



Le devenir de la littérature grise. Quelques observations.

Joachim Schöpfel

► To cite this version:

Joachim Schöpfel. Le devenir de la littérature grise. Quelques observations.. Perspectives documentaires en éducation, ENS Editions, 2006, 2006 (62). sic_00136829

HAL Id: sic_00136829

https://archivesic.ccsd.cnrs.fr/sic_00136829

Submitted on 15 Mar 2007

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

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

LE DEVENIR DE LA LITTÉRATURE GRISE

QUELQUES OBSERVATIONS

JOACHIM SCHÖPFEL

Tout cartographe, fut-il celui de l'archipel du savoir (Baltz, 2003), a besoin de quelques repères pour ses dessins et explorations.

QUELQUES REPERES

Il existe plusieurs définitions de la littérature grise, la plus connue étant celle dite « de Luxembourg », discutée et approuvée lors de la 3^e conférence internationale sur la littérature grise en 1997 : « [Grey literature is] *that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers* »¹.

La définition met déjà en évidence deux caractéristiques majeures de ce type d'information : d'une part son universalité et son ubiquité, d'autre part la difficulté de l'identifier et d'y accéder par les circuits commerciaux de l'édition classique. De plus, la définition de Luxembourg est suffisamment vague pour montrer combien il est difficile de déterminer précisément ce qu'est cette littérature « souterraine », « éphémère » ou « non conventionnelle ». Selon les mots de deux experts de la British Library, « *grey literature is difficult to define* » (Wood et Smith, 1993).

En fait, la notion de littérature grise regroupe traditionnellement trois catégories de documents, souvent

à petit tirage : conférences, rapports et thèses. Néanmoins, la frontière avec la littérature « blanche » ou « conventionnelle » est perméable – une partie des conférences est publiée par des éditeurs commerciaux, comme monographies, dans une collection ou une revue. Il en va de même pour certains rapports. Quant aux thèses, un certain nombre, notamment en sciences humaines et sociales, se retrouvent sur le marché de l'édition commerciale.

Pour tous les autres documents, cependant, qui échappent aux circuits de l'édition classique, l'absence de « contrôle commercial » pose un réel problème aussi bien pour l'enseignant-chercheur que pour le professionnel de l'information : où les chercher, où les trouver, comment les acquérir ? L'absence de « contrôle commercial » et de promotion (publicité) implique souvent aussi un manque de « contrôle bibliographique », c'est-à-dire que ces documents sont souvent insuffisamment signalés dans les catalogues ou les bases de données. Explorer cette partie de l'information scientifique nécessite des connaissances spécifiques des sources et circuits. Preuve, s'il en faut, est le recensement des thèses concernant la recherche en éducation, sur le site de l'INRP.

Depuis bientôt trente ans, des professionnels de l'information – documentalistes, bibliothécaires, chercheurs et enseignants – ont contribué à l'étude de la littérature grise, et il existe un corpus assez riche d'articles et, depuis 1993, de communications issues des conférences internationales sur la littérature

1. « [La littérature grise est] *ce qui est produit par toutes les instances du gouvernement, de l'enseignement et la recherche publique, du commerce et de l'industrie, sous un format papier ou numérique, et qui n'est pas contrôlé par l'édition commerciale* » (traduction J. S.).

grise². Plusieurs pays européens ont créé en 1985 une association pour l'identification et la diffusion de cette littérature (EAGLE). Dans les années 1990, la France a connu des initiatives au niveau national et interministériel – dont GRISELI et LIGRIA (cf. Comberousse, 1995 ; Desmichel, 1998) – qui ont abouti notamment à la mise en place de deux « guichets nationaux » d'accès aux rapports (la Documentation française et l'INIST-CNRS). D'autres pays, comme les États-Unis et le Royaume Uni, ont désigné des « centres de dépôt et de diffusion » (ex. : The British Library) ou mis en place des portails d'accès aux rapports scientifiques (ex. : GrayLit³ pour les agences fédérales américaines).

Pour les sciences de l'éducation, le portail de la veille de l'INRP⁴ référence depuis 2004 environ 1 000 colloques, rapports et thèses, dont une partie est accessible via des liens sur les sites ou documents en ligne. La base de publications du DIPF⁵, quant à elle, contient presque 1 400 références, dont 170 (12 %) sont classées sous la rubrique « autres », parmi lesquelles on trouve des actes de colloques, des rapports, etc.

Notre étude ne prétend pas à une synthèse exhaustive ; elle se limite à quelques observations sur l'importance de cette information et sur son évolution dans l'environnement numérique. Nous ajouterons des remarques sur le problème du signalement et de la diffusion, puis nous terminerons par plusieurs pronostics et questions ouvertes.

DE L'IMPORTANCE (RELATIVE) DE LA LITTÉRATURE GRISE

La littérature grise occupe une place à part entière dans la diffusion de l'information scientifique et technique (cf. Sondergaard *et al.*, 2003), et les professionnels insistent sur son importance. Il y a surtout deux raisons à cela : les résultats de recherche sont souvent plus détaillés dans les rapports, thèses ou conférences que dans les revues, et ils y sont diffusés jusqu'à 18 mois avant d'être publiés ailleurs (cf. Wood et Smith, 1993).

Nous avons évoqué l'ubiquité de cette littérature. SIGLE, la base de données européenne de littérature

2. <http://www.textrelease.com>
3. <http://www.osti.gov/graylit/index.html>
4. <http://www.inrp.fr/vst>
5. <http://www.dipf.de/publikationen.htm>

grise, a été alimentée depuis les années 1980 par des organismes de plusieurs pays de l'Union européenne. Ses références couvrent tous les domaines scientifiques, des sciences humaines et sociales (40 %) aux mathématiques et à la physique (13 %), des sciences médicales et biologiques (12 %) aux sciences de l'ingénieur (18 %).

Mais comment les chercheurs utilisent-ils réellement cette littérature ? Un moyen pour évaluer cet usage est d'analyser les citations dans leurs publications. Pour chiffrer cette utilisation, nous avons exploité deux sources différentes : une étude franco-néerlandaise en cours depuis 2004 et, pour l'intérêt de notre propos, une analyse des bibliographies de la revue *Perspectives documentaires en éducation*.

ÉTUDES SCIENTOMETRIQUES (ANALYSE DE CITATIONS)

Citons d'abord quelques résultats d'une étude franco-néerlandaise (Schöpfel *et al.*, 2005 ; Farace *et al.*, 2006) qui a analysé 64 articles scientométriques publiés entre 1987 et 2005, totalisant plusieurs dizaines de milliers de références⁶. Voici la part de documents gris dans les publications de différents domaines scientifiques.

Domaine	Part de la littérature grise
Sciences de la Terre	14 %
Biologie	5-13 %
Médecine vétérinaire	6 %
Psychiatrie (toxicomanie)	1 %
Psychologie	3 %
Sciences de l'ingénieur	39-42 %
Sciences de l'économie	9-17 %
Sociologie	7-9 %
Sciences de l'éducation	14-19 %

Tableau 1 : La part de la littérature grise dans les publications scientifiques

⁶ Echantillon basé sur une recherche systématique dans des bases de données, bibliographies etc. d'analyses de citations contenant des données statistiques sur des documents gris.

D'après ces analyses, l'importance relative de la littérature grise dépend pour beaucoup de la discipline et du sujet de la recherche, de l'approche méthodologique et des sources exploitées. Certains domaines, notamment les sciences de la vie et les sciences médicales, privilégient les vecteurs conventionnels de diffusion (revues), tandis que d'autres, comme l'agriculture, l'aéronautique et plus généralement les sciences de l'ingénieur, s'appuient davantage sur des ressources grises.

En particulier, les administrations publiques et les laboratoires de recherche publics et industriels produisent beaucoup de matériel gris, surtout des rapports et « working papers », souvent avec une diffusion interne et/ou restreinte (cf. par exemple Ullah *et al.*, 2004).

PERSPECTIVES DOCUMENTAIRES EN EDUCATION (ANALYSE DE CITATIONS)

Pour avoir une idée plus précise de l'usage des ressources grises dans la documentation en sciences de l'éducation, nous avons étudié quelques articles publiés récemment dans la revue *Perspectives documentaires en éducation*. Les résultats sont hétérogènes.

Citons d'abord le cas de quatre articles publiés dans le n° 59 de 2003, qui totalisent 372 références bibliographiques. La part des citations qui renvoient à des documents gris varie entre 12 % (Lüdke, 2003), 37 % (Duarte, 2003), 45 % (Landesmann, 2003) et 56 % (Ndoye, 2003). Cette dernière étude – une synthèse documentaire sur le rendement scolaire en Afrique – est symptomatique : la plupart des sources utilisées par Ndoye sont des documents produits par des organismes nationaux ou internationaux (ministères, universités, ONU, banque mondiale, etc.), qui échappent à toute diffusion commerciale et qui, trois ans plus tard, sont probablement difficiles à trouver et à obtenir.

Ces premiers résultats corroborent l'étude d'Auricombe (2001) sur les recherches et usages de l'information documentaire par des chercheurs et enseignants-chercheurs sur la formation du CNAM, Auricombe (2001) qui souligne l'importance de la littérature grise. En effet, 21 % de ses propres références relèvent clairement de cette catégorie (ex. : mémoires de l'INTD et de l'ENSSIB).

De façon plus globale, nous avons analysé les différentes études et notes du n° 60 de *Perspectives documentaires en éducation* de 2003. 15 des 198 sources citées

par les auteurs (soit 7 %) font référence à des conférences, des séminaires, des textes juridiques et des mémoires de thèses. Les autres citations, comme les références de la « bibliographie courante » de ce numéro, renvoient à des ouvrages et des articles de revues.

À titre comparatif, citons encore deux études relatives au domaine des sciences de l'éducation, publiées dans d'autres revues :

- Okiy (2003) a analysé 4 012 références de 70 thèses soutenues en Nigéria et a identifié 14 % de sources grises (thèses, rapports, conférences...);
- une autre analyse (Beile *et al.*, 2004) portant sur 1 842 références issues de thèses soutenues aux États-Unis aboutit quant à elle à un pourcentage de documents gris de plus de 19 %.

Source	Thématique	Part de la litt. grise
Lüdke, 2003	Chercheur Brésil (itinéraire personnel)	12 %
Duarte, 2003	TIC (repères bibliographiques)	37 %
Landesmann, 2003	Enseignants Mexique (itinéraire personnel)	45 %
Ndoye, 2003	Rendement scolaire Afrique (repères bibliographiques)	56 %
Auricombe, 2001	Pratique de chercheurs CNAM	21 %
PDE n° 60	Divers	7 %
Okiy, 2003	Thèses Nigéria	14 %
Beile <i>et al.</i> , 2004	Thèses US	19 %

Tableau 2 : La part de la littérature grise dans différentes études en sciences de l'éducation

En résumé, la part relative de la littérature grise dans l'information scientifique varie beaucoup en fonction du domaine scientifique mais aussi du sujet, de la méthodologie, voire de l'origine géographique. En sciences de l'éducation, cette littérature joue un rôle

considérable, correspondant à une moyenne entre 10 % et 20 % des sources exploitées⁷.

TYOLOGIE : DU PAPIER A L'ELECTRONIQUE

Revenons maintenant à la définition de la littérature grise. Nous rappelions plus haut que ce terme englobait traditionnellement les rapports, les conférences et les thèses. Voyons de plus près de quoi il s'agit réellement. Les plus de 800 000 références de la base SIGLE se répartissent comme suit.

Type de document	Part des références
Rapports	62,7 %
Thèses	31,7 %
Conférences	2,3 %
Fichiers (données)	2,1 %
Traductions	0,9 %
Autres	0,3 %

Tableau 3 : La part de différents types de documents dans la base SIGLE

Les rapports occupent de loin la première place parmi les différents types de littérature grise. Mais il y a rapport et rapport... La catégorie recouvre une grande variété de documents très différents : rapports institutionnels, rapports annuels ou d'activité, rapports de projets ou d'étude, rapports techniques, rapports publiés par un ministère, rapports d'un laboratoire ou d'une équipe de recherche... Certains rapports sont diffusés par des organismes publics (ex. ADEME, CERTU, Documentation française, Commission européenne), d'autres ont un statut confidentiel, sont protégés ou circulent d'une manière restreinte, comme par exemple les rapports techniques de laboratoires industriels. Certains rapports sont volumineux, avec des annexes statistiques, d'autres se limitent à quelques pages.

Quant aux autres catégories, les études scientométriques (cf. Farace *et al.*, 2006) offrent une vaste pa-

7. À titre comparatif : notre article cite environ 30 % de références « grises ».

noplie de ressources grises : à part des thèses et conférences, on y trouve pêle-mêle manuscrits non-publiés, bulletins d'information, recommandations et normes, brevets, notes techniques, données et statistiques, présentations, communications personnelles, « *working papers* », cahiers de laboratoires, préprints, travaux universitaires, « *lecture notes* », etc.

Malgré leur grande diversité, tous ces documents ont un point commun : ils contiennent des informations scientifiques et techniques uniques et significatives qui souvent ne sont pas publiées ailleurs. L'absence de signalement et de diffusion adéquate pose donc, nous le disions déjà, un réel problème à la communication scientifique.

Ceci étant, Internet est en train de modifier la donne. Pas seulement à cause du changement de comportement des utilisateurs (cf. par exemple l'étude récente sur les enseignants-documentalistes par Le Roux, 2005), mais aussi et surtout parce que de plus en plus les documents gris sont publiés sur le Web. Néanmoins, ce nouveau vecteur de publication ne veut pas nécessairement dire qu'il y a davantage de documents gris, comme souligne une étude issue du centre d'information en sciences sociales allemand (Artus, 2005). Sans accroître le nombre de publications, le passage du papier au numérique modifie par contre fondamentalement les modes d'accès et de diffusion et renforce le caractère éphémère et volatil de la littérature grise.

DES PREPRINTS AUX ARCHIVES OUVERTES

Il n'est pas nécessaire ici de décrire le mouvement vers l'accès libre à l'information scientifique (« *open access* ») qui se cristallise depuis 1994 autour des initiatives en faveur d'archives ouvertes, et il suffit, à titre d'introduction, d'orienter le lecteur vers quelques ressources récentes (Chanier, 2004 ; Fily, 2005 ; Aubry et Janik, 2005 ; guide d'A. Mahé, sur le site de l'URFIST Paris⁸ ou le site de veille de l'INIST⁹).

Néanmoins, le cas du premier serveur de préprints, ArXiv, mis en place par P. Ginsparg à Los Alamos en 1991, est intéressant pour notre étude¹⁰. Il s'agissait d'un mode de diffusion des résultats de recherche

8. <http://www.ext.umpc.fr/urfist/ArchiveOuverte/OA.html>

9. <http://www.inist.fr/openaccess>

10. <http://arxiv.org>

organisé par et pour des physiciens, sans médiation, qui échappait totalement aux circuits commerciaux de l'édition scientifique. Dans ce sens, ce serveur (qui contient aujourd'hui plus de 360 000 documents) correspondait parfaitement à la définition de la littérature grise.

Cependant, l'affaire est plus complexe. Il n'y aurait pas eu de « preprints » sans « prints », autrement dit, pas d'ArXiv sans revues scientifiques. Visant uniquement la diffusion rapide et immédiate des résultats, P. Ginsparg ne mettait pas en question l'édition commerciale, ne proposant pas d'alternative à l'évaluation par les pairs (« peer review ») qui était toujours assurée par la chaîne éditoriale classique. En fait, ArXiv réalisait une sorte de symbiose entre littérature grise et édition traditionnelle, liée au fonctionnement spécifique d'une communauté scientifique très particulière, celle de la physique nucléaire.

La création des premières archives ouvertes au sein du CNRS par Franck Laloë, physicien à l'ENS, à partir de 1998, suivait globalement la même logique¹¹, prônant la rapidité et la gratuité d'une communication scientifique directe entre chercheurs du même domaine.

Mais la crise des revues scientifiques (cf. Chartron, 2002 ; Keller, 2001) et l'appropriation des NTIC par les professionnels de l'information ont contribué à transformer ce mode de diffusion en modèle alternatif à l'édition scientifique. En France, HAL¹² du CCSD – 19 000 documents début 2006 (cf. Lardy, 2006) – est en train de devenir la première archive institutionnelle à vocation nationale, avec un double objectif :

- économique : proposer une alternative moins onéreuse à la publication scientifique, en exerçant une pression croissante sur les éditeurs STM ;
- administratif : faciliter le contrôle de la production scientifique des différents organismes de recherche (EPST, universités) par le recensement et l'évaluation des publications des chercheurs via une base centrale.

Vu la diversité méthodologique et herméneutique des différentes communautés scientifiques, on peut se demander si ce qui fonctionne depuis quinze ans en physique le fera aussi bien dans d'autres domaines, en particulier à cause des « contraintes induites par les modalités d'évaluation de la recherche » (cf. Prévot

11. <http://ccsd.cnrs.fr/>

12. <http://hal.ccsd.cnrs.fr>

2005 pour les sciences de la Terre). Une étude comparative de la bibliothèque nationale du Portugal – mathématiques *versus* agriculture – va dans le même sens (Ramalho Correia et de Castro Neto, 2002).

Quelle est la place de la littérature grise dans cet environnement émergent ? Le premier répertoire international d'archives ouvertes, le *Directory of Open Access Repositories* (OpenDOAR) mis en place par les universités de Nottingham et de Lund¹³, recense 349 sites différents dont 20 en France (février 2006). OpenDOAR indexe douze catégories de documents gris : rapports, thèses, conférences et preprints, mais aussi « *working papers* », « *learning objects* » (surtout des cours universitaires), présentations PowerPoint, mémoires d'étudiants et résultats de recherche (*data sets*). Comme ce répertoire ne permet pas de recherche fédérée dans les archives, il est impossible d'obtenir une répartition globale des documents selon cette typologie. Voici donc seulement le nombre d'archives ouvertes qui contiennent différentes catégories de littérature grise.

Type de documents	Nombre d'archives	% des sites du OpenDOAR
Thèses	211	60 %
Rapports	146	42 %
Conférences	146	42 %
Preprints	89	26 %
Mémoires d'étudiants	72	21 %
<i>Working papers</i>	66	19 %

Tableau 4 : La présence de différents types de documents dans les archives du OpenDOAR

À première vue, ces chiffres semblent prouver que la littérature grise est relativement bien représentée dans les archives ouvertes. 60 % des sites recensés contiennent des thèses, et l'on trouve des conférences et des rapports dans plus de 40 % des archives ouvertes. Mais il faut relativiser. En analysant la description des archives, on constate que le nombre de sites qui sont explicitement dédiés à cette littérature et qui facilitent la recherche et l'identification est bien moins élevé :

13. <http://www.opendoar.org/>

on trouve seulement 45 sites pour les thèses (13 %), 30 pour les rapports (9 %) et seulement 8 pour les conférences (2 %). Souvent, d'après un constat de la 7^e conférence internationale sur la littérature grise en 2005, ces documents sont en fait « noyés » dans la masse d'articles déposés dans les archives et/ou difficiles à identifier. Ainsi, le développement des archives ouvertes ne change *a priori* pas grand chose à la situation de ces ressources de littérature grise.

Même s'il ne s'agit pas d'une archive au sens stricte, le site du British Education Index¹⁴ est intéressant. Son *Internet Resources Catalogue* contient 4 250 liens (octobre 2005) dont 3 100 vers des documents individuels. Parmi ces liens se trouvent plus de 2 000 conférences, rapports, « *governmental publications* », etc., même si l'identification des documents gris est plus ou moins aléatoire à cause d'une indexation plutôt sommaire.

Terminons le chapitre par quelques remarques sur la situation en France. Même si 9 archives, dont HAL, contiennent des actes de congrès, elles les signalent au même niveau que les articles (communications individuelles) sans proposer de fonctionnalités spécifiques pour ce type d'information (regroupement par conférence, etc.). Il en va de même pour les rapports : bien sûr, plusieurs archives en disposent – surtout des archives institutionnelles : IFREMER, Institut Jean Nicod, MSH-Alpes. Mais, à notre connaissance, il n'existe à ce jour qu'un seul projet (LARA de l'INIST-CNRS, basé sur DSpace) exclusivement développé pour le dépôt de rapports scientifiques et techniques (cf. Stock *et al.*, 2006).

Quant aux thèses, la situation est différente et plus avancée, ce qui est probablement dû à trois facteurs : l'environnement international (NLTLD, conférences ETD), le dispositif national autour de l'ABES (SUDOC, projet STAR) et l'enjeu scientifique des thèses (cf. Le Hénaff et Thiolon, 2005). Ceci étant, aujourd'hui encore, le paysage français est caractérisé par une certaine hétérogénéité autour d'archives institutionnelles, thématiques et consacrées aux thèses (cf. Cyberthèses, PASTEL, TEL) qui comptabilisent ensemble seulement une petite partie de la production nationale (cf. Paillassard *et al.*, 2005). Néanmoins cette situation, peu satisfaisante comparée à d'autres pays, évoluera certainement dans les mois à venir vers un recensement et un dépôt plus

systématique des thèses électroniques françaises sur les sites de l'ABES et du CCSD.

COMMENT AMELIORER LE « CONTROLE BIBLIOGRAPHIQUE » ?

Nous remarquons plus haut que l'absence de « contrôle commercial » de la littérature grise implique un manque de « contrôle bibliographique », c'est-à-dire que ces documents sont souvent insuffisamment signalés dans les catalogues ou les bases de données. Ceci ne veut pas dire qu'il n'existe pas de normes ou de recommandations pour le catalogage des rapports, conférences, thèses, etc. Mais, contrairement aux revues et aux livres, l'absence d'enjeu commercial a contribué à un « succès » très relatif des règles énoncées.

Aujourd'hui encore, les modes de signalement des différents types de littérature grise dépendent davantage des choix des organismes qui produisent, collectent ou diffusent ces documents que d'un standard national ou international (ISO, ISBD, AFNOR, etc.). L'échec de la norme internationale pour la numérotation des rapports est symptomatique – après avoir été pendant plusieurs années le seul pays avec une agence ISRN active (INIST), la France n'a pu qu'accepter l'abandon de cette norme par l'ISO. Seule exception (encore) au niveau national, les thèses soutenues devant les universités françaises : la création de la base Téléthèses et, plus tard, le catalogage dans le SUDOC ont imposé un format bibliographique uniforme.

Au niveau européen, l'alimentation de la base SIGLE a obligé les pays membres de ce réseau à entreprendre une harmonisation du signalement de leurs ressources grises autour d'un format pivot en SGML. Mais l'alimentation de SIGLE a été arrêtée en avril 2005, et le réseau – l'association EAGLE – est en liquidation. Résultat prévisible : chaque organisme reprendra ses propres modes et règles de signalement, sans aucun contrôle bibliographique homogène.

Le développement rapide de l'Internet avec son foisonnement de ressources en ligne exerce un double effet sur le contrôle bibliographique (cf. Artus, 2005) : d'une part, le « *wild growth* » favorise le déclin de l'application de normes formelles et qualitatives, avec le risque de voir devenir la littérature grise « *even more greyish* ». D'un autre côté, ce risque a accéléré dans tous les pays la prise de conscience de la néces-

¹⁴ <http://www.leeds.ac.uk/bei>

sité de définir quelques données minimales – les métadonnées – pour cadrer le signalement des documents numériques.

Citons à titre d'exemple les projets pour adapter le Dublin Core aux rapports (Jeffery *et al.*, 2002) ou aux thèses ; en France, l'initiative des « Thèses Electroniques Françaises » (TEF) travaille sur un ensemble de métadonnées et un schéma XML unique¹⁵. Un autre exemple : la « *Text Encoding Initiative* » (TEI¹⁶) pour élaborer et vise à élaborer et à recommander, au niveau international, des normes communes de balisage de textes indépendantes de l'évolution des systèmes informatiques, et la création en 2005 par le CNRS et l'INRIA d'un centre support TEI pour l'Europe à Nancy, autour des trois structures ATILF, LORIA et INIST, qui se chargera probablement aussi du problème de la littérature grise.

Le problème de la pauvreté et de la diversité du signalement des documents déposés dans les archives institutionnelles a conduit le JISC anglais, début 2006, à charger une commission d'étudier l'interopérabilité de ces archives et à qualifier leurs ressources afin d'en faciliter l'identification par l'utilisateur final (« JISC Scoping Repository Study »).

Un dernier exemple : à l'issue de la 7^e conférence internationale sur la littérature grise, en 2005, P. de Castro et S. Salinetti, de l'Istituto Superiore di Sanità à Rome, ont pris l'initiative de créer un comité international de la littérature grise, afin d'élaborer des recommandations pour la production et la diffusion de rapports scientifiques et techniques (« *Nancy Style* »). Une première version de ce document sera publiée sur le Web¹⁷ dans les prochaines semaines (GLISC, 2006) ; des traductions italiennes et françaises sont en préparation.

Cet exemple est d'une certaine manière symptomatique de la question du contrôle bibliographique de la littérature grise : il s'agit en effet d'un problème international, pour lequel des réponses nationales n'offrent que des solutions partielles, contribuant de ce fait à la diversification des données et à la difficulté d'identification et de localisation des documents. En absence d'enjeu commercial, l'amélioration du signalement dépendra toujours de l'initiative et de la bonne volonté des organismes producteurs et/ou diffuseurs et de quelques professionnels engagés et convaincus.

¹⁵ <http://www.abes.fr/abes/documents/tef/index/html>

¹⁶ <http://www.tei-c.org>

¹⁷ <http://www.glisc.info>

ACCES ET DIFFUSION

Identifier, localiser et obtenir de la littérature grise n'est souvent pas facile – ceci tient à la nature même de cette littérature. Pendant vingt ans, la base SIGLE a proposé une solution au niveau de l'Union européenne dans la mesure où les organismes partenaires étaient obligés de tenir les documents signalés à disposition de l'utilisateur final, via des services de prêt ou de fourniture de documents.

Aujourd'hui, l'utilisateur est confronté à une multitude de sites, d'archives, de catalogues et de bases qui rendent la recherche de l'information « grise » souvent fastidieuse, voire prohibitive, sans parler du problème linguistique.

Face à l'évolution rapide de l'IST dans l'environnement numérique, tous les grands centres traditionnels de collecte et de diffusion de littérature grise, comme la British Library, l'Institut Canadien d'Information Scientifique et Technique, la TIB Hannover ou l'INIST, ont commencé à développer des services d'accès libre à ces documents, en particulier pour les thèses et les préprints, mais aussi pour les rapports, etc. (cf. Boukacem-Zeghmouri et Schöpfel, 2006).

Par contre, on constate toujours l'absence – du moins au niveau national ou européen – de portails ou d'outils de recherche dédiés à ce type de documents. Même l'éditeur Elsevier a commencé à indexer des thèses électroniques dans son moteur de recherche *Scirus*. L'initiative des membres de l'association EAGLE visant à la création d'un métamoteur de recherche spécifique aux collections européennes de littérature grise est restée à ce jour au stade de projet (cf. Schöpfel, 2006). En attendant, l'utilisateur n'a pas le choix : soit il cherche lui-même avec les moyens du bord (par exemple avec des moteurs de recherche généralistes comme Google, Scirus etc.), soit il continue d'avoir recours aux services de recherche des organismes traditionnels (SCD, BU, INIST).

L'AVENIR DE LA LITTÉRATURE GRISE

Nous avons commencé notre article par quelques repères pour comprendre l'objet « littérature grise ». En guise de conclusion, osons quelques pronostics et posons quelques questions.

Il est certain que :

– la littérature grise ne disparaîtra pas mais gardera sa place à côté de l'édition commerciale ;

- les frontières entre gris et blanc (commercial) sont en train de devenir de plus en plus diffuses et fluctuantes, en particulier dans l'environnement du mouvement vers l'accès libre à l'IST (cf. Banks, 2006) ;
- la part des documents gris publiés sur le Web augmentera rapidement ;
- Internet favorisera la diversification des ressources grises (résultats bruts, notes et commentaires personnels, messages, blogs, cours, etc.).

Il est probable que :

- le contrôle bibliographique de la littérature grise posera toujours un problème, malgré une tendance vers la normalisation des documents numériques ;
- les archives ouvertes proposeront des services et fonctionnalités plus adaptés pour une partie de la littérature grise, non seulement pour les préprints mais aussi pour les thèses et les rapports ;
- certains organismes – surtout publics mais aussi privés (Elsevier, Google...) – développeront des outils et services pour mieux exploiter les gisements de ressources grises sur le Web.

En revanche, il est peu probable que :

- la recherche et collecte de la littérature grise deviennent aussi faciles que pour les revues et livres de l'édition traditionnelle ;
- les nouveaux outils de collecte, de dépôt et d'archivage rendent la littérature grise beaucoup moins éphémère et volatile que par le passé.

Pour les professionnels de l'information et de la documentation, la littérature grise restera un défi et continuera d'offrir un domaine de recherche et d'action intéressant, au moins pour cinq axes (cf. Stock et Schöpfel, 2004).

Besoin d'une nouvelle définition

La définition traditionnelle de la littérature grise doit être approfondie et complétée par une analyse précise des nouveaux modes d'accès et de diffusion, suivant le constat de Mackenzie Owen (1997) que « *Grey does not imply any qualification [but] is merely a characterization of the distribution mode* »¹⁸.

Besoin d'un modèle économique

Malgré l'absence d'un « contrôle commercial », la collecte, la diffusion et la recherche de littérature grise

18. « [Le terme] *Gris ne contient aucune qualification [mais] caractérise uniquement le mode de diffusion* » (traduction J. S.).

a un prix qui peut être bien plus élevé que pour un article ou un livre. À ce jour, il n'existe aucun modèle économique dans ce domaine, et il manque une analyse en termes d'investissements, de coûts directs et indirects, de prix d'acquisition, etc.

Besoin d'encadrer l'archivage

Les NTIC facilitent l'archivage de n'importe quelle ressource, avec une forte incitation exercée par le mouvement de l'« *open access* ». Cependant, la question du « qui doit archiver quoi, où, quand et pour combien de temps » reste à ce jour largement sans réponse. Compte tenu des aspects politiques et surtout financiers liés à cette question, il y a un intérêt à trouver rapidement des réponses, ne fut-ce que pour une partie des ressources grises.

Besoin d'une nouvelle « chaîne de valeur »

Aux Pays-Bas, Rosendaal (2004) étudie depuis quelques années le processus de ré-appropriation de l'édition par les universités, en insistant sur la modification en profondeur de la « *value chain* » (chaîne de valeur ajoutée) des publications scientifiques. La question de l'évaluation et de la qualité des publications scientifiques sera un des grands enjeux de l'évolution de l'IST. L'impact des NTIC sur les circuits non commerciaux est complexe et à ce jour peu exploré – un vaste champ de recherche.

Besoin de clarification des aspects juridiques

Volontairement, notre étude a complètement édulcoré la question de la propriété intellectuelle de la littérature grise. Néanmoins, le statut juridique des ressources grises et le droit de leur exploitation (dépôt, archivage, diffusion, etc.) reste un (autre) enjeu majeur pour l'avenir de cette forme d'IST. L'environnement légal en France et au niveau international évolue rapidement, et il manque à ce jour une analyse documentaire qui intègre les aspects juridiques et les enjeux économiques de la littérature grise.

Il reste donc beaucoup de travail pour le cartographe qui veut s'aventurer dans cette partie de l'archipel du savoir...

Joachim Schöpfel
Édition numérique et fourniture de documents,
INIST-CNRS
Président d'EAGLE

REFERENCES BIBLIOGRAPHIQUES

ARTUS H.M. (2005). « Old WWWine in New Bottles ? Developments in electronic information and communication : structural change and functional inertia ». *The Grey Journal*, vol. 1, n° 1, p. 9-16.

AUBRY C. et JANIK J. (2005). *Les archives ouvertes : enjeux et pratiques*. Paris : ADBS.

AURICOMBE S. (2001). « Recherches et usages de l'information documentaire : analyse des pratiques des chercheurs et enseignants-chercheurs sur la formation du CNAM ». *Perspectives Documentaires en Éducation*, n° 52, p. 61-69.

BALTZ C. (2003). « Quand la documentation se réveillera... » *Documentaliste*, vol. 40, n° 2, p. 148-153.

Banks M.A. (2006). « Towards a Continuum of Scholarship : The Eventual Collapse of the Distinction Between Grey and non-Grey Literature ? ». In : D. FARACE et J. FRANTZEN (éd.). *GL7 Conference Proceedings. Seventh International Conference on Grey Literature : Open Access to Grey Resources*. Nancy, 5-6 December 2005. Amsterdam : TextRelease, p. 65-70.

BEILE P.M., BOOTE D.N. et KILLINGSWORTH E.K. (2004). « A microscope or a mirror ? A question of study validity regarding the use of dissertation analysis for evaluating research collections ». *The Journal of Academic Librarianship*, vol. 30, n° 5, p. 347-353.

BOUKACEM-ZEGHMOURI C. et SCHÖPFEL J. (2006). « Access and Document Supply : A Comparative Study of Grey Literature ». In : D. FARACE et J. FRANTZEN (éd.). *GL7 Conference Proceedings. Seventh International Conference on Grey Literature : Open Access to Grey Resources*. Nancy, 5-6 December 2005. Amsterdam : TextRelease, p. 186-193.

CHANIER T. (2004). *Archives ouvertes et publication scientifique : comment mettre en place l'accès libre aux résultats de la recherche ?* Paris : L'Harmattan.

CHARTRON G. (éd.) (2002). *Les chercheurs et la documentation numérique*. Paris : Éditions du Cercle de la Librairie.

COMBEROUSSE M. (1995). « Les nouvelles technologies au service de la littérature grise (GRISELI) ».

Bulletin des Bibliothèques Française, vol. 40, n° 2, p. 51-53.

DESMICHEL C. (1998). « Pour un meilleur accès aux rapports administratifs non publiés ». *Documentaliste*, vol. 35, n° 2, p. 97-100.

DUARTE C.L. (2003). « TIC et documentation scolaire ». *Perspectives Documentaires en Éducation*, n° 59, p. 51-55.

FARACE D.J., FRANTZEN J., SCHÖPFEL J., STOCK C. et BOEKHORST A.K. (2006). « Access to Grey Content : An Analysis of Grey Literature Based on Citation and Survey Data ». In : D. FARACE et J. FRANTZEN (éd.). *GL7 Conference Proceedings. Seventh International Conference on Grey Literature : Open Access to Grey Resources*. Nancy, 5-6 December 2005. Amsterdam : TextRelease, p. 194-203.

FILY M.F. (2005). « Introduction au concept d'archive ouverte ». <http://archivesic.ccsd.cnrs.fr>

GLISC – GREY LITERATURE INTERNATIONAL STEERING COMMITTEE (2006). *Guidelines for the production of scientific and technical reports : how to write and distribute grey literature*. Rome : Istituto Superiore di Sanità.

JEFFERY K.G., LOPATENKO A. et ASSERSON A. (2002). « Comparative study of metadata for scientific information : The place of CERIF in CRISs and scientific repositories ». In : W. ADAMCZAK et A. NASE (éd.). *Proceedings CRIS 2002 6th International Conference on Current Research Information Systems*. Kassel : University Press, p. 77-86.

KELLER A. (2001). « Future development of electronic journals : a Delphi survey ». *The Electronic Library*, vol. 19, n° 6, p. 383-396.

LANDESMANN M. (2003). « Identités et trajectoires universitaires de professeurs mexicains ». *Perspectives Documentaires en Éducation*, n° 59, p. 9-18.

LARDY J.P. (2006). « HAL – Hyper Articles en Lignes ». Lyon : URFIST. <http://urfist.univ-lyon1.fr/hal-en-profondeur.pdf>

LE HENAFF D. et THIOLON C. (2005). « Gérer et diffuser des thèses électroniques : un choix politique pour un enjeu scientifique ». *Documentaliste*, vol. 42, n° 4-5, p. 272-280.

LE ROUX L. (2005). « Circulation de l'information et « navigation communautaire chez les enseignants-documentalistes ». In : *Colloque du SIF « Les institutions*

éducatives face au numérique ». Paris, Carré des Sciences, 12-13 décembre 2005.

LÜDKE M. (2003). « Une sociologue brésilienne au croisement de diverses cultures de recherche ». *Perspectives Documentaires en Éducation*, n° 59, p. 19-28.

MACKENZIE OWEN J. (1997). « The Expanding Horizon of Grey Literature ». *GL3 Conference Proceedings. Third International Conference on Grey Literature*. <http://eprints.rclis.org/archiv/00002596>.

NDOYE A.K. (2003). « Les déterminants du rendement scolaire des filles en Afrique subsaharienne : une revue de littérature ». *Perspectives Documentaires en Éducation*, n° 59, p. 97-113.

OKIY R.B. (2003). « A citation analysis of Éducation dissertations at the Delta State University, Abraka, Nigeria ». *Collection Building*, vol. 22, n° 4, p. 158-161.

PAILLASSARD P., SCHÖPFEL J. et STOCK C. (2005). « How to get a French doctoral thesis, especially when you aren't French ». *Publishing Research Quarterly*, vol. 21, n° 1, p. 73-93.

PREVOT M. (2005). « La publication scientifique à accès libre : de l'idéal aux modalités concrètes. Application aux sciences de la terre ». *Bulletin de Liaison de la Société Française de Minéralogie et Cristallographie*, vol. 17, n° 2, p. 23-30.

RAMALHO CORREIA A.M. et DE CASTRO NETO M. (2002). « The role of eprint archives in the access to, and dissemination of, scientific grey literature: LIZA – a case study by the National Library of Portugal ». *Journal of Information Science*, vol. 28, n° 3, p. 231-241.

ROOSENDAL H. E. (2004). « Driving change in the research and HE information market ». *Learned Publishing*, vol. 17, n° 1, p. 1-6.

SCHÖPFEL J. (2006). « MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE ». In : D. FARACE et J. FRANTZEN (éd.). *GL7 Conference Proceedings. Seventh International Conference on Grey Literature : Open Access to Grey Resources*. Nancy, 5-6 December 2005. Amsterdam : TextRelease, p. 34-39.

SCHÖPFEL J., STOCK C., FARACE D.J. et FRANTZEN J. (2005). « Citation Analysis and Grey Literature : Stakeholders in the Grey Circuit ». *The Grey Journal*, vol. 1, n° 1, p. 31-40.

SONDERGAARD T. F., ANDERSEN J. et HJORLAND B. (2003). « Documents and the communication of scientific and scholarly information. Revising and updating the UNISIST model ». *Journal of Documentation*, vol. 59, n° 3, p. 278-320.

STOCK C. et SCHÖPFEL J. (2004). « Grey Literature in an Open Context : From Certainty to New Challenges ». In : D. FARACE et J. FRANTZEN (éd.). *GL5 Conference Proceedings. Fifth International Conference on Grey Literature : Grey Matters in the World of Networked Information*. Amsterdam, 4-5 December 2003. Amsterdam : TextRelease, p. 199-202.

STOCK C., ROCKLIN E. et CORDIER A. (2006). « LARA : Open Access to Scientific and Technical Reports ». In : D. FARACE et J. FRANTZEN (éd.). *GL7 Conference Proceedings. Seventh International Conference on Grey Literature : Open Access to Grey Resources*. Nancy, 5-6 December 2005. Amsterdam : TextRelease, p. 87-93.

ULLAH M.F., KANWAR S.S. et KUMAR P. (2004). « A quantitative analysis of citations of research reports published by National Institute of Hydrology, Roorkee ». *Annals of Library and Information Studies*, vol. 51, n° 3, p. 108-115.

WOOD D.N. et SMITH A.W. (1993). « SIGLE : A Model for International Co-operation ». *Interlending & Document Supply*, vol. 21, n° 1, p. 18-22.

Tous les sites ont été visités entre janvier et mars 2006.

Nous remercions Chérifa Boukacem-Zeghmouri (Lille-3) pour ses commentaires critiques lors de la rédaction de cette étude.