papers/1/> > (consulté le 24 février 2010).

Heuer Rolf-Dieter, Holtkamp Annette et Mele Salvatore, “Innovation in scholarly communication: vision and projects from high-energy physics”. *Information Services & Use*, vol. 28, 2003, pp. 83-96, arXiv : 0805.2739.

INSPIRE. [En ligne] < <http://inspirehep.net/> > (consulté le 24 février 2010).

Kreitz Patricia A. et Brooks Travis C., “Subject access through community partnerships: a case study”, *Science & Technology Libraries*, vol. 24, 2003, pp. 153-172, arXiv : physics/0309027.

Kunz Paul *et al*, *The Early World Wide Web at SLAC*.

[En ligne] < <http://www.slac.stanford.edu/history/earlyweb/history.shtml> > (consulté le 24 février 2010).

Mele Salvatore, Dallman David, Vigen Jens et Yeomans Joanne, “Quantitative analysis of the publishing landscape in high-energy physics”, *Journal of High-Energy Physics*, vol. 12, 2006, S01, arXiv : cs.DL/0611130.

SCOAP³, site Web. [En ligne] < <http://www.scoap3.org/> > (consulté le 24 février 2010).

SPIRES, site Web. [En ligne] < <http://www.slac.stanford.edu/spires> > (consulté le 24 février 2010).

par Anna Wojciechowska

+++++

PRATIQUES
DOCUMENTAIRES
ET PRATIQUES
D'AUTO-ARCHIVAGE
DES MATHÉMATIENS
ET INFORMATIENS
EN FRANCE

+++++

+++++
PRATIQUES DOCUMENTAIRES ET
PRATIQUES D'AUTO-ARCHIVAGE
DES MATHÉMATICIENS ET
INFORMATIENS EN FRANCE
+++++

Récemment, de nombreuses études se sont penchées sur la manière dont les usagers appréhendent le nouveau mode de fonctionnement de la publication scientifique. Au niveau des pratiques documentaires, elles montrent l'évolution des comportements des usagers : évolutions des pratiques de recherche d'information, des pratiques de lecture, d'échanges et de publications.

Elles portent surtout sur les pays anglo-saxons [Allen, 2005 ; JISC, 2005 ; Swan, 2005a et 2005b ; Swan et Brown, 2005].

La taille de leurs échantillons est très variable, de moins de vingt personnes pour certaines études, jusqu'à plusieurs centaines. L'enquête la plus riche en information est celle de Swan et Brown [2005] qui montre les pratiques des chercheurs de quinze disciplines.

Nous présentons dans cet article les résultats partiels de deux enquêtes menées auprès des mathématiciens et des informaticiens en France concernant leurs pratiques documentaires et pratiques d'auto-archivage. La première enquête a eu lieu en 2005 et la deuxième en 2007. Elles se sont poursuivies par des entretiens avec un certain nombre de participants choisis d'après leurs réponses au questionnaire de manière à constituer un échantillon de profils différents. Ces enquêtes nous ont donné l'occasion de voir le changement dans le temps des comportements des usagers et de le comparer avec les études anglo-saxonnes.

Nous avons recueilli des données auprès des enseignants-chercheurs et doctorants en mathématiques et informatique *via* des bibliothèques du réseau RNBM¹.

Nous estimons à 2200 le nombre de personnes qui en furent destinataires. 128 personnes ont répondu en 2005 et 190 en 2007. Le nombre de réponses ne constitue pas d'échantillon représentatif (6 % et 8,6 %), mais

1. Réseau national des bibliothèques de mathématiques (RNBM) : < <http://www.rnbm.org> >.

cela arrive habituellement aux enquêtes qui ne sont pas organisées par la hiérarchie scientifique (les études anglo-saxonnes).

Le questionnaire a été composé de trente-six questions réparties en quatre parties portant respectivement sur :

- la recherche de l'information ;
- les publications ;
- l'auto-archivage ;
- la connaissance des revues en accès libre.

PARTICIPANTS

Des réponses incomplètes n'ont pas toujours permis de connaître avec précision la répartition des répondants entre mathématiciens et informaticiens en 2005. En 2007, 16 % des participants étaient des informaticiens et 84 %, des mathématiciens. Parmi les personnes qui ont participé aux enquêtes, la majorité était composée d'enseignants-chercheurs et de chercheurs CNRS.

La productivité d'un chercheur n'est pas la même au début, au milieu ou à la fin de sa carrière. L'utilisation qu'il fait des nouvelles technologies pour accéder à l'information n'est pas non plus la même. D'où l'intérêt d'informations concernant l'âge des participants pour tenter de dégager des différences dans les pratiques en fonction de ce critère. Dans notre échantillon, presque 70 % en 2005 et 60 % en 2007 des participants ont moins de 40 ans.

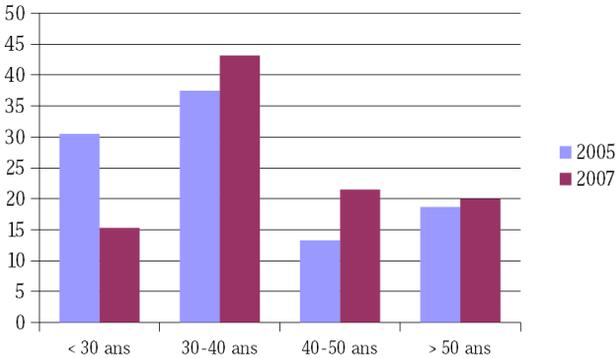


Figure 1 : participation selon l'âge

RECHERCHE DE L'INFORMATION

Les chercheurs sont aussi bien auteurs que lecteurs des articles. Pour cette raison, la première partie des questions posées concernait les chercheurs lecteurs et leur pratique de recherche de l'information scientifique (et des sources de cette information) nécessaire à leur travail. Il s'agit surtout des recherches de références bibliographiques et de textes intégraux d'articles, aussi bien récents qu'anciens.

À la question *Où obtenez-vous les articles dont vous avez besoin ?*, les chercheurs pouvaient donner plusieurs réponses (tableau 1).

Tableau 1

Source des articles	2005	2007
Bibliothèque du laboratoire	75,8 %	80,5 %
Bibliothèque universitaire	27,3 %	20,5 %
Bases de données ²	56,3 %	52,1 %
ScienceDirect ³	39,8 %	42,6 %
Sites des éditeurs	33,6 %	29 %
SpringerLink ⁴	45,3 %	47,9 %
Inist ⁵	5,5 %	23,7 %
Autres	32 %	36,8 %

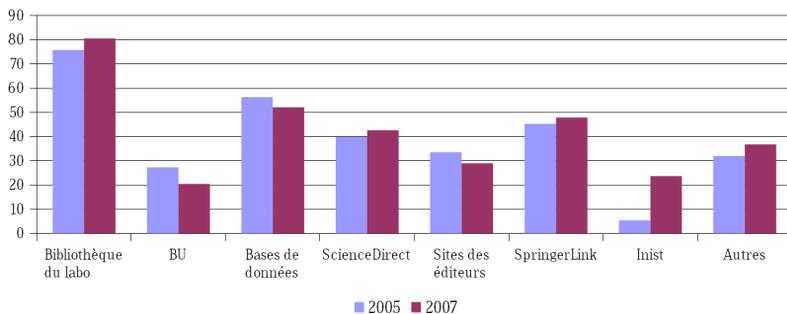


Figure 2. Sources des articles

- Surtout MathSciNet (< <http://e-math.ams.org/mathscinet/> >) et Zentralblatt Math (< <http://www.emis.de/ZMATH/> >), les bases de données payées par les bibliothèques. Emis : The European Mathematical Information Service (< <http://www.emis.de/index.html> >).
- ScienceDirect : journaux en texte intégral d'Elsevier, l'accès payant dans le cadre du consortium Couperin, < <http://www.sciencedirect.com/> > et < <http://www.couperin.org/> >.
- SpringerLink : l'accès payant aux journaux en texte intégral de Springer, Birkhauser et Kluwer : < <http://www.springerlink.com/> >.
- Institut de l'information scientifique et technique (Inist) : < <http://www.inist.fr/> > ; les portails BiblioSciences et BiblioStic.

Même si 80 % des personnes interrogées (en 2007) trouvent les articles (ou leurs références) dans la bibliothèque de recherche de leur laboratoire, déjà 52 % citent les bases de données comme sources d'information, et les journaux en texte intégral (accessibles par les abonnements payants de la bibliothèque) sont de plus en plus consultés : 47 % des chercheurs trouvent les articles dans SpringerLink et 43 % dans ScienceDirect.

Depuis la création des BiblioSciences et BiblioStic, les portails de l'Inist, ces accès aux journaux sont beaucoup utilisés : 5,5 % en 2005 et 24 % en 2007.

Parmi d'autres sources d'articles données par 32 %, il y a des pages et des sites Web personnels, arXiv, Hal, Hopf Archive, MathDoc, Cedram, Emani (Electronic Mathematics Archives Network Initiative), WDM (World Digital Mathematics Library), des sites de conférences, d'autres abonnements en ligne, comme JSTOR⁶, CiteSeer⁷, ou le contact direct avec les auteurs.

Parmi les sites proposés comme points d'accès aux articles en texte intégral en accès libre, les chercheurs en majorité ont choisi arXiv (75 %) et Google (64 %). Le nombre de personnes qui consultent Hal a augmenté. Parmi d'autres réponses, on peut trouver : les pages personnelles, the Digital Bibliography & Library Project (base de données bibliographiques en informatique) ou Numdam⁸.

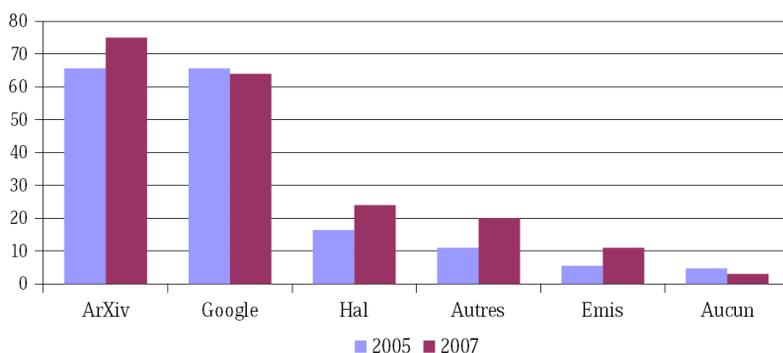


Figure 3. Accès aux articles en texte intégral en accès libre

6. JSTOR (The Scholarly Journal Archive): < <http://www.jstor.org/> >.

7. CiteSeer (Scientific Literature Digital Library): < <http://citeseer.ist.psu.edu/> >.

8. Numdam (fonds ancien en mathématiques numérisé) : < <http://www.numdam.org/> >.

Les mathématiciens âgés de plus de 30 ans utilisent plus souvent arXiv que les mathématiciens âgés de moins de 30 ans et l'ensemble des informaticiens préfèrent, quant à eux, utiliser Google.

Google permet, en effet, de trouver les pages personnelles, les pages des laboratoires ou des bibliothèques qui possèdent les serveurs locaux où les publications scientifiques sont stockées, ce moteur permet aussi d'accéder directement aux articles. L'utilisation d'arXiv concerne surtout l'accès aux prépublications électroniques, même si on y trouve aussi des articles publiés.

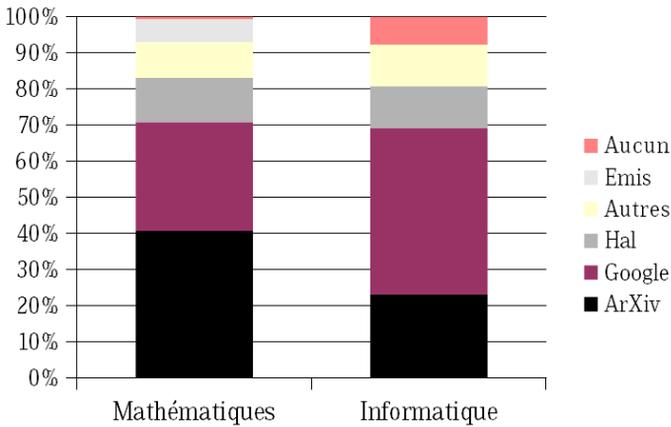


Figure 4. Accès aux articles par domaine en 2007

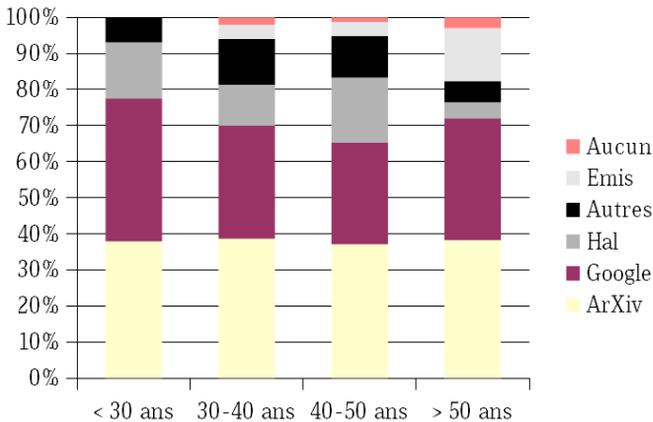


Figure 5. Accès aux articles par âge en 2007

La question concernant les points d'accès aux prépublications électroniques a confirmé que les chercheurs ont une bonne connaissance de Hal et d'arXiv, qui sont plus souvent interrogés en 2007 (75 %) qu'en 2005 (58 %), même si 76 % d'entre eux accèdent cependant à ces preprints électroniques par l'intermédiaire de pages personnelles. L'accès par les sites des laboratoires ou des bibliothèques a été signalé par 20 % des répondants, ce qui signifie qu'il existe des dépôts locaux.

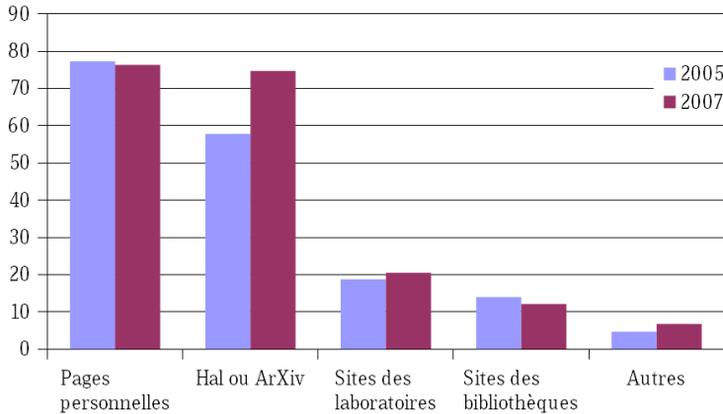


Figure 6. Accès aux prépublications en ligne

Les réponses obtenues à la question *Pouvez-vous accéder facilement aux articles dont vous avez besoin pour votre travail ?* montrent qu'en 2007, 90 % des chercheurs (80 % en 2005) obtiennent la plupart des articles qui leur sont nécessaires.

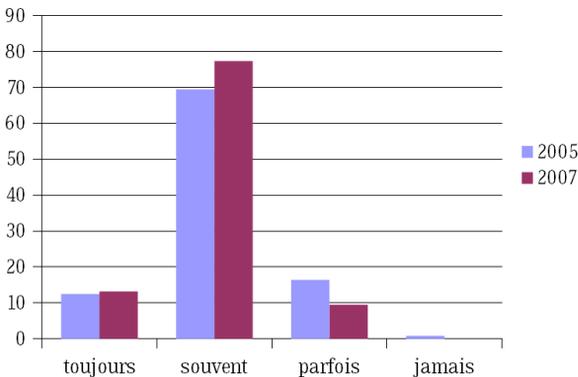


Figure 7. Facilité d'accès aux articles

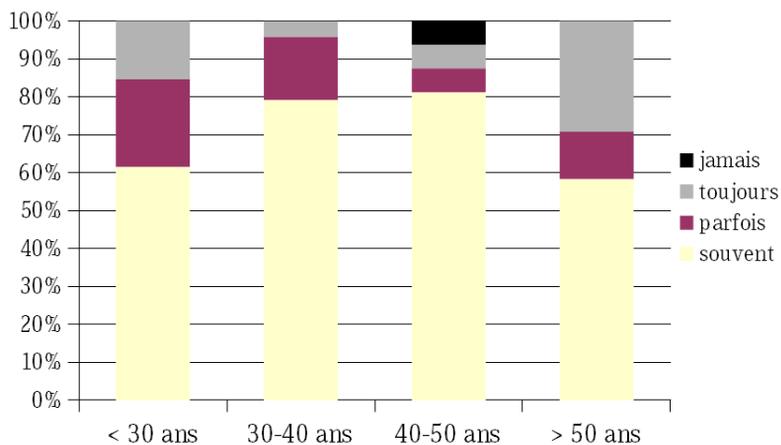


Figure 8. Facilité aux articles par âge en 2005

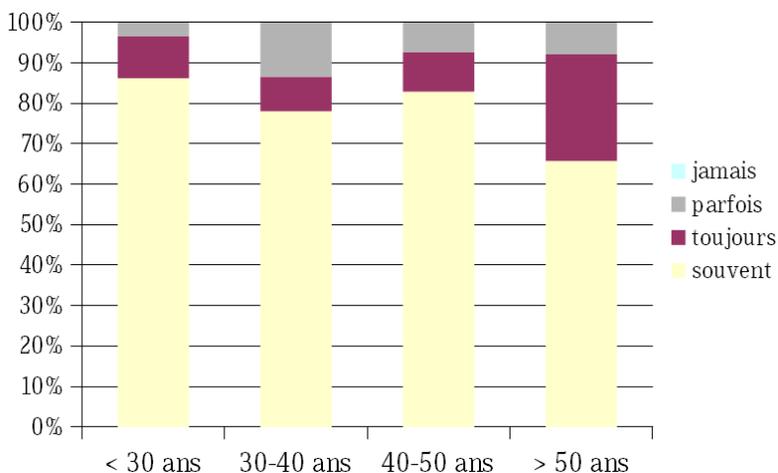


Figure 9. Facilité aux articles par âge en 2007

Nous avons voulu savoir si cette facilité d'accès aux articles est obtenue grâce à l'aide du personnel des bibliothèques. À la question *Faites-vous appel à un(e) bibliothécaire lors de vos recherches documentaires ?*, les réponses (tableau 2) ont été consternantes pour les documentalistes.

Tableau 2

Demande d'aide	2005	2007
Toujours	0 %	0 %
Souvent	5,5 %	74 %
Parfois	58,6 %	58,4 %
Jamais	35,9 %	34,2 %

Nous savons que les articles accessibles en texte intégral sont plutôt récents (en général à partir de 1995). Il existe aussi de plus en plus d'articles anciens numérisés *a posteriori* dans le cadre de divers projets locaux, nationaux ou internationaux, qui sont aussi en ligne en accès libre.

Si les chercheurs n'ont pas forcément besoin d'aide pour leurs recherches documentaires, c'est justement parce que 50 % d'entre eux (presque 40 % en 2005) utilisent les articles qui ont été publiés pendant les dix dernières années, c'est-à-dire ceux qui sont en majorité disponibles en texte intégral en ligne ou dans les bibliothèques en version papier.

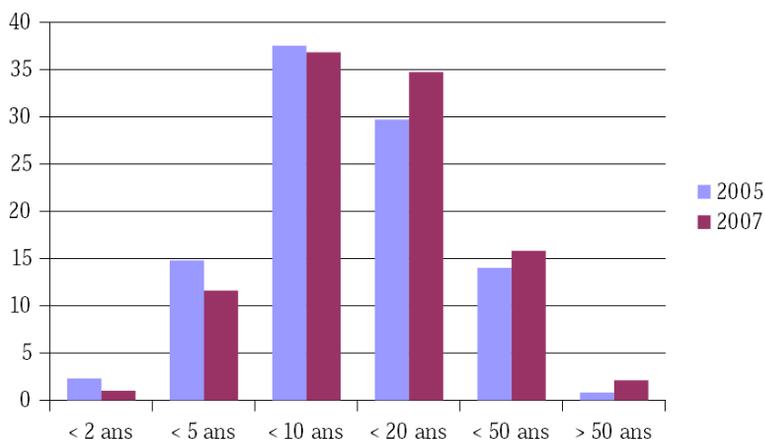


Figure 10. Âge des articles consultés

Les publications électroniques sont consultées de plus en plus (presque 80 % des personnes interrogées en 2007, 60 % en 2005) et au moins une fois par semaine ou plus (tableau 3). Comme indiqué précédemment, il s'agit d'articles publiés pendant les dix dernières années.

Tableau 3

Fréquence de consultation des articles électroniques	2005	2007
Une fois par semaine ou plus	57,8 %	77,9 %
Une fois par mois ou plus	31,3 %	18,4 %
Une fois par trimestre ou plus	10,2 %	3,7 %
Pas de réponse	0,7 %	0 %

Presque tous (93 % en 2005 et 96 % en 2007) consultent les preprints en ligne, mais, comme nous allons le voir plus loin, ils sont moins nombreux à déposer leurs prépublications en ligne en accès libre.

PUBLICATIONS

+++++
 Nous allons nous pencher à présent sur le volume de publications des chercheurs, sur leur façon de déposer ces publications dans les archives ouvertes, sur le nombre et les types de dépôts qu'ils effectuent.

Une publication peut avoir la forme d'une prépublication, c'est-à-dire d'un texte finalisé, mais non encore publié ou en cours de validation, ainsi que celle d'un article déjà validé par un comité scientifique et publié. Nous avons voulu savoir combien d'articles en moyenne un chercheur publie par an.

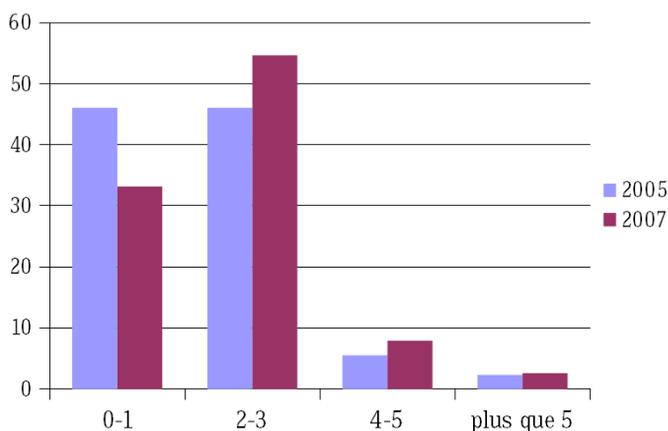


Figure 11. Nombre de publications par an

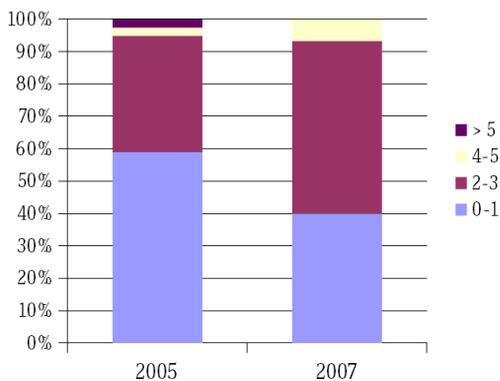


Figure 12. Publications < 30 ans

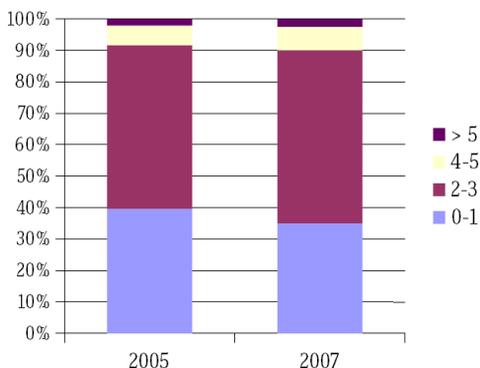


Figure 13. Publications de 30-40 ans

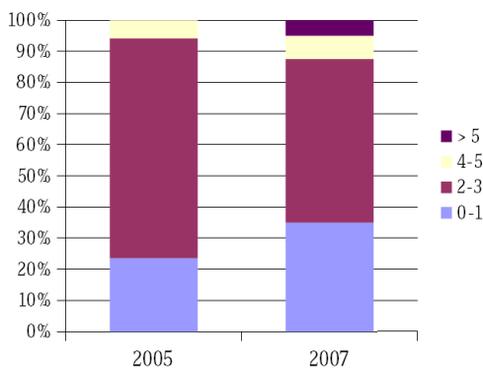


Figure 14. Publications de 40-50 ans

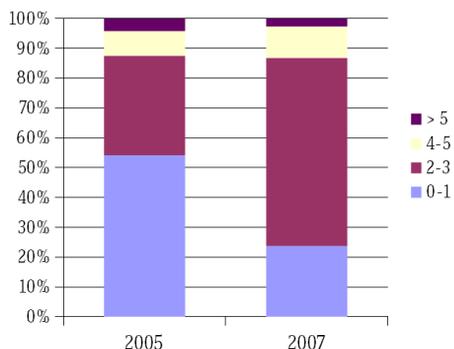


Figure 15.
 Publications > 50 ans

En 2005, 46 % des chercheurs interrogés publient au maximum un article par an, mais il s'agit ici surtout des doctorants ou jeunes enseignants (la participation de cette tranche d'âge – moins de 30 ans – était faible en 2007 et correspondait à 33 %). 46 % des personnes interrogées déclarent en publier 2 ou 3 par an (55 % en 2007).

La publication d'un article correspond généralement à une avancée de travaux de recherche. C'est sur la base de ces articles que le chercheur est évalué et financé. Pour les chercheurs qui ont répondu aux questionnaires (chacun pouvait donner plusieurs réponses), la priorité est de communiquer les résultats de leur recherche à la communauté scientifique, ensuite (beaucoup plus loin) d'avancer dans leur carrière (surtout les jeunes de moins de 30 ans), d'acquérir du prestige personnel dans son domaine.

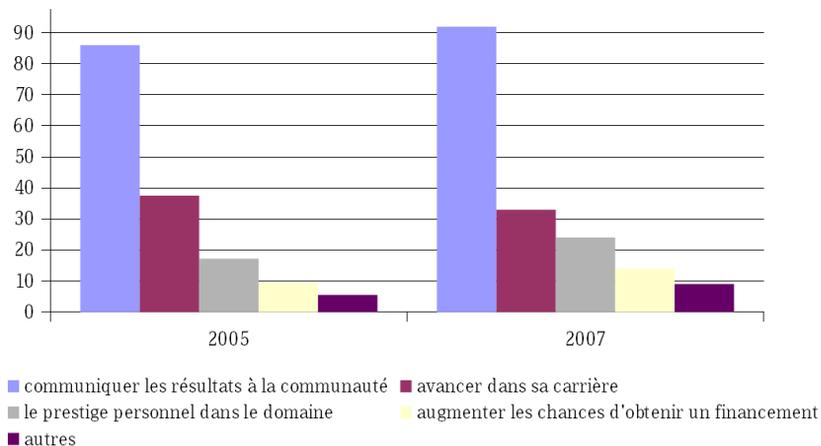


Figure 16. Objectifs de publication

Les autres réponses (curiosité scientifique, volonté de reconnaissance des résultats, volonté de préciser la formulation des résultats obtenus, etc.) représentent un peu moins de 10 %.

LES CHERCHEURS ET LES ARCHIVES OUVERTES

+++++

Si presque la moitié des chercheurs participant à l'enquête en 2005 disent connaître le terme archives ouvertes, en 2007, ils sont déjà 75,8 % (80 % des mathématiciens et 60 % des informaticiens). Il s'agit ici surtout de chercheurs âgés de 40 à 50 ans. Seulement 40 % des jeunes (moins de 30 ans) ont répondu positivement à cette question.

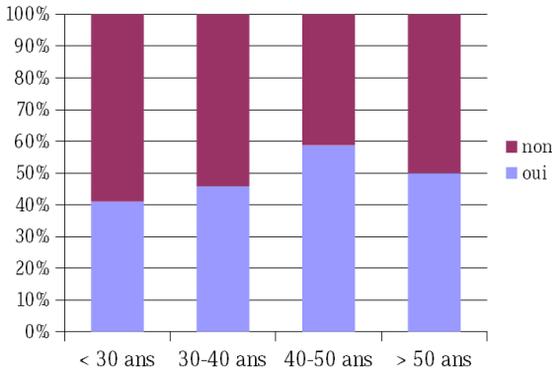


Figure 17. Connaissance des archives ouvertes en 2005

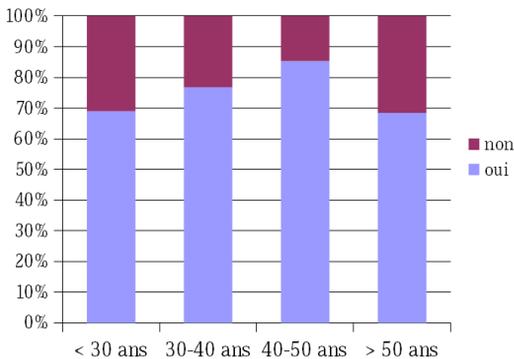


Figure 18. Connaissance des archives ouvertes en 2007

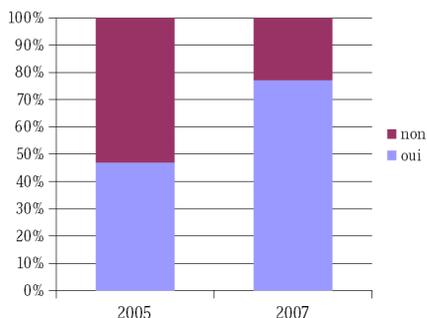


Figure 19.
 Connaissance des archives
 ouvertes

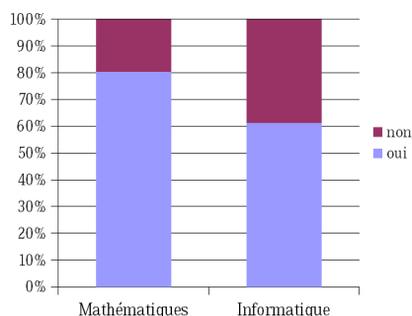


Figure 20.
 Connaissance des archives
 ouvertes en 2007

Nous avons voulu savoir comment les chercheurs ont appris l'existence des archives ouvertes, quelles sont leurs motivations pour effectuer les dépôts, et ce qu'ils savent sur le droit d'auteur et sur les revues en accès libre.

Le questionnaire proposait une liste des sources probables d'information sur les archives ouvertes (tableau 4). Les chercheurs pouvaient donner plusieurs réponses (ou aucune).

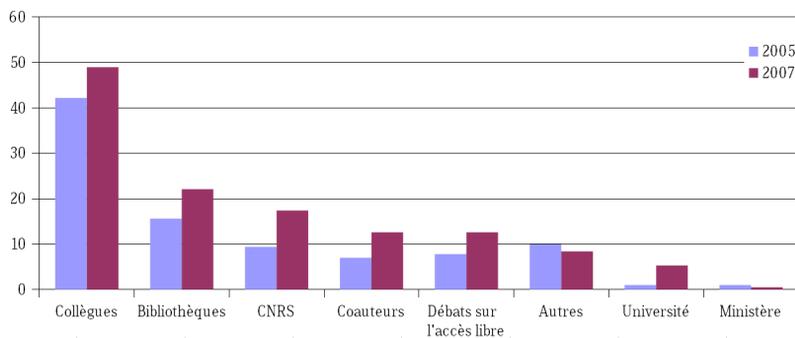


Figure 21. Sources d'information sur les archives ouvertes.

Tableau 4

Source d'information sur les archives ouvertes	2005	2007
Grâce aux collègues	42,2 %	49 %
Des informations de la bibliothèque	15,6 %	22,1 %
Autres	10 %	8,4 %
Des informations du CNRS	9,4 %	17,4 %
Les débats sur l'accès libre	7,8 %	12,6 %
Grâce aux coauteurs	7 %	12,6 %
Des informations de l'université	0,8 %	5,3 %
Des informations du ministère	0,8 %	0,5 %

Les collègues constituent la source d'information sur l'existence des archives ouvertes institutionnelles pour le plus grand nombre de chercheurs. Nous pouvons aussi constater que la communication institutionnelle a été améliorée en 2007, et le débat sur l'accès libre a attiré plus d'auditoire (surtout parmi les chercheurs âgés de plus de 50 ans).

Autres sources signalées :

- ce questionnaire ;
- les informations de l'Inria⁹ ;
- les informations du laboratoire ;
- les informations de MathDoc.

Nous avons voulu connaître les motivations des chercheurs pour les dépôts dans les archives ouvertes. En 2007, 16 % n'ont pas répondu à cette question. Parmi ceux qui déposent leurs publications dans les archives ouvertes institutionnelles, 55 % en 2005 et 79 % en 2007 le font par principe (afin de donner l'accès aux résultats de leur recherche au plus grand nombre de personnes¹⁰) et 38 % (25 % en 2005) parce que les archives ouvertes existent (ce qui constitue une façon simple de valider et/ou publier leur travail¹¹) (tableau 5).

9. Institut national de recherche en informatique et en automatique.

10. Les précisions recueillies lors des entretiens réalisés après les enquêtes.

11. Les précisions recueillies lors des entretiens réalisés après les enquêtes.

Tableau 5

Motivation	2005	2007
Principe des archives ouvertes	54,7 %	78,6 %
Citations des articles en accès libre plus fréquentes que des articles publiés dans les journaux payants	7,8 %	6,6 %
Encouragement de la bibliothèque et de l'administration	6,3 %	8,6 %
Encouragement des collègues et coauteurs	11 %	11,3 %
Existence des archives ouvertes dans mon domaine	25 %	37,8 %
Autres	11 %	8 %

Parmi d'autres motivations signalées :

- obligation institutionnelle ;
- garder une trace ;
- protéger les résultats ;
- enregistrer la date d'un travail non publié ;
- communication rapide des résultats de recherche ;
- rendre accessible aux bibliothèques pauvres, par réciprocité.

EXPÉRIENCE D'AUTO-ARCHIVAGE

+++++

L'auto-archivage consiste à déposer un document électronique sur un site Web pouvant être consulté gratuitement par tous. Il est effectué afin d'optimiser la visibilité de la recherche et l'accessibilité aux travaux qui en rendent compte.

- Il existe plusieurs façons d'auto-archiver un article (pré- ou post-publication) : le chercheur peut déposer une copie de son article sur un site Web (personnel ou du laboratoire) ;
- dans des archives ouvertes institutionnelles (Hal, par exemple) ;
- ou dans des archives ouvertes par domaine (arXiv, par exemple).

Les publications scientifiques sont en grande partie déposées en ligne par les auteurs et/ou les coauteurs (tableau 6).

Tableau 6

Auteur du dépôt des publications	2005	2007
Vous-même	74,2 %	87,4 %
Votre coauteur	10,9 %	20,5 %
Le secrétariat du laboratoire ou la bibliothèque	11,7 %	9,5 %
Personne	5,5 %	5,2 %

Les dépôts effectués par les secrétariats des laboratoires ne contiennent pas de textes intégraux et cela concerne surtout les informaticiens. Il s'agit ici des dépôts des notices bibliographiques.

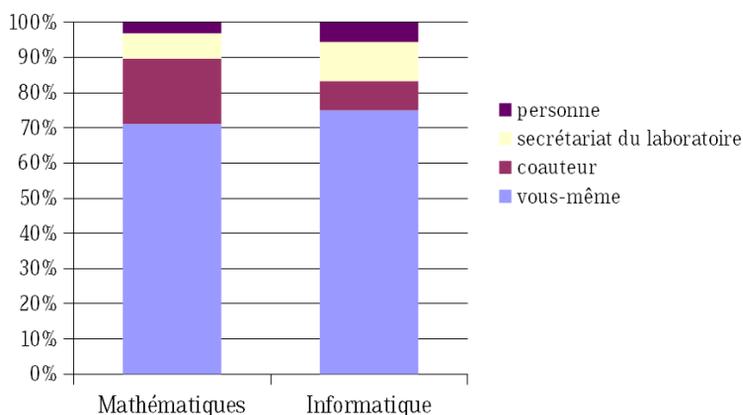


Figure 22. Auteur des dépôts en 2007

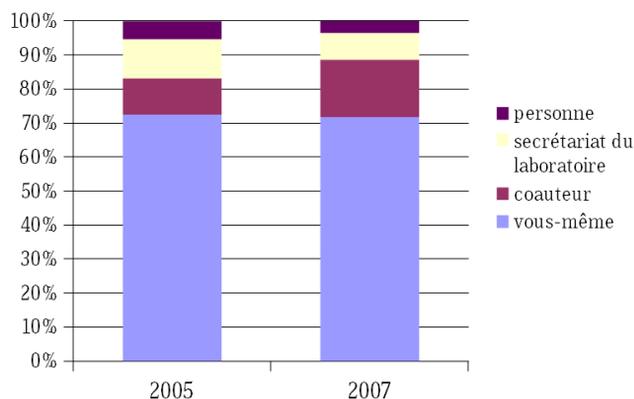


Figure 23. Auteur des dépôts

La question *Quel genre de publications déposez-vous dans les archives ouvertes ?* a permis de faire le point sur le type de publications déposées le plus souvent en ligne (tableau 7).

Tableau 7

Types de publications déposées	2005	2007
Articles référencés	64,8 %	69,5 %
Actes de colloques	22,7 %	20 %
Prépublications	77,3 %	87,4 %
Rapports techniques	13,3 %	10 %
Chapitres de livres	6 %	7,9 %
Dissertation, habilitation, thèse	23,4 %	16,8 %
Cours	18 %	21,1 %
Exercices	18 %	14,7 %
Autres	2,3 %	1 %

En 2005, 65 % des personnes interrogées ont déposé en ligne des articles publiés et 77 % des prépublications. En 2007, ils sont respectivement 70 % et 87 %. Parmi d'autres types de publications, on peut trouver des errata et des fichiers de conférences à projeter (type PowerPoint).

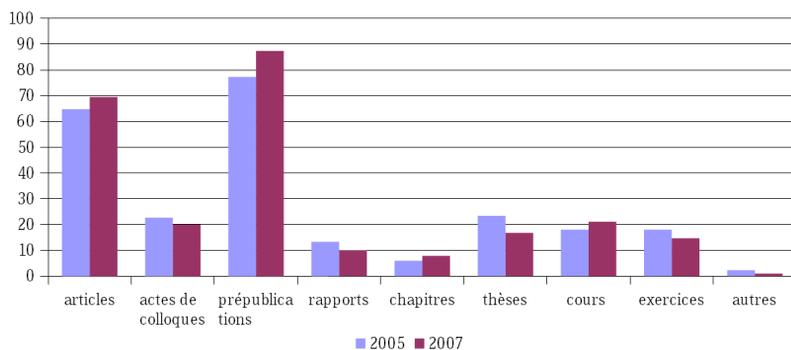


Figure 24. Publications déposées dans les archives ouvertes

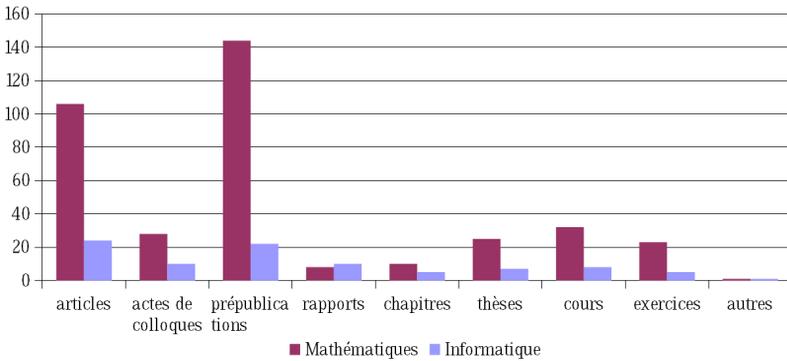


Figure 25. Types de dépôts par domaine en 2007

Les prépublications sont déposées surtout par les mathématiciens. Les informaticiens auto-archivent en majorité les actes de colloques, les rapports techniques et les chapitres de livres.

82 % des chercheurs interrogés (73 % en 2005) ne craignent pas le pillage ni l'usage abusif des prépublications en ligne, ce qui correspond (presque) à la proportion de déposants de preprints dans les archives ouvertes.

Une partie des chercheurs ont déclaré le dépôt des articles sur les sites Web personnels et cela concerne toutes les tranches d'âge. Parmi ceux qui n'ont déposé aucun article, la majorité ne possède pas de pages Web personnelles.

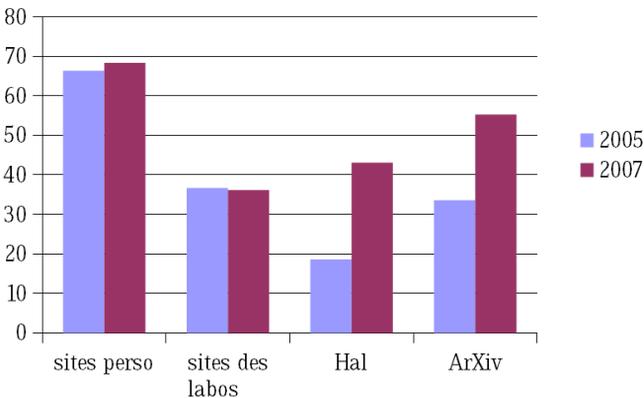


Figure 26. Au moins une prépublication déposée

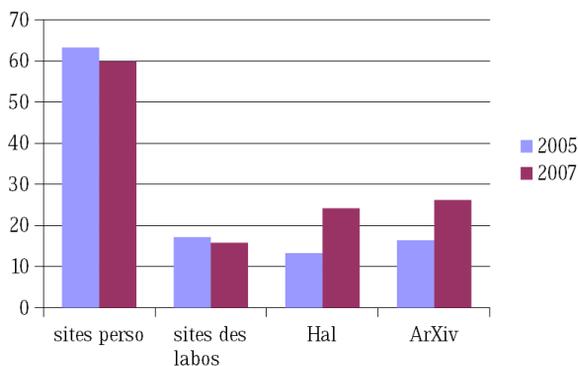


Figure 27. Au moins un article déposé

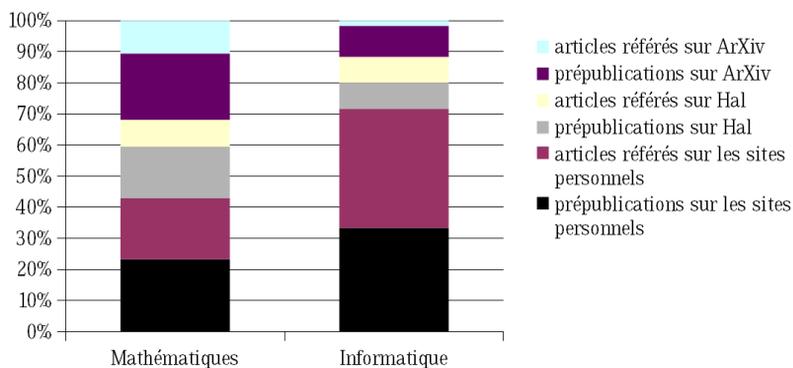


Figure 28. Les dépôts sur le Web par domaine en 2007

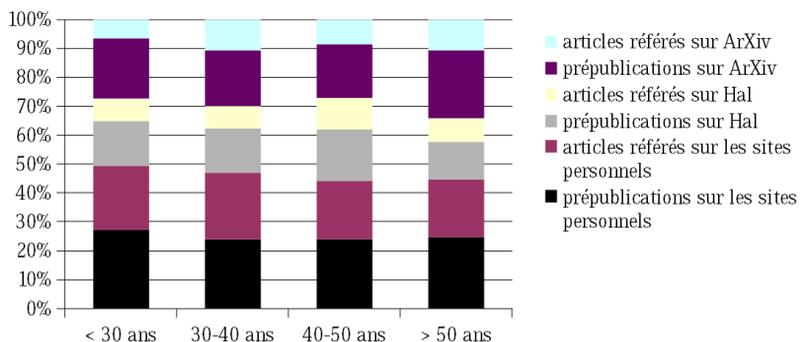


Figure 29. Les dépôts sur le Web par âge en 2007

Il est impossible de comparer le nombre d'articles publiés par un chercheur par an avec le nombre de dépôts effectués pendant les trois dernières années, car souvent les chercheurs déposent aujourd'hui les publications plus anciennes.

Les personnes interrogées pouvaient donner plusieurs réponses, ce qui veut probablement dire que ceux qui ont archivé des articles sur leurs sites Web l'ont également fait (en partie) dans des archives institutionnelles. Les résultats globaux n'indiquent donc pas le vrai niveau de l'auto-archivage, d'autant moins que les articles peuvent être déposés par d'autres personnes (le coauteur, le secrétariat de son laboratoire, etc.) ou tout simplement le chercheur n'a pas donné de réponse.

Nous pouvons constater que le nombre de dépôts en ligne a augmenté en 2007, surtout dans Hal ou arXiv (les chercheurs âgés de plus de 50 ans sont les moins nombreux à déposer dans Hal et les plus nombreux à déposer dans arXiv). Les archives ouvertes (Hal et arXiv) sont utilisées surtout par les mathématiciens. Les informaticiens, en majorité, auto-archivent leurs articles sur leurs pages Web personnelles.

Ensuite, nous avons voulu connaître l'opinion des chercheurs sur l'ergonomie de Hal et d'arXiv (tableau 8).

Tableau 8

Facilité de dépôt	Dans Hal		Dans arXiv	
	2005	2007	2005	2007
Très facile	4 %	9,5 %	16,4 %	20,5 %
Facile	15,6 %	26,8 %	25,8 %	36,3 %
Un peu difficile	4 %	7,4 %	7,8 %	8,4 %
Difficile	2 %	6,3 %	2,3 %	3,2 %
Sans opinion	74,4 %	50 %	47,7 %	31,6 %

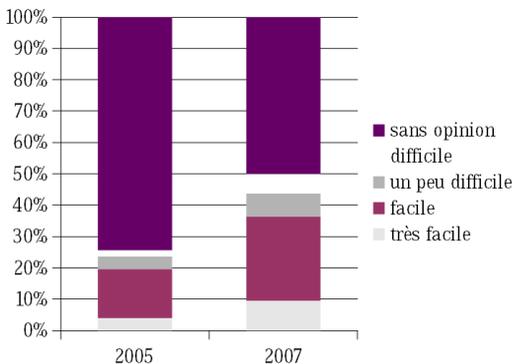


Figure 30.
Facilité d'utiliser Hal

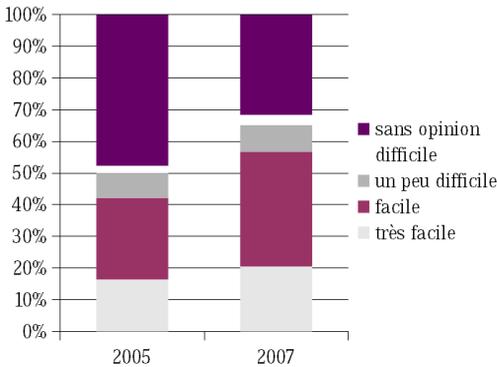


Figure 31.
 Facilité d'utiliser ArXiv

En ce qui concerne le temps nécessaire pour effectuer un dépôt dans Hal, nous avons remarqué qu'il a diminué : le premier dépôt prend moins de 30 minutes pour 34,5 % en 2005 et 42,4 % en 2007, et le dépôt suivant moins de 15 minutes pour 18,5 % en 2005 et 30,5 % en 2007.

Dans arXiv, le premier dépôt prend moins de 30 minutes pour 41,3 % en 2005 et 44,5 % en 2007, et le dépôt suivant moins de 15 minutes pour 20 % en 2005 et 31 % en 2007.

LES CONNAISSANCES DES JOURNAUX EN ACCÈS LIBRE

La publication dans des revues électroniques librement consultables constitue la deuxième forme du libre accès à l'information scientifique et augmente significativement la visibilité des auteurs.

À la question *Avez-vous soumis un manuscrit dans un journal en accès libre pendant les trois dernières années ?*, seulement 17 % (22 personnes) en 2005 et 15 % (29 personnes) en 2007 donnent une réponse positive. Il s'agit, à 90 %, de mathématiciens.

Les chercheurs ont été invités à indiquer les raisons pour publier dans un journal libre (tableau 9).

Tableau 9

Raisons de publier dans un journal libre	2005	2007
1. L'accès libre pour tous les lecteurs.	63,6 %	66,7 %
2. Ce journal est prestigieux dans mon domaine.	36,4 %	40 %
3. Ce journal est édité par les gens que je connais.	27,3 %	20 %
4. Je me sens concerné par les frais des abonnements de ma bibliothèque.	22,7 %	43,3 %

5. La publication des articles est plus rapide.	22,7 %	36,7 %
6. Je suis contre les publications dans des journaux commerciaux.	27,3 %	10 %
7. J'étais attiré par l'éditeur ou le comité éditorial.	18,2 %	23,3 %
8. Le lectorat est plus large que dans le cas d'un journal normal (commercial).	9 %	7 %
9. Je pense que mon article va être cité plus souvent.	4,5 %	6,7 %
10. Autres.	4,5 %	3,4 %
11. J'étais influencé par mes collègues.	4,5 %	0 %
12. J'étais influencé par mon laboratoire.	0 %	0 %

Parmi les chercheurs invités à indiquer les raisons pour publier dans un journal en accès libre, ceux qui le font sont généralement motivés par le principe même d'accès libre (64 % en 2005 et 67 % en 2007) et par une bonne réputation de ce journal dans leur domaine (36 % en 2005 et 40 % en 2007).

Nous avons aussi demandé aux personnes qui n'ont pas publié dans des revues libres d'indiquer les raisons de ne pas l'avoir fait (tableau 10).

Tableau 10

Raisons de ne pas publier dans un journal libre	2005	2007
1. Je ne connais pas assez bien les journaux libres dans mon domaine pour y déposer un article.	72 %	66,4 %
2. Les journaux libres dans mon domaine ne sont pas considérés comme assez prestigieux.	10 %	21,7 %
3. Je suis contre le principe auteur-payeur.	7 %	17,5 %
4. Je n'ai pas trouvé de journaux libres dans mon domaine.	4 %	10,5 %
5. Je publie toujours dans les mêmes journaux et je suis satisfait.	3 %	7,7 %
6. Ma décision a été influencée par mes collègues.	2 %	1,4 %
7. Le lectorat d'un journal libre est moins important.	2 %	1,4 %
8. Autres.	2 %	7,7 %
9. Ma décision a été influencée par mon laboratoire.	1 %	0 %
10. La publication des articles est moins rapide que dans des journaux traditionnels.	0 %	0,7 %
11. Je n'ai pas trouvé de financement pour publier dans un journal libre.	0 %	1,4 %

Le pourcentage a été calculé par rapport aux 143 personnes (100 en 2005) qui ont déclaré ne pas avoir publié dans des journaux libres.

Les raisons principales de ne pas publier dans un journal en accès libre sont que les chercheurs ne connaissent pas de journaux libres dans leur domaine (72 % en 2005 et 66 % en 2007) et, dans une moindre mesure, qu'ils craignent que les journaux libres de leur domaine ne soient pas considérés comme assez prestigieux (10 % en 2005 et 22 % en 2007) ou qu'ils sont contre le principe auteur-payeur (7 % en 2005 et 17,5 % en 2007). Autre raison : les articles refusés.

Les revues en accès libre sont mal connues et pourtant il en existe en mathématiques plus de 100 titres (< <http://www.doaj.org/> >). Le modèle économique « auteur-payeur » qui est appliqué dans d'autres pays n'a pas la faveur en France et, probablement, il est mal connu.

Nous avons enfin voulu connaître les intentions des chercheurs par rapport aux publications dans des journaux libres. D'où la question : *Si vous n'avez jamais publié dans un journal libre, est-ce que vous envisagez de le faire ?* (tableau 11).

Tableau 11

Publication future dans un journal libre	2005	2007
Très probablement	12,5 %	12,1 %
Probablement	21,9 %	21,1 %
Non	4,7 %	6,8 %
Je ne sais pas	38,3 %	44,7 %
Pas de réponse	22,6 %	15,3 %

34,4 % des chercheurs envisagent de publier dans un journal libre (les mêmes chiffres pour 2005 et 2007) dans le futur et environ 40 % ne savent pas encore s'ils vont le faire (surtout les chercheurs âgés de plus de 50 ans).

COMPARAISON AVEC LES ÉTUDES ANGLO-SAXONNES

+++++
Nos enquêtes nous ont donné l'occasion de voir l'évolution des comportements des usagers dans le temps et permettent une comparaison avec les études anglo-saxonnes.

L'enquête la plus riche en information est celle de Swan et Brown qui montre les pratiques des chercheurs de 15 disciplines. Il faut néanmoins tenir compte du fait que les données de l'enquête de Swan et Brown concernent l'année 2004 et, dans ce cas, nous pouvons les comparer seulement avec les résultats de notre enquête de 2005.

Ensuite, les questions posées aux chercheurs ne sont pas tout à fait les mêmes, ce qui ne permet toujours pas de faire des comparaisons. Ainsi, il n'est pas évident de pouvoir comparer les réponses données par une communauté avec les réponses obtenues de l'ensemble de 15 disciplines.

La comparaison de la communauté des mathématiciens et des informaticiens de nos enquêtes avec toutes les communautés scientifiques (confondues) chez Swan et Brown

+++++

Sources d'information sur les archives ouvertes

Dans nos enquêtes, les collègues sont la source principale d'information sur l'existence des archives ouvertes institutionnelles pour 42 % des mathématiciens et informaticiens.

Dans l'enquête de Swan et Brown, pour les chercheurs de tous les domaines confondus, les collègues (22 %) et le suivi des débats sur l'accès libre (21 %) constituent les sources principales d'information sur les archives ouvertes.

Motivations des chercheurs pour les dépôts dans les archives ouvertes

Pour les chercheurs interrogés dans nos enquêtes (55 %), ainsi que pour ceux interrogés par Swan (67 %), c'est le principe même des archives ouvertes qui est donné comme la motivation principale.

Objectifs de publication

Communiquer les résultats à la communauté est l'objectif principal de publication dans toutes les enquêtes (pour 86 % dans nos enquêtes et 92 % chez Swan et Brown : tous les domaines confondus).

Mais, en ce qui concerne les autres objectifs, il y a de grandes différences dans les réponses données : avancer dans la carrière (80 % chez Swan et Brown contre 37,5 % dans nos enquêtes) ; le prestige personnel dans le domaine (76 % chez Swan et Brown contre 17 % dans nos enquêtes) ; et augmenter les chances d'obtenir un financement (64 % chez Swan et Brown contre 9 % dans nos enquêtes).

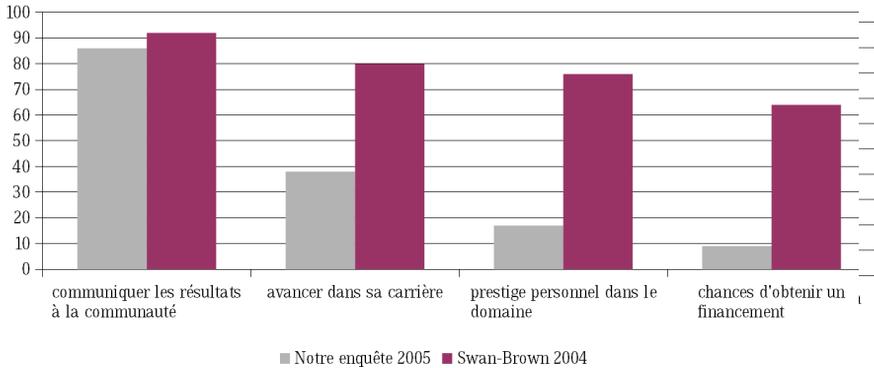


Figure 32. Objectifs de publication

Cette disparité entre les réponses dans ces deux enquêtes mériterait certainement une étude approfondie permettant d'expliquer une éventuelle spécificité de la communauté des mathématiciens et informaticiens français par rapport aux scientifiques anglo-saxons.

Droit d'auteur sur le dernier article publié¹²

Chez Swan et Brown (tous les domaines confondus), les chercheurs sont plus nombreux à garder le droit d'auteur (35 % par rapport au 5,5 % dans nos enquêtes) et plus nombreux à demander la permission pour auto-archiver (17 % contre 8 % dans nos enquêtes).

La comparaison de la communauté des mathématiciens et des informaticiens de nos enquêtes avec la même communauté et celle des sciences sociales dans l'enquête de Swan et Brown

+++++

Tableau 12

	Notre enquête (2005)	Swan et Brown : mathématiciens et informaticiens (2004)	Swan et Brown : sciences sociales (2004)
Types de publications déposées :			
- articles évalués	65 %	26 %	42 %
- prépublications	77 %	36 %	32 %
- rapports techniques	13 %	21 %	34 %

12. Selon le droit anglo-saxon, la notion de « copyright » ne recouvre que la part patrimoniale du droit d'auteur : < <http://www.law.cornell.edu/uscode/17> >.

Raisons pour publier dans des revues en accès libre :			
– le principe même de l'accès libre	64 %	73 %	77 %
– la bonne réputation du journal	36 %	45 %	36 %
– la publication des articles est plus rapide	23 %	50 %	50 %
– le lectorat est plus large	9 %	50 %	59 %
– l'article va être cité plus souvent	5 %	40 %	50 %
Raisons principales pour ne pas publier dans des revues en accès libre :			
– ne savent pas ce qu'est une revue en accès libre	72 %	26 %	37 %
– n'ont pas trouvé de journaux libres dans leur domaine	4 %	22 %	29 %

Types de publications déposées

Les résultats de nos enquêtes montrent que les chercheurs en mathématiques et informatique déposent plus d'articles référés et de prépublications, mais moins de rapports techniques que leurs collègues anglo-saxons (tableau 12).

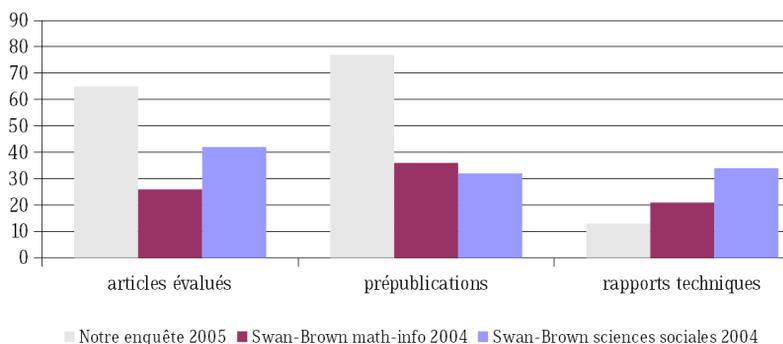


Figure 33. Types de publications déposées

Publications dans des journaux ouverts

Généralement, les chercheurs qui publient dans des journaux ouverts sont motivés par le principe même de l'accès libre et par une bonne réputation de ce journal dans leur domaine (tableau 12).

Mais nous avons remarqué aussi des différences : chez Swan et Brown, les chercheurs sont très nombreux à penser que la publication des articles est plus rapide, que le lectorat est plus large que dans le cas d'un journal commercial et que l'article va être cité plus souvent.

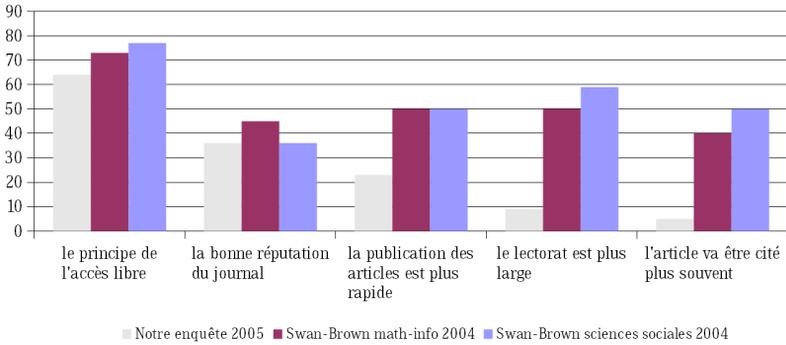


Figure 34. Raisons pour publier dans des revues en accès libre

Les raisons principales de ne pas publier dans un journal en accès libre ne sont pas tout à fait les mêmes (tableau 12). Dans nos enquêtes, essentiellement, c'est parce que ces chercheurs ne savent pas ce que c'est qu'une revue en accès libre. Chez Swan et Brown, essentiellement, les chercheurs n'ont pas trouvé de journaux libres dans leur domaine.

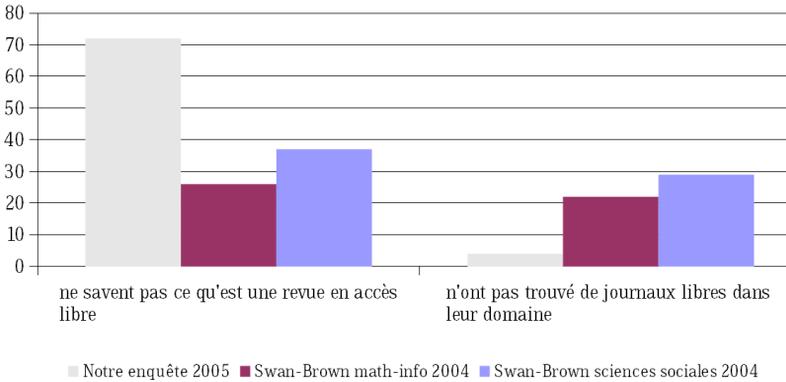


Figure 35. Raisons principales pour ne pas publier dans des revues en accès libre

SYNTHÈSE DES ENQUÊTES

+++++

Ces enquêtes nous ont donné l'occasion de voir le changement dans le temps des comportements des usagers en ce qui concerne l'évolution de leurs pratiques de recherche d'information, des pratiques de lecture,

d'auto-archivage ainsi que l'évolution d'intégration des dépôts dans les archives ouvertes.

Malgré la taille réduite de l'échantillon, l'analyse comparative des données des enquêtes 2005 et 2007 permet de faire quelques remarques.

Recherche de l'information

+++++

Dans leur majorité, les chercheurs trouvent dans les bibliothèques les articles (ou leurs références) nécessaires à leur travail, mais ils consultent souvent les journaux en ligne en texte intégral. Même si 80 % des personnes interrogées trouvent les articles (ou leurs références) dans la bibliothèque du laboratoire, déjà 52 % citent les bases de données comme source d'information, et les journaux en texte intégral sont de plus en plus consultés : 47 % des chercheurs trouvent les articles dans SpringerLink et 43 % dans ScienceDirect.

32 % donnent d'autres sources d'articles, comme les pages Web personnelles, arXiv, Hal, MathDoc¹³, Cedram¹⁴, JSTOR¹⁵, CiteSeer¹⁶, ou le contact direct avec les auteurs.

Depuis la création des portails de l'Inist¹⁷, ces accès aux journaux sont beaucoup utilisés : 5,5 % en 2005 et 24 % en 2007.

L'objectif des chercheurs quand ils interrogent la littérature scientifique est de trouver le texte intégral de l'article dont ils ont besoin. Les 56 % qui utilisent des bases de données, comme MathSciNet ou Zentralblatt, regrettent qu'elles contiennent surtout des notices bibliographiques, et que l'accès au texte intégral soit très limité.

Pour accéder aux articles en texte intégral en accès libre, les chercheurs en majorité utilisent arXiv (pour accéder aux prépublications électroniques) et Google (pour la recherche des pages Web personnelles). Les mathématiciens âgés de plus de 30 ans utilisent plus souvent arXiv que les mathématiciens âgés de moins de 30 ans, et l'ensemble des informaticiens préfèrent, quant à eux, utiliser Google.

13. MathDoc : < <http://math-doc.ujf-grenoble.fr/> >.

14. Centre de diffusion de revues académiques mathématiques (Cedram) :
< <http://www.cedram.org/> >.

15. JSTOR (The Scholarly Journal Archive): < <http://www.jstor.org/> >.

16. CiteSeer (Scientific Literature Digital Library): < <http://citeseer.ist.psu.edu/> >.

17. Institut de l'information scientifique et technique (Inist) : < <http://www.inist.fr/> >.

Les articles disponibles en ligne en texte intégral les plus consultés (au moins une fois par semaine) ont été publiés pendant les dix dernières années. Les possibilités croissantes d'accès en ligne permettent une amélioration considérable de l'accès à l'information scientifique publiée.

PUBLICATIONS

+++++
55 % des chercheurs publient entre 2 et 3 articles par an. La majorité les archive sur des pages Web personnelles et 38 % déclarent le faire depuis au moins cinq ans.

Une partie des chercheurs qui déposent des articles sur leurs sites Web personnels le font également (en partie) dans des archives institutionnelles. Les chercheurs déposent en 2007 plus d'articles dans Hal (13 % en 2005 et 24 % en 2007) ou arXiv (16 % en 2005 et 26 % en 2007), mais ils déposent toujours beaucoup plus d'articles sur leurs pages Web personnelles (60 %).

Les personnes qui ont déjà effectué des dépôts de publications dans Hal ou arXiv trouvent facile l'utilisation de ces outils et précisent qu'il leur faut moins de 30 minutes pour le premier dépôt et moins de 15 minutes pour le dépôt suivant. Le nombre de ces personnes a fortement augmenté en 2007. Les chercheurs âgés de plus de 50 ans sont les moins nombreux à déposer dans Hal et les plus nombreux à déposer dans arXiv. Les archives ouvertes (Hal et arXiv) sont utilisées surtout par les mathématiciens.

Les auteurs affirment être sensibilisés au libre accès et à l'existence des archives ouvertes institutionnelles surtout grâce à la communication entre eux.

AUTO-ARCHIVAGE

+++++
La première enquête montre encore une très faible participation aux dépôts dans les archives ouvertes, une réticence de la part des chercheurs liée aux problèmes techniques et sociaux (les habitudes du modèle traditionnel de la publication, une absence de communication institutionnelle, des contextes locaux).

Les résultats de la deuxième enquête montrent déjà une appropriation des techniques par les usagers, l'influence (parfois une pression) de la part des autorités. Nous pouvons voir une progression des usages et l'intégration des pratiques des dépôts.

Si les pratiques d'archivage commencent à faire partie des coutumes des chercheurs en mathématiques et informatique, il s'agit ici du « réflexe Google » : ils déposent leurs publications sur leurs sites personnels en s'assurant que l'on parvient à les retrouver grâce au moteur de recherche le plus utilisé, même si l'utilité des archives ouvertes institutionnelles est déjà un peu mieux comprise.

L'évolution actuelle du mouvement des archives ouvertes demande une plus grande visibilité des informations officielles provenant du CNRS ou du ministère. La pratique a démontré que chaque fois qu'une information concernant Hal est transmise, le nombre de dépôts augmente ; à l'inverse, pendant les périodes de « silence », une baisse des dépôts est remarquée.

REVUES EN ACCÈS LIBRE

+++++

Les sources des articles en accès libre ne sont toujours pas très bien connues – comme les journaux libres.

Même si, selon une étude menée par le Centre for Information Behaviour and the Evaluation of Research (CIBER)¹⁸, de plus en plus de scientifiques publient dans des journaux en libre accès, l'appropriation de ce nouveau mode de communication demeure lente.

En attendant, il faudra diffuser plus d'information concernant des revues en accès libre, par exemple l'adresse de « Journal Info »¹⁹ (des informations sur le type d'accès, le coût ou encore sur les titres alternatifs en accès libre pour 690 titres en mathématiques). Il s'agit en grande partie d'une mission pour les bibliothèques de diffuser des informations provenant des éditeurs de ces journaux. Nous avons aussi remarqué une réticence envers le modèle de l'auteur-payeur.

COMPARAISON AVEC LES ÉTUDES ANGLO-SAXONNES

+++++

Notre étude semble montrer que les sources d'information sur les archives ouvertes, les motivations des chercheurs et leur objectif principal pour déposer leurs articles en accès libre sont les mêmes.

18. CIBER : < http://www.ucl.ac.uk/ciber/ciber_2005_survey_final.pdf >.

19. Journal Info : < <http://jinfo.lub.lu.se/jinfo?func=subject&slid=223> >.

Mais lorsque l'on s'intéresse aux objectifs secondaires, cela permet de mettre en évidence des spécificités de la communauté des mathématiciens et informaticiens français.

Cette comparaison confirme encore une fois que les chercheurs de nos enquêtes méconnaissent tant les aspects juridiques de la publication scientifique que les revues en accès libre.

RÉFÉRENCES BIBLIOGRAPHIQUES

+++++

Allen James, *Interdisciplinary Differences in Attitudes Towards Deposit in Institutional Repositories*, 2005.

[En ligne] < <http://eprints.rclis.org/archive/00005180/> >.

JISC *Disciplinary Differences Report*, 2005.

[En ligne] < http://www.jisc.ac.uk/uploaded_documents/Disciplinary%20Differences%20and%20Needs.doc >.

Swan Alma, *JISC Open Access Briefing Paper. Technical Report*, JISC, HEFCE, 2005a.

[En ligne] < <http://eprints.ecs.soton.ac.uk/11005/> >.

Swan Alma, *Open Access Self-Archiving: an Introduction. Technical Report*, JISC, HEFCE, 2005b.

[En ligne] < <http://eprints.ecs.soton.ac.uk/11006/> >.

Swan Alma et Brown Sheridan, "Authors and open access publishing", *Learned Publishing*, vol. 17, n° 3, 2004, pp. 219-224.

[En ligne] < <http://eprints.ecs.soton.ac.uk/11003/> >.

Swan Alma et Brown Sheridan, "Open access self-archiving: an author study", *Technical Report, External Collaborators, Key Perspectives Inc.* 2005.

[En ligne] < <http://eprints.ecs.soton.ac.uk/10999/> >.

Wojciechowska Anna, *Archives ouvertes : état des lieux et pratiques dans les domaines des cas des mathématiques et de l'informatique*, Thèse SIC, Marseille, Université Aix-Marseille III, 2008, p. 263.

par Ilham Derfoufi

+++++

PRATIQUES
NUMÉRIQUES DE
CHERCHEURS :
REFLET DE LA
DISCIPLINE, L'EXEMPLE
DES SCIENCES DE
L'ÉDUCATION

+++++

+++++

PRATIQUES NUMÉRIQUES DE CHERCHEURS : REFLET DE LA DISCIPLINE, L'EXEMPLE DES SCIENCES DE L'ÉDUCATION

+++++

Inspiré de ma thèse intitulée *Information scientifique pour la recherche en éducation : analyse comparée de l'offre et des usages entre la France et le Royaume-Uni*, cet article aborde la question de l'influence de la discipline sur les pratiques numériques des chercheurs en éducation. L'objectif de notre analyse des usages était d'identifier les facteurs favorisant ou décourageant l'usage de l'information numérique. Des enquêtes nous ont ainsi permis de vérifier que la nature de l'offre ainsi que l'âge et la culture technique des chercheurs influençaient cet usage. Dans le contexte français, le non-usage répond à une inadéquation de l'offre numérique essentiellement anglophone pour un public maîtrisant peu l'anglais mais également à une faible motivation à la formation (fin de carrière, mauvaise conception et langue des outils de recherche), ce qui entraîne une faible culture documentaire numérique. Toutefois, nos enquêtes ont révélé l'existence d'un autre facteur majeur d'usages : la discipline.

La connaissance de la discipline permet de comprendre les pratiques informationnelles de ses chercheurs. Il s'agit donc d'explorer le domaine scientifique, de comprendre son organisation intellectuelle et la manière dont se fait la recherche en procédant à une analyse de domaine. Cette approche est celle adoptée par de nombreux chercheurs en sciences de l'information : Hjørland et Albrechtsen, Talja et Maula, Fry, Palmer, Bates, et bien d'autres... L'étude des disciplines est aussi la base de la théorie du sociologue anglais Richard Whitley.

Nous proposons dans un premier temps de donner un aperçu de la discipline et de la démarche scientifique dans ce domaine. Ensuite, sur la base des enquêtes réalisées en France et au Royaume-Uni, quelques pratiques informationnelles de chercheurs dans cette discipline seront décrites. Enfin, ces données permettront l'analyse de domaine, un nouveau cadre d'interprétation des pratiques informationnelles.

LES SCIENCES DE L'ÉDUCATION, UNE DISCIPLINE ?

+++++

Nées en 1967 après bien des débats sur leur constitution, les sciences de l'éducation parviennent à s'établir comme une véritable discipline universitaire. Cependant, leur reconnaissance institutionnelle par certains organismes de recherche comme le CNRS n'est pas encore acquise [Plaisance et Vergnaud, 2005 ; Guibert, 2006]. Au moins deux facteurs y contribuent. D'abord, leur unité : les sciences de l'éducation marquent dès le début du xx^e siècle une rupture avec « la » science de l'éducation et confirment dès lors des approches plurielles à l'étude des objets d'éducation. Constituent-elles une véritable discipline ou seulement des *appendices* d'autres disciplines [Mialaret, 2006] ? La diversité des approches, des méthodologies utilisées, des objets étudiés sème le doute sur leur unité qui, de fait, menacerait leur identité [Guibert, 2006]. Certains pourtant s'appuient sur leur pluralité pour démontrer la nécessité de leur unité : les disciplines s'unissent pour étudier un même objet. « L'unité du champ s'organise autour d'un objet d'étude pluriel » [Chatelanat *et al.*, 2004, p. 5]. Les sciences de l'éducation donnent l'image d'un champ dispersé dont l'identité est encore vague. Leur légitimité est également mise en cause. Elles ont été décrites d'« habits d'Arlequin », d'« agrégats morcelés », et tantôt associées à des sciences de « contrebande » ou de « contrefaçon », tantôt à des sciences « au rabais », des sciences « illégitimes » et même de « fausses » sciences [Marchive, 1998]. Le refus de légitimité serait le fruit de leur relative jeunesse [Mondo, 2003], elles n'auraient pas encore fait leurs preuves [Guibert, 2006]. Outre encore les critiques qu'elles reçoivent des politiques, des enseignants et du grand public, ce sont les scientifiques qui leur refusent une légitimité scientifique. De nombreux chercheurs¹ ont évoqué la question de la scientificité des sciences de l'éducation [Hadji et Baillé, 1998]. Le débat est riche et passionnant, retenons simplement que la discipline semble encore tourmentée, peu stabilisée, fragile et en quête d'une véritable reconnaissance institutionnelle et scientifique.

La démarche scientifique

+++++

La démarche scientifique en sciences de l'éducation est rarement décrite. « La finalité principale de la recherche en sciences de l'éducation est de

1. Yves Chevallard, Marc Bru, Michel Develay, Charles Hadji, Nadine Charbonnel, Marie Duru-Bellat, Alain Mingat, Jacques Baillé, etc.

nous permettre de mieux connaître, de mieux expliquer, de mieux comprendre les faits et les situations d'éducation qui nous entourent. En fait c'est augmenter, développer, enrichir et/ou préciser notre savoir actuel en ce domaine. » [Mialaret, 2006]. Pour atteindre cette finalité, le chercheur doit rassembler un maximum d'informations sur les techniques d'observation et de description des situations éducatives afin de permettre leur compréhension maximale car les objets de recherche, que sont les situations observées, sont complexes. À la différence des disciplines des sciences de la nature qui cherchent à « expliquer »² un phénomène, les sciences humaines cherchent d'abord à « comprendre » une situation avant de l'expliquer. En éducation, « l'intuition dans l'interprétation des résultats » a son importance. « Le chercheur doit être capable de dire qu'à un certain moment de sa recherche, tel ou tel résultat lui a donné l'intuition que ce résultat pouvait être expliqué de telle ou telle façon, et que cette intuition a donné lieu à un nouveau maillon de la chaîne expérimentale » [Mialaret, 2004]. Cependant, pour obtenir un consensus sur l'explication d'une situation observée, l'ensemble des observateurs doivent accepter les méthodes d'évaluation des différentes variables relatives à cette situation. Ainsi, plusieurs techniques de recherche sont utilisées, certaines reflètent les traditions de travail des disciplines mères (philosophie, psychologie, sociologie), mais leur adoption pour l'étude d'une situation d'éducation doit être discutée puis approuvée par l'ensemble du groupe de recherche, ce qui entraîne souvent, selon Gaston Mialaret, des divergences parmi les chercheurs. La démarche scientifique en sciences humaines se caractérise également par l'incertitude des résultats de la recherche. En effet, à l'inverse des disciplines des sciences exactes où la *quasi-certitude des résultats* d'une expérimentation est garantie, les sciences humaines ne peuvent proposer une « vérité » absolue. L'infinité des variables qui constituent une situation d'éducation fait *qu'elle ne se reproduit jamais à l'identique*.

Outre les nombreuses contraintes qui se posent au chercheur lors de l'étude d'une situation d'éducation, la collaboration avec ses partenaires semble aussi problématique. Il collabore avec les praticiens car sans l'accord de ces derniers les expérimentations ne peuvent se faire, puis avec d'autres chercheurs de sa discipline ou d'autres disciplines des sciences de l'éducation. La littérature rapporte abondamment les conflits

2. Il s'agit du point de vue de Wilhelm Dilthey (philosophe, sociologue et historien allemand) cité par Gaston Mialaret [2006].

qui peuvent naître du premier type de collaboration mais beaucoup moins du second, celui de la collaboration interdisciplinaire.

Des sciences entre pluridisciplinarité et transdisciplinarité

++++
Les sciences de l'éducation seraient à la fois multidisciplinaires, pluridisciplinaires, transdisciplinaires et interdisciplinaires. Ces notions décrivent les types de relations qui existent ou pourraient exister entre les disciplines [Benoist *et al.*, 1983]. L'idéal en sciences de l'éducation serait que toutes ces relations soient présentes : « On connaît la distinction entre la multidisciplinarité – correspondant à la multiplicité des contenus et des méthodes d'exploitation du fait éducatif –, la pluridisciplinarité – correspondant au découpage d'un objet commun en fonction des divers éclairages scientifiques –, l'interdisciplinarité – supposant des incursions dans des disciplines voisines de la sienne –, et la transdisciplinarité enfin qui en cherchant à "débusquer les invariants" permet d'aller au-delà des disciplines de départ. Là est sans doute le chemin. » [Marmoz, 1988, p. 85]. Voyons de plus près ce que toutes ces notions signifient.

La pluridisciplinarité

Les sciences de l'éducation ont souvent été considérées comme le *champ d'application des diverses disciplines* [Marmoz, 1988]. Bien que cette image soit quelque peu réductrice, elle véhicule bien cette notion de pluridisciplinarité. « Les sciences de l'éducation précisément parce qu'elles importent une partie de leurs concepts, de leurs systèmes de référence et de leurs méthodologies à partir des acquis respectifs des différentes sciences humaines et sociales sont nécessairement pluridisciplinaires et (ou) interdisciplinaires. La complexité de leur objet les y oblige. » [Ardoino, In : Marmoz, 1988, p. 83]. C'est donc la nature de l'objet, le fait éducatif, et plus précisément sa complexité qui d'emblée induisent la pluridisciplinarité et l'interdisciplinarité dans le domaine de l'éducation. Nous reviendrons sur l'interdisciplinarité. La notion de pluridisciplinarité est néanmoins employée avec beaucoup de prudence. Certains insistent sur le fait que la pluridisciplinarité des sciences de l'éducation ne doit pas faire oublier leur unité. « Il est très bien de cerner un objet de différentes façons, encore faut-il être certain d'avoir cerné le même objet et que la jonction entre les analyses se fasse. » [Marmoz, 1988, p. 84]. L'indépendance dans les analyses des questions éducatives des différentes disciplines est aussi évoquée par Michel Develay : « Chaque discipline

conduit son projet individuel » de même que le danger de se contenter de cette seule approche pluridisciplinaire, « Ce risque d'aveuglement de chaque spécialité est celui de l'imperméabilité des disciplines les unes aux autres et celui d'une parcellisation des territoires à l'infini, conduisant à une cécité aux grands buts de l'activité humaine centrée sur la question éducative. » [Develay, 2004, p. 56]. La pluridisciplinarité peut également s'appliquer au sein d'une même discipline. Chaque discipline des sciences de l'éducation peut avoir différentes approches pour étudier un fait éducatif. La psychologie de l'éducation, par exemple, serait « écartelée entre des approches cognitivistes et relationnelles » qui pourraient entraîner l'éclatement du domaine [Develay, 2004, p. 56]. Mais la pluridisciplinarité ne doit pas refléter qu'une *juxtaposition de travaux épars* [Marmoz, 1988]. « Il ne s'agit pas de les [les contributions des disciplines] juxtaposer mais bien de passer d'une pluridisciplinarité à une interdisciplinarité. »³

L'interdisciplinarité

Correspondant au concept de *multiréférentialité* de Jacques Ardoino⁴, la notion d'interdisciplinarité implique la recherche de la complémentarité entre les différentes approches disciplinaires dans la compréhension d'un fait éducatif. « L'ambition de l'analyse multiréférentielle est de conjuguer ces savoirs hétérogènes⁵ dans le but d'appréhender la compréhension de situations éducatives [...] » [Develay, 2004, p. 57]. La complémentarité permet une plus grande richesse des savoirs qui vient des *débordements et transgressions interdisciplinaires*⁶ [Marmoz, 1988]. La recherche sur un fait éducatif est enrichie par les différentes approches disciplinaires. Develay donne ainsi l'exemple de « recherches de sociologues de l'éducation sur l'établissement scolaire qui se sont bonifiées d'approches de didacticiens ». Mialaret évoque l'interdisciplinarité avec l'introduction de l'ordinateur dans le système éducatif qui a créé de nouvelles situations d'éducation et dont les problèmes génèrent de nouveaux champs d'investigation pour la discipline mère. Il parle aussi de deux modalités d'interdisciplinarité : celle « de coopération » qui va consister à unifier « les méthodes, les techniques et les processus d'interprétation des faits », mais qui posera problème au moment de l'interprétation des faits « puisque

3. Avanzini Guy, « Philosophie de l'éducation », in Marmoz, 1988, p. 66.

4. Ardoino Jacques, 1977, in Develay, 2004.

5. L'auteur parle des savoirs du psychologue, de l'économiste et de l'historien.

6. Les termes en italique sont empruntés à Jacques Ardoino.

les approches vont s'entrecroiser » ; puis celle « de création » qui est envisagée comme un second niveau. Les interprétations ne seront plus juxtaposées mais « constitueront un nouvel objet d'analyse scientifique, considéré essentiellement sous l'angle de très nombreuses interrelations qui se développent entre tous les partenaires, et comme l'émergence d'une nouvelle réalité humaine spécifique au domaine de l'éducation ». Ce niveau d'interdisciplinarité idéal dans la recherche nous semble proche de la *transdisciplinarité*.

La transdisciplinarité

Toutes les définitions de la transdisciplinarité convergent vers le même objectif, à savoir que l'ensemble des disciplines doivent parvenir à avoir un paradigme commun. Tom Burton Bottomore⁷ distingue « interdisciplinarité » et « transdisciplinarité » : « La première implique la rencontre et la coopération entre deux disciplines ou plus, chacune de ces disciplines apportant (au niveau de la théorie ou de la recherche empirique) ses propres schémas conceptuels, sa façon de définir les problèmes et ses méthodes de recherche. En revanche, la seconde implique que contact et coopération entre diverses disciplines ont lieu surtout du fait que ces disciplines ont fini par adopter un même ensemble de concepts fondamentaux ou quelques éléments d'une même méthode de recherche, pour parler de manière plus générale, le même paradigme ». Une autre définition montre bien ce niveau supérieur de l'interdisciplinarité : « La transdisciplinarité est le champ de l'échange de schèmes abstraits des dialogues interdisciplinaires locaux dès qu'il y a généralisation ou possibilité d'exportation de schèmes ou de relations conceptuelles » [Benoist⁸, 1983]. Il semble, cependant, qu'en sciences humaines et sociales, ce degré d'interdisciplinarité est rarement atteint. Edgar Morin et Massimo Piatelli-Palmarini⁹ estiment que s'il est possible de développer une certaine forme d'interdisciplinarité, la transdisciplinarité est un objectif bien plus difficile à atteindre. La recherche en sciences de l'éducation sera donc davantage pluridisciplinaire ou interdisciplinaire, et il faudra en tenir compte pour étudier les pratiques informationnelles des chercheurs.

7. Tom Burton Bottomore était un sociologue marxiste anglais.

8. Jean-Marie Benoist était un philosophe, écrivain et universitaire français.

9. Edgar Morin est chercheur et directeur émérite au CNRS, Massimo Piatelli-Palmarini est spécialiste des sciences cognitives et directeur de recherche au Massachusetts Institute of Technology à Boston. Ces auteurs ont apporté leur contribution dans *Interdisciplinarité et sciences humaines* [Benoist *et al.*, 1983].

PRATIQUES ET DISCIPLINE

+++++

Les pratiques informationnelles ont été analysées à travers deux séries d'enquêtes (quantitatives et qualitatives) en France et au Royaume-Uni en 2007. La comparaison a révélé des pratiques communes mais divergentes également. En sciences de l'éducation, les chercheurs viennent d'horizons divers : philosophie, sociologie, histoire, économie... Comme les sciences de l'information, les sciences de l'éducation sont une discipline jeune où des méthodes de travail aussi différentes que celles des sciences de la nature et des sciences humaines et sociales se rencontrent. Ghislaine Chartron [2003] avait d'ailleurs évoqué cette particularité des sciences jeunes dans son article « Éléments pour une approche comparée de la publication scientifique ». Selon elle, certains chercheurs favorisent la publication dans des revues nationales, voire internationales, alors que d'autres, comme les historiens, privilégient la monographie. Ces orientations se reflètent sur les pratiques de lecture.

L'objectif de nos enquêtes était de sonder les chercheurs sur leur environnement informationnel d'une manière générale et plus spécifiquement sur certains composants de cet environnement : ressources consultées, services utilisés, etc. Nous évoquerons ici quelques résultats pertinents pour l'analyse de domaine. D'abord, l'environnement informationnel des chercheurs : si, en France, les bibliothèques en constituent une bonne partie, au Royaume-Uni, cet environnement est essentiellement numérique. Le contexte documentaire est, en effet, sensiblement différent. La documentation de recherche utilisée par une majorité des chercheurs francophones est principalement de langue « française », et une grande partie de cette documentation est sous format papier (ouvrages, revues, rapports officiels, données de terrain). Le chercheur est donc souvent contraint de se déplacer en bibliothèque pour pouvoir la consulter. Celle utilisée par les chercheurs anglophones – des revues principalement – est, à l'inverse, en majeure partie disponible en ligne. Cependant, en France, le déplacement en bibliothèque répond aussi au rapport affectif qu'entretiennent les chercheurs avec les ouvrages, particulièrement, ainsi qu'au besoin de faire appel à des intermédiaires (bibliothécaire, documentaliste). À l'opposé, le chercheur anglophone est très autonome. La visite des bibliothèques implique le *butinage* (*browsing*) dans les rayons. Cette pratique est très répandue dans le contexte francophone et se trouve transposée dans l'environnement numérique, le *butinage* devient alors la *navigation* sur les sites Web (portails de bibliothèques, sites institutionnels, sites de conférences,

etc.) pratiquée autant par les francophones que par les anglophones. Un intérêt a également été apporté aux outils utilisés pour la recherche d'information. En France, le catalogue de la bibliothèque, de par sa simplicité, est un outil extrêmement utilisé. Pour la même raison, Google rencontre un franc succès. Il permet de rechercher les dernières parutions d'un auteur, l'actualité d'un thème... Pour les articles, la recherche se fait soit par les bases de données bibliographiques, soit en utilisant les index ou plateformes de revues (Web of Knowledge, Cairn...), surtout au Royaume-Uni. Les chercheurs y ont accès par le portail de la bibliothèque mais disposent aussi d'un accès direct sur leurs intranets. Outre la navigation sur Internet, la recherche directe est le mode le plus couramment utilisé par les chercheurs anglophones. L'utilisation des bases de données est beaucoup moindre. Pour les chercheurs francophones, l'anglais et la complexité de ces outils, comparés à Google par exemple, en sont les raisons principales. Francis, la base la plus citée, est jugée trop compliquée : « Je ne me sens pas très à l'aise avec ce type d'outil, il y a des manières plus simples de trouver les références d'un article, par exemple Google [...] » déclare un chercheur. La conception est aussi mise en cause tant en France qu'au Royaume-Uni : « ... on dirait que ces bases ne sont pas faites par des gens du domaine, avec les mots clefs que je saisis, je ne trouve jamais rien d'intéressant alors je finis par abandonner. » ; *"The one I use sometimes is Psychinfo that has really good search in terms of keyword searching that might not be keywords."*¹⁰. En effet, la recherche dans certaines disciplines (psychologie, philosophie) se ferait par concepts. Certains chercheurs francophones aborderont le problème de la langue pour justifier leur découragement. Beaucoup d'interfaces de bases proposées sont en anglais, et la recherche se fait sur des contenus en anglais ; pour la plupart, ces outils ne leur seraient d'aucune utilité. Sur le plan de la collaboration, les chercheurs ont admis avoir peu de contacts avec leurs collègues tant dans un cadre interdisciplinaire qu'à l'international du fait de la barrière de la langue. Les nouvelles technologies n'y changent rien apparemment. Cette faible collaboration entre chercheurs avait déjà été évoquée par Chartron [2003] en matière de modes de publication et de travail pour les économistes et les sociologues à la différence des psychologues qui collaborent davantage. Ces trois disciplines faisant partie

10. Extraits d'entretiens. « Celle [base de données] que j'utilise parfois est Psychinfo, qui est bien conçue en termes de possibilité de recherche par mots clés, lesquels peuvent ne pas être des mots clés. » [Nous traduisons les citations en anglais dans cet article.]

des sciences de l'éducation, la collaboration des chercheurs dépendrait alors de leur domaine de spécialité. Parmi les « nouvelles » ressources proposées, nous avons interrogé les chercheurs sur les archives ouvertes. Elles sont assez peu connues, principalement dans l'espace francophone, mais leur connaissance n'implique pas vraiment leur utilisation : plus de 70 % déclarent ne pas faire de dépôts. Sur ce point, les réactions des deux communautés de chercheurs convergent, reflétant des inquiétudes liées à la validité des contenus : « il faudrait expliquer les tenants et les aboutissants de cette démarche ». La pression à publier dans des revues prestigieuses est un autre frein au Royaume-Uni : *“there is a great pressure in universities to publish in peer reviewed journals or in high reputation publishers. It is almost a waste of time to think about open access.”*¹¹ S'ils ne déposent pas de publications dans les archives ouvertes, ils ne les consultent pas davantage. Ainsi que le souligne Chartron, les processus de production de la publication et de l'usage de l'information sont très liés : « Les sources d'information privilégiées par le chercheur sont généralement celles où il cherchera à valoriser ses travaux, à publier ». Cet éclairage sur la discipline, sa culture et sa communauté va nous permettre à présent de développer une analyse de domaine.

L'approche analytique de domaine pour les sciences de l'éducation

+++++

Sur la base de la démarche scientifique en sciences de l'éducation et ses caractéristiques, dont nous avons tenté de donner un aperçu, nous proposons de faire une analyse de domaine selon la théorie de Whitley, dans un premier temps, puis en référence aux travaux de Bates, Talja et Maula, ainsi que Fry pour tenter de comprendre les pratiques informationnelles des chercheurs de cette discipline.

L'analyse de domaine selon la théorie de Whitley

Whitley s'intéresse aux structures intellectuelle et sociale des disciplines. Il se base sur deux dimensions d'identité culturelle, *mutual dependance* (MD) et *task uncertainty* (TU), pour expliquer les différences culturelles des disciplines. Le premier aspect du concept de MD fait référence aux relations qui s'établissent entre les chercheurs d'une discipline. Whitley

11. « Il y a une grande pression dans les universités pour que les chercheurs publient dans des revues à comité de lecture ou des revues prestigieuses de grands éditeurs, c'est presque une perte de temps de penser à publier en *open access*. »

indique que la respectabilité d'une discipline dans la communauté scientifique dépend du type de relations qu'entretiennent les chercheurs entre eux au moment de faire une recherche. La reconnaissance et la respectabilité supposent l'unité du groupe disciplinaire, les chercheurs doivent adhérer à des idées, des procédures, des résultats d'une manière uniformisée. L'autre aspect de ce concept évoque à nouveau l'unité concernant le jugement des chercheurs sur la valeur d'une recherche. Les critères de qualité doivent être les mêmes pour l'ensemble des individus qui examineront une recherche. Ainsi, selon la théorie de Whitley, plus le groupe disciplinaire sera uni, plus le niveau de MD sera élevé. Appliqué à des domaines multidisciplinaires et/ou interdisciplinaires, ce concept de MD montre toute la difficulté des relations qui peuvent s'établir entre des chercheurs issus d'horizons divers [Jun, 1983]. La pluridisciplinarité et l'interdisciplinarité dans les sciences de l'éducation génèrent des relations conflictuelles comme nous l'avons évoqué. L'image davantage pluridisciplinaire du domaine accentuerait les conflits. Il y aura donc surtout des juxtapositions de travaux, d'idées, de méthodes de recherche. « D'une part, la communauté de chercheurs existe bien [...] mais elle n'est pas [...] organisée autour d'un paradigme commun (ni même dominant). La diversité des problématiques, des méthodologies, des origines (disciplinaires) même des chercheurs est telle que les sciences de l'éducation continuent d'apparaître, aujourd'hui encore, davantage selon le modèle de "l'auberge espagnole" que selon le modèle d'une véritable "communauté scientifique", structurée autour d'un paradigme dominant [...] » [Mondo, 2003]. Bien que leur unité soit souvent revendiquée [Marmoz, 1988 ; Mialaret, 2004 ; Develay, 2004], les sciences de l'éducation donnent l'image d'un champ éclaté [Mondo, 2003 ; Develay, 2004]. Quant à assurer un cadre commun d'évaluation de la recherche dans un champ ou un domaine pluridisciplinaire et/ou interdisciplinaire, cela semble bien difficile : « Le pire problème, sans doute pour le chercheur engagé dans l'aventure interdisciplinaire, est notre système d'évaluation par les pairs. En l'absence de commission *ad hoc*, les commissions par disciplines ne peuvent évaluer un travail de recherche interdisciplinaire qu'au prix du jugement séparé de plusieurs groupes d'experts. L'évaluation reste ainsi pluri-, mais jamais interdisciplinaire [...] » [Kourilsky, 2002]. Une reconnaissance commune par les pairs est donc assez rare. Whitley indique "*diversity of cognitive concerns within a discipline leads to a decreasing degree of MD, which impedes development*

of peer recognition"¹² [Fry, 2006, p. 308]. En sciences de l'éducation, il est difficile d'assurer une certaine unité dans le groupe tant au niveau des procédures de travail que dans l'interprétation des résultats. Le niveau de MD est donc faible. Si le concept de MD semble davantage concerner la posture de la communauté autour d'un projet de recherche, celui de TU fait référence plus précisément à la coordination du travail de recherche et à la production de résultats fiables, les deux concepts étant cependant intimement liés. La multiplicité des techniques de recherche issues des différentes disciplines et les divergences qui peuvent survenir au moment de faire un choix pour analyser une situation d'éducation attestent d'un problème de coordination dans la communauté. Bien qu'il soit souvent fait référence à la nécessité de parvenir à un accord avant de démarrer une recherche, on évoque aussi les divergences liées à ces accords. Un autre aspect du TU concerne la fiabilité des résultats. Mialaret décrit bien cette incertitude qui s'introduit dans l'interprétation des situations d'éducation. Ainsi, à l'inverse des disciplines des sciences exactes où la quasi-certitude des résultats est garantie, tel n'est pas le cas pour les sciences de l'éducation. Le niveau de TU est donc élevé. Whitley indique également que le nombre des institutions impliquées dans la recherche est un indicateur du niveau de TU pour la communauté. Ainsi, quand la recherche est financée par plusieurs organismes, le niveau de *strategic uncertainty* (SU), une composante du TU, est plus élevé. C'est précisément le cas des sciences de l'éducation, où plusieurs organismes sont impliqués dans la recherche et son financement [Plaisance et Vergnaud, 2005]. Le niveau de SU est également influencé par le public auquel s'adresse la recherche. Plus le public sera restreint, plus le niveau de SU sera faible. En sciences de l'éducation, la pluralité des publics fait écho à celle des différentes disciplines.

L'analyse de domaine pour les sciences de l'éducation semble bien confirmer la théorie de Whitley qui ne traite pas directement les pratiques informationnelles mais en suggère quelques idées. Dans des domaines multi- ou interdisciplinaires, l'absence de coordination, l'insuffisance de dialogue entre les disciplines impliquent le recours à diverses références disciplinaires, d'où des systèmes d'information peu élaborés. Les répondants à nos enquêtes souhaitaient avoir un accès unifié aux ressources, déplorant ainsi leur trop grande dispersion. Sur un autre registre, l'aspect de manque de reconnaissance dans une discipline pluridisciplinaire,

12. « la diversité de facteurs cognitifs au sein d'une discipline entraîne un degré décroissant de dépendance mutuelle (DM) qui empêche toute reconnaissance par les pairs ».

auquel nous avons déjà fait référence, est développé par Fry alors qu'elle analyse les collaborations des chercheurs en géographie socioculturelle. Elle indique que les notions de confiance et de reconnaissance sont importantes pour cette communauté au point d'influencer leur utilisation des nouvelles technologies dans un cadre collaboratif mais également pour la dissémination de leurs productions [Fry, 2006]. À leur image, les chercheurs des sciences de l'éducation collaborent peu et n'utilisent pas les réseaux de recherche en ligne (groupes de discussion, blogs, etc.) et nombreux sont les réfractaires aux archives ouvertes si les dépôts n'ont pas été publiés ou évalués par les pairs. Si des pratiques informationnelles peuvent être générées par l'identité culturelle des disciplines, comme le suggère Whitley, certaines s'expliquent par d'autres facteurs disciplinaires.

L'analyse de domaine selon les modèles de Bates, et de Talja et Maula

Bates, Talja et Maula se sont basées sur d'autres facteurs liés au domaine pour expliquer les techniques de recherche d'information et l'usage des revues numériques et des bases de données dans diverses disciplines. Bates indique que la *taille du domaine* et son *niveau de dispersion* influenceraient les stratégies de recherche [Fisher *et al.*, 2005]. La taille du domaine fait référence au nombre de documents pertinents pour un domaine donné. Plus le domaine sera important, plus les documents pertinents seront nombreux. "*Research areas with high numbers of topically relevant materials are best searched by browsing*"¹³ [Fisher *et al.*, 2005] du fait de la dispersion du domaine. Cette hypothèse semble se vérifier en sciences de l'éducation où le *browsing* est privilégié tant dans l'environnement traditionnel (butinage) que dans l'environnement numérique (navigation). Cette technique est d'ailleurs commune aux deux communautés de chercheurs. Le *browsing* confirme aussi la théorie de Chang¹⁴ selon laquelle elle serait utilisée pour découvrir les ressources d'un domaine que l'on maîtrise peu, une éventualité dans un champ pluridisciplinaire.

L'orientation vers l'ouvrage ou la revue a été étudiée par Kling et Covi, et Walsh et Bayma [Fisher *et al.*, 2005]. Les techniques de recherche utilisées par les chercheurs dépendraient de l'importance accordée à l'ouvrage ou à la revue, comme ressource privilégiée dans une discipline.

13. « Les domaines de recherche où le nombre de documents pertinents est important sont mieux explorés par la technique du *browsing* ».

14. "Chang's Browsing". In Fisher *et al.*, 2005.

Ainsi, les disciplines où l'article est le principal canal de publication feront davantage appel à la recherche directe. Inversement, les disciplines où les ouvrages ont davantage de prestige feraient des recherches par auteur. L'ouvrage est davantage utilisé en France et permet de comprendre pourquoi la recherche par auteur (*via* le catalogue) est fréquente. Néanmoins, notre enquête qualitative a révélé deux types de comportements, probablement liés à la disponibilité de l'offre numérique dans les régions étudiées : au Royaume-Uni, où la revue électronique est très utilisée, la recherche directe est privilégiée *via* des plateformes multidisciplinaires (Web of Knowledge) ; en France, les revues électroniques francophones étant rares, on utilisera davantage le *browsing* en alternance avec la recherche par auteur.

Talja et Maula [2003, p. 676] se sont intéressées à l'usage des revues numériques et des bases de données : "*E-journals and databases are likely to be used more heavily in fields in which topical relevance is the primary relevance criterion and less in fields in which paradigmatic relevance is the primary relevance criterion.*"¹⁵ Ainsi, les disciplines où le critère de pertinence serait lié au paradigme feraient un usage beaucoup moins important des revues électroniques et des bases de données. Nos entretiens avec les chercheurs ont montré que l'utilisation des bases de données dans certaines disciplines était problématique parce que les ressources n'y sont pas indexées par concepts. Concernant l'usage des revues, du fait de contextes d'offre différents, l'hypothèse est difficilement vérifiable.

En conclusion, il s'avère que pratiques et discipline sont bel et bien liées dans la recherche. L'analyse de domaine a permis d'illustrer la complexité des systèmes d'information dans des champs multidisciplinaires grâce notamment à la théorie de Whitley. Les différentes hypothèses que nous avons tenté de vérifier sur notre terrain ont permis de comprendre pourquoi certaines techniques de recherche, tel le *browsing*, avaient du sens dans un champ multidisciplinaire et pourquoi des ressources, telles les archives ouvertes ou les bases de données, rencontraient peu de succès en sciences de l'éducation. Par ailleurs, nous avons pu constater que le dialogue des disciplines était rare. Autant dans le contexte français que dans le contexte anglais, les collaborations interdisciplinaires et internationales semblent être limitées, du fait surtout d'approches de recherche

15. « Les revues électroniques et les bases de données seraient plus utilisées dans les domaines où la pertinence liée au sujet (de recherche) est majeure et moins utilisées dans ceux où la pertinence liée au paradigme prime. »

différentes¹⁶. Nous pensons que ces caractéristiques expliquent le non-usage des réseaux de recherche en ligne. Comme Fry, nous pensons qu'il est probable que les chercheurs de domaines pluridisciplinaires, comme les sciences de l'éducation, où l'organisation du travail est variable d'un champ à un autre, feront davantage appel à la communication informelle traditionnelle (conférences, colloques, etc.) pour coordonner le travail collaboratif et utiliseront les canaux formels pour disséminer leurs travaux et obtenir ainsi la reconnaissance de leurs pairs.

RÉFÉRENCES BIBLIOGRAPHIQUES

+++++

Ardoino Jacques, *Éducation et politique. Propos actuels sur l'éducation, pour un projet dans une perspective socialiste*, Paris, Gauthier-Villars, 1977.

Bates M. J., "Learning about the information seeking of interdisciplinary scholars and students", *Library Trends*, vol. 45, 1996, pp. 155-164.

Benoist Jean-Marie et al (dir.), *Interdisciplinarité et sciences humaines*. Paris, Unesco, 1983.

Brent David, « Trois attitudes dans la recherche en éducation... », *Revue des sciences de l'éducation*, 2005, vol. 31, n° 2, pp. 397-416.

Chartron Ghislaine, « Éléments pour une approche comparée de la publication scientifique », in *Forum universitaire. La communication scientifique en quatre dimensions*, Colloque annuel du centre de coopération interuniversitaire franco-québécoise, ministère des Finances, de l'Économie et de la Recherche, Montréal, Archives nationales du Québec, 2003.

[En ligne] < http://archivesic.ccsd.cnrs.fr/sic_00000435/fr/ > (consulté le 13 mai 2008).

Chatelanat Gisèle, Moro Christiane et Saada-Robert Madelon, *Unité et pluralité des sciences de l'éducation : sondages au cœur de la recherche*, Bern, Berlin, Bruxelles, Peter Lang, 2004.

Develay Michel, *Propos sur les sciences de l'éducation. Réflexions épistémologiques*, Issy-les-Moulineaux, ESF Éditeur, 2004.

16. De l'avis d'un enseignant-chercheur à l'Institute of Education à Londres.

Fisher K. E. *et al*, *Theories of Information Behavior*, Medford (NJ), ASIS&T, 2005.

Fry Jenny, "Scholarly research and information practices: a domain analytic approach", *Information Processing & Management*, vol. 42, 2006, pp. 299-316.

Fry Jenny et Talja Sanna, "The intellectual and social organization of academic fields and the shaping of digital resources", *Journal of Information Science*, vol. 33, n° 2, 2007, pp. 115-133.

Guibert Pascal (dir.), *Initiation aux sciences de l'éducation*, Paris, Vuibert, 2006.

Hadji Charles et Baillé Jacques, *Recherche et éducation. Vers une nouvelle alliance : la démarche de preuve en 10 questions*, Bruxelles, Paris, De Boeck Université, 1998.

Hjørland Birger, "Domain analysis in information science: eleven approaches—traditional as well as innovative", *Journal of Documentation*, vol. 58, n° 4, 2002, pp. 422-462.

Kling Rob et McKim Geoffrey W., "Not just a matter of time: field differences and the shaping of electronic media in supporting scientific communication", *Journal of the American Society for Information Science*, vol. 51, n° 14, 2000, pp. 1306-1320.

Kourilsky François (dir.), *Ingénierie de l'interdisciplinarité, un nouvel esprit scientifique*, Paris, L'Harmattan, 2002.

Marchive Alain, *Les sciences de l'éducation pour l'ère nouvelle*, numéro spécial, *Identité et constitution des sciences de l'éducation*, 1998, vol. 31, n° 1-2.

Marmoz Louis, *Les sciences de l'éducation en France : histoire et réalités*, Issy-les-Moulineaux, EAP, 1988.

Mialaret Gaston, *Les méthodes de recherche en sciences de l'éducation*, Paris, Presses universitaires de France, 2004, collection « Que-sais-je ? ».

Mialaret Gaston, *Sciences de l'éducation : aspects historiques, problèmes épistémologiques*, Paris, Presses universitaires de France, 2006, collection « Quadrige ».

Monjo Roger, *Réflexions sur l'épistémologie des sciences de l'éducation : le cas français*, 2005.

[En ligne] <http://alor.univ-montp3.fr/cerfee/article.php3?id_article=332> (consulté le 13 mars 2008).

Palmer Carole L. et Neumann Laura J., "The research work of interdisciplinary humanities scholars: exploration and translation", *Library Quarterly*, n° 1, vol. 72, 2002, pp. 85-117.

Plaisance Éric et Vergnaud Gérard, *Les sciences de l'éducation*, Paris, La Découverte, 2005.

Talja Sanna et Maula Hanni, "Reasons for the use and non-use of electronic journals and databases: a domain analytic study in four scholarly disciplines", *Journal of Documentation*, vol. 59, n° 6, 2003, pp. 673-691.

Ui Jun, « Étude de quelques problèmes posés par l'environnement », *Interdisciplinarité et sciences humaines*, Paris, Unesco, 1983, pp. 325-342.

Walsh John P. et Bayma Todd, "Computer networks and scientific work", *Social Studies of Science*, vol. 26, 1996, pp. 661-703.

Whitley Richard, *The Intellectual and Social Organization of the Sciences*, Oxford, Clarendon Press, 2000.

*par Claire Warwick, Melissa Terras, Isabel Galina, Paul Huntington et
Nikoleta Pappa*

traduit par Laure Manceau

+++++

BIBLIOTHÈQUES, RESSOURCES
D'INFORMATION ET
UTILISATEURS DE RESSOURCES
ÉLECTRONIQUES DANS
LES SCIENCES HUMAINES

+++++

+++++

BIBLIOTHÈQUES, RESSOURCES D'INFORMATION ET UTILISATEURS DE RESSOURCES ÉLECTRONIQUES DANS LES SCIENCES HUMAINES¹

+++++

En 2005, l'université de Bangor décide qu'elle peut se passer de huit de ses bibliothécaires spécialisés car, à l'ère de Google, et avec la menace qui pèse sur les budgets, elle a du mal à justifier le financement de médiateurs pour aider les usagers de la bibliothèque à trouver les ressources nécessaires à leur travail [Curtis, 2005]. Cette action peut sembler extrême, mais elle suit pourtant un mode de pensée qui envahit graduellement le monde de l'information numérique jusqu'à sa conséquence logique : si de vastes quantités d'informations sont disponibles sur le Web, de quelle utilité sont encore les professionnels de l'information ?

Lorsqu'au milieu des années 1990, l'influence d'Internet a commencé à se faire ressentir, certains auteurs ont éprouvé la nécessité de se demander si le bibliothécaire, qui n'était plus utile à l'utilisateur pour effectuer des recherches dans des systèmes tels que Dialog², tenait toujours un rôle [Cronin, 1998 ; Fourie, 1999 ; Gellman, 1996]. Dans le rapport [Brown et Swan, 2007] du Research Information Network (RIN) et du Consortium of Research Libraries (CURL) sur l'usage des bibliothèques de recherche par les chercheurs, le rôle du médiateur n'est pas mentionné comme futur rôle potentiel pour les bibliothécaires. Cette étude démontre cependant que les bibliothécaires continuent d'accomplir tout un ensemble d'activités que l'on pourrait décrire comme de l'intermédiation, dans le sens où ils conseillent les usagers – que ce soit de manière informelle ou par

1. Le projet LAIRAH est financé par l'AHRC-ICT Strategy Scheme. Les auteurs témoignent une reconnaissance particulière à trois autres projets du Strategy Scheme pour leur esprit de collaboration dans la réalisation de ces recherches, et pour leur avoir bénéficié de leurs résultats. Ces projets sont : Gathering Evidence (université de Bristol), RePAH (universités de Sheffield et de De Montfort), et Peer Review and Evaluation of Digital Resources in the Arts and Humanities (Institute of Historical Research, université de Londres). N.D.T. : cet article a été publié sous le titre "Library and information resources and users of digital resources in the humanities", dans *Program: Electronic Library and Information Systems*, vol. 42, n° 1, 2008, pp. 5-27.

2. [En ligne] < <http://www.dialog.com/> > (consulté le 13 juillet 2010).

le biais de formations – sur des questions de création, d'utilisation et de gestion des ressources numériques, ou sur des sujets plus traditionnels, tels que la propriété intellectuelle et la citation. En dépit de l'absence de preuves solides pour l'étayer, il existe également une thèse qui a de plus en plus cours tant dans le secteur commercial que dans celui de l'information, selon laquelle la technologie devrait rendre les gens plus productifs et permettre ainsi des économies [Lin et Shao, 2006]. On pourrait donc se dire par extension qu'une augmentation des investissements dans les ressources numériques permettrait de dépenser moins dans les bibliothèques et les archives classiques, et dans le personnel qui y officie, comme le démontre l'exemple cité plus haut.

Le présent article fournit des preuves qui vont à l'encontre de telles visions. Notre étude de l'usage des ressources numériques par les chercheurs en sciences humaines apporte des preuves flagrantes de l'importance continue des ressources d'information tant classiques que numériques. En nous appuyant sur des preuves quantitatives, nous défendons l'importance des institutions de l'information, des bibliothécaires, des archivistes et des professionnels de la documentation qui facilitent l'accès aux ressources et en assurent la qualité, même à l'ère de ce que Gregory Crane a décrit en 2006 comme celle de la numérisation de millions de livres [Crane, 2006].

LE PROJET LAIRAH

+++++

Le projet LAIRAH³, basé à la School of Library Archive and Information Studies (SLAIS) de l'University College London (UCL), est une étude de 15 mois, réalisée entre juin 2005 et septembre 2006 dans le but de découvrir ce qui influence la durabilité et l'utilisation à long terme des ressources numériques dans les sciences humaines à travers l'analyse et l'évaluation de leur utilisation en temps réel. Cette étude a été financée par l'Arts and Humanities Research Council (AHRC) ICT Strategy Scheme⁴. Les résultats de ces recherches devraient donc avoir une influence sur la prochaine politique d'au moins l'un des principaux organismes de financement du Royaume-Uni. Nous pouvons ainsi espérer que notre travail

3. Log analysis of Internet resources in the arts and humanities (LAIRAH), [en ligne] < <http://www.ucl.ac.uk/slais/research/circah/lairah/> > (consulté le 13 juillet 2010).

4. [En ligne] < <http://www.ahrcict.rdg.ac.uk> > (consulté le 13 juillet 2010). ICT: Information and Communication Technology.

sur l'importance des ressources d'information influencera les décisions prises quant à leur avenir financier. Ceci est également important dans la mesure où l'AHRC est l'organisme qui participe au financement de la formation des futurs bibliothécaires et archivistes, par le biais de bourses d'études pour les masters dans les universités du Royaume-Uni.

Les objectifs du projet de recherche se définissent ainsi :

- déterminer l'échelle d'utilisation des ressources numériques dans les sciences humaines en ayant recours à l'analyse de *logs* des portails du Humbul, du Artifact et du Arts and Humanities Data Service ;
- déterminer si les ressources utilisées ont des points communs ;
- mettre en lumière les zones de bon fonctionnement et les aspects du projet qui pourraient être améliorés pour une meilleure utilisation et plus de durabilité.

Au cours du projet, le Humbul et l'Artifact ont fusionné pour devenir Intute : Arts and Humanities Service⁵.

TRAVAUX PRÉCÉDENTS DANS CE DOMAINE

+++++

Recherche d'information en sciences humaines

+++++

De récents et utiles travaux sur les besoins en information et le comportement des universitaires de sciences humaines en matière de recherche d'information ont été effectués par Barrett [2005], Talja et Maula [2003], Green [2000], Herman [2001] et Ellis et Oldman [2005]. Des études qui ont fait école, menées par Stone [1982] et Watson-Boone [1994], ont démontré que les usagers des sciences humaines ont besoin d'un large éventail de ressources, en termes d'époque et de genre. Cela reste vrai dans un environnement numérique, où ils continuent d'avoir besoin de matériel imprimé, ou même de manuscrits, tout autant que de ressources électroniques, qui, de par leur nature, peuvent laisser supposer un matériel bien plus foisonnant que ce dont se servent les chercheurs [British Academy, 2005]. Marcia Bates a analysé les activités pratiquées par les chercheurs en sciences humaines dans un environnement numérique utilisant

5. [En ligne] < <http://www.intute.ac.uk/artsandhumanities/> >. Suite à des restrictions budgétaires, le portail Intute n'est plus mis à jour depuis juillet 2011. Le site est toujours disponible à l'adresse < <http://www.intute.ac.uk/> > (consulté le 5 décembre 2011).

le système Dialog, antérieur à l'existence du Web [Bates, 1996]. Le mode d'interrogation du système rendait difficile la tâche des utilisateurs ne bénéficiant pas de formation spécifique, et c'est la raison pour laquelle les recherches étaient généralement menées par des professionnels de l'information. La recherche sur le Web est donc très différente de celle décrite par Bates, puisqu'elle se fait par le biais d'une interface d'utilisation graphique et qu'elle ne requiert aucune ou peu d'intervention d'un professionnel de l'information.

Un thème majeur de la littérature consacrée aux usagers des sciences humaines est qu'ils ne ressemblent pas à ceux des sciences⁶ et des sciences sociales, bien que nombre de concepteurs de ressources électroniques les envisagent de la sorte [Bates, 2004]. Les chercheurs en sciences humaines seront plus enclins à utiliser ce que David Ellis appelle le « chaînage » et à suivre les références trouvées dans d'autres ouvrages [Ellis et Oldman, 2005]. Cette pratique cependant va à l'encontre de la recherche par mots clés, qui tend à être la norme pour les systèmes d'information, ce qui a donc été considéré comme une preuve que les techniques des chercheurs en sciences humaines sont plutôt pauvres [Chu, 1999]. Dès le milieu des années 1980, Stephen Wiberley avait montré que les chercheurs en sciences humaines construisaient leur recherche sur des termes bien définis, mais que ces termes étaient très différents de ceux utilisés par les scientifiques, étant plus enclins, par exemple, à inclure des noms de personnes ou de lieux [Wiberley, 1983, 1988].

Lehmann et Renfro [1991] ainsi que Wiberley [2000] suggèrent que les universitaires du domaine des sciences humaines sont réceptifs à la technologie tant qu'elle est synonyme d'économie suffisante de temps ou d'efforts. Les travaux de Dalton et Charnigo [2004] et ceux de Whitmire [2002] ont également démontré que les chercheurs en sciences humaines qui ont recours aux ressources numériques ont tendance à être exigeants sur la qualité des documents et sont capables d'élaborer une stratégie de recherche complexe, en fonction de leur degré de formation.

Plus récemment, des études se sont orientées vers le travail de chercheurs officiant dans des disciplines spécialisées : Talja et Maula [2003] ainsi qu'Ellis et Oldman [2005] ont étudié les chercheurs en littérature ; Dalton et Charnigo [2004], Anderson [2004], et Duff et ses collègues [Duff et Cherry, 2001 ; Duff *et al.*, 2004] ont étudié ceux en histoire. Selon ces

6. N.D.T. : le terme « sciences » dans le texte fait référence aux disciplines des sciences techniques et médicales (STM) par opposition aux sciences sociales et aux sciences humaines.

auteurs, le comportement informationnel des universitaires dans des disciplines individuelles est suffisamment différent pour constituer un objet d'étude en soi, et une étude des sciences humaines prises dans leur globalité comporte un risque de sur-généralisation. Nous avons cependant suivi la méthodologie de l'étude la plus récente, menée par la British Academy (abordée plus loin), qui traite du large éventail de sujets en sciences humaines. Nous avons également choisi un vaste échantillon de disciplines afin d'obtenir une image représentative des usages en sciences humaines dans un délai relativement restreint. Toutefois, si nous obtenons un financement pour approfondir nos recherches, nous avons l'intention de mener une étude plus ciblée, focalisée sur les domaines de l'histoire et de la littérature anglaise.

Rapports sur l'utilisation des technologies d'information et de communication en sciences humaines

+++++

Depuis 2005, les organismes de financement du Royaume-Uni soutenant les arts et les sciences humaines cherchent également à enquêter sur l'état des besoins et usages des ressources numériques en sciences humaines. L'étude qu'a menée la British Academy en 2005 sur la recherche en arts, sciences humaines et sciences sociales, a conclu que les universitaires faisaient un bon usage d'une grande diversité de ressources numériques. L'échantillon sélectionné pour l'étude était relativement petit et tendait à surreprésenter les universitaires « seniors », en termes d'âge et de poste occupé (l'échantillon comportait de nombreux professeurs). Il faut souligner qu'en dépit de la croyance selon laquelle l'enthousiasme pour la technologie serait l'apanage des plus jeunes, le rapport faisait état d'une large utilisation des ressources numériques par l'échantillon. Selon les auteurs, dans un avenir prévisible, les ressources utilisées en recherche seront tant imprimées sur papier que numériques, et donc les ressources numériques ayant le plus de valeur sont celles qui, comme les catalogues de bibliothèques en ligne, permettent de localiser d'autres sources d'information de n'importe quel format. Le rapport défend donc l'idée que ces ressources secondaires devraient bénéficier d'une numérisation en priorité.

En même temps que le projet LAIRAH, trois autres projets ont été commandés par l'AHRC pour recueillir des données sur l'utilisation des ressources des technologies de l'information et de la communication (TIC) dans les sciences humaines :

1. Le projet Research in Portals in the Arts and Humanities (RePAH)⁷, basé à l'université de Sheffield, d'août 2005 à septembre 2006.
2. Évaluation par les pairs et analyse des ressources numériques pour les arts et les sciences humaines menées par l'Institute of Historical Research (IHR)⁸ d'octobre 2005 à septembre 2006.
3. Gathering Evidence⁹ : recherche sur l'usage actuel des TIC et les besoins futurs en arts et sciences humaines (AHRC-ICT), à l'université de Bristol, de fin 2005 à septembre 2006.

Tous ces projets ont partagé leurs résultats et comparé leurs données. Les résultats que nous présentons plus loin établissent donc des comparaisons spécifiques avec ceux de ces projets et avec ceux de l'étude menée par la British Academy en 2005.

Les trois rapports de l'ICT Strategy ont mis au jour un enthousiasme très répandu pour le recours aux ressources numériques, même si, encore une fois, les échantillons sont quelque peu orientés. Dans tous les cas, les données ont été recueillies au moyen de questionnaires et de réunions de groupes, dont les participants ont été recrutés *via* des sites Web ou des listes de discussion. Cela peut laisser supposer que les participants pouvaient avoir de fortes chances d'être enthousiasmés par les ressources numériques [Huxley *et al.*, 2007, p. 19]. Le projet Gathering Evidence a mis au jour un enthousiasme similaire pour les outils d'aide à la recherche ; cependant, ses participants utilisaient aussi des ressources primaires en ligne, telles que des textes électroniques. Comme le projet RePAH, ils ont découvert que les participants auraient apprécié accéder à davantage de ressources primaires sous forme numérique. Selon RePAH, l'universitaire typique du domaine des sciences humaines refuse de consacrer du temps et des efforts à l'apprentissage nécessaire pour se servir de nouveaux outils informatiques [Brown *et al.*, 2007, p. 8] et il est évident, à la lecture du rapport de Gathering Evidence, que l'usage qui est fait des ressources numériques demeure à un niveau relativement rudimentaire. Bien que les universitaires déclarent que les TIC ont eu pour effet de transformer leur méthode de recherche, les activités qu'ils mettent en avant incluent un

7. [En ligne] < <http://repah.dmu.ac.uk> > (consulté le 13 juillet 2010).

8. [En ligne] < <http://www.history.ac.uk/digit/peer/index.html> > (consulté le 13 juillet 2010).

9. [En ligne] < <http://www.ilrt.bris.ac.uk/projects/project?search=AHRC-ICT> > (consulté le 13 juillet 2010).

accès facilité aux revues en ligne, la possibilité de publier des articles sur le site du département, et un accès plus pratique à tout un ensemble de textes numérisés, tel que Early English Books Online [Huxley *et al.*, 2007, p. 7]. De telles activités n'auront rien de révolutionnaire pour des spécialistes des ressources informatiques, mais elles sont de toute évidence très estimées par les universitaires.

Toutefois, le projet IHR a dévoilé des doutes largement répandus quant à la façon de juger de la valeur académique des ressources directement produites sur le Web – et a donc mis au jour une demande de moyens d'évaluation par les pairs [IHR, 2006]. Ce besoin apparaît également dans un rapport récent de la Modern Language Association (MLA) d'Amérique, qui montre un manque d'expérience dans de nombreuses disciplines en termes d'estimation de la valeur académique des documents numériques, 40% des participants allant même jusqu'à dire qu'ils ne savaient pas comment estimer la valeur d'un article revu par un pair sous forme électronique.

La littérature indique donc que les universitaires ont adopté les ressources numériques au sens large du terme avec un enthousiasme manifeste. Les données sont pourtant utilisées de manière traditionnelle, et l'on note un net refus de s'engager plus avant sur la valeur académique de nouvelles publications numériques et d'apprendre de nouvelles techniques. Toutefois, nous ne connaissons aucune étude ayant utilisé des méthodes quantitatives, en particulier l'analyse de *logs* (décrite plus bas), pour mesurer les niveaux d'utilisation des ressources numériques en sciences humaines. Notre étude se concentre également non pas sur les ressources en général, mais sur le type de ressources numériques qui est le plus utile aux chercheurs. Notre étude est également la première à établir une comparaison entre les préférences exprimées par les utilisateurs et les preuves quantitatives de ce qu'ils utilisent dans les faits.

MÉTHODES

++++
L'étude LAIRAH a eu recours à une combinaison de techniques quantitatives et qualitatives. Pour les besoins de cet article, nous nous attacherons aux résultats dérivés de deux mesures quantitatives : l'analyse approfondie de *logs* et un questionnaire. Dans une phase ultérieure de notre recherche, nous avons également créé deux ateliers et animé des entretiens avec des producteurs de ressources numériques.

Analyse approfondie de *logs*

+++++

Nous avons utilisé l'analyse approfondie de *logs* afin d'évaluer les niveaux d'utilisation des ressources numériques dans le domaine des arts et des sciences humaines. Cette technique a largement été employée par le Centre for Information Behaviour and the Evaluation of Research (CIBER)¹⁰ et dans d'autres domaines tels que l'information médicale et l'édition commerciale (voir, par exemple, Huntington *et al.* [2002]). Ceci nous a permis de dégager des récurrences dans l'utilisation des ressources numériques dans les sciences humaines et d'identifier une sélection de ressources utilisées et non utilisées.

Toutes les plateformes de documentation numérique ont la possibilité de générer des *logs* qui fournissent un historique automatique et en temps réel de leur utilisation. Ils représentent l'empreinte numérique des utilisateurs, et en les analysant, il est possible d'identifier leur comportement en matière de recherche d'information.

Les *logs* sont capables de nous renseigner sur le type de personnes qui utilisent les services. Leur avantage est qu'ils offrent des preuves abondantes et fiables sur l'utilisation qui est faite des ressources. Les *logs* enregistrant l'utilisation de tous ceux qui entrent dans le système, il est possible de suivre de près le comportement de millions de personnes dans le monde. Ils possèdent non seulement une taille et une portée sans égales, mais constituent un enregistrement direct et immédiatement accessible de ce que ces personnes ont fait – et non ce qu'elles ont annoncé, ce qu'elles auraient peut-être fait, ce qu'elles se sont senties obligées de faire, ce qu'elles pensaient faire. Les données ne sont pas filtrées et représentent le comportement de l'utilisateur ; elles sont un complément des importantes données contextuelles obtenues en entrant en contact avec de vrais utilisateurs et en étudiant leurs expériences et leurs intérêts.

Les données concernant les *logs* d'un serveur sont des enregistrements des pages Web visionnées. Ces enregistrements sont le fruit d'une demande de l'ordinateur du client et offrent un enregistrement des pages fournies par le serveur Internet à l'ordinateur du client. Le serveur enregistre l'adresse Internet de l'ordinateur du client. Ces adresses suivent un numéro de protocole Internet (IP) et sont associées à des informations concernant le nom de domaine (DNS, *Domain Name Server*). Les données

10. [En ligne] < <http://www.ucl.ac.uk/ciber/ciber.php> > (consulté le 13 juillet 2010). Cette adresse ramène à cette nouvelle page : < <http://www.ucl.ac.uk/dis/> >, le site du Department of Information Studies de l'University College London (UCL).

du DNS renseigné sur le nom de l'organisation, son type (universitaire ou commercial, par exemple), le pays d'enregistrement. Voici un extrait de fichier *log* :

```
66.XXX.XXX.XX - - [24/Feb/2005:00:07:12 + 0000] "GET /deposit/  
depintro.htm HTTP/1.1"  
200 318 "http://ahds.ac.uk/copyrightfaq.htm"
```

Dans cet exemple :

- (66.XXX.XXX.XX) est l'adresse IP (chaque X représentant un chiffre masqué pour des questions de confidentialité). C'est une adresse anonyme qu'utilisent les ordinateurs pour envoyer et recevoir correctement des données sur Internet. Dans les fichiers *logs* originaux, les X sont bien sûr remplacés par des chiffres, grâce auxquels on peut identifier des ordinateurs individuels. Ces adresses ont été utilisées pour les besoins de notre analyse, mais ont ensuite été supprimées, afin qu'aucun ordinateur ne puisse être identifié à partir des résultats publiés.
- (24/Feb/2005:00:07:12 + 0000) est un dateur qui certifie le jour et l'heure d'envoi du fichier en réponse à la demande du client.
- (GET/deposit/depintro.htm) enregistre le fichier envoyé au client et les répertoires où le fichier est stocké sur le serveur.
- (HTTP/1.1) correspond à l'enregistrement de la communication entre le serveur et le client en version hypertexte.
- (200) est le champ du statut et indique si la requête était correcte et si un fichier a été envoyé.
- (318) donne la taille en octets du fichier envoyé.
- (http://ahds.ac.uk/copyrightfaq.htm) est le *log* du référent et donne l'adresse du dernier site visité par le client.

Les informations sont stockées sous forme de fichiers texte ASCII en format compressé. Pour cette étude, les *logs* archivés de Humbul représentent environ 150 Mo, soit 20 % d'un disque compact. Les données concernant l'adresse DNS et le numéro IP ne peuvent être utilisées pour identifier l'utilisateur [Albitz et Liu, 2006]. Pour préserver l'anonymat plus avant, les *logs* que nous avons analysés ont été purgés de toutes données personnelles.

Nous avons utilisé les *logs* des trois principaux portails numériques de sciences humaines au Royaume-Uni :

1. Arts and Humanities Data Service (AHDS)
2. Humbul
3. Artifact

Dans le cas d'AHDS et de Humbul, nous avons pu analyser un an de données, grâce au logiciel SPSS. Mais dans le cas d'Artifact, beaucoup moins de données étaient disponibles, en raison du manque de moyens techniques des fournisseurs de ce service pour la maintenance de leurs propres *logs*. Les données d'Artifact ont été disponibles lorsqu'il a fusionné avec Humbul, mais nous n'avions que trois mois de données et cela arrivait relativement tard dans la vie du projet. En conséquence, pour les besoins de cet article, nous nous concentrerons sur les résultats des *logs* de Humbul et d'AHDS. Idéalement, nous aurions aimé utiliser les *logs* des serveurs de projets individuels impliquant les sciences humaines et le numérique. Toutefois, la récolte de données des trois portails a en soi été un processus de longue haleine, et faire de même à partir de projets individuels aurait été impossible étant donné nos délais.

Limites de la méthode

Les données de *logs* ont leurs limites. Bien qu'elles puissent indiquer depuis quel pays l'utilisateur accède au site, s'il utilise un fournisseur d'accès à Internet (FAI) commercial, ou s'il vient d'une institution universitaire, de telles données peuvent être trompeuses. Les *logs* suggéraient qu'un nombre anormalement élevé d'utilisateurs étaient américains, alors que le questionnaire nous indiquait que seuls 15 % des utilisateurs étaient basés en dehors du Royaume-Uni. C'est en partie dû au fait que beaucoup de FAI du Royaume-Uni, tels que BT Internet, sont enregistrés aux États-Unis. C'est également en partie pour cela que nous avons toujours recours à des questionnaires comme éléments de comparaison avec les *logs*.

Il est également impossible d'identifier un étudiant ou un professeur qui travaille chez lui, puisqu'il est très probable qu'ils se connectent aux ressources numériques de l'université *via* un FAI commercial. Nous avons également conscience que les ordinateurs des universités sont accessibles au public et donc utilisés par des usagers multiples. Nous devons donc juger du moment où la session de l'utilisateur se terminait, en nous basant sur des périodes d'inactivité entre les périodes d'utilisation, mais nous ne pouvions dire avec certitude si ce n'était pas la même personne revenue après une pause-café, par exemple. Ceci est toutefois moins important pour notre étude, car nous étions à la recherche de tendances parmi de vastes quantités de données sur les utilisateurs, et nous n'essayions pas de

suivre le cheminement d'un individu à travers une ressource numérique dans ce cas précis.

Enfin, les *logs* nous renseignent sur les pages vues, mais ne nous disent pas si elles ont été lues, ou si l'utilisateur a été satisfait de ce qu'il y a trouvé. C'est pour cela que nous avons poursuivi par une recherche qualitative sur l'opinion des chercheurs en sciences humaines sur les ressources numériques [Warwick *et al.*, 2008]. Il est également probable que des utilisateurs se comportent différemment s'ils accèdent à une ressource directement *via* Google ou s'ils utilisent un portail. Cette hypothèse sera testée dans la seconde étape de notre recherche, si nous en obtenons le financement.

Questionnaire

+++++

Afin d'établir un élément de comparaison avec les données des *logs*, nous avons mis au point un questionnaire sur les sites Web de l'AHDS et de Humbul, ainsi que sur le projet RePAH, portant sur l'utilisation des ressources. Cette méthode a été utilisée à plusieurs reprises par l'équipe de chercheurs de CIBER, puisque leur expérience a démontré qu'il pouvait exister des différences entre les ressources que les utilisateurs disent avoir consultées et le comportement enregistré par les fichiers *logs*. Comme le soutiennent aussi Harley et Henke [2007], l'utilisation combinée de l'analyse de *logs* et de questionnaires en ligne permet aux chercheurs d'obtenir un aperçu aussi large que possible de l'utilisation des sites Web.

Notre méthode diffère donc de celle d'autres études, telles que le projet Gathering Evidence évoqué plus haut. Nous n'avions pas en tête de publier le questionnaire, nous ne visions pas non plus un échantillon représentatif de la population des chercheurs en sciences humaines au Royaume-Uni. Nous voulions simplement comparer les données des *logs* des sites étudiés avec ce que les utilisateurs pensaient faire sur ces sites, et l'opinion qu'ils en avaient. Ceux qui ont répondu au questionnaire sur le site de RePAH sont probablement des personnes qui avaient vu des présentations du projet ou en avaient entendu parler. Il est donc probable que les données recueillies *via* le questionnaire surreprésentent l'opinion de ceux qui sont intéressés par l'utilisation des ressources numériques et des professionnels de la documentation, puisqu'ils avaient déjà trouvé de tels portails. Toutefois, obtenir un échantillon vraiment représentatif de l'utilisation universitaire des ressources numériques en sciences humaines n'est pas aisé, puisque ceux qui sont suffisamment intéressés pour

répondre à tout type de sondage tendent toujours à être des enthousiastes du numérique. Néanmoins, nous avons demandé à tous les participants de renseigner leur statut parmi les catégories suivantes : étudiant de deuxième cycle, étudiant de troisième cycle, universitaire, professionnel de la documentation, et amateur intéressé ; ceci afin d'avoir une impression générale de la mesure dans laquelle les réponses que nous obtiendrions seraient typiques de la population générale.

Pour obtenir une vue d'ensemble aussi complète que possible, il est donc important de comparer les données des deux méthodes de collecte, *logs* et questionnaires. En plus de cela, nous avons comparé nos résultats à ceux des questionnaires des projets de l'IHR et des TIC à l'université de Bristol.

RÉSULTATS

+++++
Contre toute attente, il a été très difficile d'estimer les niveaux d'utilisation des ressources en termes absolus. Notre recherche s'est étalée sur une période qui a coïncidé avec des changements importants dans le mode de fonctionnement des portails : la fusion de Humbul et Artifact qui sont devenus le service Intute Arts and Humanities ; et la centralisation d'AHDS. AHDS a également procédé à des changements majeurs dans la fonctionnalité de son site Web central, permettant aux usagers de se relier aux ressources elles-mêmes et non seulement aux métadonnées. Il est également possible qu'un nombre croissant de visiteurs accèdent aux ensembles de données d'AHDS *via* les fournisseurs d'accès.

Le projet RePAH a mis au jour qu'au cours de l'étude, 7643 ressources distinctes ont été visitées *via* Humbul, sur un total de 11680 qui étaient disponibles au public lorsque la fusion a eu lieu. Cela suggère que 36 % des données du site de Humbul ont été négligées au cours de notre étude, bien que nous soyons dans l'incapacité de prouver que personne n'y ait eu accès. Il est également probable que les ressources soient visitées directement, par exemple *via* des moteurs de recherche tels que Google, en tapant l'adresse URL, ou en utilisant des signets, et non *via* les portails thématiques. Il est également bon de rappeler que certaines publications papier de spécialités des sciences humaines ne sont jamais consultées, fait que confirment les faibles tirages généralement accordés aux monographies en sciences humaines. Même en sciences, ce sont en moyenne 27% des articles qui ne sont jamais cités, chiffre qui se monte à 44,52% en science informatique [Science Watch, 1999].

Quoi qu'il en soit, dans le cas de la publication de revues ou d'ouvrages, un éditeur commercial prend le risque financier et vendra une revue ou un livre à une bibliothèque, sans tenir compte du fait qu'il soit lu ou cité. Dans le cas des ressources numériques en sciences humaines, une part importante du financement public est gâchée si une ressource n'est pas utilisée. Nos résultats présentés ci-dessous visent donc à accroître notre connaissance de la manière dont réagissent les utilisateurs face à de telles ressources et à partager les bonnes pratiques pouvant permettre de s'assurer que les ressources numériques créées à l'avenir auront un maximum de chances d'être utilisées.

Analyse de logs

+++++

Les logs de Humbul

Les logs de Humbul indiquaient entre 2 000 et 4 000 connexions quotidiennes au site Web le week-end et entre 6 000 et 8 000 en semaine. La majorité des utilisateurs étaient originaires du Royaume-Uni ou des États-Unis – même si ce chiffre est exagéré du fait que des utilisateurs se connectent *via* un FAI basé aux États-Unis. Par exemple, BTopeworld.com¹¹ est un fournisseur britannique enregistré en tant que compagnie américaine. Le tableau 1 montre la répartition par pays d'usage du Humbul.

L'histoire est le sujet le plus populaire et environ un quart (27,1 %) des sujets s'y rapportent. D'autres sujets populaires comprennent l'anglais (16,9 %), la religion (6,5 %), les sciences humaines générales (6,2 %), et la philosophie (5,1 %), comme le montre le tableau 2.

Les logs nous ont indiqué quels domaines de premier niveau étaient le plus souvent visités. Si un utilisateur décidait de consulter une ressource, les logs enregistraient le site visité et donnaient l'adresse du site ainsi que le répertoire d'où était issue la ressource. Environ 11,5 % des documents étaient vus par des utilisateurs qui cliquaient directement sur la ressource. Tout au long de l'année, 7 463 ressources distinctes ont été vues *via* le site de Humbul. Nous avons choisi les 40 sites les plus visités pour la poursuite de notre étude, comme le montre le tableau 3.

Au total, 26 sites du tableau 3 pourraient être labellisés « ressources de références », dans la mesure où ils sont destinés aux bibliothèques,

11. Devenu BT Total Broadband, < <http://www.productsandservices.bt.com/> > (consulté le 5 décembre 2011).

archives, recueils de textes électroniques ou éditeurs, que ce soit au Royaume-Uni ou à l'international. Pour pousser nos recherches, nous avons extrait des détails des sous-répertoires appartenant aux universités britanniques dans cette liste : en ordre de popularité, Édimbourg, Sheffield, UCL, Greenwich et la School of Advanced Study (université de Londres). Presque toutes les pages de cette dernière étaient consacrées aux pages Web des centres de recherche thématiques, tels que le Warburg Institute, l'Institute of English Studies et l'Institute of Classical Studies. Il y avait des liens vers des recueils numériques, tels que ceux du Warburg Institute et de l'Institute of English Studies, mais beaucoup de liens dirigeaient simplement vers les pages des centres de recherche, leur bibliothèque ou leur forum d'étudiants.

Tableau 1 : répartition géographique des utilisateurs de Humbul

Pays / région	Utilisateurs
États-Unis	50,1%
Royaume-Uni	29,9%
Europe occidentale	7,6 %
Europe de l'Est	3,6 %
Asie	2,7%
Australie / Océanie	2,3 %
Amérique du Nord	2,2 %
Autres	1,7%

Tableau 2 : répartition des sujets vus

Sujets	Consultation
Histoire	27,1%
Anglais	16,9 %
Autres	15,5 %
Religion	6,5 %
Sciences humaines gén.	6,2 %
Philosophie	5,1 %
Langues classiques	5 %
Linguistique	4,7 %
Archéologie	4,5 %
Américain	4,2 %
Français	4,2 %

Les ressources les plus populaires à la School of Advanced Study (avec 2 % ou plus du total des occurrences) sont regroupées dans le tableau 4.

Le grand nombre d'occurrences pour les pages Web des centres de recherche, ainsi que pour les ressources numériques spécifiques, suggère que beaucoup d'utilisateurs consultent la page Web avant une visite, mais il ne s'agit pas d'un substitut à la visite du centre en elle-même. De la même façon, nombre de personnes consultent le site Internet d'un musée avant de s'y rendre pour se renseigner sur son actualité, mais très peu considèrent cette démarche comme une alternative aux collections réelles [Marty, 2007].

Trois des ressources les plus populaires à Édimbourg (29% au total) étaient le Centre for the History of the Book (deuxième), le *Dictionary of the Older Scots Tongue* (cinquième) et l'*Edinburgh Journal of Gadda Studies*, comme le montre le tableau 5 (les deux derniers sites ne donnent pas accès à la ressource, mais simplement des renseignements sur cette dernière).

À l'université de Sheffield, six ressources analogues figuraient dans les logs. *Assemblage* (une revue d'archéologie) était classée au deuxième rang, si on additionne les visites sur la page d'accueil et celles sur un numéro particulièrement populaire. Elle est suivie par The Association for Low Countries, *CAPRA* (une revue d'archéologie), The Centre for the English Cultural Tradition, l'International Bande Dessinée Society, et la Hegel Society of Great Britain, comme l'indique le tableau 6 (bien que chacun de ces projets n'ait reçu que moins de 2% des occurrences et se soit donc placé assez bas dans le classement).

Toutefois, ces ressources constituaient un plus faible pourcentage des occurrences totales (12%), ce qui n'est pas surprenant étant donné la solide réputation de Sheffield dans la production de ressources numériques en sciences humaines.

Les logs de Humbul indiquent donc que malgré sa fonction de portail consacré avant tout aux recherches spécialisées, de nombreux utilisateurs qui ont cliqué vers ces ressources l'ont fait pour accéder aux ressources, centres et revues.

Tableau 3 : les 40 sites de ressources les plus visités via Humbul entre juin 2005 et septembre 2006

URL du site	Nombre de visites	%
http://www.bbc.co.uk/	4 166	1,5
http://www.wsu.edu/	2 473	0,9
http://www.geocities.com/	1 969	0,7

http://www.nd.edu/	1 517	0,6
http://ads.ahds.ac.uk	1 216	0,4
http://www.bl.uk/	1 047	0,4
http://www.arts.ed.ac.uk/	1 042	0,4
http://www.pbs.org/	1 031	0,4
http://www.emule.com/ (ensemble de textes)	936	0,3
http://memory.loc.gov/ammem/index.html (bibliothèque du Congrès)	836	0,3
http://www.fordham.edu/	813	0,3
http://www.shef.ac.uk/	811	0,3
http://www.channel4.com/	789	0,3
http://www.newadvent.org/ (site de référence catholique)	713	0,3
http://www.llgc.org.uk/ (Bibliothèque nationale du pays de Galles)	680	0,3
http://www.spartacus.school/ (référence historique)	659	0,2
http://www.luminarium.org/ (référence d'études médiévales)	659	0,2
http://etext.lib.virginia.edu (Virginia E-text Center)	649	0,2
http://www.cambridge.org/home (Cambridge University Press)	643	0,2
http://www.ucl.ac.uk/	636	0,2
http://www.iwm.org.uk/ (Imperial War Museum)	624	0,2
http://www.loc.gov/ (bibliothèque du Congrès)	614	0,2
http://ccat.sas.upenn.edu/ARIT/	606	0,2
http://www.gre.ac.uk/	599	0,2
http://www.archives.gov.on.ca/ (archives de l'Ontario)	575	0,2
http://www3.oup.co.uk (Oxford University Press)	573	0,2
http://www.archives.gov/ (Archives nationales américaines)	563	0,2
http://www.accd.edu/	560	0,2
http://www.nationalarchives.gov.uk/ (Archives nationales britanniques)	559	0,2
http://www.georgetown.edu/	546	0,2
http://www.hti.umich.edu/ (Humanities Text Initiative)	540	0,2
http://www.sas.ac.uk/ (School of Advanced Study, London University)	536	0,2
http://www.kb.nl/ (Bibliothèque nationale des Pays-Bas)	520	0,2
http://etext.virginia.edu (Virginia E-text Center)	506	0,2
http://www.bu.edu/ et http://www.stoa.org/ (textes classiques)	504	0,2
http://history.hanover.edu/	503	0,2
raven.cc.ku.edu	499	0,2
http://learningcurve.pro.gov.uk (matériel pédagogique d'histoire)	490	0,2
http://www.17thc.us/ (Collected materials colonial New England)	485	0,2
	479	0,2
		12,6

Logs d'AHDS

Au cours de notre étude, il y a eu en moyenne entre 1 000 et 3 000 visites par jour sur le site central d'AHDS ; de mars à août 2005, ce chiffre a augmenté pour atteindre une fourchette de 5 000 à 8 000 visites quotidiennes. Le profil national est similaire à celui de Humbul, bien qu'abstraction faite des domaines commerciaux, il s'avère que 86 % des utilisateurs sont du Royaume-Uni, comme le montre le tableau 7.

Tableau 4 : ressources les plus populaires sur le nom de domaine de la School of Advanced Study¹²

URL	Fréquence	%
http://www.sas.ac.uk/irs/ (Institute of Germanic and Romance Studies)	341	16,23037
http://www.sas.ac.uk/ies/centre.htm (Institute of English Studies)	243	11,56592
http://www.sas.ac.uk/commonwealthstudies/research/bdeep.html (Projet sur les documents anglais à la fin de l'Empire)	189	8,995716
http://www.sas.ac.uk/warburg/ (Warburg Institute)	138	6,568301
http://www.sas.ac.uk/warburg/mnemosyne/DigitalCollections.htm (collections de la bibliothèque numérique)	129	6,139933
http://www.sas.ac.uk/IRS/CWWF/Index.htm	128	6,092337
http://www.sas.ac.uk/commonwealthstudies (Institute of Commonwealth Studies)	119	5,66397
http://www.sas.ac.uk/ies/williamsharp.htm (Institute of English Studies, archive numérique William Sharp)	92	4,378867
http://www.sas.ac.uk/icls/imaginesit/Default.htm (Institute of Classical Studies – Imagines Italicæ collection)	91	4,331271
http://www.sas.ac.uk/icls/Hellenic/ (Hellenic Society)	89	4,236078
http://www.sas.ac.uk/aristotelian_society/	74	3,522132
http://www.sas.ac.uk/icls/pgforum/ (forum de diplômés)	66	3,141361
http://www.sas.ac.uk/ies (Institute of English Studies)	64	3,046168
http://www.sas.ac.uk/icls/library/libhome.htm (Bibliothèque d'études classiques)	64	3,046168
http://www.sas.ac.uk/warburg/mnemosyne/entrance.htm (bibliothèque du Warburg Institute)	59	2,808187
http://www.sas.ac.uk/igs/ (Institute of Germanic and Romance Studies)	58	2,76059

12. Ces adresses sont celles utilisées à l'époque de l'étude, néanmoins l'adresse centrale reste la même : < <http://www.sas.ac.uk/> > (consulté le 5 décembre 2011).

http://www.sas.ac.uk/ilas/ (Institute for the Study of the Americas)		2,712994
http://www.sas.ac.uk/ies/cmpps/Projects/Sharp/ (Institute of English Studies, archive numérique William Sharp)	44	2,094241

Tableau 5 : ressources les plus populaires sur le nom de domaine d'arts.edinburgh¹³ (plus de 2 % des occurrences)

URL	Fréquence	%
http://www.arts.ed.ac.uk/scothist/courses/eurowitchhunt/ (The European Witch Hunt)	931	40,69056
http://www.arts.ed.ac.uk/chb/index.html (Centre for the History of the Book)	312	13,63636
http://www.arts.ed.ac.uk/witches/index.html (The Survey of Scottish Witchcraft)	268	11,71329
http://www.arts.ed.ac.uk/finart/rome.html (Rome project)	212	9,265734
http://www.arts.ed.ac.uk/dost/ (Dictionary of the Older Scots Tongue)	192	8,391608
http://www.arts.ed.ac.uk/europgstudies/rprojects/avant-garde/ (Avant Garde Project)	181	7,910839
http://www.arts.ed.ac.uk/gadda/ (Journal of Gadda Studies)	166	7,255245

L'AHDS est une organisation qui archive la production numérique des projets de recherche. Ainsi l'on s'attendrait à ce que la plupart des utilisateurs y accèdent afin d'effectuer des recherches dans ces données archivées en vue d'une réutilisation dans leur recherche, et non pour se connecter à d'autres ressources. Toutefois, il s'est dégagé le schéma suivant, confirmé par les résultats de notre questionnaire : de nombreuses pages Web qui présentaient un lien sortant sur les centres de l'AHDS concernaient, la plupart du temps, le dépôt de matériaux ou des informations sur les droits d'auteur, comme le montre l'exemple sur l'archéologie dans le tableau 8.

Il semblerait donc que le dépôt de données soit plus fréquent que la réutilisation. Nous avons toutefois été surpris du nombre de liens sortants qui dirigeaient vers des ressources informationnelles, même dans les logs de l'AHDS. Dans la section histoire, par exemple, on peut également voir des liens pointant vers des ressources, qui sont des ensembles de données

13. Ces adresses sont celles utilisées à l'époque de l'étude, néanmoins l'adresse centrale reste la même : < <http://www.arts.ed.ac.uk/> > (consulté le 5 décembre 2011).

génériques, telles que des données de recensement ou encore des cartes historiques, comme le montre le tableau 9.

Tableau 6 : ressources les plus populaires sur le nom de domaine de l'université de Sheffield¹⁴

URL	Fréquence	%
http://www.shef.ac.uk/f/frenchfilmstars/home.html (French Film Stars Project)	260	6,537591
http://www.shef.ac.uk/french/research/gide.html (André Gide Editions Project)	231	5,808398
http://www.shef.ac.uk/uni/academic/A-C/bakh/ (Bakhtin Project)	211	5,305507
http://www.shef.ac.uk/uni/academic/N-Q/phil/AHRB-Project/index.html	193	4,852904
http://www.shef.ac.uk/assem/ (Assemblage)	165	4,148856
http://www.shef.ac.uk/uni/academic/A-C/bakh/sociolinguistics.html (Bakhtin project, sociolinguistique)	159	3,997988
http://www.shef.ac.uk/uni/academic/A-C/biblst/DJACurrres/Postmodern2/Dictionary.html (dictionnaire d'hébreu classique)	155	3,89741
http://www.shef.ac.uk/english/language/quantling/index.html (linguistique quantitative)	151	3,796832
http://www.shef.ac.uk/uni/projects/ptpdlp/ (cours en ligne Pathways to Philosophy)	112	2,816193
http://www.shef.ac.uk/japan2001/ (Waka for Japan 2001)	105	2,640181
http://www.shef.ac.uk/uni/academic/A-C/archst/research/process/cp01.html (Architecture, research process module)	101	2,539603
http://www.shef.ac.uk/alcs/ (Association of Low Country Studies)	97	2,439024
http://www.shef.ac.uk/assem/4/ (Assemblage, n° 4)	92	2,313301
http://www.shef.ac.uk/p/partonopeus/contents.htm (Partonopeus of Blois Project)	92	2,313301
http://www.shef.ac.uk/hri/ (Humanities Research Institute)	88	2,212723
http://www.shef.ac.uk/archaeology/research/madagascar/ (Tombs, Landscape and Society in Southern Madagascar)	82	2,061856

Cela sous-entend que même si les utilisateurs sont conscients que l'AHDS archive une grande quantité de ressources de recherche spécialisées, la

14. Ces adresses sont celles utilisées à l'époque de l'étude, néanmoins l'adresse centrale reste la même : < <http://www.shef.ac.uk/> > (consulté le 5 décembre 2011).

majorité d'entre eux sont eux-mêmes producteurs ou sont, une fois encore, en quête de vastes collections de références. Nous trouvons bel et bien des mentions de projets de recherche individuels *via* l'AHDS, mais ils apparaissent très peu souvent. Cela tendrait à indiquer que les universitaires sont tout disposés à archiver leurs recherches, mais moins prompts à réutiliser des données ou des ressources créées par d'autres universitaires. Cette impression est étayée par le travail que nous avons effectué subséquemment lorsque nous avons tenté de réintroduire les ressources négligées par les universitaires en sciences humaines. Étant donné qu'ils se sont de toute évidence habitués aux niveaux élevés de contenu et de fourniture de données pratiqués par les organisations commerciales, les bibliothèques, les archives et les musées, et à leur facilité d'accès, les participants ont souvent été déçus par la qualité des ressources universitaires. D'un autre côté, ils avaient le sentiment de pouvoir se fier à des ressources produites par une organisation de documentation, l'Imperial War Museum, au vu de la réputation de l'organisation en matière de qualité de ses données dans le monde analogique [Warwick *et al.*, 2008].

Tableau 7 : répartition par pays (DNS) des sessions d'utilisateurs pour AHDS – instituts universitaires seulement

Pays	Utilisateurs
Royaume-Uni	85,9%
États-Unis	6,9%
Australie	3,8%
Autres	2,2%
Allemagne	1,2%

Tableau 8 : pages « archéologie » consultées sur AHDS

Pages	Consultation
Autres	27,1%
Dépôt de travaux	7,1%
Guides	5,8%
Formats des travaux à déposer	5,4%
Équipe	4,9%
Gestion de projets	4,1%
Évaluation	3,8%
Copyright – introduction	3,7%
Roches	3,6%

Tableau 9 : pages « histoire » consultées via AHDS

Pages	Consultation
Cartes historiques	11,1%
Projets	10,7%
Personnel	10,6%
Statistiques de recensement	5,1%
CHCC	4,6%
Recensement XIX ^e siècle	4,6%
Accès aux CHCC	2,6%
A-propos	2,3%
Collections spéciales	2,3%

N.B. : CHCC signifie *contemporary and historical census collections* (recensements actuels et passés).

Résultats du questionnaire

+++++

Données démographiques

Nous avons reçu 149 réponses au questionnaire sur une période de 4 mois. 85% des participants étaient du Royaume-Uni, les étrangers les plus représentés étant les Américains des États-Unis, les Canadiens et les Australiens. Le tableau 10 met en lumière le statut de ces participants. La catégorie « Autres » est la plus importante ; elle inclut des participants non basés au Royaume-Uni, des universitaires à la retraite, du personnel du service informatique et des chercheurs amateurs intéressés par le sujet. Il en ressort néanmoins que la majorité des participants étaient parties prenantes dans des travaux universitaires, qu'ils soient professeurs, assistants ou étudiants. C'était peut-être prévisible, dans la mesure où ces portails sont destinés aux acteurs de l'enseignement supérieur au Royaume-Uni. Les résultats indiquent que toutes les disciplines couvertes par le domaine de l'AHRC (évoquées plus haut) étaient représentées de façon à peu près égale, et qu'un tiers des participants disaient avoir entrepris une recherche multidisciplinaire. Cette donnée démographique signifie que notre échantillon peut être comparé aux études menées dans la littérature, citées plus haut, malgré le fait que notre échantillon était davantage « de commodité » que destiné à représenter fidèlement la population universitaire du Royaume-Uni.

Tableau 10 : statut des participants

Statut	% des réponses totales (n = 149)
Autres	20
Chercheur indépendant	19
Conférencier/universitaire	19
Assistant dans l'enseignement supérieur	14
Étudiant-chercheur en 3 ^e cycle	13
Chercheur postdoctorant	8
Étudiant en 3 ^e cycle	7

L'importance des ressources informationnelles a immédiatement été visible à la lecture des réponses au questionnaire. Comme le rapport de la British Academy l'avait souligné en 2005, et contrairement à certaines idées reçues, les universitaires des sciences humaines ne sont pas des « technophobes » hostiles aux nouvelles technologies et préférant fréquenter les bibliothèques et les archives en quête de sources imprimées. En effet, 149 de nos participants se sont déclarés enthousiasmés par l'utilité des ressources numériques : 87% utilisaient Internet quotidiennement – 48 % pendant plus de quatre heures par jour –, 81 % se désignaient eux-mêmes comme des utilisateurs assidus des ressources numériques, 88 % étaient d'accord ou entièrement d'accord pour dire qu'ils ne pouvaient pas faire leur travail sans ressources numériques, et 77 % étaient d'accord ou entièrement d'accord avec le fait que les ressources numériques avaient changé leur manière de travailler. Ces résultats sont comparables à ceux de l'IHR et de l'université de Bristol, évoqués plus haut.

Les ressources numériques les plus utiles

Afin de ne pas trop influencer les utilisateurs, nous avons décidé de ne pas donner de définition des ressources numériques. Mais afin de comprendre quelle était celle des utilisateurs, nous leur avons demandé d'énumérer leurs trois ressources préférées, autrement dit celles qu'ils jugeaient le plus utiles à leurs recherches. Leurs réponses ont donné une majorité écrasante de ressources très génériques, comparables en version papier à des textes de référence ou même à une bibliothèque ou à des archives. Comme l'indique le tableau 11, un très large éventail de ressources et de sites ont été mentionnés, mais le site Internet de la bibliothèque de l'université arrive loin devant, avec 13 % d'utilisateurs la citant comme

ressource la plus importante. En comparaison, Google n'a recueilli que 4 % des votes.

« Autres » ressources – collections d'information et de références

Comme le montre le tableau 11, la catégorie la plus importante de ressources est celle intitulée « Autres ». Le tableau 3 fait état des détails de toutes les ressources mentionnées. Cependant, la majorité d'entre elles pourraient entrer dans ce que l'on pourrait appeler les ressources informationnelles ou de références, telles que les bibliothèques, les archives et les portails thématiques, qu'ils soient financés par l'argent public ou commerciaux. Ce sont, par exemple, la British Library, la National Library of Scotland, les National Archives, JSTOR, l'AHDS ou Humbul, SOSIG, Literature Online (LION) et le Dictionary of National Biography (DNB). Des centres thématiques spécialisés tels que Palatine¹⁵ (pour la danse, le théâtre et la musique) ont également été mentionnés, ainsi que des portails de documentation créés à titre privé tels que Voice of the Shuttle¹⁶ et le Online Reference Book for Mediaeval Studies¹⁷, ou encore des bibliothèques thématiques en ligne comme Perseus Digital Library¹⁸ et la Royal Historical Society Bibliography¹⁹.

Les participants au questionnaire n'ont cité que quatre projets de recherche financés par l'État britannique :

1. Old Bailey Online²⁰ [Shoemaker, 2005].
2. Practice as Research in Performance (PARIP)²¹.
3. Powys Digital History Project, produit par le service des archives du Powys²² [Reid, 2000].
4. Le site Photographic Exhibitions in Britain, basé à la National Gallery of Canada, également en partie financé par l'AHRC²³.

15. [En ligne] < <http://www.palatine.ac.uk> > (consulté le 13 juillet 2010).

16. [En ligne] < <http://vos.ucsb.edu> > (consulté le 13 juillet 2010).

17. [En ligne] < <http://the-orb.net> > (consulté le 13 juillet 2010).

18. [En ligne] < <http://www.perseus.tufts.edu> > (consulté le 13 juillet 2010).

19. N.D.T. : la Bibliography of British and Irish Studies (BBIH) a remplacé la Royal Historical Society (RHS) Bibliography en janvier 2010. [En ligne] < <http://www.history.ac.uk/projects/bbih> > (consulté le 5 décembre 2011).

20. [En ligne] < <http://www.oldbaileyonline.org> > (consulté le 13 juillet 2010).

21. [En ligne] < <http://www.bristol.ac.uk/parip> > (consulté le 13 juillet 2010).

22. [En ligne] < <http://history.powys.org.uk> > (consulté le 13 juillet 2010).

23. [En ligne] < <http://peib.dmu.ac.uk/> > (consulté le 13 juillet 2010).

Tableau 11 : ressources numériques les plus intéressantes

Autres	64 %
Bibliothèque universitaire	13 %
Google	8 %
AHDS / Humbul	4 %
Web of Knowledge	3 %
Média d'information	2 %
Espaces d'apprentissage virtuel universitaires	2 %

Figuraient également deux projets de recherche financés par les États-Unis : le site Internet de Child Language Data Exchange (CHILDES)²⁴ et la Perseus Digital Library²⁵. On peut souligner que tous les sites mentionnés ci-dessus sont aussi des ressources de références qui rassemblent ou numérisent une importante quantité d'informations afin que des universitaires d'horizons variés puissent les consulter, et non pas des sites qui produisent les résultats d'un projet de recherche original. Deux d'entre eux ont également été produits par une bibliothèque et un centre d'archives. Bien entendu, cela ne signifie pas que les participants n'ont jamais recours à des ressources numériques spécialisées, puisque nous les avons seulement interrogés sur celles qu'ils utilisaient le plus souvent, mais de toute évidence, ils n'utilisent pas de telles ressources aussi fréquemment que les sites qui concentrent une multitude de documents, qu'ils soient physiques ou virtuels. Ces résultats sont également étayés par une étude menée dans le cadre d'un projet jumeau dans notre département – User-Centred Interactive Search with Digital Libraries²⁶ – qui a mis au jour une préférence similaire pour les ressources génériques chez les universitaires en sciences humaines consultés [Rimmer *et al.*, 2006].

Domaines thématiques

+++++

Lorsque les données sont réparties entre les domaines thématiques selon lesquels l'AHRC organise ses axes de recherche, les mêmes schémas se dégagent. Les ressources les plus spécialisées ne sont mentionnées qu'une seule fois et ont donc été étiquetées « Autres ». La bibliothèque

24. [En ligne] < <http://childes.psy.cmu.edu> > (consulté le 13 juillet 2010).

25. [En ligne] < <http://www.perseus.tufts.edu/> > (consulté le 13 juillet 2010).

26. [En ligne] < <http://www.ucl.ac.uk/annb/ucis.html> > (consulté le 13 juillet 2010).

universitaire demeure la ressource la plus populaire dans tous les domaines sauf deux : langues classiques, histoire ancienne et archéologie d'une part, et arts et médias visuels d'autre part. Dans ces deux domaines, c'est Google qui est cité comme ressource principale dans le travail. Toutefois, l'on pourrait avancer que dans le cas des langues classiques, la bibliothèque réelle a été remplacée par la bibliothèque en ligne, puisque la Perseus Digital Library, un ensemble de ressources classiques qui comprend textes, images et matériel de réalité virtuelle, s'avère être très populaire.

Néanmoins, dans les autres disciplines, les ressources informationnelles comptent pour la moitié des ressources identifiées. Si un consensus existe en matière de ressources utiles, c'est sur les collections d'informations.

Bien que la bibliothèque universitaire demeure d'une importance primordiale, d'autres bibliothèques, telles que la Bodleian à Oxford ou la British Library, sont citées par des disciplines spécialisées. Toutes les autres ressources mentionnées plus d'une fois sont de vastes regroupements de références tels que le DNB Online, JSTOR, Early English Books Online (EEBO), LION, Lexis Nexis, Grove Online, Répertoire international de littérature musicale, et Westlaw. Les sites d'information en ligne sont également d'importantes ressources numériques dans plusieurs disciplines. Toutes les mentions qui précèdent sont bien entendu des services commerciaux, et notre étude qualitative a démontré par la suite que les utilisateurs se sont très vite habitués aux niveaux élevés concernant l'exactitude du contenu, la mise à jour et le design de l'interface que les produits commerciaux se doivent d'offrir. Il est également important de souligner que les universitaires accèdent à ces vastes sites d'information *via* le site Internet de leur université, ce qui renforce le statut de la bibliothèque en tant que fournisseur d'information de haute qualité, en laquelle les universitaires ont confiance.

Comparaison avec les données de l'IHR

++++
Nous avons comparé nos données sur les domaines thématiques avec celles de l'étude menée par l'IHR, qui était d'accord pour poser la même question. Bien qu'ils aient choisi de ne pas permettre aux utilisateurs d'inclure des ressources génériques telles que Google et les catalogues de bibliothèques en ligne, les données étaient très similaires aux nôtres sur la mise en avant des ressources d'information comme le matériel de

recherche le plus précieux. Probablement en raison d'une prédominance d'historiens dans cette étude, il y avait un consensus beaucoup plus marqué sur les ressources les plus utiles. Bien que le questionnaire ait été mis au point par l'IHR, British History Online était l'une des ressources les plus populaires, et parmi celles citées à plusieurs reprises figuraient EEBO, LION, Eighteenth Century Collections Online (ECCO). Là encore, les autres ressources mentionnées tendaient à être des collections de références, produites par des bibliothèques ou des archives, ou des collections commerciales auxquelles les participants avaient accès *via* la bibliothèque de l'université. Et là aussi les ressources produites par des universitaires étaient en minorité, mais l'une des rares à être citées quelquefois était le Old Bailey Online, un projet remarquable pour sa popularité parmi les universitaires, que nous avons étudié en détail dans notre étude de cas suivante.

Dans cette section, nous avons démontré que les résultats du questionnaire indiquent une tendance générale des participants à considérer les ressources numériques comme moyen le plus utile d'accès aux ressources d'information. Ils préfèrent les portails d'informations (l'équivalent de la bibliothèque ou des archives) aux ressources de recherche spécialisées (l'équivalent d'une monographie d'un texte littéraire pour une étude primaire). La quantité de ressources qui tombent dans la catégorie « Autres » sous-entend également qu'il existe un très large éventail de ressources utilisées, et très peu de consensus sur la plus utile d'entre elles. Il faut également souligner le fait suivant, et peut-être s'en inquiéter : les universitaires semblent ne pas avoir recours aux ressources créées pour eux par d'autres universitaires, préférant celles créées par des producteurs commerciaux ou des spécialistes de la documentation dans des bibliothèques ou des archives.

DÉBAT

+++++

Collections d'informations

+++++

Les preuves récoltées *via* le questionnaire et l'analyse des *logs* suggèrent donc que les utilisateurs de ressources numériques en sciences humaines ont une haute estime des ressources informationnelles. Comme l'a avancé l'équipe du projet RePAH, la plupart des utilisateurs des sciences humaines ne font pas confiance aux collections présélectionnées ou pré-analysées et préfèrent prendre leurs propres décisions quant aux données

qu'ils trouvent à partir de grandes collections de ressources [Brown *et al.*, 2007, p. 22]. Une préférence similaire pour l'exhaustivité plutôt que la précision a été remarquée chez les historiens par Dalton et Charnigo [2004] et Duff *et al.* [2004]. Cela peut contribuer à expliquer pourquoi nous avons remarqué une nette préférence pour de larges collections de références numériques au détriment de projets de recherche spécialisés. Quelle qu'en soit la raison, cette préférence est en tout cas indéniable.

Les ressources informationnelles physiques demeurent très importantes. Cela démontre l'importance des structures universitaires traditionnelles pour les chercheurs en sciences humaines. Les ressources numériques n'ont pas encore remplacé les bibliothèques et les archives. Au lieu de cela, elles peuvent être envisagées, ainsi que les ressources Internet qu'elles produisent, comme un moyen de découvrir d'autres ressources. Ainsi, un universitaire consulte la page d'un centre de recherche pour se renseigner sur la date d'un séminaire, ou un historien cherche sur le Web les horaires d'ouverture d'un bureau d'archives local avant de s'y rendre. Non seulement les bibliothèques universitaires représentent le premier point d'accès aux ressources numériques pour beaucoup d'utilisateurs, mais les bibliothèques et les centres d'archives, nationaux ou spécialisés, sont également très estimés et largement utilisés. Cela va dans le sens de récentes recherches suggérant que les acteurs des sciences humaines ont toujours besoin de ressources génériques, traditionnelles, et qu'ils apprécient rencontrer les dépositaires du savoir face à face [Barrett, 2005]. Il sera important de prendre ces préférences et ces comportements en considération lorsque de nouvelles ressources seront conçues plutôt que d'essayer de remplacer le tangible par le numérique. Car comme l'ont démontré Anne Adams et Angela Sasse en 1999, si les ressources informationnelles sont conçues pour aller à l'encontre des pratiques des utilisateurs, il y a de fortes chances pour qu'elles soient évitées.

Il se peut également que les universitaires aient tendance à utiliser de grandes collections de références parce qu'ils sont habitués à leur mode de fonctionnement. Les revues en ligne connaissent un franc succès car bien que le mode d'accès à un article soit électronique, la façon dont cette information est utilisée est très habituelle. La plupart des gens impriment l'article choisi et le lisent [Liu et Stork, 2000]. De la même façon, il est très probable que l'usage des données offertes par DNB Online soit similaire à celui de la version papier, et les participants à l'étude IHR Peer Review ont admis qu'ils avaient tendance à citer la version imprimée, par exemple, des comptes rendus du Old Bailey, alors qu'ils ont eu recours à la

version électronique [IHR, 2006, p. 30]. Cela démontre que de telles ressources, étant habituelles, n'exigent pas d'efforts d'utilisation particuliers, à l'inverse des nouveaux logiciels d'analyse de données.

Le rôle des bibliothèques

+++++

Toutefois, une autre explication pour l'utilisation des ressources informationnelles peut se trouver dans le lien, une fois encore, à la bibliothèque universitaire, considérée par les universitaires comme une ressource numérique vitale. Selon notre étude, ils utilisent la page Web de la bibliothèque comme un portail d'accès à d'autres ressources, qu'il s'agisse de vastes collections de références ou de liens vers d'autres pages. Dans une étude à part, nous avons eu des difficultés à trouver des ressources numériques spécialisées en ce qui concerne la recherche en sciences humaines, que ce soit à partir de la page d'accueil d'un département ou la bibliothèque universitaire – par « spécialisées » nous entendons l'équivalent numérique des monographies, publiées au terme d'une recherche financée [Pappa *et al.*, 2006]. Cela peut contribuer à expliquer pourquoi une si grande partie des résultats utilisés sont justement des collections d'informations. L'accès à ces collections est en général payant et se fait par la bibliothèque, et il est possible que ces grandes collections d'informations vers lesquelles les bibliothécaires dirigent les utilisateurs *via* un lien hypertexte soient les seules connues des bibliothécaires, même ceux spécialistes d'une discipline. Ainsi, les utilisateurs tendent-ils à suivre les liens qu'on leur propose, et si ces derniers n'incluent pas de ressources numériques spécialisées, ils n'entreprendront pas de recherche plus approfondie pour en trouver.

Bien que les bibliothécaires spécialisés puissent tout à fait connaître des livres et revues qui existent dans leur domaine, ils ne sont peut-être pas aussi à jour sur les ressources numériques spécialisées et les logiciels d'analyse. À l'UCL SLAIS, un module sur les ressources numériques en sciences humaines²⁷ fournit une formation de ce type aux nouveaux diplômés ; toutefois, rester à la page est plus difficile pour les professionnels en milieu de carrière. Mais il pourrait s'agir d'un domaine où des cours de formation continue voient le jour.

27. [En ligne] < <http://www.ucl.ac.uk/slais/teaching/modules/instg008/> > (consulté le 13 juillet 2010).

Les ressources d'information et le financement

+++++

La préférence des utilisateurs pour les ressources d'information génériques au détriment des ressources spécialisées a plusieurs conséquences. Le rapport de la British Academy suggère qu'étant donné leur préférence pour ce qu'ils appellent les ressources secondaires, telles que les catalogues de bibliothèques, il faudrait donner la priorité à la numérisation de telles données, en préférence à celle des ressources primaires [British Academy, 2005]. D'autres projets d'ICT Strategy ont mis au jour le souhait de davantage de ressources numériques, mais comme le rapport de la British Academy l'exprime clairement, même avec le plus optimiste des programmes de numérisation, la plupart des ressources en sciences humaines resteront probablement telles quelles pendant encore de longues années.

Nos recherches montrent en effet l'importance constante de l'objet tangible et des centres de recherche, bibliothèques et archives réels. Cependant, nos résultats indiquent également que les chercheurs en sciences humaines trouvent les grandes collections de références tout aussi utiles que les outils d'aide à la recherche. En termes de priorités de financement, cela signifie donc qu'à l'heure actuelle, les projets qui rassemblent ces vastes collections de ressources, qu'elles soient génériques ou thématiques, sont bienvenus et ont des chances d'être très utilisés. Ce qui a tendance à avantager les ressources commerciales ou celles des bibliothèques et des archives, puisque ces dernières tendent à numériser des collections entières, sans tenir compte uniquement des éléments qui pourraient être utiles aux chercheurs. Il apparaît que tant que la qualité du matériel est bonne, c'est tout ce que les universitaires demandent. Cela n'exclut pas le financement de projets de recherche de moindre envergure, plus spécialisés, offrant des éléments plus sélectifs. Il y a toutefois peu de chances pour que de telles ressources attirent un grand nombre d'utilisateurs. Les organismes de financement devront alors se poser la difficile question de savoir si le degré d'utilisation doit devenir un critère de financement des projets de recherche, ou si de tels projets doivent être considérés comme érudition pure dont l'utilisation ultérieure n'est pas envisagée. Mais cela pose ensuite des questions tout aussi épineuses sur la façon d'archiver de tels travaux au sein d'une organisation telle que l'AHDS.

CONCLUSION

+++++

Nous avons débuté cet article par l'évocation d'une bibliothèque universitaire qui pensait que les ressources numériques pouvaient remplacer les bibliothèques « physiques » et les professionnels de l'information en tant qu'intermédiaires. Fort heureusement pour l'avenir de la profession, notre étude indique que cette vision des choses est fondamentalement erronée. Tout au moins en sciences humaines, les ressources numériques n'ont pas remplacé la bibliothèque. Leur fonction de portails d'information n'aura fait qu'accroître leur importance. Loin d'être superflues, les ressources numériques exigent des bibliothécaires qu'ils endossent de nouveaux rôles. Ils sont donc devenus des fournisseurs, des producteurs, des gardiens et des intermédiaires de l'information. Ils assument à présent, en termes numériques, certains rôles remplis par des éditeurs dans le monde de l'édition papier (dans le cas, par exemple, des sites de dépôts institutionnels), et la bibliothèque est désormais envisagée comme une porte d'entrée vers d'autres collections de ressources numériques [Unsworth, 2005]. Il est donc capital que les bibliothèques et les centres d'archives soient convenablement financés, et que les dépenses des TIC ne soient pas perçues comme un pas vers le remplacement des ressources tangibles et du personnel.

Le discernement des professionnels de l'information est encore peut-être plus important. À mesure que le volume des documents numériques augmente, les utilisateurs peinent de plus en plus à se maintenir à niveau. Nos recherches ont cependant montré que les utilisateurs préfèrent avoir recours à de grandes collections de documents plutôt qu'à des ressources spécialisées. Les universitaires font confiance à leur bibliothèque pour leur fournir des liens vers des documents de qualité et pour leur servir d'accès à ces vastes collections d'information. La bibliothèque incarne donc une garantie cruciale de la bonne qualité de telles ressources, qu'elles soient de grandes collections commerciales ou des pages Web pointant vers des ressources du domaine public.

Il est évident que les acteurs professionnels des bibliothèques et des centres d'archives ont toujours un rôle très important à jouer en tant qu'intermédiaires de l'information. L'un des participants à notre questionnaire, un universitaire expert en ressources numériques, a fait la remarque suivante : « Ce que veulent les gens, de plus en plus, ce sont des conseils pour s'y retrouver dans le nombre énorme de ressources numériques, ils sont tout simplement déconcertés par la quantité d'informations disponibles

et ne savent pas quoi en faire. J'ai remarqué que les gens sont passés de "Waouh, c'est génial que vous ayez fait ça" à "Oui, c'est génial que vous ayez fait ça, mais comment est-ce que ça fonctionne avec, vous savez, l'ensemble Truc", ou "Comment est-ce que je l'intègre dans ces autres choses qui sont en train de se passer ?". Et donnez-moi juste une liste des [...] dix principales. »

C'est en effet le genre d'intermédiation dont les bibliothécaires ont toujours été en charge. Loin de rendre les compétences des professionnels de l'information inutiles, leur expertise est de plus en plus demandée et le domaine des compétences qu'on attend d'eux s'élargit.

RÉFÉRENCES BIBLIOGRAPHIQUES

+++++

Adams Anne et Sasse M. Angela, "Users are not the enemy: why users compromise security mechanisms and how to take remedial measures", *Communications of the ACM*, vol. 42, n° 12, 1999, pp. 40-46.

Albitz Paul et Liu Cricket, *DNS and BIND*, Sebastopol (Ca), O'Reilly, 5^e édition, 2006.

Anderson Ian G., "Are you being served? Historians and the search for primary sources", *Archivaria*, vol. 58, 2004, pp. 81-130.

[En ligne] < <http://journals.sfu.ca/archivar/index.php/archivaria/article/viewArticle/12479> > (consulté le 5 octobre 2007).

Barrett Andy, "The information seeking habits of graduate student researchers in the humanities", *Journal of Academic Librarianship*, vol. 31, n° 4, 2005, pp. 324-331.

Bates Marcia J., "The Getty end-user online searching project in the humanities: report n° 6: overview and conclusions", *College and Research Libraries*, vol. 57, n° 6, 1996, pp. 514-523.

Bates Marcia J., "The cascade of interactions in the digital library interface", *Information Processing and Management*, vol. 38, n° 3, 2002, pp. 381-400.

British Academy, "E-resources for research in the humanities and social sciences: a British Academy policy review", 2005.

[En ligne] < <http://www.britac.ac.uk/policy/eresources/e-resources.cfm> > (consulté le 5 octobre 2007). Summary of the Review report (PDF).

Brown Sheridan et Swan Alma. "Researchers' use of academic libraries and their services", Research Information Network and Consortium of Research Libraries, 2007.

[En ligne] < <http://www.rin.ac.uk/system/files/attachments/Researchers-libraries-services-report.pdf> > (consulté le 5 octobre 2007).

Brown Stephen, Ross Robb, Gerrard David, Greengrass Mark et Bryson Jared, "RePAH: a user requirements analysis for portals in the arts and humanities", Arts and Humanities Research Council, 2007.

[En ligne] < <http://repah.dmu.ac.uk/report/pdfs/RePAHReport-Complete.pdf> > (consulté le 5 octobre 2007).

Chu Clara M., "Literary critics at work and their information needs: a research-phases model", *Library and Information Science Research*, vol. 21, n° 2, 1999, pp. 247-273.

Crane Gregory, "What do you do with a million books?", *D-Lib Magazine*, vol. 12, n° 3, 2006.

Cronin Blaise, "Information professionals in the digital age", *International Information and Library Review*, vol. 30, n° 1, 1998, pp. 37-50.

Curtis Polly, "Bangor librarians face Internet threat", *Guardian Unlimited*, 16 février 2005.

[En ligne] < http://www.guardian.co.uk/uk_news/story/0,3604,1415830,00.html > (consulté le 5 octobre 2007).

Dalton Margaret Stieg et Charnigo Laurie, "Historians and their information sources", *College and Research Libraries*, vol. 65, n° 5, 2004, pp. 400-425.

Duff Wendy M. et Cherry Joan M., "Use of historical documents in a digital world: comparisons with original materials and microfiche", *Information Research*, vol. 6, n° 1, 2001.

[En ligne] < <http://InformationR.net/ir/6-1/paper86.html> > (consulté le 5 octobre 2007).

Duff Wendy M., Craig Barbara et Cherry Joan, "Historians' use of archival sources: promises and pitfalls of the digital age", *Public Historian*, vol. 26, n° 2, 2004, pp. 7-22.

Ellis David et Oldman Hanna, "The English literature researcher in the age of the Internet", *Journal of Information Science*, vol. 31, n° 1, 2005, pp. 29-36.

Fourie Ina, "Should we take disintermediation seriously?", *Electronic Library*, vol. 17, n° 1, 1999, pp. 9-16.

Gellman Robert, "Disintermediation and the Internet", *Government Information Quarterly*, vol. 13, n° 1, 1996, pp. 1-8.

Green Rebecca, "Locating sources in humanities scholarship: the efficacy of following bibliographic references", *Library Quarterly*, vol. 70, n° 2, 2000, pp. 201-229.

Harley Diane et Henke Jonathan, "Toward an effective understanding of website users: advantages and pitfalls of linking transaction log analyses and online surveys", *D-Lib Magazine*, vol. 13, n° 3-4, 2007.

[En ligne] < <http://www.dlib.org/dlib/march07/harley/03harley.html> > (consulté le 5 octobre 2007).

Herman Eti, "End-users in academia: meeting the information needs of university researchers in an electronic age. Part 2. Innovative information-accessing opportunities and the researcher: user acceptance of IT-based information resources in academia", *Aslib Proceedings*, vol. 53, n° 10, 2001, pp. 431-457.

Huntington Paul, Nicholas David, Williams Peter et Gunter Barrie, "Characterising the health information consumer: an examination of the health information sources used by digital television users", *Libri*, vol. 52, n° 1, 2002, pp. 16-27.

Huxley Lesly, Mullings Christine, Hodos Tamar et Jones Diane, *Gathering Evidence: Current ICT Use and Future Needs for Arts and Humanities Researchers*, London, Arts and Humanities Research Council, 2007.

Institute of Historical Research, *Peer Review and Evaluation of Digital Resources for the Arts and Humanities*, London, Arts and Humanities Research Council, 2006.

[En ligne] < <http://www.history.ac.uk/projects/peer-review> > (consulté le 13 juillet 2010).

Lehmann Stephen et Renfro Patricia, "Humanists and electronic information services: acceptance and resistance", *College and Research Libraries*, vol. 52, n° 5, 1991, pp. 403-413.

Lin Winston T. et Shao Benjamin B.M., "The business value of information technology and inputs substitution: the productivity paradox revisited", *Decision Support Systems*, vol. 42, n° 2, 2006, pp. 493-507.

Liu Ziming et Stork David, "Is paperless really more? Rethinking the role of paper in the digital age", *Communications of the ACM*, vol. 42, n° 11, 2000, pp. 94-97.

Marty Paul F., "The digital museum in the life of the user", Digital Humanities, Urbana-Champaign, University of Illinois, 2-8 juin 2007.
[En ligne] < <http://www.digitalhumanities.org/dh2007/dh2007.abstracts.pdf> > (consulté le 5 octobre 2007).

MLA, "MLA task force on evaluating scholarship for tenure and promotion", 2007.
[En ligne] < http://www.mla.org/tenure_promotion > (consulté le 5 octobre 2007).

Pappa Nikoleta, Warwick Claire, Terras Melissa et Huntington Paul, "The (in)visibility of digital humanities resources in academic contexts", Digital Humanities, Paris, Université de la Sorbonne, 5-9 juillet 2006.

Reid Gordon, "The digitisation of heritage material: arguing for an interpretative approach based on the experience of the Powys Digital History Project", *Program: Electronic Library and Information Systems*, vol. 34, n° 2, 2000, pp. 143-158.

Rimmer Jon, Warwick Claire, Blandford Ann, Gow Jeremy et Buchanan George, "User requirements for humanities digital libraries", Digital humanities, Paris, Université de la Sorbonne, 5-9 juillet 2006.
[En ligne] < <http://www.ucl.ac.uk/annb/docs/jrcwabjggbdh2006preprint.pdf> > (consulté le 5 octobre 2007).

Science Watch, "Citations reveal concentrated influence: some fields have it, but what does it mean?", Science Watch, 1999.
[En ligne] < http://archive.sciencewatch.com/jan-feb99/sw_jan-feb99_page1.htm > (consulté le 5 octobre 2007).

Shoemaker Robert, "Digital London: creating a searchable Web of interlinked sources on eighteenth century London", *Program: Electronic Library and Information Systems*, vol. 39, n° 4, 2005, pp. 297-311.

Stone Sue, "Humanities scholars: information needs and uses", *Journal of Documentation*, vol. 38, n° 4, 1982, pp. 292-313.

Talja Sanna et Maula Hanni, "Reasons for the use and non-use of electronic journals and databases: a domain analytic study in four scholarly disciplines", *Journal of Documentation*, vol. 59, n° 6, 2003, pp. 673-691.

Unsworth John, "Pubrarians and liblishers: new roles for old foes", Annual meeting of the Society for Scholarly Publishing (association de l'édition universitaire), Boston, 2 juin 2005.

Warwick Claire, Terras Melissa, Huntington Paul et Pappa Nikoleta, "If you build it will they come? The LAIRAH study: quantifying the use of online resources in the arts and humanities through statistical analysis of user log data", *Literary and Linguistic Computing*, vol. 23, n° 1, 2008, pp. 85-102.

Watson-Boone Rebecca, "The information needs and habits of humanities scholars", *Reference Quarterly*, vol. 34, n° 2, 1994, pp. 203-216.

Whitmire Ethelene, "Disciplinary differences and undergraduates' information-seeking behaviour", *Journal of the American Society for Information Science and Technology*, vol. 53, n° 8, 2002, pp. 631-638.

Wiberley Stephen E., "Subject access in the humanities and the precision of the humanist's vocabulary", *Library Quarterly*, vol. 53, n° 4, 1983, pp. 420-433.

Wiberley Stephen E., "Names in space and time: the indexing vocabulary of the humanities", *Library Quarterly*, vol. 58, n° 1, 1988, pp. 1-28.

Wiberley Stephen E., "Time and technology", *College and Research Libraries*, vol. 61, n° 5, 2000, pp. 421-431.

par Marin Dacos et Emma Bester

+++++

QUE SAVONS-NOUS
DE L'IDENTITÉ,
DES COMPORTEMENTS
ET DES ATTENTES
DES LECTEURS DE
REVUES.ORG EN
2008 ET 2009 ?

+++++

+++++

QUE SAVONS-NOUS DE L'IDENTITÉ, DES COMPORTEMENTS ET DES ATTENTES DES LECTEURS DE REVUES.ORG EN 2008 ET 2009 ?

+++++

Le Centre pour l'édition électronique ouverte (Cléo), unité mixte de recherche et de service du CNRS, de l'EHESS, de l'université de Provence et de l'université d'Avignon, développe une plateforme d'édition électronique complète pour la communauté scientifique des sciences humaines et sociales. Le cœur en est Revues.org qui, créé en 1999, est le plus ancien portail français de revues scientifiques en sciences humaines et sociales ; il développe une politique favorable au libre accès ; il présente ce jour 140 revues électroniques offrant plus de 40 000 articles en libre accès en texte intégral. Il s'accompagne de *Calenda*, le calendrier des sciences sociales (10 000 événements scientifiques), et d'*Hypothèses*, plateforme de publication de carnets de recherche en sciences humaines et sociales (2 000 billets publiés depuis 2008).

CONTEXTE, ENJEUX ET LIMITES DE L'ÉTUDE SUR LE LECTORAT DE REVUES.ORG

+++++

Avec une fréquentation en croissance constante et plus d'un million de visites mensuelles (figure 1), Revues.org offre un point d'observation sur les pratiques de lecture en ligne en sciences humaines et sociales francophones. Par ailleurs, afin de mieux répondre aux besoins et attentes des lecteurs de revues, un questionnaire bilingue français-anglais a été mis en ligne sur le portail de Revues.org fin avril 2009¹. Un premier volet de questions interroge les lecteurs sur leur connaissance et usages spécifiques de Revues.org et des services associés. Le volet suivant vise à appréhender plus largement leurs pratiques relatives aux suivi et recherche

1. Le questionnaire est actuellement actif, en ligne sur < <http://leo.hypotheses.org/1762> > (consulté le 6 mai 2009).

des textes. Le dernier ensemble de questions permet d'identifier le profil des utilisateurs du portail.

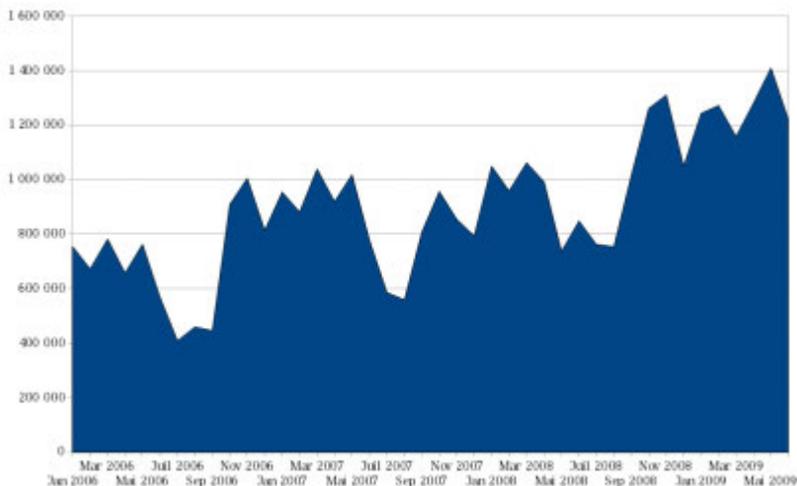


Figure 1 : visites mensuelles 2006-2009 (relevé des logs sur le serveur Revues.org de janvier 2006 à juin 2009)

Source : Cléo.

Afin de croiser les données et les approches, les premiers résultats sont comparés avec les résultats d'une enquête similaire menée par le portail Cairn fin 2008 ; 700 questionnaires complets ont été obtenus pour chacune des deux enquêtes. Ils sont également rapportés à des analyses de logs issus des serveurs de Revues.org, afin de compenser les biais d'une seule approche par questionnaire. Des différences importantes apparaissent entre les deux méthodes. Cela se remarque particulièrement dans les réponses à la question portant sur les revues consultées sur le portail (tableau 1) : la revue *Études rurales*, par exemple, très peu citée par les répondants à l'enquête, est l'une des plus visitées sur le portail et, *a contrario*, les *Cahiers du Centre de recherches historiques*, les *Cahiers d'histoire*, ou encore la *Revue de l'histoire des religions* sont souvent cités comparativement au nombre restreint de visites mesurées grâce aux logs. Plusieurs facteurs expliquent ces différences.

D'une part, la réponse à la question sur les pratiques de lecture décrit vraisemblablement autant des pratiques de lecture avérées qu'un indice de notoriété de chaque titre, voire, dans certains cas, de légitimité d'un titre bien choisi...

D'autre part, le nombre de personnes ayant répondu à l'enquête en ligne correspond à une faible part du lectorat de Revues.org. Le 2 juillet 2009, 700 réponses avaient été collectées, alors que la Lettre électronique de Revues.org avait 35 000 abonnés (figure 2) et que le site comptabilisait 1,4 million de visiteurs en mai 2009 (figure 1). La population répondante fait partie du lectorat fidèle de Revues.org, puisque 450 d'entre eux sont abonnés à la Lettre électronique (figure 9).

**Tableau 1 : « Je consulte... »
 (décompte enquête, visites mesurées / logs)**

Revue et carnets de recherche	Enquête	Logs (visites)	Remarque
Terrain	153	38 726	
L'Homme	143	25 077	
Clio. Histoire femmes et sociétés	132	43 187	
Cultures & conflits	127	39 060	
SociologieS	124	7 507	Incohérence apparente
Socio-anthropologie	122	13 023	
Cahiers de la Méditerranée	115	13 938	
EspacesTemps.net	114	37 501	
Archives de sciences sociales des religions	109	22 675	
Revue d'histoire du XIX ^e siècle	106	22 242	
Cahiers d'histoire	102	5 745	Incohérence apparente
VertigO	92	27 645	
Etudes rurales	90	10 227	Incohérence apparente
Cahiers d'études africaines	87	20 147	
Corps et culture	87	10 949	
Revue de l'histoire des religions	86	7 480	Incohérence apparente
Annuaire de l'Ecole pratique des hautes études. Section des sciences historiques et philologiques	80	3 192	
La Revue pour l'histoire du CNRS	79	9 735	
Cahiers de recherches médiévales	78	18 025	
Cahiers du Centre de recherches historiques	76	198	Incohérence apparente
Cybergeo : revue européenne de géographie	73	58 363	Incohérence apparente
<i>Il manque Nuevo Mundo, 98 000 visites...</i>			Incohérence apparente

Leurs réponses permettent cependant des premières interprétations et hypothèses sur l'identité, les comportements et les attentes des lecteurs de revues en sciences humaines et sociales. L'article rend compte des résultats de l'enquête et s'arrête, en trois temps successifs, sur la sociologie du lectorat, ses usages et pratiques documentaires dans le champ des sciences sociales et humaines, et ses attentes.

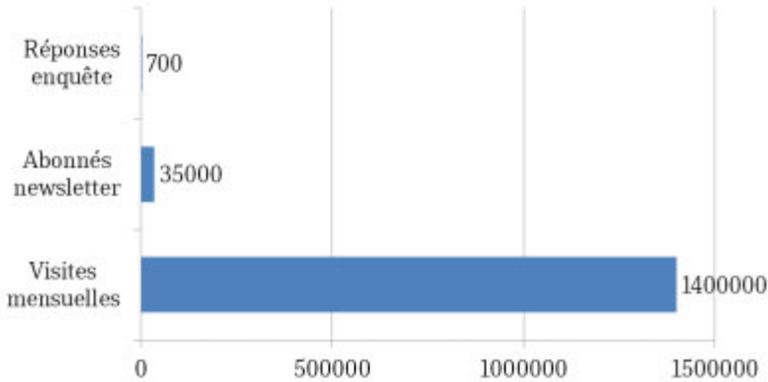


Figure 2 : 700 réponses à l'enquête en ligne sur *Revue.org*
Source : Cléo. Enquête en ligne 2009.

UN LECTORAT DE HAUT NIVEAU, LARGEMENT INTERNATIONAL

+++++
L'étude des profils fait émerger deux caractéristiques notables concernant l'origine géographique et l'appartenance socioprofessionnelle des lecteurs de *Revue.org*.

Lectures internationales

+++++
Un tiers des lecteurs répondants n'habitent pas en France : originaires principalement du Maghreb, du Canada et des pays limitrophes que sont la Suisse et la Belgique, ces lecteurs ont en commun d'être francophones, ce qui correspond à un déterminisme linguistique logique pour un portail essentiellement rédigé en français et pour une enquête rédigée en deux langues (français et anglais). Une comparaison des résultats de l'enquête avec les *logs* de fréquentation du serveur (figure 3), si elle confirme l'importance de ce bassin francophone dans l'ensemble du lectorat de *Revue.org*, fait par ailleurs apparaître les États-Unis et la Grande-Bretagne, pourtant quasiment absents de l'enquête en ligne.

Se pencher sur la question des répondants hispanophones permet de mieux cerner les limites de l'enquête. Quasiment absents de l'enquête, ils sont pourtant très nombreux à consulter les revues hispanophones, au premier rang desquelles on compte *Nuevo Mundo*. En effet, le nombre de pages accédées par les ressortissants espagnols et mexicains est comparable à celui des pages accédées via l'Italie, l'Allemagne ou le Brésil. La faiblesse des réponses s'explique sans doute par l'absence de questionnaire en espagnol, ainsi que par l'identification probablement moins forte du portail pour les revues de ce type. Dès lors, la réponse à l'enquête du portail apparaît sans doute moins intéressante.

Quoi qu'il en soit, la différence entre les statistiques de fréquentation et l'enquête qualitative en ligne met à nouveau en évidence les limites de la seule enquête par questionnaire.

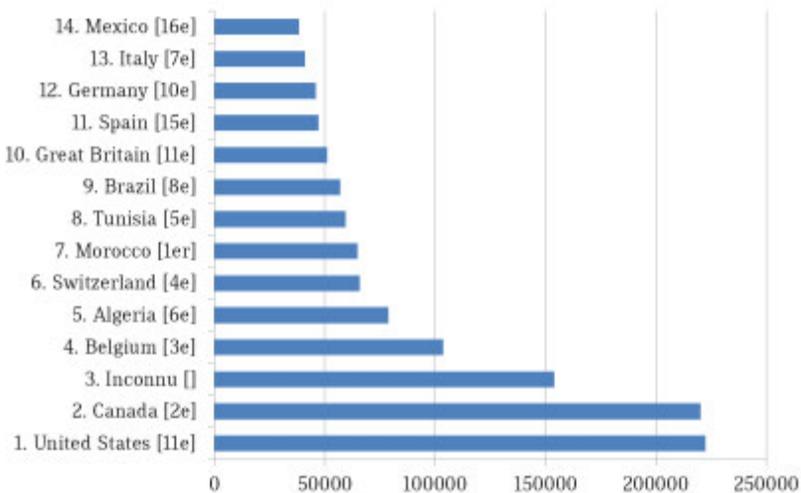


Figure 3 : pays selon les statistiques de fréquentation : comparaison statistiques de fréquentation (logs) (n.) / enquête qualitative (n°)
Source : Cléo. Enquête en ligne 2009.

Les deux approches montrent, cependant, que le lectorat du portail *Revue.org* est fortement international. C'est une des caractéristiques des pratiques scientifiques, que favorise la diffusion sur Internet, et qu'accroît la présence de revues non francophones (*Anthropology of Food*, *China Perspectives*, *Current Psychology Letters*, *Field Actions Science Report*, *Journal of the Short Story in English*, *Surveys and Perspectives Integrating*

Environment and Society...) ou multilingues sur Revues.org (*Confins, Nuevo Mundo, Cybergeo, Trace, Trivium...*).

Un public universitaire avancé

+++++

La seconde caractéristique qui ressort de l'étude des profils réside dans l'écrasante majorité de lecteurs hautement diplômés, plus de 50 % des répondants à l'enquête ayant poursuivi leurs études au-delà de la quatrième année après le baccalauréat, et un tiers des lecteurs étant titulaires d'un doctorat. Il s'agit donc d'un public de spécialistes, principalement d'enseignants, de chercheurs, d'étudiants avancés et de documentalistes ou bibliothécaires. La figure 4 présente les résultats comparés des enquêtes de Cairn et de Revues.org sur l'appartenance socioprofessionnelle des répondants. Le public des deux portails diffère ici, de façon marquante, dans la part d'étudiants de premier et second cycles : conséquente dans l'ensemble du lectorat de Cairn, elle s'avère faible pour Revues.org. Le constat est strictement inversé lorsqu'on observe les populations plus installées dans la vie. Le lectorat de Revues.org est donc plus âgé et plus professionnel. Celui de Cairn plus jeune et plus étudiant.

Ce contraste interroge sur de possibles spécificités des portails, en termes de contenu ou de fonctionnalités, à même d'intéresser des publics distincts. Mais ces données interrogent avant tout sur la connaissance que les usagers ont des ressources en sciences humaines et sociales, sur les pratiques et les modalités d'accès à ces ressources, en termes de recherche, consultation, navigation.

La question de l'intermédiation aux ressources se pose également. On peut s'interroger ici, entre autres, sur le rôle de la prescription des enseignants et sur celui de la mise en visibilité des ressources dans les espaces documentaires accédés par les étudiants de premier et second cycles. Captifs des ressources dispensées dans les bibliothèques universitaires et par les services communs de documentation, ils forment 17 % du lectorat de Cairn : les bibliothèques semblent bien jouer ici leur rôle de médiateur, de mise en relation d'un public cible avec une ressource spécifique, nécessitant abonnement. *A contrario*, le déficit de cette population dans le lectorat de Revues.org (3 %) laisse entendre une méconnaissance de cette ressource : Revues.org peut manquer ici de la recommandation des enseignants et/ou de l'intermédiation bibliothécaire. Dans ce cas, on peut s'interroger sur la place qui est faite au libre accès dans les politiques documentaires des bibliothèques universitaires, dont l'une des missions

consiste justement à acquérir des ressources pour les mettre à disposition de leurs usagers².

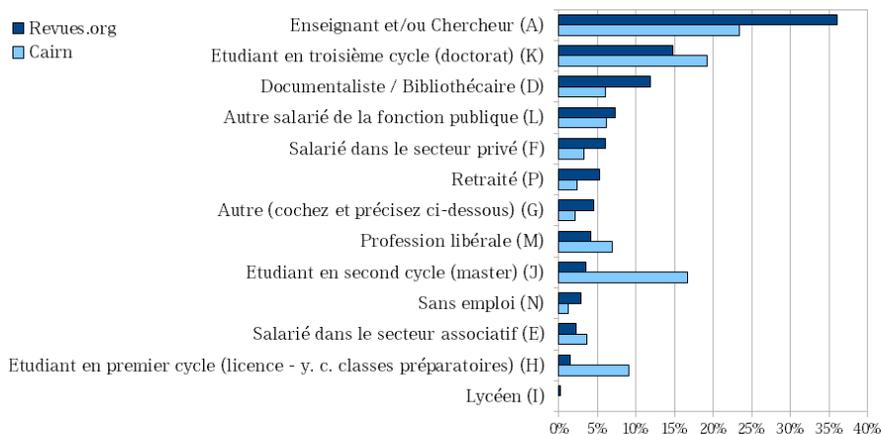


Figure 4 : profession du lectorat de Cairn et de Revues.org

Le second volet de l'enquête apporte des éléments d'information sur les modalités de recherche et de consultation des ressources en sciences humaines et sociales, qui viennent nuancer certains préjugés. L'analyse des logs de serveurs, menée en complément de l'enquête en ligne, permet par ailleurs de formuler une hypothèse intéressante sur le comportement des lecteurs de carnets de recherche sur la plateforme *Hypothèses*, en comparaison des comportements de lecture des revues sur *Revue.org*. On présentera d'abord les données de logs avant de revenir aux apports de l'enquête en ligne.

LES LECTEURS DES CARNETS MOINS NOMBREUX ET PLUS FIDÈLES QUE CEUX DES REVUES ?

+++++

Parmi les indicateurs fournis par les analyses de logs, on retient ici le nombre de visites et le nombre de visiteurs uniques. En divisant le nombre de visites d'un site sur une période d'un mois par le nombre de visiteurs

2. Cette question a fait l'objet d'une enquête de terrain (juin-septembre 2009), réalisée dans le cadre d'un stage pour l'obtention du master 2 de l'Institut national des techniques de la documentation (INTD). Les résultats sont présentés dans le mémoire de fin d'études : Bester Emma. *L'économie de l'attention pour le libre accès. Le cas de Revues.org dans les bibliothèques universitaires*, Mémoire des sciences de l'information, INTD, 2009.

uniques pour la même période, on obtient un taux moyen de consultation par visiteur. Plus le résultat de l'opération s'élève au-dessus de 1, plus le nombre de visites sur un site par un même visiteur est élevé : on peut à ce titre parler de « taux de retour » du visiteur sur le site ou « taux de fidélité ». En appliquant cette mesure aux revues de Revues.org les plus fréquentées (tableau 2), on observe des taux de retour relativement faibles (entre 1,2 pour *Nuevo Mundo* et 1,5 pour *Cultures & Conflits*) ; les temps réels de consultation et les pratiques de lecture en ligne ou hors ligne nous manquent ici pour procéder à une analyse plus fine des comportements d'usage.

Tableau 2 : fidélité du lectorat ?

	Visiteurs uniques	Visites	Score de "fidélité"
Nuevo mundo (revue)	85 000	98 000	1,15
Cybergeog (revue)	40 000	58 000	1,45
Clio (revue)	36 000	43 000	1,19
Cultures & conflits (revue)	26 000	39 000	1,5
Balkanologie (revue)	4 200	5 400	1,29
Economie publique (revue)	4 000	6 500	1,63
Culture et politique arabe (carnet)	10 000	19 000	1,9
L'édition électronique ouverte (carnet)	5 000	17 000	3,4
Evaluation (carnet)	3 000	7 500	2,5
Quanti (carnet)	2 600	5 500	2,12
Polibistro (carnet)	1 000	4 400	4,4

Selon le guide d'utilisation du logiciel AWstats³ : **“Unique visitor:** a host that has made at least 1 hit on 1 page of your Web site during the current period (1 month).” [Un visiteur qui a cliqué au moins une fois sur une page de votre site dans la période actuelle (1 mois).] **Visits:** number of visits made by all visitors. Think ‘sessions’, say a unique IP accesses a page, and then requests three others without an hour between any of the requests.” [Nombre de visites faites par l'ensemble des visiteurs. Chaque « session » renvoie à une seule adresse IP ayant accédé à une page Web et à trois requêtes dans l'heure suivant la connexion.]

3. [En ligne] < http://awstats.sourceforge.net/docs/awstats_glossary.html > (consulté le 5 décembre 2011).

Comparons les taux de fidélité des revues les plus fréquentées du portail *Revues.org* avec ceux des carnets de recherche les plus fréquentés sur *Hypothèses*, la plateforme de carnets de recherche animée par le Cléo. Si les consultations sont quantitativement plus faibles sur les blogs que sur les revues, les taux de retour sont par contre plus élevés, allant de 1,9 pour le carnet *Culture et politique arabes* à 4,4 pour le carnet *Polit'bistro*. Le lectorat des blogs apparaît donc comme plus concentré mais également plus fidèle que celui des revues. Il faut peut-être voir dans cette mesure un indicateur quantitatif plus pertinent de la réussite du modèle de carnet que la mesure qui consiste à comptabiliser le nombre et la longueur des commentaires, et qui est la plus généralement utilisée.

MODALITÉS DE RECHERCHE, DE CONSULTATION ET DE LECTURE : QUELQUES DONNÉES D'USAGE

+++++

Au-delà de Google

+++++

L'enquête en ligne fait apparaître que l'utilisation du moteur de recherche Google, si elle domine, n'écrase pas pour autant les autres stratégies de recherche chez ce public universitaire avancé. Plus de 300 répondants signalent ainsi utiliser les ressources des bibliothèques universitaires, physiques et électroniques, sur site et à distance. Ils déclarent utiliser des sites précis, directement identifiés (figure 5). C'est ainsi qu'un large panel des ressources électroniques disponibles en sciences humaines et sociales est connu et consulté par le lectorat de *Revues.org* (figure 6). Se distinguent particulièrement les portails Cairn et Persée, qui sont signalés par plus de 50 % des répondants, et la ressource anglophone JSTOR, mentionnée par 250 personnes ; une part plus restreinte de 100 à 200 répondants consulte également les archives ouvertes de HAL-SHS et accède au portail BiblioSHS.

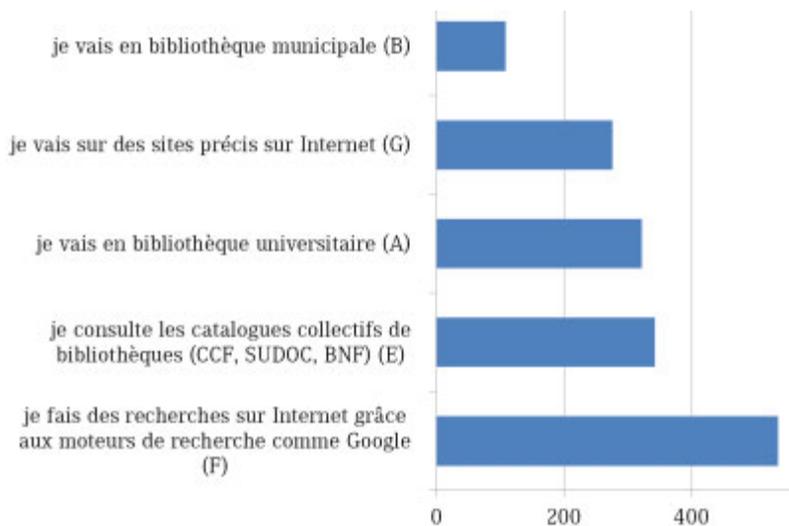


Figure 5 : « Pour trouver des ressources en sciences humaines et sociales... »

Source : Cléo. Enquête en ligne Revues.org 2009.

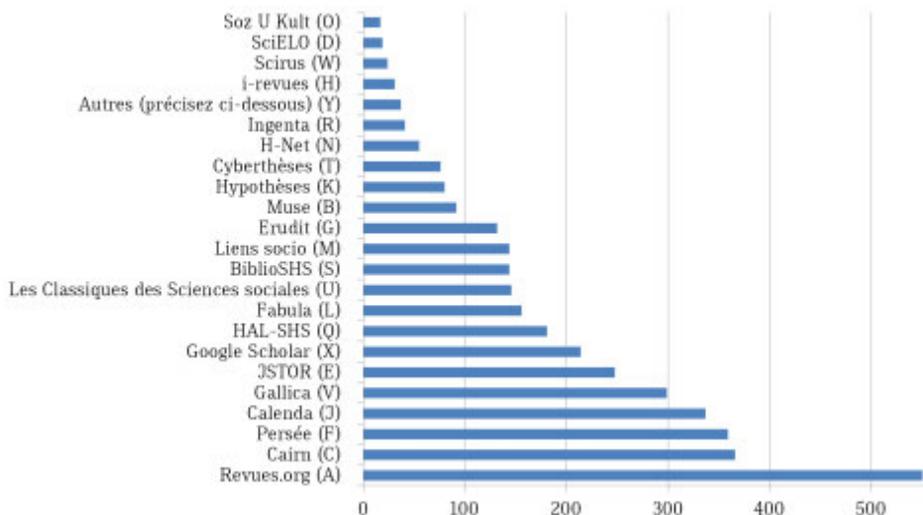


Figure 6 : « Je consulte telle et telle ressource en sciences humaines et sociales ».

Source : Cléo. Enquête en ligne 2009.

De faibles consultations en bibliothèque et de fortes consultations à domicile

+++++

Les consultations à domicile et sur le lieu de travail constituent l'essentiel des modalités géographiques d'accès aux articles de Revues.org (figure 7). Ce résultat semble logique, dans la mesure où le lectorat professionnel constitue une partie importante de la population répondante. Le travail à domicile, fréquent dans le milieu universitaire, est encouragé par le nombre d'articles de Revues.org en libre accès. La consultation en bibliothèque paraît, elle, particulièrement faible et interroge de nouveau sur la situation actuelle des bibliothèques universitaires face aux contenus numériques, particulièrement lorsqu'une partie importante de ces contenus est en libre accès.

On note également l'émergence de pratiques de consultation nomades, essentiellement sur ordinateur portable, tandis qu'un frémissement apparaît du côté des lecteurs sur terminaux téléphoniques, comme les *smartphones*. Le recul régulier, ces dernières années, du marché des ordinateurs fixes, au profit des ordinateurs portables, explique cette situation. La généralisation des accès nomades à Internet – que sont les wifi publics, les wifi partagés (FON, Free, Neuf, etc.) et les accès 3G – constitue une tendance qui paraît appelée à se renforcer, confortant ainsi la lecture nomade.

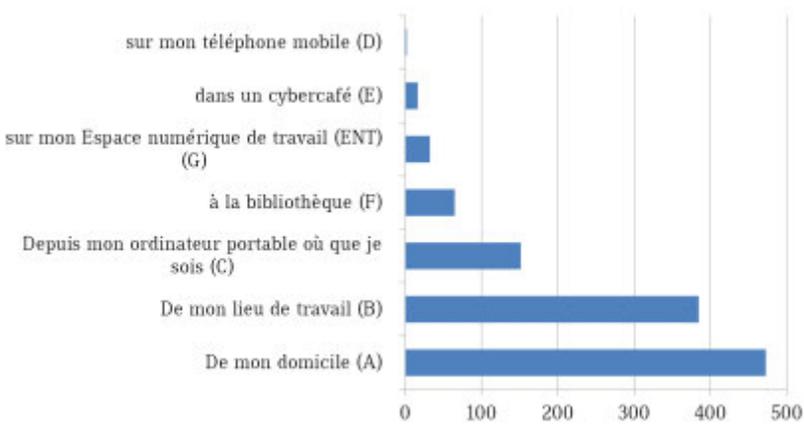


Figure 7 : « Je consulte Revues.org... »
Source : Cléo. Enquête en ligne 2009.

Forte utilisation des favoris, au détriment des flux RSS

+++++

Les modalités de consultation de Revues.org témoignent également d'un certain degré de culture Web des usagers du portail (figures 8 et 9). Moins de 100 personnes passent, en effet, par un moteur de recherche pour accéder au portail Revues.org, alors que près de la moitié des répondants ont inscrit le site Revues.org dans leurs favoris – utilisant ainsi leurs signets pour se rendre sur le portail. Mieux : 170 personnes écrivent directement l'URL du site dans la barre d'adresse.

En revanche, l'usage des flux RSS est faible puisque seuls 60 répondants les utilisent pour suivre l'actualité du portail Revues.org ; c'est le service de lettre électronique qui domine largement à cet effet (450 abonnés sur les 700 répondants). Ils sont à peine davantage (80 répondants) à utiliser les flux RSS proposés sur le calendrier de la recherche en sciences humaines et sociales, *Calenda*, ce qui pourrait néanmoins laisser supposer une plus forte adéquation entre cette offre documentaire et l'offre de service RSS. Malheureusement, l'enquête ne nous dit pas si les flux RSS de chaque revue sont plus utilisés que les flux généraux fournis par le portail. La faible utilisation des flux RSS contraste avec l'usage des signets ou la mémorisation des URL. Elle reste étonnante, dans la mesure où deux fois plus de répondants (120) déclarent utiliser des gestionnaires de références bibliographiques, ou encore des agendas électroniques en ligne (figure 10), dont l'usage ne paraît pas moins complexe que celui de flux RSS.

Dans un monde de la recherche pratiquant une veille bibliographique intensive, la très faible appropriation de la syndication de contenus reste source d'étonnement et de perplexité. La blogosphère grand public semble avoir plus rapidement adopté ces pratiques de veille. La communauté scientifique est-elle particulièrement conservatrice, privilégiant des méthodes éprouvées au détriment de nouvelles approches de veille ? Le problème se situe-t-il plutôt au niveau de l'explication, de la formation et, plus généralement, de l'offre pédagogique concernant ces nouveaux dispositifs ? La propagation rapide des univers *Netvibes* dans les bibliothèques universitaires semble, pourtant, un vecteur simple de diffusion de ces nouvelles pratiques. Notre hypothèse est qu'il y a un hiatus fort entre le monde des bibliothèques et le monde universitaire, l'un et l'autre ayant développé des compétences parallèles. L'étanchéité entre les deux mondes semble, pourtant, contre-productive.

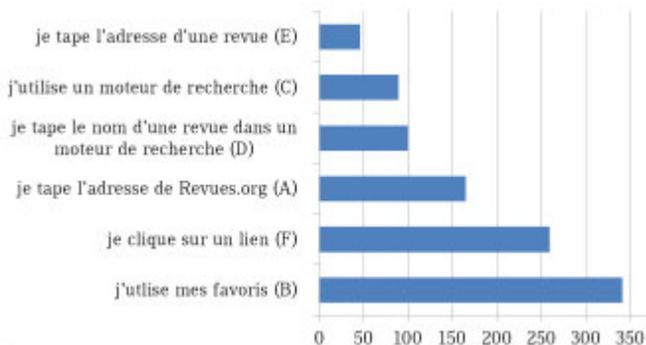


Figure 8 : « Pour aller sur Revues.org... » Source : Cléo. Enquête en ligne 2009.

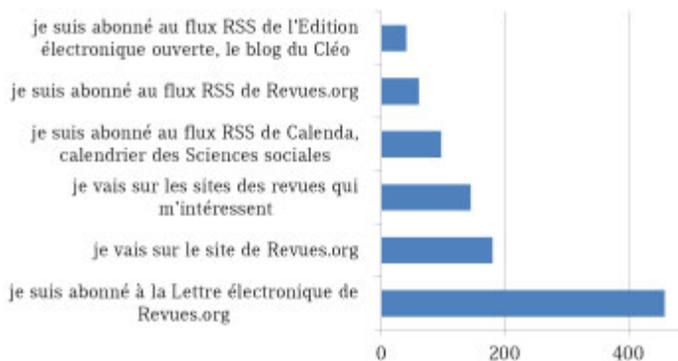


Figure 9 : « Pour suivre l'actualité de Revues.org... »
 Source : Cléo. Enquête en ligne 2009.

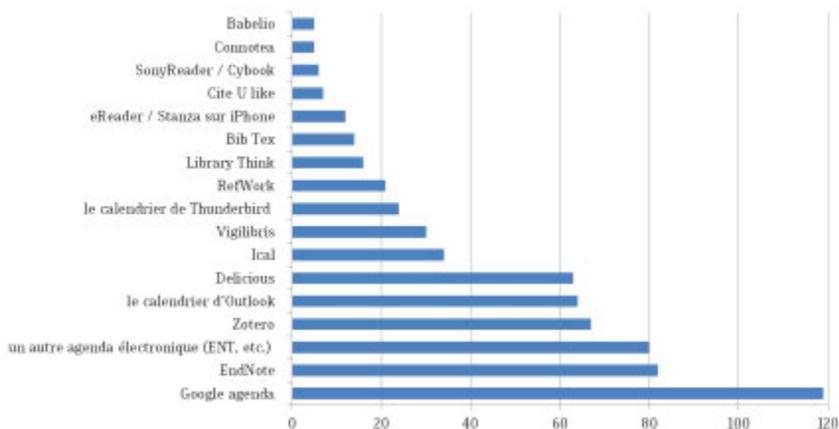


Figure 10 : « Moi-même, j'utilise... » Source : Cléo. Enquête en ligne 2009.

La lecture à l'écran domine

+++++

Un lieu commun très répandu veut que la lecture à l'écran soit inconfortable et inadaptée, particulièrement pour les textes longs. L'enquête montre pourtant le contraire (figure 11). En effet, la lecture des articles sur Revues.org a lieu majoritairement à l'écran, que ce soit directement en ligne (plus de 450 réponses) ou après sauvegarde du fichier sur ordinateur (270 réponses). La lecture après impression devient minoritaire, avec seulement 250 réponses. Cela interroge sur les motivations qui président à ces choix et sur les types de lecture adoptés en fonction des contextes (lecture rapide ou longue à l'écran, annotée, active, taille des articles, temps de lecture, aller-retour vers l'article, possibilités techniques d'impression, etc.), mais cela confirme tout de même une évolution des pratiques de lecture vers une nette domination de la lecture à l'écran.

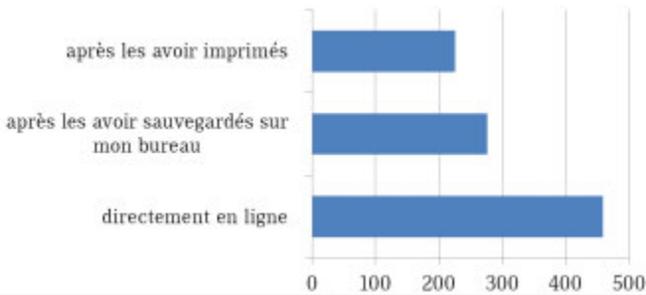


Figure 11 : « Je lis les articles de Revues.org... »
Source : Cléo. Enquête en ligne 2009.

ATTENTES DES LECTEURS DE REVUES ÉLECTRONIQUES EN SCIENCES HUMAINES ET SOCIALES

+++++

Un dernier point à aborder concerne les attentes des lecteurs de Revues.org (figure 12). Les données analysées jusqu'ici semblent montrer que les lecteurs de Revues.org ne sont pas technophobes. Ils restent cependant avant tout intéressés par des améliorations documentaires que l'on peut qualifier de traditionnelles. Les attentes portent principalement sur l'organisation et l'accès aux contenus : près de 300 répondants notent leur intérêt pour des accès thématiques aux revues et une recherche par disciplines⁴.

4. Ces fonctionnalités ont été ajoutées depuis (automne 2009).

De nouveaux produits documentaires sont également souhaités : bibliographies thématiques en ligne, collections de livres en ligne et revues de comptes rendus de livres intéressent de 200 à 300 répondants. Les outils de visualisation de l'information ou d'organisation personnalisée des contenus sont, *a contrario*, peu demandés : moins de 90 répondants ont sélectionné la possibilité de géolocaliser les événements de *Calenda* ou la mise en place d'espaces numériques personnels.

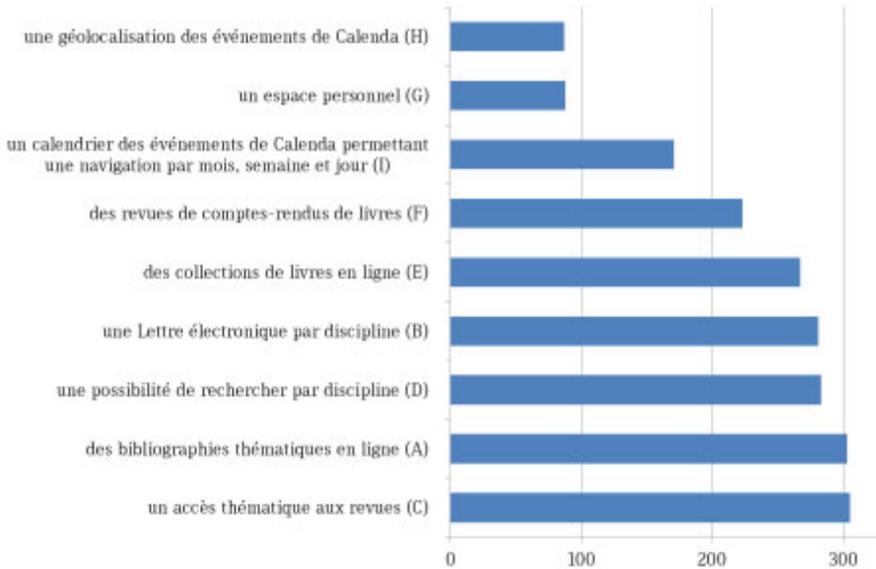


Figure 12 : « Dans les prochains mois, je souhaite voir émerger les innovations suivantes... »

Source : Cléo. Enquête en ligne 2009.

CONCLUSION : À LA RECHERCHE DU LECTEUR INCONNU

+++++

L'enquête en ligne menée par Revues.org auprès de son lectorat permet de dépasser l'approche relativement grossière de l'analyse statistique de *logs* qu'offrent les logiciels classiques de mesure de la fréquentation tels qu'AWstats. Elle permet, en effet, d'étudier qualitativement une population plus restreinte, mais sur laquelle nous disposons d'informations plus précises, particulièrement en ce qui concerne leur sociologie et leurs pratiques de lecture en ligne sur d'autres portails. Nous avons donc découvert

que Google n'était pas l'*alpha* et l'*oméga* de l'accès à Revues.org pour cette population particulière, fortement diplômée, et inscrite dans les professions du savoir. Cette population paraît fidèle à un certain nombre de sites fortement spécialisés, dont elle a un usage régulier. Sans être technophobe, cette population apparaît réticente à l'utilisation des flux RSS et n'utilise quasiment pas les bibliothèques pour accéder à Revues.org.

Cette courte étude reste très incomplète, tant la proportion des lecteurs étudiés paraît faible. Plus encore, manque une étude qualitative poussée des *logs* de Revues.org, afin d'élucider de véritables parcours de recherche et de lecture, sur un temps court ou même sur une durée moyenne. Le Centre pour l'édition électronique ouverte est prêt à accueillir et encadrer tout *data miner* qui acceptera de s'armer d'une pioche pour creuser profondément dans ses *logs*. Ces données, de plus en plus massives, ne donneront toute leur sève et tout leur sens que dans le cadre d'une approche extrêmement fine, ciblant un petit nombre d'utilisateurs repérables, et confrontant leurs pratiques réelles à une enquête orale ciblée. La mise au point d'une grille d'analyse, de *patterns* de lecture et de comportements en ligne, permettra, sans aucun doute, d'éclairer d'un jour nouveau les véritables pratiques de lecture de la littérature savante en ligne.

par Anita Beldiman-Moore

+++++

LES RESSOURCES
NUMÉRIQUES ENTRE
NOTORIÉTÉ ET
USAGES :
ENQUÊTES SUR LES
PRATIQUES ET ATTENTES
DES LECTEURS DE
LA BIBLIOTHÈQUE DE
SCIENCES PO (2005-2008)

+++++

+++++

LES RESSOURCES NUMÉRIQUES ENTRE NOTORIÉTÉ ET USAGES : ENQUÊTES SUR LES PRATIQUES ET ATTENTES DES LECTEURS DE LA BIBLIOTHÈQUE DE SCIENCES PO (2005-2008)

+++++

La bibliothèque de Sciences Po de 2000 à 2008

- de 9 000 à 11 000 lecteurs actifs / an
- de 290 000 à 255 000 prêts annuels (après développement de l'accès direct)
- 1 800 visites / jour en moyenne depuis 2004
- 620 000 volumes
- de 9 000 à 13 000 volumes acquis / an
- de 4 300 à 2 800 périodiques courants

LE DÉVELOPPEMENT DES COLLECTIONS DE RESSOURCES NUMÉRIQUES

+++++

Les ressources numériques ont pris dès 1999 une place croissante dans les collections de la bibliothèque de Sciences Po avec les premiers abonnements à des bases de données et la mise en place en 2001 d'un accès à des revues électroniques par le biais d'Ebsco et de JSTOR.

Cette politique d'acquisition volontariste est illustrée par l'évolution des dépenses en ressources numériques, passées de 6 % des dépenses documentaires totales de la bibliothèque en 2000 (35 000 €) à 21 % en 2008 (185 000 €) (figure 1).

Dans le même temps, le développement de l'accès aux revues en ligne (de 0 à 13 000 revues électroniques sur la période étudiée) a permis une rationalisation des abonnements aux périodiques en version imprimée qui sont passés de 4 300 à 2 800.

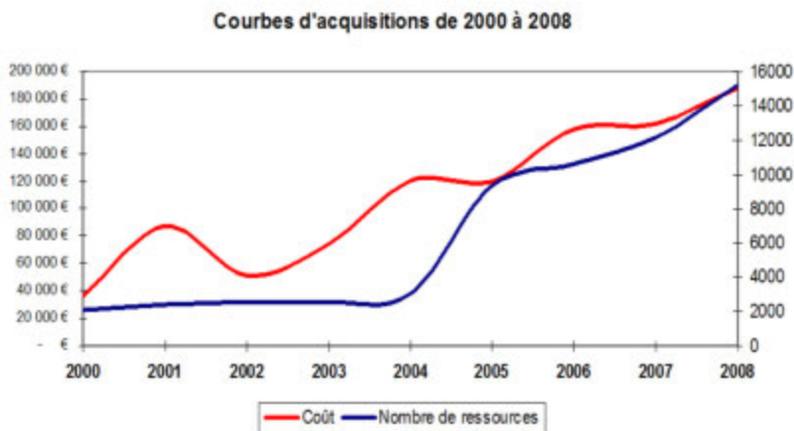


Figure 1 : courbes d'acquisition en coût et en nombre des ressources numériques de 2000 à 2008

L'ÉVALUATION DES USAGES ET DES COLLECTIONS

+++++

Dans ce contexte, et très vite, s'est posée la question de l'évaluation de l'usage de ces ressources mais le décompte des consultations reste l'apanage des fournisseurs. Non seulement les algorithmes de calcul ne sont jamais explicités mais, selon les éditeurs, les ressources, voire les années, les chiffres fournis font état de « sessions », de « recherches » et plus tard, à partir de 2006, de « téléchargements », rendant toute comparaison systématique difficile, voire impossible, ou en tout état de cause incomplète¹.

Le processus de recueil de ces données est également très variable, soit par l'utilisation d'un outil en ligne dans le module d'administration de la base de données concernée, soit par e-mail avec ce que cela implique de relance et de mise à jour des contacts. En aucune façon, la bibliothèque de Sciences Po, à l'instar des autres bibliothèques abonnées, n'a « la main » sur le recueil, le calcul, la périodicité et la granularité de ces statistiques de consultation.

À ce panorama aléatoire, souvent incomplet mais néanmoins essentiel dans l'évaluation des collections, il était donc nécessaire de trouver un complément d'information. L'enquête auprès du public déjà envisagée

1. L'initiative Counter représente néanmoins une avancée du fait de sa plus grande adoption par les éditeurs, comme nous le verrons dans la dernière partie de cet article.

pour la mesure de la satisfaction et des usages des lecteurs de la bibliothèque est apparue en 2004 comme un bon moyen d'obtenir un point de vue différent et complémentaire sur la consultation de ces ressources.

L'enquête 2005

++++
Commandée au cabinet JLM Conseil et développée grâce au logiciel Modalisa², une première enquête a été initiée en 2004. Très complète, elle couvrait tous les usages et pratiques à la bibliothèque et abordait le sujet des ressources électroniques dans deux questions :

- « Utilisez-vous les ressources numériques (bases de données, revues en ligne, cédérom) proposées par la bibliothèque de Sciences Po ? » Avec comme choix de réponse unique : « Oui, au moins une fois par semaine » ; « Oui, au moins une fois par mois » ; « Oui, au moins une fois par an » ; « Non, jamais ».
- « Parmi ces bases de données, cochez toutes celles que vous connaissez et toutes celles que vous utilisez... » Question suivie de la liste de nos 22 bases de données de l'époque, avec la possibilité de répondre « Connue » et/ou « Utilisée ».

Le questionnaire, sous la forme d'une page Web, fut administré en février 2005 *in situ* par des vacataires postés dans chaque salle de la bibliothèque selon un roulement établi sur trois semaines³. Il fut ainsi rassemblé 690 réponses qui correspondaient à 9 % des lecteurs actifs de la bibliothèque recensés en 2004.

Cette première enquête offrit une vision assez surprenante de l'usage des ressources numériques disponibles à la bibliothèque de Sciences Po. En effet, 46 % des lecteurs interrogés déclaraient ne *jamais* utiliser ces ressources (figure 2).

Étaient alors concernés dans ces 46 % les étudiants du Programme international et des « autres » filières (respectivement 56 et 57 % d'entre eux ont déclaré ne jamais utiliser ces ressources), ainsi que les enseignants et les doctorants ou masters dits de recherche (respectivement 34 et 37 % d'entre eux).

2. Logiciel développé par la société Kynos, plus d'informations à l'adresse : < <http://www.modalisa.com> > (consulté le 14 novembre 2011).

3. Un mailing fut également lancé sur la fin de cette période auprès des enseignants-chercheurs de Sciences Po, eu égard à leur peu de représentation dans la première base de réponses recueillies. Ce mailing rassembla 64 réponses, soit 9 % du total final des réponses.

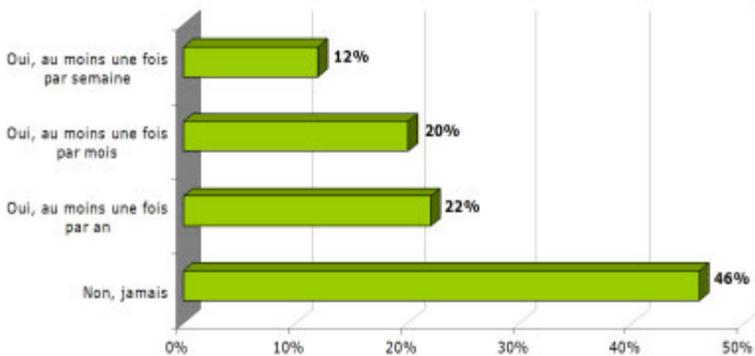


Figure 2 : réponses en 2005 à la question « Utilisez-vous les ressources numériques (bases de données, revues en ligne, cédérom) proposées par la bibliothèque de Sciences Po ? »

Les réponses à la seconde question sur les ressources numériques (figure 3) ont permis d'éclairer quelque peu ce constat plutôt décourageant. En effet, en comparant les ressources déclarées connues et celles utilisées, l'on s'est aperçu que ces ressources étaient très peu connues dans leur ensemble. À l'exception notable de l'*Encyclopædia Universalis*, connue de 45 % des répondants et utilisée par près de 20 % d'entre eux, seules 5 autres ressources étaient connues de plus de 20 % du public interrogé (13 ressources au total dépassaient les 10 % de public *connaisseur*).

L'enquête 2008

+++++

Fort des enseignements précieux de cette première expérience d'enquête, Michel Gardette, alors directeur de la bibliothèque de Sciences Po, a décidé d'inscrire cette démarche dans le temps et de réitérer une enquête similaire à deux ou trois ans d'intervalle.

En 2007, il fut donc procédé à l'acquisition du logiciel Modalisa, et un groupe de travail s'est attelé à la mise à jour et aux modifications du questionnaire pour une administration prévue en janvier 2008.

La gageure, comme pour toute enquête dont les résultats sont destinés à une analyse diachronique aussi bien que synchronique, a consisté en l'introduction dans le questionnaire de modifications conjoncturelles tout en gardant une structure similaire se prêtant à une comparaison dans le temps.

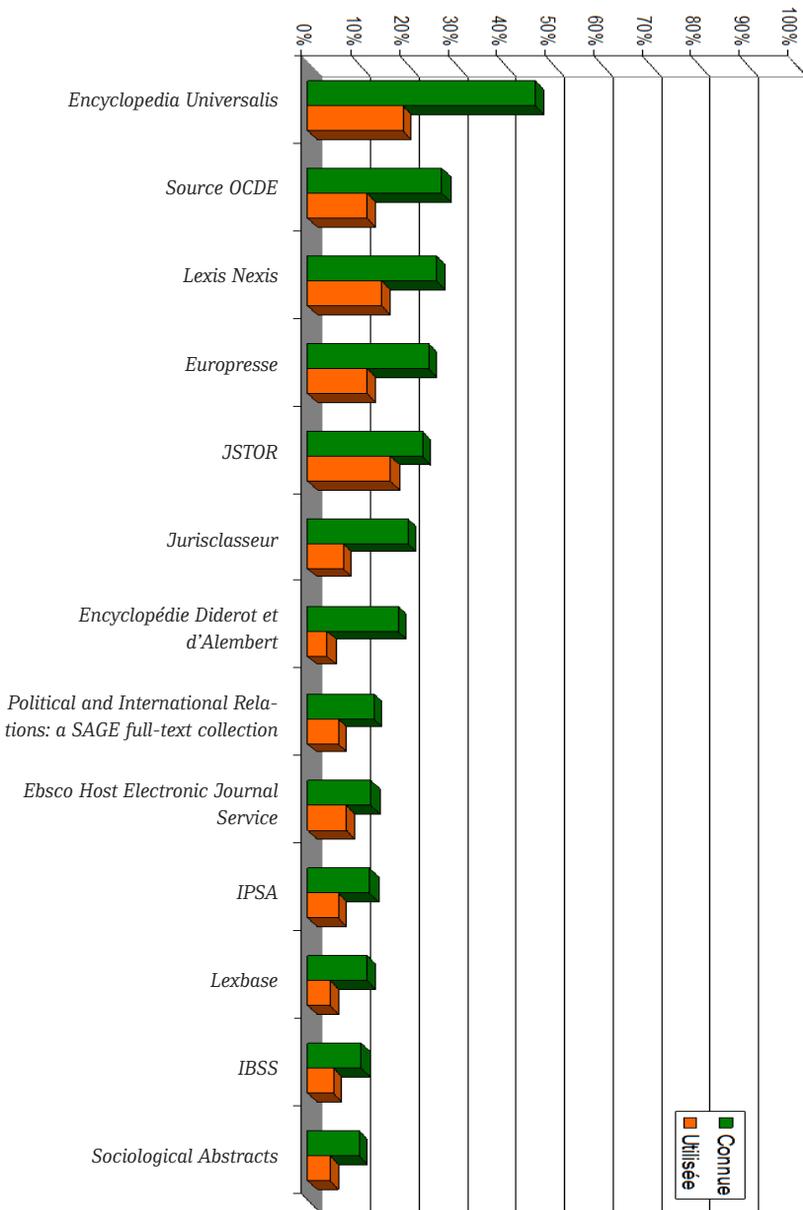


Figure 3 : réponses en 2005 à la question « Parmi ces bases de données, cochez toutes celles que vous connaissez et toutes celles que vous utilisez... » (bases connues par plus de 10 % des répondants)

La première question a donc été maintenue en l'état et la deuxième question simplement mise à jour avec la liste des bases de données, au nombre de 42, disponibles en 2007.

Enfin, une troisième question a été ajoutée sur l'accès distant à ces ressources, le service ayant été mis en place au dernier trimestre 2007 :

- « Vous êtes-vous déjà connecté à ces bases de données via l'accès distant (hors campus de Sciences Po) quand cela était possible ? » Avec comme choix de réponse unique : « Oui » ou « Non ».

Le questionnaire fut adressé le 3 janvier 2008 par e-mail à 14 000 personnes (étudiants, enseignants, chercheurs, salariés – à l'exception de ceux de la bibliothèque et des lecteurs extérieurs). Le 1^{er} février 2008, l'enquête était close et 2 257 réponses avaient été enregistrées, soit 16 % des envois et 21 % des lecteurs actifs recensés en 2007.

Les réponses obtenues montrent de façon flagrante que ces ressources sont bien plus utilisées qu'en 2005 puisque seuls 17 % de nos lecteurs sont encore des non-utilisateurs (figure 4).

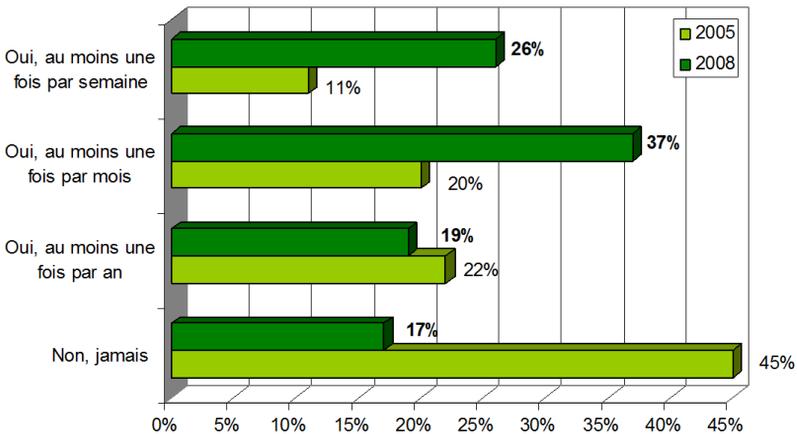


Figure 4 : réponses comparées 2005-2008 à la question « Utilisez-vous les ressources numériques (bases de données, revues en ligne, cédérom) proposées par la bibliothèque de Sciences Po ? »

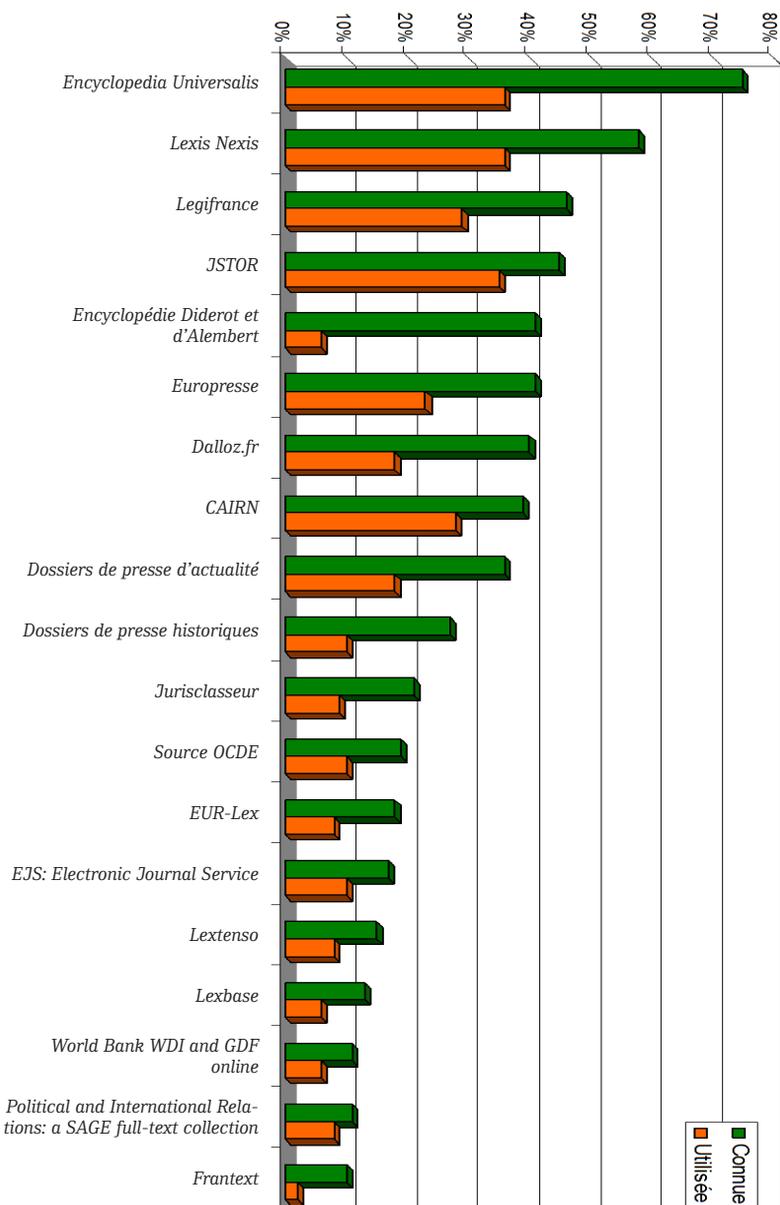


Figure 5 : réponses 2008 à la question « Parmi ces bases de données, cochez toutes celles que vous connaissez et toutes celles que vous utilisez... » (bases connues par plus de 10 % des répondants)

Les enseignants représentent toujours 14 % des non-utilisateurs déclarés là où les doctorants et masters dits de recherche ne sont plus respectivement que 2 et 7%. De même au sein des filières, les étudiants du Programme international et des « autres » filières sont toujours les moins utilisateurs (24 à 27% d'entre eux n'utilisent jamais ces ressources).

Cette hausse de la consultation des ressources numériques se fait aussi bien sur le campus que hors les murs puisque l'accès distant (objet de la troisième question) est adopté par plus de 51% des répondants.

Une fois encore, les réponses à la deuxième question jettent une lumière intéressante sur la progression de ces nouveaux usages ou l'adoption de ces nouveaux outils (figure 5). Les ressources sont plus utilisées parce que plus connues.

Malgré l'accroissement significatif du nombre de ressources, les utilisateurs semblent s'y retrouver puisque 19 bases de données sont connues de plus de 10% du public ayant répondu et certaines dans des proportions nettement plus importantes qu'en 2005 (7 d'entre elles sont connues de 40% et plus des répondants). Leur utilisation est toujours conditionnée par leur notoriété⁴. Pourtant certaines surprises peuvent être relevées. Ainsi, la base de données « Business Source premier » (BSP), pourtant parmi les plus consultées selon les statistiques du fournisseur, n'apparaît ici que dans le peloton de queue des 17 bases de données situées entre 9 et 5% de notoriété (figure 6).

Une telle différence pose les limites de ce type d'enquête déclarative. En effet, BSP est une base qui offre l'accès à des revues en texte intégral, et les utilisateurs peuvent être amenés à consulter ces revues sans savoir précisément qu'il s'agit de cette base de données.

En creusant plus finement la distinction entre notoriété et usages, il apparaît que la comparaison des taux d'utilisation (ou pourcentage d'utilisateurs parmi ceux qui déclarent connaître la ressource) des enquêtes 2005 et 2008 (figure 7) est remarquablement similaire surtout pour les ressources les plus connues, ce qui tendrait à faire penser que leur public demeure structurellement le même.

4. À l'exception de l'*Encyclopédie* de Diderot et d'Alembert très connue mais peu utilisée.

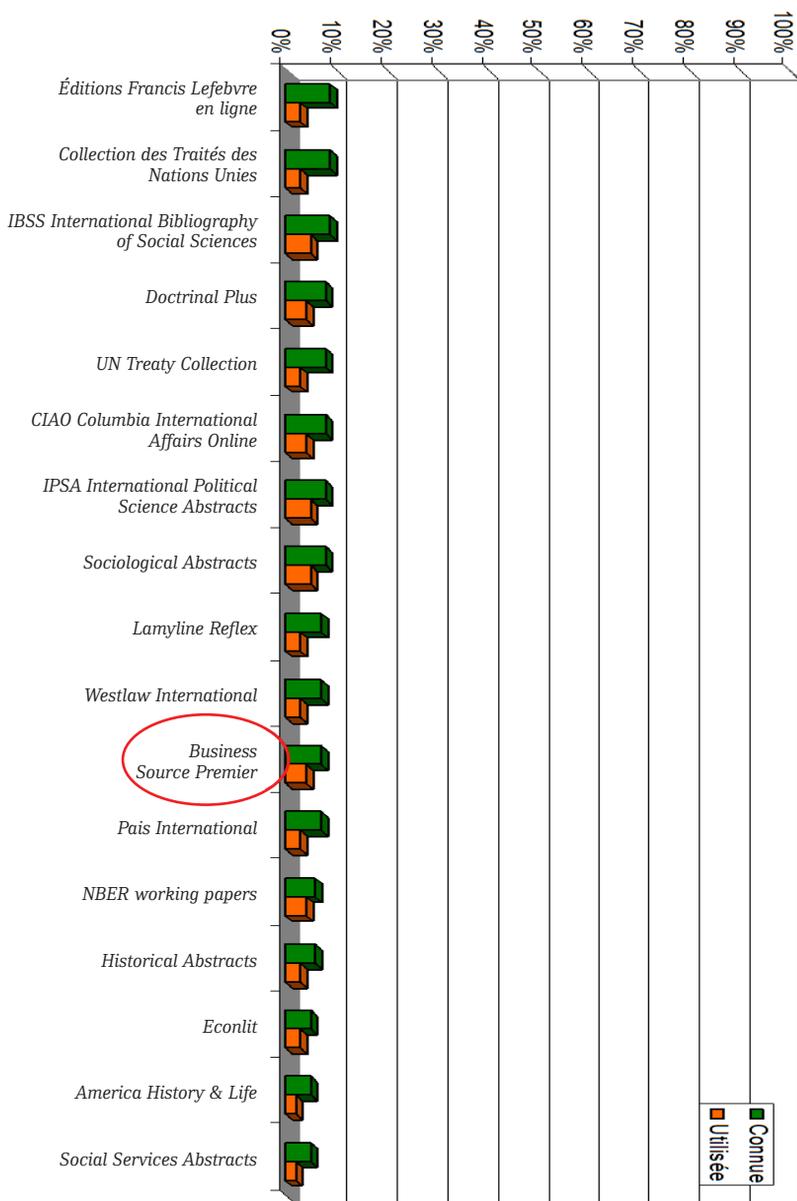


Figure 6 : réponses 2008 à la question « Parmi ces bases de données, cochez toutes celles que vous connaissez et toutes celles que vous utilisez » (bases connues par 9 à 5 % des répondants)

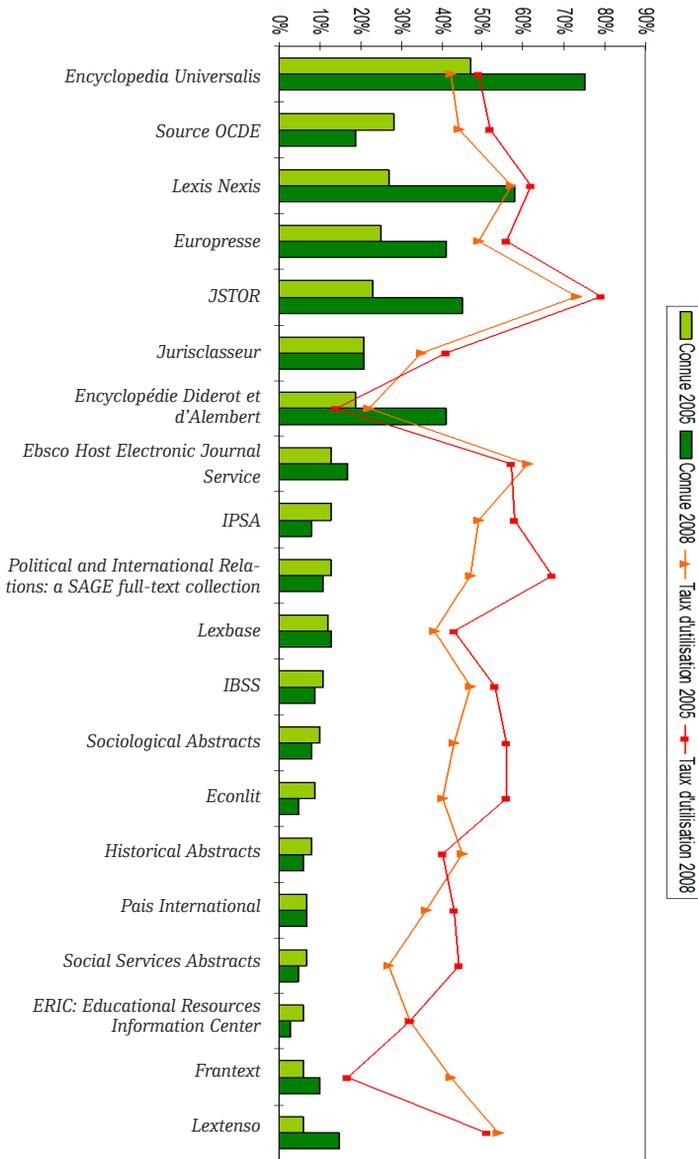


Figure 7 : réponses comparées 2005-2008 à la question « Parmi ces bases de données, cochez toutes celles que vous connaissez et toutes celles que vous utilisez » (en histogramme le pourcentage de notoriété et en courbe le pourcentage d'utilisation quand la ressource est connue)

Les enseignements des deux expériences d'enquête

+++++

Les enquêtes de public ont apporté à la bibliothèque de Sciences Po un angle de vue tout à fait original et riche d'enseignements sur les usages des ressources numériques proposées. Leur utilisation est, comme nous l'avons vu, essentiellement conditionnée par la notoriété de ces ressources et si elle s'avère évoluer en termes de volume de consultation au fil des ans, elle reste éminemment stable dans sa structure tant par ressource que par population cible.

L'offre, l'accès et les services que la bibliothèque met en place autour de ces ressources, comme nous le verrons plus loin, ont aussi leur rôle dans cette évolution positive qui complète les statistiques quantitatives fournies par les éditeurs.

Une telle démarche atteint cependant ses limites sous certains aspects. Ainsi, la deuxième question de l'enquête concernant les ressources numériques⁵ est désormais trop « lourde » pour être administrée. Avec une offre autour de 30 à 40 items listés, l'utilisateur perd patience et tend à ne plus répondre (figure 8).

D'autant plus qu'en 2008-2009, outre les bases de données listées ci-dessus, la bibliothèque a commencé à développer une collection d'e-books.

En 2010, la bibliothèque se propose de réitérer son enquête « Usages » en introduisant un certain nombre d'évolutions. Cette question en particulier devra faire l'objet d'une reformulation adéquate afin de garder tout le bénéfice de la subtilité apportée par la distinction entre notoriété et usages tout en déjouant les pièges d'une liste fastidieuse.

Les autres résultats de ces enquêtes ont également mis en exergue la primauté du rôle prescripteur des enseignants dans l'usage de la bibliothèque. Il semble judicieux dès lors de centrer tout ou partie d'une future enquête sur cette population jusque-là assez évasive⁶.

5. « Parmi ces bases de données, cochez toutes celles que vous connaissez et toutes celles que vous utilisez. »

6. 64 enseignants-chercheurs ont répondu en 2005, 120 en 2008 sur les près de 2 000 – vacataires et titulaires confondus – que compte Sciences Po.

Parmi ces bases de données, cochez toutes celles que vous utilisez ou connaissez		
	Connue	Utilisée
Encyclopaedia Universalis	<input type="checkbox"/>	<input type="checkbox"/>
Encyclopédie Diderot et d'Alembert	<input type="checkbox"/>	<input type="checkbox"/>
Frantext	<input type="checkbox"/>	<input type="checkbox"/>
Europresse	<input type="checkbox"/>	<input type="checkbox"/>
Lexis Nexis	<input type="checkbox"/>	<input type="checkbox"/>
Dossiers de presse historiques	<input type="checkbox"/>	<input type="checkbox"/>
Dossiers de presse d'actualité	<input type="checkbox"/>	<input type="checkbox"/>
CAIRN	<input type="checkbox"/>	<input type="checkbox"/>
EJS : Electronic Journal Service	<input type="checkbox"/>	<input type="checkbox"/>
JSTOR	<input type="checkbox"/>	<input type="checkbox"/>
Dalloz.fr	<input type="checkbox"/>	<input type="checkbox"/>
Doctrinal Plus	<input type="checkbox"/>	<input type="checkbox"/>
Éditions Francis Lefebvre en ligne	<input type="checkbox"/>	<input type="checkbox"/>
EUR-Lex	<input type="checkbox"/>	<input type="checkbox"/>
Juriscasseur	<input type="checkbox"/>	<input type="checkbox"/>
Lamyline Reflex	<input type="checkbox"/>	<input type="checkbox"/>
Legifrance	<input type="checkbox"/>	<input type="checkbox"/>
Lexbase	<input type="checkbox"/>	<input type="checkbox"/>
Lextenao	<input type="checkbox"/>	<input type="checkbox"/>
Collection des Traités des Nations Unies	<input type="checkbox"/>	<input type="checkbox"/>
UN Treaty Collection	<input type="checkbox"/>	<input type="checkbox"/>
Westlaw International	<input type="checkbox"/>	<input type="checkbox"/>
Business Source Premier	<input type="checkbox"/>	<input type="checkbox"/>
Chelem	<input type="checkbox"/>	<input type="checkbox"/>
Econlit	<input type="checkbox"/>	<input type="checkbox"/>
NBER working papers	<input type="checkbox"/>	<input type="checkbox"/>
Regional Business News	<input type="checkbox"/>	<input type="checkbox"/>
SourceOCDE	<input type="checkbox"/>	<input type="checkbox"/>

Terminé

Figure 8 : copie d'écran du questionnaire Web de 2008 pour la question « Parmi ces bases de données, cochez toutes celles que vous connaissez et toutes celles que vous utilisez »

UNE ÉVOLUTION DES USAGES ET SERVICES

+++++

Cette poussée encourageante de l'utilisation des ressources numériques de la bibliothèque tient sans doute pour beaucoup à l'évolution de la place même du numérique durant ce laps de trois ans. Nous avons vu, dans une première partie, l'évolution en nombre de ressources, elle s'accompagne également d'une évolution en termes d'usages. Entre les deux moments des enquêtes, à savoir 2004 et 2008, la bibliothèque de Sciences Po enregistrait un nombre relativement stable de prêts de documents (entre 250 000 et 255 000 prêts à l'année) pour une hausse de près de 600 %

du nombre de recherches dans les bases de données⁷. De la même façon, entre 2006 et 2008, le nombre de documents téléchargés a augmenté de 73 %.

La bibliothèque elle-même a accompagné cette évolution en redessinant complètement, en septembre 2007, l'architecture et le design du menu des bases de données sur son site Web : < <http://bibliotheque.sciences-po.fr/fr/rechercher/bdd/index.html> >, leur donnant par là même une plus grande visibilité. Visibilité accrue encore par la mise en place de l'accès distant en octobre 2007, comme nous l'évoquions plus haut. Cet accès distant concernait, au moment de l'enquête 2008, 70 % des ressources mises à disposition par la bibliothèque.

Parallèlement, la formation assurée par la bibliothèque et comprise dans le cursus des étudiants de 1^{re} année a intégré dès 2007 la recherche documentaire en ligne avec une approche méthodologique et des exercices pratiques sur une demi-douzaine de bases de données.

DES PISTES POUR UNE ÉVALUATION À VENIR

+++++

L'évaluation des ressources numériques représente désormais un pan incontournable de l'évaluation des collections dans toute bibliothèque et entre dans le calcul de nombreux indicateurs de performance. À la réponse statistique des éditeurs, la bibliothèque de Sciences Po, avec ses deux enquêtes, a pu apporter un éclairage original sur la base d'un constat déclaratif concernant la notoriété et l'usage de ces ressources.

De nouvelles pistes s'ouvrent dans le domaine des données quantitatives d'usage avec l'adoption de la « norme » Counter⁸ et la généralisation des statistiques sur les téléchargements qui permettent désormais des comparaisons plus pointues au cœur même de la consultation des ressources numériques.

Parallèlement, les données d'usage collectées par les consortiums de type Couperin permettent aux bibliothèques adhérentes de pouvoir replacer le volume de leurs propres consultations dans un contexte national et de trouver là une nouvelle source de pondération des statistiques émanant des fournisseurs.

7. 72 000 recherches en 2004 sur les 22 bases de données disponibles alors, contre 430 000 en 2008 pour 42 bases de données.

8. En 2009, 43 % des fournisseurs qui livrent des statistiques se conforment à ces préconisations. En 2008, cela se montait plutôt à 25 %.

L'autre piste de développement de l'évaluation de ces ressources réside dans l'analyse de *logs*. Cette solution, mise en œuvre notamment par l'Institut de l'information scientifique et technique (Inist) sur ses portails de bases de données et développée ailleurs dans le cadre de laboratoires de recherche, demeure cependant coûteuse en ressources informatiques et humaines. Inspirées par ces expériences, les équipes de la bibliothèque de Sciences Po réfléchissent à cette piste au moins en ce qui concerne les ressources en accès distant. Pour les autres ressources numériques, hors même le coût informatique et humain évoqué plus haut, la solution purement technique reste à inventer.

Enfin, sur un plan plus qualitatif et dans un contexte d'approche directe de notre public, nous envisageons la constitution de groupes d'utilisateurs pour des entretiens collectifs et des débats sur l'offre de la bibliothèque, son développement et les pratiques de ses utilisateurs.

En définitive, la difficulté même de quantifier l'usage de ces ressources « nouvelles » dont la place ne cesse de croître dans les pratiques comme dans les collections nous a obligés à inventer de nouvelles stratégies d'évaluation, de nouvelles solutions techniques et méthodologiques qui offrent un éclairage croisé et somme toute plus complet et plus fin de la consultation des ressources numériques.

par Caroline Bruley et Delphine Merrien

+++++

UN PANORAMA DES
PRATIQUES ET USAGES
DES E-BOOKS AU SEIN
DES ÉTABLISSEMENTS
COUPERIN

+++++

UN PANORAMA DES PRATIQUES ET USAGES DES E-BOOKS AU SEIN DES ÉTABLISSEMENTS COUPERIN

Le travail de la « Cellule e-Book » (CeB) de Couperin se structure autour de deux grands axes : la coordination des négociations e-books au sein du consortium et la veille sur les thématiques et problématiques liées aux e-books. Actuellement, les deux angles d'étude du groupe sont les pratiques et usages des e-books en bibliothèque universitaire (BU) et le cartable électronique en université.

Côté usages et pratiques, nous en sommes au début de nos investigations, c'est donc un premier bilan qui est présenté ici, en grande partie illustré par l'enquête menée en mars 2009 par Sophie Brezel sur les e-books dans les bibliothèques d'enseignement supérieur (77 réponses reçues). Nous nous sommes également appuyés sur un dossier de master SIB, intitulé « Quelle place pour l'offre de livres électroniques en bibliothèques ? »¹. Enfin, les articles sur les livres électroniques en bibliothèque universitaire rédigés par Laurent Jonchère ont été une source d'information importante.

Commençons par le problème qui se pose d'emblée lorsqu'on parle d'e-books : la définition. « Le terme d'e-book recouvre tout ouvrage publié dans un format électronique (PDF, HTML, format propriétaire...). Techniquement, il s'agit d'un fichier ou d'un ensemble de fichiers. L'e-book est dématérialisé par opposition au livre sur support papier et exige une machine active pour le rendre lisible sous forme d'une succession de mots alignés sur une "page" ou audible par synthèse vocale. »² Par e-book, nous entendrons ici deux réalités :

- *Le livre dématérialisé, son contenu* : à ce niveau, la question de la frontière entre un livre électronique et d'autres

1. Billaud Marjolaine, Dolhem Perrine et Martinez Chloé. « Quelle place pour l'offre de livres électroniques en bibliothèque ? », Dossier de master Sciences de l'information et des bibliothèques (SIB), 1^{re} année, ensib, janvier 2009.

2. Éduscol, « Livre électronique, livre numérique (e-book) ». [En ligne] < <http://eduscol.education.fr/dossier/livrelec> > (consulté le 10 novembre 2011).

ressources électroniques (bases de données, encyclopédies, revues) se pose de façon aiguë...

- *La machine appelée communément « reader »*, livrel ou tablette de lecture, qui est l'une des façons de consulter ce contenu.

Qu'en est-il du positionnement des e-books ? Depuis quelques années, le marché du livre électronique tend à se déplacer vers l'Europe et à se concentrer sur les bibliothèques et particulièrement les bibliothèques universitaires. Ces établissements sont pour la plupart en phase d'expérimentation, étape indispensable pour mesurer la satisfaction des utilisateurs, tant pour les éditeurs et les diffuseurs que pour les professionnels de la documentation. Même quand la bibliothèque n'en possède pas encore, les e-books sont un objet de réflexion, d'études ou de tests. Ces expérimentations sont encore trop peu nombreuses et hétérogènes pour que l'on puisse entamer une étude globale, et les bibliothèques n'ont pas suffisamment de recul pour disposer de statistiques parlantes et évaluer les usages qui en sont faits. Néanmoins, des tendances commencent à se faire jour. C'est ce que nous allons mettre en évidence au travers de l'étude des pratiques et des usages des e-books en BU, du côté des bibliothécaires puis du côté du public. Nous terminerons en tentant une visée prospective des usages et pratiques en fonction des actions en cours ou à mener pour améliorer les supports, les contenus et l'accès à l'information.

PRATIQUES ET USAGES DES E-BOOKS EN BU : DU CÔTÉ DES BIBLIOTHÉCAIRES

+++++

Les choix documentaires effectués par les bibliothèques vont avoir un impact sur l'offre proposée et le public visé : de cette façon, elles ciblent et orientent les usages. Il nous semble donc important de commencer par là notre étude.

On note un nombre important de bibliothèques numériques au sein des BU ; à l'inverse, la consultation de tablettes de lecture se fait davantage en bibliothèque publique. Cela s'explique par le fait que les bibliothèques numériques sont bien adaptées aux besoins des étudiants et des chercheurs en permettant la convergence de nombreux types de documents et de supports en un seul point, en proposant un choix documentaire spécialisé

et en permettant d'identifier l'accès à la lecture grâce aux adresses IP³. Les expériences de prêts de tablettes de lecture sont actuellement tentées en BU, mais de façon plus marginale.

Les bibliothèques numériques

+++++

L'offre en livres électroniques

Quels usages font aujourd'hui les BU des ressources proposées par les éditeurs et agrégateurs ?

Un usage massif : la majorité des établissements (67 sur 77⁴) a des e-books et la plupart de ceux qui n'en ont pas encore y réfléchissent. L'offre tend à se développer davantage dans les bibliothèques d'enseignement supérieur et de façon plus marquée dans les disciplines scientifiques et juridiques ainsi que dans les établissements pluridisciplinaires. Depuis 2007, l'évolution est flagrante : l'ERE⁵ 2007 indiquait que seulement 48 établissements sur 161 proposaient des e-books (hors encyclopédies type *Encyclopædia Universalis* et corpus d'ouvrages anciens comme Frantext, Early English Books Online [EEBO]...) ; parmi ces établissements, 29 avaient une seule ressource, 10 en avaient 2, 5 en proposaient 3, et 4 en avaient plus de 3. Ces établissements ont dépensé 429 497 € pour les e-books sur un total de 25 001 420 € dépensés pour la documentation électronique en France, tous établissements confondus.

Des pratiques d'appropriation diversifiées : deux formules d'acquisition – package ou titre à titre – se partagent le marché. Les modalités d'accès dépendent de l'offre : dans l'idéal, accès illimité par reconnaissance d'adresses IP et/ou de proxy ; mais encore souvent, et surtout pour les plateformes françaises, on trouve un accès authentifié (connexion au Lightweight Directory Access Protocol [LDAP]) avec système de prêt à durée limitée. Les BU font plus appel à des éditeurs qu'à des agrégateurs

-
3. Billaud Marjolaine, Dolhem Perrine et Martinez Chloé. « Quelle place pour l'offre de livres électroniques en bibliothèque ? », *op. cit.*
 4. Les données chiffrées ainsi indiquées sont celles présentées par Sophie Brezel dans son rapport d'analyse de l'enquête sur les e-books dans les bibliothèques d'enseignement supérieur (mars-avril 2009).
 5. Enquête sur les ressources électroniques acquises par les services communs de la documentation (SCD), les services interétablissements de coopération documentaire (SICD) et les bibliothèques d'établissements d'enseignement supérieur. Cette enquête est menée annuellement par la Mission de l'information scientifique et technique et du réseau documentaire du ministère de l'Enseignement supérieur et de la Recherche. Elle a pour objectif d'établir un recensement aussi exhaustif que possible des actions menées par les établissements en matière d'acquisition de ressources électroniques payantes spécialisées.

alors que les autres établissements (écoles, instituts) sont dans un rapport inverse. Les éditeurs les mieux représentés sont : Elsevier, Springer, Wiley et l'Encyclopædia Universalis. Les agrégateurs les mieux représentés sont Numilog de façon massive, puis Cyberlibris et OCLC NetLibrary. En 2007, toujours selon l'ERE, c'est Numilog qui tenait le haut du pavé (20 établissements abonnés), puis venaient Elsevier (15), NetLibrary (10) et Springer (10). Cela dénote une offre très orientée vers les premiers cycles. Et de fait, avec les e-books, les BU voient l'occasion d'équilibrer un peu la balance des ressources électroniques, jusqu'ici quasi exclusivement réservées aux enseignants-chercheurs. Ce sont des ressources pour les étudiants des premiers cycles et en français qui sont privilégiées par les établissements, d'autant que certains projets nationaux ou régionaux, comme le Plan d'aide à la réussite en licence, permettent de trouver des fonds pour financer ces ressources. Ce lectorat est renforcé par l'utilisation de plus en plus importante des e-books dans les technologies de l'information et de la communication pour l'enseignement (TICE) et l'enseignement à distance. Le rapport du Joint Information Systems Committee (JISC) parle d'« un potentiel considérable dans l'enseignement supérieur et le secteur de la formation continue pour le développement des collections de livres électroniques ». L'idée d'un « cartable électronique » fait son chemin.

Plusieurs types de bibliothèques électroniques peuvent cohabiter en fonction du public visé, et au sein de chacune, on peut trouver des e-books gratuits, acquis par abonnement à durée limitée, acquis définitivement ou encore numérisés par l'établissement. C'est le cas à Lille 1 où, depuis 2000, une bibliothèque numérique gratuite est développée sur une plateforme logicielle open source DSpace. En février 2006, une bibliothèque électronique payante (Numilog) venait s'ajouter à cette offre et, désormais, l'établissement propose Orion, un service en ligne qui permet une interrogation unique et simultanée des ressources électroniques du service commun de la documentation avec accès au texte intégral.

L'intégration des e-books au sein des collections : quelles pratiques ?

Dans la majorité des cas, les acquisitions de livres électroniques se font en complément des livres papier. Il est à noter qu'en lettres et sciences humaines ainsi qu'en sciences de la vie, on procède significativement plus par substitution que dans les autres disciplines. L'apparition des e-books dans l'offre des bibliothèques n'a pas remis en cause ni modifié les politiques d'acquisition et documentaire des établissements. L'acquisition des livres électroniques se fait en parallèle de l'acquisition des autres

supports : cela reste un produit émergent qui n'a pas encore trouvé sa place au même titre que le papier dans les politiques documentaire et d'acquisition des établissements. Ce sont d'ailleurs souvent deux services différents qui gèrent monographies papier et livres électroniques, comme à la Bibliothèque interuniversitaire de médecine (fusionnée en janvier 2011 avec la Bibliothèque interuniversitaire de pharmacie pour former la Bibliothèque interuniversitaire de santé), où les livres électroniques sont gérés par le service des périodiques (en charge des périodiques électroniques et des bases de données) et non par le service des monographies. Il y a d'un côté le papier et de l'autre l'électronique. Cette place des e-books peut être d'autant plus difficile à délimiter que, dans le cas d'acquisitions ou d'abonnements par package, le nombre de titres est disproportionné en volume par rapport aux autres collections (électroniques ou papier). C'est ce que l'on explique à Paris 3 : « nous avons acheté les bases ECCO [Eighteenth Century Collections Online] et EEBO qui contiennent à elles seules près de 300 000 titres. Cela interdit de répondre de façon sensée à la question de la substitution / complémentarité [par rapport au papier] »⁶. Une exception est à noter à Lille 1, où la politique d'acquisition des e-books tient compte de différents critères : disciplines enseignées à Lille 1 au niveau licence, nombre de prêts dans les disciplines, achat de nouveaux exemplaires.

Même constat pour les e-books gratuits : pas de réelle politique d'intégration et une faible représentation au sein des collections (moins de 10 % dans 33 établissements sur 45). Certaines bibliothèques ont fait le choix de ne pas proposer d'offre gratuite mais de se limiter aux ressources acquises par l'établissement. Quand elle existe, l'offre gratuite est la plus large possible pour une légère majorité des établissements (14) et complémentaire de l'offre payante pour les autres. Au sein de cette offre gratuite, on retrouve un certain nombre de bibliothèques numériques publiques telles que Gallica, Medic@ ou encore la bibliothèque électronique de Lisesieux.

Côté signalisation des e-books au sein des collections, les documents sont plutôt bien signalés et le plus souvent *via* différents outils :

- Les catalogues (OPAC, Sudoc, catalogue maison) : des améliorations sont à mener côté Sudoc, les procédures de localisation automatique des e-books n'ayant pas encore été

6. Témoignage recueilli par Sophie Brezel pour son rapport d'analyse de l'enquête sur les e-books dans les bibliothèques d'enseignement supérieur.

établies par l'ABES. Mais le bénéfice est immédiat : à Montpellier 3, depuis leur signalement dans l'OPAC, une forte hausse des consultations des e-books a été constatée.

- Les plateformes d'accès propres à l'institution (portail documentaire, espace numérique de travail [ENT] de l'université).
- Les outils électroniques de communication (listes sur le site Web, liste A to Z, blog).
- Les interfaces d'éditeurs.

Pour promouvoir ces ressources, les trois quarts des établissements font des efforts de communication. Les modes de communication employés sont majoritairement électroniques, le premier support cité est Internet, les formations à l'utilisation des ressources documentaires proposées par la bibliothèque arrivant en deuxième position, suivies des courriers internes. Des initiatives de communication originales sont tentées sur des blogs ou *via* des marque-pages insérés dans les ouvrages prêtés. Certains établissements décident cependant de ne pas communiquer autour des livres électroniques, estimant qu'ils sont encore trop peu présents dans les collections et qu'aucune politique documentaire n'en formalise l'usage ; or, il est difficile de communiquer autour d'un concept qui existe mais ne fait pas encore l'objet d'une politique bien définie.

Les tablettes de lecture

+++++

Il a été signalé en introduction de cette partie que, comparée aux bibliothèques numériques, la présence des tablettes de lecture au sein des établissements est plutôt marginale et, qu'à ce niveau, on était surtout en *phase expérimentale* et de test (8 sur 77). Néanmoins, les BU sont nombreuses à penser à l'acquisition de tablettes : la BU droit lettres de Pau réfléchit actuellement à l'achat de tablettes numériques avec bibliothèque intégrée (tous les titres utiles aux L1 de droit, aux L2 de maths, etc.). Lyon 1 envisage, de son côté, le prêt de tablettes pour les prochaines années mais attend une évolution des technologies et des contenus associés.

Pour les établissements qui se sont lancés, le matériel proposé est limité à *trois modèles* : Reader de Sony, Cybook de Booken et iLiad d'iRex. Les résultats liés à cette expérience de prêt de tablettes montrent que le matériel n'est pas encore suffisamment performant et que la tablette n'est pas encore entrée dans les usages des étudiants, si bien qu'il y a, selon la

BU d'Angers, une « nécessité absolue d'assurer une médiation auprès des usagers ».

À ce jour, seule la BU pluridisciplinaire d'Angers a dressé un *premier bilan*⁷ pour le prêt de ses tablettes de lecture. Sa démarche est originale par rapport à ses consœurs : elle n'a pas d'offre e-books, estimant que le contenu n'est pas assez satisfaisant et/ou trop lié aux plateformes des éditeurs. Par contre, elle propose des tablettes de lecture à ses lecteurs depuis bientôt un an : 10 Cybook (Booken Gen 3) ont été acquis par la BU et mis à la disposition du public en septembre 2008 ; et un partenariat a été monté avec Sony France pour la mise en circulation en mai 2009 de 10 PRS-505 (Sony). Le lectorat des 3 BU est visé. Ce livrel est prêté comme un livre (équipé d'un code-barres, réservable, prêté pour une durée de 21 jours), avec un pré-chargement de Publie.net (90 textes de littérature contemporaine jamais publiés sur support papier). L'emprunteur charge ce qu'il veut sur le livrel (formats DOC, PDF, HTML, PRC). Les tablettes sont remises à zéro après restitution. Pour promouvoir le projet, la BU a mis en place des séances de présentation dans le hall et fait de la publicité sur le blog de la BU (bua'bloc), *via* mailing, et en déposant des affichettes en banque de prêt. Le constat est qu'il faut communiquer encore plus et surtout accompagner l'utilisateur dans l'utilisation de l'outil tout en essayant de convaincre les prescripteurs (enseignants-chercheurs mais aussi bibliothécaires). Par ailleurs, cette expérience soulève un certain nombre de questions : quelle place pour le bibliothécaire dans le circuit de diffusion du document numérique ? Quel rôle pour les bibliothèques dans la constitution des contenus ? Quelle position de l'institution bibliothèque dans les nouveaux modèles économiques ? Autant de questions qui se posent aussi pour les e-books « contenus » et auxquelles nous tenterons de répondre plus loin. Avant cela, il convient de compléter ce tableau des pratiques et usages en s'attachant aux usagers.

7. Ce paragraphe reprend la présentation qu'Olivier Tacheau a effectuée pour la Journée sur le livre électronique à la bibliothèque de l'université libre de Bruxelles, le 24 avril 2009. [En ligne] < <http://www.slideshare.net/tacheau/les-livrels-la-bu-dangers> > (consulté le 29 juin 2009).

PRATIQUES ET USAGES DES E-BOOKS EN BU : DU CÔTÉ DES USAGERS

++++
Ce premier bilan sera forcément limité : les établissements pointent le manque de recul dont ils disposent pour mesurer les usages des livres électroniques. Cela est dû à plusieurs facteurs :

- Très peu d'enquêtes d'usage et de satisfaction (6 sur 51) ont été menées.
- On note des problèmes liés aux dispositifs techniques d'accès aux e-books, ne permettant pas d'avoir des statistiques d'usage significatives. « Nos statistiques fournies par Numilog ne sont pas représentatives de l'usage : du fait du proxy, nous n'avons qu'un seul usager identifié, et elles ne donnent que les prêts en cours. » (un bibliothécaire du SCD d'Orléans).
- Le signalement insuffisant de ces ressources dans les catalogues locaux conduit à une faible consultation des e-books.
- Enfin, se pose un problème de temporalité : « Notre offre d'e-books s'est élargie en 2009, par conséquent, les statistiques ne sont pas encore exploitables. » (Institut national des sciences appliquées, Lyon). Et quand ils disposent de chiffres, les établissements manquent d'éléments de comparaison. Il faudrait avoir des chiffres sur plusieurs années. L'offre, qui est plutôt riche et diversifiée aujourd'hui, ne l'était pas il y a encore deux ans.

Pratiques et usages liés aux livres électroniques

++++
Qui utilise les e-books ? Du côté des disciplines, l'usage des e-books dans les sciences exactes est particulièrement important. Plus de 30 % des participants à l'enquête de Sophie Brezel ont déclaré que l'utilisation des e-books dans les sciences (chimie, physique et biologie) a été très large. Concernant le niveau d'études des utilisateurs, l'offre des BU étant, comme on l'a vu, clairement orientée vers les étudiants de premier cycle, ce sont majoritairement des étudiants des niveaux L et M qui consultent ces ressources.

Comment les e-books sont-ils utilisés ? On utilise les e-books plutôt qu'on ne les lit, comme un outil de référence, plus approprié à une consultation rapide et à une recherche d'information qu'à une lecture

suivie et approfondie (5 à 10 minutes de connexion en moyenne). Le livre électronique semble plus adapté à un usage par « tranches » (concept du *slice'n'dice* : découpage numérique du livre en chapitres ou unités de contenu) qu'au mode traditionnel de consultation en texte intégral⁸. Lors de la table ronde « e-book, le mot du futur ? », parrainée par Elsevier et organisée par l'université d'Aberdeen (Écosse) le 17 mai 2009, la question des usages a largement été abordée. Les lecteurs utilisent les multiples fonctionnalités proposées par les e-books : zoomer, chercher dans le texte, parfois annoter, renvoyer grâce aux notes de bas de page vers un autre livre, insérer dans sa bibliographie la référence bibliographique de l'e-book, mémoriser la page où s'est arrêtée la lecture, etc. La lecture est rapide et efficace : on ne lit pas tout un livre, on veut la réponse dont on a besoin ; ainsi, 20 % des usagers balaient les pages, 80 % lisent des extraits en se fixant pendant plus d'une minute. La recherche et le balayage représentent 25 % de l'usage d'un livre. Les chapitres d'e-books les plus consultés introduisent ou résument un thème. Les services demandés par les usagers aux bibliothèques sont l'accès immédiat à l'information, n'importe quand et de n'importe où, pouvoir télécharger et pouvoir copier / coller.

Ce qui change par rapport aux périodiques électroniques :

- *Les publics* : les e-books proposés par les BU s'adressent plutôt à des étudiants de premier cycle alors que les périodiques électroniques visent les étudiants avancés (inscrits en master ou en doctorat) et les chercheurs.
- *Les pratiques de lecture* : pour les e-books, il s'agit plus de consultation ou de recherche ponctuelle et ciblée. L'utilisation en serait plus marginale que celle des périodiques. Les e-books sont utilisés comme des ouvrages de référence, ils se consultent par partie ou par chapitre mais rarement en intégralité.
- *Le contenu* : les e-books choisis par les BU françaises sont moins spécialisés que les revues. Il s'agit d'une offre francophone à usage pédagogique (manuels). L'offre est beaucoup plus limitée que pour les périodiques électroniques. Et la notion d'actualité y est moins présente que dans les revues.

Une tendance est à noter : l'effacement de la notion de type de document (revue, e-book, etc.) au profit d'une plateforme qui proposerait

8. Jonchère Laurent. « Quel avenir pour le livre électronique dans les bibliothèques universitaires françaises ? ». *Documentaliste – Sciences de l'information*, 2005, vol 42, n° 1, pp. 26-30.

indifféremment tous ces documents. C'est ce que proposent de plus en plus de plateformes d'éditeurs (Wiley, Elsevier, Springer), qui sont communes aux périodiques et aux livres électroniques. Les utilisateurs n'y distinguent pas nécessairement les livres électroniques des revues, ce sont pour eux des ressources en ligne : ils vont piocher indifféremment l'information dont ils ont besoin dans des e-books, des périodiques, des bases de données, la distinction par type de ressource ne fait pas sens pour eux. Il en va de même dans le cas d'un métamoteur de recherche, qui requête en même temps des sources de types différents et rapatrie de la même façon tous les résultats.

Ce qui change par rapport au papier : un rapport publié en janvier 2008 par l'University College London sur les comportements des futurs chercheurs (ceux de la « génération Google », nés après 1993) insiste sur les nouvelles pratiques qu'ont adoptées les jeunes usagers des bibliothèques virtuelles pour la recherche et la lecture en ligne⁹. « *Les étudiants d'aujourd'hui maîtrisent tous de façon "native" le langage numérique des ordinateurs, des jeux vidéo et de l'Internet* » (Marc Prensky¹⁰). Ils ont tendance à ne parcourir qu'une partie infime des documents consultés et passent autant de temps à leur recherche qu'à la consultation effective du document recherché. On est très loin de la lecture d'un document papier : ici, « la lecture consiste à "scanner" le titre, le résumé et les informations visuellement mises en relief plutôt qu'à faire une lecture linéaire du document ». Il semblerait même que « ces utilisateurs privilégient la lecture en ligne pour s'épargner une lecture du document papier, qui ne permet pas cette photographie rapide des éléments clés ». Ces « *digital natives* ont l'habitude de penser en mode multitâche, ils préfèrent les graphiques aux textes, les accès aléatoires, les hyperliens, le zapping au déroulement linéaire d'une même tâche des heures durant et, finalement, sont plus performants en réseau » (Marc Prensky).

9. Nous reprenons ici certains points exposés par Cécile Touitou dans son article : « Les nouveaux usages des générations Internet ». *Bulletin des bibliothèques de France*, 2008, t. 53, n° 4, pp. 67-70. Les passages entre guillemets sont extraits de cet article.

10. Marc Prensky est un consultant américain dans le domaine des apprentissages et de la pédagogie, incluant particulièrement les nouvelles technologies. Cécile Touitou cite son article : "Digital natives, digital immigrants", *On the Horizon*, vol. 9, n° 5, octobre 2001.

Pratiques et usages liés aux tablettes

++++
Revenons au bilan du SCD d'Angers. Il a enregistré 54 prêts en 6 mois dont 14 prolongations. Les prêts sont de 20 jours en moyenne, avec un taux de rotation de 50% (54 prêts sur 90 créneaux potentiels). L'outil a de gros progrès à faire : fragilité matérielle de l'objet, pas assez ergonomique, non tactile, non annotable, pas en couleur, non connecté (pas de dispositif GSM, wifi, Bluetooth), sans parler des formats limités, des contenus insuffisants et du coût élevé.

À l'université de Toulon, un prêt de livrel – PRS-505 (Sony) – de longue durée (6 semaines) a également été lancé auprès de lecteurs volontaires qui devaient laisser leur avis. Un professeur agrégée en arts plastiques à l'IUT Services et réseaux de communication de Saint-Raphaël, spécialisée en esthétique des produits et culture de l'image, a laissé un témoignage intéressant. Elle loue l'ergonomie de l'outil : bonne luminosité, lecture prolongée possible, options de zoom appréciables (mais néanmoins insuffisantes), manipulation aisée des différents menus, marqueurs facilitant la prise de notes ultérieure, outil manipulable et discret. Par contre, la lecture non linéaire de gros ouvrages est très difficile à cause d'une navigation page par page : l'outil gagnerait à permettre un appel de page direct. Problème également de certains formats non acceptés. « Simplicité d'installation et de chargement de contenu, l'e-book est un outil qui complète le travail effectué sur l'ordinateur, nous dit-elle, d'une part par la qualité de son écran qui ne fatigue pas à la lecture, d'autre part par sa taille réduite qui permet de le consulter en toute situation. L'e-book est un avantage par rapport au support papier dans le sens où l'on peut très rapidement enrichir ou nettoyer la base de consultation des documents. »

PRATIQUES ET USAGES : QUELLES PERSPECTIVES ?

++++
Le bilan que l'on peut tirer de ces premiers résultats sur les pratiques et usages des e-books en BU est que nous sommes encore dans une phase d'expérimentation. L'intérêt est que rien n'est figé et que les BU peuvent œuvrer avec les éditeurs et les agrégateurs pour affiner et cibler leur offre, en travaillant tout particulièrement sur les points développés ci-après.

Mieux répondre aux besoins des utilisateurs

+++++

Et si l'on parlait des besoins réels des étudiants pour bâtir une offre de livres électroniques, tant du point de vue des supports que des contenus ?
Le livre électronique se conçoit en effet comme un ensemble de services. « Il apparaît nécessaire d'évoluer vers une convergence des supports, car les étudiants sont déjà suréquipés en objets numériques (iPods, clés USB, portables, etc.), et de généraliser les moyens d'accès à distance aux ressources, pour des usagers de moins en moins présents sur les campus. Le manuel de cours reste pour ces derniers un outil fondamental ; il faudrait leur donner la possibilité de reproduire numériquement les usages du papier (surlignage, annotations, résumés), tout en leur proposant de nouvelles fonctionnalités (version audio, partage de données entre utilisateurs, etc.) »¹¹.

Enrichissement du contenu : comme il est clair que les étudiants de premier cycle sont le public visé par les e-books en BU, il ne tient qu'aux éditeurs et agrégateurs d'étoffer leur offre avec des titres de niveau licence et en français. Les ouvrages académiques francophones et les manuels d'enseignement font encore cruellement défaut. « Notre offre de livres électroniques dépend largement du marché éditorial académique francophone, encore très restreint. Il s'agit pour nous prioritairement d'abonder l'acquisition de manuels en exemplaires multiples pour le niveau L. Or, l'offre de manuels dans les disciplines d'enseignement est encore très faible. » (un bibliothécaire chargé des ressources électroniques à l'université de Reims Champagne-Ardenne).

Faciliter l'accessibilité aux ressources

+++++

Améliorer le fonctionnement des plateformes de consultation et de prêt d'e-books : un gros travail reste à faire de ce côté-là, en particulier pour les agrégateurs français type Numilog. Le système de prêt d'e-books conduit à l'instauration d'un parcours complexe et verrouillé pour l'utilisateur. Et les systèmes de protection DRM (*digital rights management*) ne permettent pas un usage satisfaisant (interopérable et rétro-compatible) des e-books pour les lecteurs. Pour favoriser les usages d'e-books en BU, il faut être en

11. Intervention d'Henri Isaac, maître de conférences en gestion, université Paris-Dauphine (auteur du Rapport « Université numérique », rendu en janvier 2008) : « Livres électroniques : pour une compréhension des besoins étudiants », lors de la 4^e Journée Couperin sur le livre électronique.

mesure de proposer une offre homogène, lisible et ouverte en termes de nombre et de type d'accès.

Améliorer l'ergonomie des plateformes de consultation et de prêt d'e-books : plusieurs usagers ont jugé les plateformes « un peu fouillis » et indiqué avoir du mal à s'y retrouver. Par ailleurs, d'une plateforme à l'autre, les différences sont grandes, il n'y a aucune homogénéité qui faciliterait la prise en main. Intégrer le fonctionnement complet de chaque plateforme est long et fastidieux et freine les usages.

Fonctionnement des tablettes de lecture : elles ne sont pas encore assez performantes pour une lecture confortable, en situation de nomadisme notamment. Des évolutions technologiques sont attendues dans les années à venir pour améliorer la qualité de lecture sur écran. Deux étudiants du Miage (Méthodes informatiques appliquées à la gestion des entreprises) de Paris 5 ont exposé leur vision d'un « reader étudiant » : un outil nomade, autonome, peu encombrant, permettant d'accéder à tout moment aux ressources en ayant la possibilité d'en enrichir le contenu, de partager des commentaires et des annotations avec d'autres utilisateurs.

Créer les conditions d'un service performant

+++++

Communication autour des e-books : faire connaître aux usagers l'existence de la ressource est la condition *sine qua non* d'un usage généralisé. Il faut tendre vers la meilleure visibilité possible des titres dans le catalogue en ligne ou sur le site Web de l'institution et une campagne de promotion et d'information efficacement menée (création d'événements comme un jury littéraire ou un concours d'écriture). Les environnements numériques de travail sont également des outils stratégiques de promotion et de développement du livre électronique.

Modèle économique : le modèle économique proposé par les fournisseurs n'est pas toujours compatible avec les financements dont bénéficient les bibliothèques. Le modèle de l'abonnement ne convient pas aux établissements qui inscrivent leurs acquisitions de livres électroniques dans le cadre d'un projet d'investissement. « La solution de l'abonnement (par exemple Cyberlibris) ne nous satisfait pas. Les subventions que nous pouvons obtenir auprès de collectivités territoriales ou par l'université dans le cadre du Plan d'aide à la réussite en licence excluent les abonnements du périmètre des projets (s'agissant du financement de projets d'investissement, le subventionnement est limité aux achats d'ouvrages). » (un bibliothécaire chargé des ressources électroniques à l'université Reims)

Champagne-Ardenne). Par ailleurs, les offres éditoriales doivent correspondre à des tarifs réalistes, pour éviter d'aboutir à une crise semblable à celle des périodiques en ligne dans le secteur académique, et accepter d'évoluer pour répondre au mieux aux besoins et attentes d'un public universitaire.

Adopter une logique de niche : le SCD d'Angers préconise de ne pas hésiter à se lancer dans un projet de prêt de livrels mais avec une logique de niches : sélection / agrégation de titres payants et/ou gratuits dans un domaine précis, sélection d'articles de presse ou d'iconographie, prêt de corpus thématiques gratuits sur carte SD, mise en place de bornes ergonomiques de téléchargement, expérimentation pédagogique avec les enseignants, numérisation de documents locaux...

Travailler main dans la main avec les acteurs du marché : face à ce marché en devenir, la plupart des éditeurs ont adopté une position d'attente et semblent disposés à modifier les conditions de leurs prestations en fonction des retours d'expériences et des observations formulées par les utilisateurs. Les bibliothèques se retrouvent donc dans une relative position de force vis-à-vis des éditeurs et agrégateurs, auxquels il ne revient pas de définir seuls les conditions d'un marché encore en formation. Le constat de cette relation d'interdépendance entre les deux parties impose la nécessité d'un réel travail de collaboration¹². Par ailleurs, au sein même des bibliothèques, il est nécessaire d'impliquer le personnel : le SCD d'Angers insiste sur la nécessité de former des bibliothécaires veilleurs, sélectionneurs et rédacteurs, et d'intégrer les problématiques du livrel dans la chaîne traditionnelle de la bibliothèque tout en repensant les interactions entre l'établissement et son public ainsi que les modes de coopération (inter)nationale pour le partage de la valeur ajoutée.

CONCLUSION

+++++

Au terme de cette étude, il apparaît donc clairement que, même si l'introduction des e-books en BU est récente et encore en phase expérimentale et même si, par conséquent, les pratiques et usages qui en découlent sont en évolution constante et ont été peu étudiés, quelques tendances fortes ressortent :

12. C'est pour cadrer ce travail que la CeB a rédigé plusieurs documents à destination des négociateurs et des éditeurs, en particulier l'offre idéale. [En ligne] < http://www.couperin.org/IMG/pdf/offre_ideale_ebooks_2009.pdf > (consulté le 29 juin 2009).

- Les bibliothèques numériques sont bien plus représentées en BU que le prêt de tablettes de lecture, encore très marginal.
- Les publics cibles sont les étudiants de premier cycle et les contenus visés sont des ouvrages et des manuels en français adaptés au niveau L. La piste des TICE et de l'enseignement à distance est à suivre pour exploiter au mieux les e-books à l'université.
- L'offre de contenu en BU croît régulièrement depuis 5 ans, néanmoins les e-books sont encore peu intégrés dans les politiques d'acquisition et documentaire, ils restent en marge des collections traditionnelles.
- On utilise les e-books plutôt qu'on ne les lit, comme un outil de référence, plus approprié à une consultation rapide et à une recherche d'information (5 à 10 minutes de connexion en moyenne) par tranches, qu'à une lecture linéaire suivie et approfondie.

« Puisque l'électronique et le papier se distinguent si bien par leurs usages, peut-être le temps est-il venu pour les éditeurs et les diffuseurs de se départir du modèle traditionnel de la bibliothèque pour accentuer cette différence et exploiter au maximum les avantages d'un format numérique »¹³ : le livre électronique doit se concevoir comme un ensemble de services.

Nous concluons avec Laurent Jonchère, qui indique qu'une intégration réussie du livre électronique dans les bibliothèques universitaires françaises dépend d'un certain nombre de facteurs, parmi lesquels :

- la qualité des contenus offerts ;
- leur adéquation aux programmes enseignés ;
- un accès facile et étendu aux documents ;
- une bonne visibilité des titres dans l'OPAC ou sur le site Web de l'institution ;
- une campagne de promotion et d'information efficacement menée ;
- une possible intégration des ressources dans des programmes d'enseignement à distance.

13. Jonchère Laurent. « Quel avenir pour le livre électronique dans les bibliothèques universitaires françaises ? », *op. cit.*, p. 26.

+++++
LISTE DES AUTEURS
+++++

Anita Beldiman-Moore

Responsable marketing et infocentre,
Bibliothèque de Sciences Po

Emma Bester

Doctorante, Conservatoire national
des arts et métiers (CNAM), Équipe
Dispositifs d'information et de
communication à l'ère du numérique
(DICEN)

Assistante de documentation,
Centre pour l'édition électronique
ouverte (Cléo)

Caroline Bruley

Conservateur des bibliothèques,
responsable de la bibliothèque des
sciences et de la cellule informatique
documentaire du SCD de l'université
Jean Monnet Saint-Étienne

Ghislaine Chartron

Professeur, Conservatoire national des
arts et métiers (CNAM), responsable
de l'Institut national des techniques
documentaires (INTD)

Marin Dacos

Ingénieur de recherche, CNRS,
directeur du Centre pour l'édition
électronique ouverte (Cléo)

Ilham Derfoufi

Docteur en sciences de l'information
Traductrice scientifique, spécialisée
en sciences de l'information

Benoît Epron

Maître de conférences,
directeur de la recherche, ensib

Jenny Fry

Lecturer, Department of Information
Science,
Loughborough University

Isabelle Galina

Honorary lecturer, Department of
Information Studies, University
College London

Anne Gentil-Beccot

Bibliothécaire, CERN (Genève)

Paul Huntington

Research Associate, UCL Department
of Information Studies, University
College London (until 2009)

Annaïg Mahé

Maître de conférences,
Urfist de Paris / École nationale des
chartes

Delphine Merrien

Conservateur des bibliothèques,
Mission innovation et numérique de
l'ensib – Observatoire du numérique
dans l'enseignement supérieur

Nikoleta Pappa

Human Resources Information Officer,
University College London

Melissa Terras

Doctor of Engineering Science,
Co-director of UCL Centre for Digital
Humanities, Reader in Electronic
Communication, Department of
Information Studies, University
College London

Anna Wojciechowska

Ingénieur de recherche, CNRS,
responsable de la bibliothèque
du Centre de mathématiques et
informatique (CMI),
université d'Aix-Marseille, Marseille

Claire Warwick

Professor of Digital humanities, head
of the Department of Information
Studies,
University College London

Secrétariat d'édition:
Silvia Ceccani

Mise en page:
Alexandre Bocquier

Conception graphique:
atelier Perluette, 69001 Lyon.
< <http://www.perluette-atelier.com> >

dépôt légal: 1^{er} semestre 2012

Pour suivre l'actualité de nos publications : < <http://www.enssib.fr/presses> >.