

documentation

La recherche sur internet de documents de littérature grise en santé publique

CHANTAL RONDET-GRELLIER
DOMINIQUE VILLEBRUN

La littérature grise est définie par l'Agence française de normalisation (Afnor) comme « *tout document dactylographié ou imprimé, souvent à caractère provisoire, reproduit et diffusé à un nombre d'exemplaires inférieur au millier, en dehors des circuits commerciaux de l'édition et de la diffusion* ». Dans le *Medical Subject Headings* (MESH), vocabulaire d'indexation de la base de données bibliographiques Medline¹, différentes rubriques recouvrent ce que l'on appelle la littérature grise : "Congresses", "Government publications", "Laboratory manuals", "Academic dissertations" et "Technical reports". Il s'agit donc d'une littérature à faible tirage, à caractère spécialisé, de diffusion limitée et interne à l'organisme qui la produit. Elle échappe aux circuits habituels de l'édition et n'est pas soumise au dépôt légal. Cette littérature a, de fait, toujours été difficile à localiser pour les professionnels de l'information. Dans le domaine de la santé publique, la littérature grise est principalement représentée par trois types de documents : les rapports émanant d'organismes et d'institutions, les mémoires et les thèses, et les données (chiffrées ou qualitatives) issues d'études et d'enquêtes.

RECENSEMENT ET DIFFUSION DE LA LITTÉRATURE GRISE

Les politiques de recensement et de diffusion de la littérature grise varient selon les pays.

■ Grâce aux bases de données bibliographiques, la recherche de documents primaires, tels les articles et les ouvrages, ne pose plus véritablement de difficultés aux documentalistes.

■ En revanche, toute une catégorie de documents échappe aux circuits de collecte et de diffusion classiques.

■ Ces documents, dits de "littérature grise" ont toujours été difficiles à recueillir, mais aujourd'hui de nouveaux moyens permettent aux professionnels d'y accéder.

■ En France, dans les années 90, bien avant l'introduction d'internet et des documents électroniques, les pouvoirs publics ont pris conscience de la nécessité de promouvoir l'accès à la littérature grise. En l'absence de circuits spécifiques de collecte, la question du recensement et du traitement de ces documents s'est posée. Un dispositif de signalement et de collecte auprès des administrations a alors été mis en place par la Commission de coordination de la documentation administrative (CCDA). Ces administrations ont été incitées à signaler et déposer les documents à caractère :

• **scientifique et technique** à l'Institut de l'information scientifique et technique du CNRS (Inist). Ils sont répertoriés dans les bases Pascal/Francis ;

• **administratif et politique** auprès de la Documentation française. Ces documents sont réunis dans la base "Logos" mise en place par la Banque d'information politique et d'actualité (Bipa).

Ces deux organismes ont été chargés du recensement et de la diffusion auprès du public.

Après une phase expérimentale lancée auprès de quelques admi-

nistrations volontaires, le projet a été élargi à l'ensemble des ministères. Il est devenu opérationnel et permanent en 1994 sous le nom de "Ligria" (Littérature grise des administrations). Au départ, seule une diffusion sous forme "papier" ou disquette des rapports collectés était prévue, mais l'ouverture des sites internet de l'Administration, comme celui de la Documentation française, et l'utilisation des techniques de numérisation ont permis, depuis, de mettre à la disposition du public ces documents sur le web².

■ À l'échelon européen, depuis 1980 a été créée une base intitulée "SIGLE" (*System for Information on Grey Literature in Europe*), qui rassemble la littérature grise produite dans les pays membres. Cette démarche s'inscrit dans le cadre d'un réseau européen d'agences nationales. L'Inist est l'agence responsable chargée d'alimenter cette base pour la partie française et d'attribuer l'ISRN (*International Standard Report Number*), numéro international normalisé unique (comme l'*International Standard Book Number* pour les monographies), de façon à constituer un fichier national ►

MOTS CLÉS

- Documentation
- Information
- Internet
- Littérature grise
- Recherche
- Santé publique

NOTES

1. Medline : <http://disc.vjf.inserm.fr:2010/basismesh/mesh.html>
2. Desmichel C., Pour un meilleur accès aux rapports administratifs non publiés, *Documentaliste-Sciences de l'information*, 1998 ; 35 ; 2 : 97-100
3. Office of Scientific and Technical Information : <http://www.osti.gov/bridge/>
4. Bibliothèque de la New York Academy of Medicine : <http://www.nyam.org/library/greylit/whatis.shtml>
5. NTIS : <http://www.ntis.gov/>
6. Foenix-Riou B., Recherche et veille sur le web visible et invisible. Agents intelligents, annuaires sélectifs, interfaces des grands serveurs, portails thématiques, Éd. Tec & Doc, 2001
7. Netsources, 2001, 33

► des documents des organismes producteurs.

■ **Aux États-Unis**, un guichet unique d'accès aux rapports des agences fédérales existe depuis 1990 – le *National Technical Information Service* (NTIS) –, mais comme la fourniture gratuite de documents par internet a fortement diminué ses ressources et que différentes agences fédérales ont quitté cet organisme, il n'existe plus de référencement unique et normalisé, donc plus de signalement exhaustif.

Certaines agences, comme l'*Office of Scientific and Technical Information*³ dans le domaine de l'environnement, ont pris l'option du tout électronique et mettent en ligne gratuitement le texte intégral de leurs rapports. On peut également signaler l'initiative de la Bibliothèque de la *New York Academy of Medicine*⁴, qui recense ce type de documents et fournit un accès en ligne aux fonds de tout un ensemble d'institutions dans le domaine de la santé. Quant au NTIS⁵, il continue à fonctionner de façon plus classique: le signalement est effectué sur internet, mais la fourniture du document est payante et se fait sous forme "papier", CD ou microfiche.

La situation actuelle est donc

paradoxalement, les documents disponibles grâce à internet et à la numérisation sont de plus en plus nombreux, mais ils sont, en revanche, difficilement repérables du fait de la multiplicité des sources. En conséquence, l'obstacle majeur pour les documentalistes réside dans la localisation (ou l'existence) de ces documents.

COMMENT TROUVER DES DOCUMENTS DE LITTÉRATURE GRISE SUR INTERNET ?

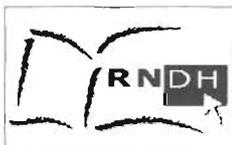
En principe, la littérature grise est difficile à trouver à l'aide des moteurs de recherche généralistes. En effet, ceux-ci n'indexent que la partie "visible" du web, qui est composée de sites en accès libre offrant des pages reliées entre elles, soit, en 2001, environ 4 milliards de pages. C'est dans l'autre partie, dit "web invisible" que l'on aura le plus de chance de trouver ce genre de littérature. Dans le "web invisible", se trouvent des pages dynamiques générées par un moteur de recherche (dans ce cas, seule la première page sera indexée par les moteurs classiques et les documents que nous recherchons se trouveront dans les pages suivantes), des sites où il est nécessaire de s'identifier

(gratuits ou non), des serveurs classiques (Dialog, Datatar, etc.), ainsi que des bases de données et des sites non référencés.

Notons que le volume d'information dans le "web invisible" est très nettement supérieur à celui du "web visible" et que cette information y est de bonne qualité et à 95 % en accès libre⁶.

Dans la méthode employée pour la recherche de ces documents, il faut tenir compte de cette spécificité. Pour remédier à ces difficultés, il est donc conseillé de procéder de la façon suivante :

- **la première démarche** consiste à visiter les sites généralistes qui recensent les rapports et les sites institutionnels de santé publique ;
- **afin d'affiner la recherche, et selon le thème étudié**, il faut ensuite rechercher les sites clés de ce domaine, où l'on trouve des rapports, des enquêtes, des thèses. Par exemple, pour les toxicomanies, on peut citer le site de la MILDIT et le site de la Fondation de la recherche sur les toxicomanies (canadien). Pour trouver ces sites fédérateurs, il est nécessaire de passer par un annuaire sélectif ou un site portail ;
- **une troisième méthode** consiste, enfin, à rechercher des pages de liens dans le domaine traité⁷. ■



Rubrique réalisée avec le Réseau national des documentalistes hospitaliers (RNDH)

LES AUTEURS

Chantal Rondet-Grellier, Centre d'Expertise Collective Inserm SC14 Paris (75),
Dominique Villebrun, Inserm U500 "Épidémiologie des maladies chroniques et du vieillissement", Montpellier (34)

■ Répertoire de sites généralistes et institutionnels

1 - Sites français recensant de la littérature grise mais ne fournissant pas de documents en texte intégral :

- **Banque de données en santé publique (BDSP)** : <http://www.bdsp.tm.fr>

Accès à des rapports, soit en associant au mot recherché le terme "littérature grise", soit en limitant la recherche à la catégorie de publications "rapport". À noter que la BDSP développe actuellement la mise en ligne de documents en texte intégral pour les personnes abonnées.

- **Article@ Inist** : <http://services.inist.fr/public/fre/consilt.htm>

Dans cette base, sont référencés des rapports, des ouvrages, des conférences, ainsi que des thèses en sciences.

- **SUDOC** : <http://corail.sudoc.abes.fr/>
Catalogue des thèses soutenues en France et en cours. À noter également, sur le site de l'INIST, <http://www.inist.fr/> la possibilité d'accéder à "Thesa", banque de données signalant les sujets de thèse en cours dans les grandes écoles.

- **FNORS** : <http://www.fnors.org/default.htm>
Liste des publications publiées par les Observatoires régionaux de la santé (ORS) depuis 1995. L'adresse de l'ORS permet de se procurer le document cité.

2 - Sites français sur lesquels l'accès à des

documents en texte intégral est possible :

- **La Documentation française** : <http://www.ladocumentationfrancaise.fr>
Bibliothèque des rapports publics (liste par ancienneté et moteur de recherche) ; collection des discours publics ; agenda des missions en cours confiées par le gouvernement.
- **Ministère de la Santé** : <http://www.sante.gouv.fr/index.htm>
Documents en texte intégral dans les rubriques "Recherche, études et statistiques" et "Publications". Voir également le site <http://www.social.gouv.fr>
- **Portail du ministère et des agences sanitaires** : <http://www.sante.fr/>

Voir en particulier le site de l'Anaes, qui contient de nombreux documents produits par cette agence (recommandations pour la pratique clinique, conférences de consensus, guides méthodologiques...).

- **Institut Pasteur** : <http://www.pasteur.fr/externe.html>. En texte intégral, livres, thèses, rapports, actes de congrès dans les rubriques "Information scientifique", "Livres en ligne".

- **Credes** : <http://www.credes.fr/>
Documents en ligne : données chiffrées, publication "Questions d'économie de la santé", supports d'intervention ou de cours, documents de travail...

- **Haut comité de la santé publique** : <http://hcsp.ensp.fr/hcspi/explore.cgi/accueil>
L'ensemble des rapports thématiques produits par le HCSP est disponible en texte intégral.

- **Inserm** : <http://www.inserm.fr>
Serveur d'information scientifique
<http://disc.vjf.inserm.fr:2010>

Des liens sont faits avec des documents trouvés sur le web. Certains rapports d'expertise collective de l'Inserm sont en texte intégral sur le serveur d'information scientifique.

- **Institut de veille sanitaire (InVS)** : <http://www.invs.sante.fr/>

Rapports et publications de l'Institut.

- **Caisse nationale d'Assurance maladie (Cnam)** : <http://www.ameli.fr>

Dans la rubrique "Assuré", annuaire des professionnels de santé ; dans la rubrique "Connaître l'assurance maladie", chiffres et repères, actualité statistique, publications médicales et études.

- **Université Lyon 2** : <http://theses.univ-lyon2.fr/> et **Université Lille 1** : <http://bibliotheques.univ-lille1.fr/grisemine>
Deux expériences de mise en ligne de thèses.

3 - Sites étrangers

- **GrayLit** : <http://www.osti.gov/graylit>
Texte intégral des fonds de littérature grise de plusieurs agences fédérales : *Department of Energy, Department of Defense, NASA...*

- **Center for Disease Control and Prevention (CDC)** : <http://www.cdc.gov/>
Enquêtes et données épidémiologiques aux États-Unis. Rapports en ligne.

- **National Institute of Health (NIH)** : <http://www.nih.gov/> (USA). Cet institut gouvernemental, équivalent du ministère de la Santé français, regroupe 27 instituts qui ont chacun leur site, sur lesquels se trouvent des rapports ou autres documents en ligne.

• Organisation Mondiale de la Santé

(OMS) : <http://www.who.int/>
Les publications de l'OMS (rapports, articles, littérature grise) sont recensées sur le site de la bibliothèque Wholis : <http://www.who.int/library/database/index.en.shtml>.

■ Annuaire et portails

1 - Francophones

- **Le Catalogue et index des sites médicaux francophones (CISMEF)** : <http://www.chu-rouen.fr/cismef/>
Il permet de rechercher des sites par index alphabétique (mots-clés du Mesh), thématique et par type de ressources (organismes et institutions). Accès au texte intégral dans le moteur de recherche avancée soit par type de ressource en cherchant les rapports techniques, soit en associant la rubrique "rapport technique" avec un mot-clé du Mesh.

- **La Bibliothèque de l'Université de Montréal** : <http://www.bib.umontreal.ca/>
Les ressources recensées sont regroupées par sujet/discipline.

- **Promedical.net** : <http://www.promedical.net/ClubPro/index.html>. Édité par une équipe constituée de médecins, professionnels de la documentation et spécialistes de projets web, ce répertoire de sites est organisé par disciplines médicales, portails généralistes, sites institutionnels, industries et métiers de la santé, établissements de soins...

- **ANASYS** : <http://www.anasys.org/>
Portail des sites de santé publique : l'Association nationale des praticiens conseils pour la promotion de l'analyse du système de santé a élaboré un guide d'utilisation des ressources de l'internet en médecine et santé publique, régulièrement actualisé.

2 - Anglophones

- **OMNI (Organising Medical Networked Information)**, portail organisé en réseau et animé par l'Université de Nottingham, <http://omni.ac.uk/>. Recherche à partir d'un moteur ; il est ensuite possible de restreindre à des sous-rubriques.

- **La WWW Virtual Library**, <http://vlib.org/>
Un des plus anciens annuaires du web créé par l'inventeur du format html. Il fonctionne en réseau et est organisé par grands domaines, puis par zones géographiques et sous-rubriques thématiques.

- **Social Science Information Gateway**

(SOSIG) : <http://sosig.ac.uk> :

Portail anglais en sciences sociales organisé par grandes rubriques. Couvre différents types de ressources : documents, bases de données et annuaire de centres de recherche et d'organismes. Lien vers le texte intégral des documents recensés.

■ Recherche de pages de liens

Il est possible en premier lieu de consulter des sites référençant des liens, comme celui de la Bibliothèque Interuniversitaire de Médecine (<http://www.bium.univ-paris5.fr/>) ou des Mutuelles de France (<http://www.mutuelles-de-france.fr>). La recherche des pages de liens se fera ensuite par l'intermédiaire de moteurs de recherche.

- **Les pages de liens** ont souvent dans leur URL le mot "bookmark", "signet", "lien", "link", et l'on pourra alors limiter dans l'écran d'accueil la recherche aux pages URL en ajoutant aux termes de la requête l'opérateur inurl : directement suivi des mots "bookmark", "signet", "lien", "link".

Exemple : *aging inurl:bookmark*. Pour Google, il s'agira de faire la recherche au singulier et au pluriel, le moteur n'utilisant pas la troncature. Dans AltaVista, pour combiner les différentes variantes du mot lien, on utilisera la grille de recherche avancée.

- **Il est possible également de limiter la recherche** aux pages contenant le thème recherché dans le texte d'un lien. Dans Google, il faudra, dans la recherche avancée, utiliser la rubrique "Emplacement" et dans Altavista, sur l'écran d'accueil, utiliser l'opérateur anchor, suivi du mot recherché.

- **Enfin, on peut, lorsque l'on connaît un site pertinent** dans le domaine concerné, récupérer les pages ayant un lien dirigeant vers cet URL qui seront, elles aussi, certainement en rapport avec l'objet de la recherche. Il faut dans ce cas utiliser l'opérateur *link* dans Google, Altavista et Alltheweb.

Dans tous les cas, il est important de visiter régulièrement les sites que l'on a recensés et de procéder à des mises à jour. Il faut signaler le site *Googlert* (<http://www.googlert.com>), service de veille de Google, qui propose l'envoi par mail, selon une périodicité décidée par l'utilisateur, des nouveaux résultats de requêtes définies au préalable par celui-ci.