



Grisemine, une bibliothèque numérique de littérature grise universitaire.

Marie-France Claerebout, Jean-Bernard Marino

► **To cite this version:**

Marie-France Claerebout, Jean-Bernard Marino. Grisemine, une bibliothèque numérique de littérature grise universitaire.. Jan 2004. sic_00000892

HAL Id: sic_00000892

https://archivesic.ccsd.cnrs.fr/sic_00000892

Submitted on 23 Jan 2004

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

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

Grisemine, une bibliothèque numérique de littérature grise universitaire

Marie-France CLAEREBOUT, Jean-Bernard MARINO

*Service commun de la documentation, Université des sciences et technologies de Lille
BP 155 – 59653 Villeneuve d'Ascq cedex (France)
marie-france.claerebout@univ-lille1.fr, jean-bernard.marino@univ-lille1.fr*

*Traduction de la communication faite au 5^e colloque international sur la littérature grise (GL5),
Amsterdam, 4-5 décembre 2003.*

RÉSUMÉ

Le glissement progressif du support imprimé vers le numérique, sous l'impulsion des éditeurs commerciaux, aboutit à des produits documentaires finis, disponibles directement sur le poste de l'utilisateur final, sans médiation. Cette évolution risque de réduire progressivement la bibliothèque universitaire à un rôle de relais passif finançant l'accès à l'information mais exclu de fait de la chaîne de traitement et de diffusion de l'information.

De sa capacité d'adaptation dépend sa survie : c'est pourquoi nous avons tenu à créer sans retard une bibliothèque numérique consacrée à la littérature grise, et baptisée Grisemine. Nous y avons transposé notre savoir-faire de professionnels de traitement de l'information soucieux de la qualité de l'indexation, du catalogage et de l'organisation des collections.

Créée fin 2001, Grisemine s'enrichit régulièrement de documents d'enseignement et de recherche : thèses, cours, communications, rapports... dans les différentes disciplines abordées par notre université.

La réaction favorable des auteurs contactés, tout comme la fréquentation croissante de notre site, (<http://bibliotheques.univ-lille1.fr/grisemine>), confirment le bien-fondé de notre démarche. Reste à atteindre une taille critique garantissant crédibilité et pérennité. Nous envisageons pour cela de développer la coopération avec d'autres acteurs engagés dans la production et la diffusion de la littérature grise.

MOTS-CLÉS

Bibliothèque numérique, littérature grise, travaux universitaires, contenus en ligne, chaîne documentaire, mutations professionnelles.

LE CONTEXTE

Depuis plus d'un siècle, notre bibliothèque universitaire s'est efforcée d'acquérir pour ses lecteurs non seulement des livres et revues commerciales mais aussi de la littérature grise imprimée. Cette dernière est constituée en grande partie de thèses (60.000 environ) provenant d'échanges nationaux et internationaux entre bibliothèques. S'y ajoutent des documents provenant d'institutions ou de chercheurs : documents de travail, actes de colloques, rapports officiels, photocopiés de cours, sujets d'examens,...

Dans un monde en évolution ...

À présent, l'électronique offre de nouveaux modes de diffusion du savoir via la Toile. Les grands éditeurs commerciaux en font un usage efficace aboutissant à la lecture directe des grandes revues de recherche sur l'écran du lecteur final.

Mais ce mouvement s'accompagne d'une inflation des coûts d'abonnements aux revues et conduit à reconsidérer le rôle de la littérature grise. C'est ainsi que nous assistons à une évolution dans les modes de diffusion de l'information savante, engagée par les auteurs et encouragée par l'usage croissant de la Toile. Avant leur parution dans les grandes revues de recherche, des articles sont proposés en avant-première à la communauté, sur des sites personnels, sur des sites d'institutions, ou dans des bases d'archives ouvertes de pré-publications.

Doit-on en conclure que, si la diffusion du savoir sous forme imprimée passe par la bibliothèque, sa diffusion sous forme numérique passe directement par le réseau, laissant la bibliothèque hors circuit ? Les nouvelles technologies vont-elles, dans le monde de la communication savante, remplacer bibliothèques et bibliothécaires ?

...transposons notre savoir-faire

Il convient, en bibliothèque universitaire, de s'interroger : sommes-nous en passe d'être exclus du processus de diffusion du savoir ? La réponse dépend de nous. Il s'agit de montrer à la communauté universitaire tout le parti qu'elle peut tirer de nos compétences documentaires, de notre savoir-faire.

Une première étape consiste à continuer de fournir l'information. Notre vocation ne change pas, que les documents se présentent sous forme imprimée ou numérique : il s'agit de collecter, cataloguer, ordonner, conserver et diffuser dans un esprit de service public. À l'image de notre bibliothèque physique située au centre du campus, la bibliothèque numérique de l'université trouve sa place au centre de l'université en ligne, en offrant les services documentaires indissociables des services d'enseignement et de recherche.

D'où le double aspect de la logistique documentaire en ligne sur le site du « service commun de la documentation » : d'une part permettre l'accès aux revues et documents électroniques commerciaux, d'autre part ouvrir l'accès à la littérature grise, sachant que cette dernière constitue un irremplaçable réservoir de connaissances pour l'enseignement et la recherche. Nos compétences professionnelles nous permettent de construire des collections organisées (classées et indexées) de littérature grise avec des outils éprouvés, soit de véritables bibliothèques numériques. C'est ce que nous avons commencé de faire avec Grisemine.

UNE PREMIÈRE ÉTAPE : GRISEMINE

Nous avons créé la bibliothèque numérique Grisemine à la fin de 2001. Conçue comme une mine de littérature grise, elle a grandi lentement car nos moyens sont limités mais sûrement, en raison de la pertinence de son contenu.

Au-delà de l'ambition d'affirmer notre présence dans les processus de diffusion du savoir, nous visons plusieurs buts :

- offrir plus de visibilité à la production universitaire de langue française ;
- développer de nouveaux fonds ;
- offrir à nos lecteurs existants – étudiants et chercheurs – un accès plus diversifié à la recherche ;
- gagner de nouveaux lecteurs hors nos murs par le biais de services distants ;
- contribuer à développer un complément gratuit à l'édition commerciale.

Les contenus en ligne

Dans cette première étape, nous nous limitons à la littérature grise en français. Le poids respectif attribué à chaque discipline et à chaque type de document est déterminé d'après notre plan de développement des collections.

Comment procédons-nous ?

Nous acceptons des textes de tous pays, dès qu'ils correspondent aux disciplines de notre université : sciences, techniques, sciences sociales.

Nous nous intéressons à trois types de documents produits par les universités, les institutions d'enseignement ou de recherche, les organismes publics :

- premièrement, des documents de recherche ayant déjà fait l'objet de communication : thèses et mémoires, communications de colloques, articles de revues (revues à faible diffusion de sociétés qui souvent ne disposent pas encore de version électronique) ;
- deuxièmement, des documents de recherche inédits : notes, documents de travail, pré-publications ;

- troisièmement, des documents d'enseignement : cours, supports de travaux dirigés, exercices, sujets d'examens.

Fonctionnalités

Grisemine répond aux critères de fonctionnement de toute bibliothèque : collecte, catalogage, indexation, conservation, organisation de la recherche et de la diffusion des textes. Notons d'ailleurs que certains documents sont à la fois disponibles en ligne et sur nos rayonnages (thèses de notre université par exemple).

Nous utilisons les outils documentaires actuels et les nouvelles technologies pour offrir le texte intégral au format électronique via une interface de recherche sur l'Internet.

L'utilisateur a ainsi la possibilité de :

- feuilleter la bibliothèque par sujet ou par discipline ;
- faire une recherche dans des champs comme le titre, les auteurs, les mots clés, la table des matières, le résumé, ...
- faire une recherche dans le texte intégral ;
- imprimer ou télécharger librement les textes.

Cette possibilité d'impression et de navigation à distance dans le texte apparaît comme l'un des nouveaux services que peut apporter la bibliothèque.

Aspects scientifiques et juridiques

Une question revient souvent : « Le texte que vous me demandez est déjà disponible sur mon site ; pourquoi donc ne pas vous contenter d'un lien ? ». Une telle approche est envisageable, mais ce n'est pas la nôtre : notre logique de bibliothécaires nous conduit à privilégier des procédures assurant un accès fiable, unifié et de qualité au contenu documentaire.

Nous procurer une copie de chaque texte pour construire notre propre bibliothèque représente certes une démarche plus complexe et dévoreuse de temps. Cependant la maîtrise de toute la chaîne est une garantie de pérennité car nous évitons le risque de perdre de précieux documents par simple rupture de liens. C'est aussi une garantie d'homogénéité et de qualité.

Chaque auteur coopérant avec Grisemine doit appartenir à une institution d'enseignement ou de recherche reconnue et ses documents sont étiquetés au nom de l'établissement. La simplicité de ce contrôle permet une diffusion rapide mais nous pouvons au besoin envisager la mise en œuvre d'une validation par comité de lecture.

Avant le traitement de tout nouveau document, nous sollicitons l'accord du détenteur des droits de reproduction et de diffusion. Dans le cas d'un texte

préalablement publié (dans le circuit commercial ou non), il faut le plus souvent s'adresser à l'éditeur, auquel l'auteur a cédé l'intégralité de ses droits. Il est surtout notable que bien des auteurs sont très peu informés des questions de propriété intellectuelle.

Aspects techniques

Nous collectons la littérature grise tant sous forme imprimée que sous forme électronique. Beaucoup de documents étant à présent disponibles sous forme numérique native, la diffusion en est facilitée à coût réduit.

Les textes sont diffusés au format PDF image + texte, et protégés contre la copie et l'extraction de contenu afin d'en garantir un usage sûr et loyal. Nous ajoutons en marge, sous forme de signets, la table des matières qui sert ainsi d'outil de navigation pour atteindre directement la partie souhaitée du texte. Grâce au format image, la version imprimée à distance est identique à l'original. Quant au texte caché, il permet la recherche plein texte dans le document.

Nous étudions également la possibilité de diffuser les textes au format HTML, sous réserve de l'autorisation des auteurs.

Description catalographique

L'accroissement du nombre de documents d'une collection impose une structuration, que la collection soit imprimée ou électronique. Notre interface de recherche sur Internet¹ dérive des outils et méthodes de description catalographique : chaque texte électronique est lié à un jeu de métadonnées balisées en MarcXML. Créé par la Bibliothèque du Congrès², MarcXML allie la rigueur du format Marc, sa pertinence à la description catalographique, aux qualités d'interopérabilité offertes par XML.

Le maintien de normes et standards internationaux dans les champs de métadonnées apporte une garantie complémentaire d'interopérabilité. Chaque texte reçoit une double indexation systématique : l'une, locale (classification par grandes disciplines), utilisée pour la recherche sur notre site et l'autre, internationale (classification Dewey), utilisable à des fins d'échanges ou de mise en commun de métadonnées.

De même, dans un double souci de pertinence des résultats et d'interopérabilité, nous usons de mots-matières de type Rameau. Pour élargir l'éventail de termes pertinents, la notice descriptive de chaque texte comprend le résumé et les parties principales de la table des matières, et ces champs sont utilisés par l'outil de recherche.

¹ Littérature grise en texte intégral H<http://bibliotheques.univ-lille1.fr/grisemineH>

² MarcXML H<http://www.loc.gov/standards/marcxml/H>

Dans le cas où le texte à été repris d'un autre site Internet, nous incluons l'adresse d'origine dans le signalement. Par souci de coopération, nous prévoyons également un lien vers ce site.

PROSPECTIVE

Le retour d'expérience de Grisemine

Le bilan, au bout de deux ans, est positif. Certes, la croissance de Grisemine est lente : d'une part les effectifs du projet sont réduits, d'autre part le temps de collecte et surtout de traitement sont longs du fait de la richesse des métadonnées saisies. Le rythme d'accroissement moyen est de 30 textes par mois et la base atteint un volume d'un peu plus de 750 documents. Malgré ces chiffres encore modestes, la fréquentation du site augmente régulièrement : une centaine d'accès par jour à la fin de 2003.

Une importante communication a été nécessaire pour faire connaître notre site. À présent, les utilisateurs viennent plus spontanément le consulter, parfois même les textes sont consultés directement depuis d'autres sites qui les signalent. Nous constatons par ailleurs que les accès ne proviennent pas uniquement de pays francophones, puisqu'en 2003 des connexions ont été captées depuis 74 pays différents. Ceci nous encourage à collecter et à diffuser à terme dans d'autres langues sous-représentées sur la Toile, dans un souci d'équitable diversité linguistique...

Qui sont nos utilisateurs et que recherchent-ils ?

Il faut noter que les utilisateurs n'appartiennent pas tous à des institutions d'enseignement ou de recherche. Nous avons été contactés par plusieurs entreprises en quête de littérature grise universitaire dans certaines sous-disciplines bien particulières.

Nous observons également un vif intérêt pour les cours que nous mettons en ligne : ils représentent 11% du fonds mais plus de 30% des connexions. C'est pourquoi nous avons dernièrement révisé notre plan de développement en vue d'accroître la proportion de cours dans la bibliothèque numérique.

Et les auteurs ?

Les auteurs contactés acceptent en général de bonne grâce de nous confier leurs textes, même si leur accord n'est parfois pas suivi d'effet et nécessite un rappel. D'où un volant non négligeable d'heures de travail consacrées aux contacts, pointages et relances.

D'autre part, une nouvelle tendance se fait jour depuis quelques mois : de plus en plus d'auteurs nous proposent spontanément leurs textes. Nous voyons là un bon indicateur de crédibilité et de reconnaissance.

Se rapprocher des auteurs

Le bibliothécaire voit ses tâches évoluer et se diversifier : la structuration des contenus constitue un nouveau défi qui l'amène à intervenir en amont de la chaîne documentaire. Bien que la littérature grise soit par essence hétérogène, certains documents peuvent être calibrés d'une façon homogène dès leur création, en particulier les thèses, mémoires, communications de colloques, rapports.

Notre université est l'une des premières en France à s'engager dans le projet national de structuration des thèses à des fins de diffusion électronique et de conservation³. Le but est d'étiqueter le texte de façon à permettre une traduction automatique dans un format XML où seront identifiés chaque élément de la page de titre et chaque partie composant le texte.

Les doctorants se voient donc recommander l'usage d'une feuille de style. La bibliothèque se charge, en coordination avec l'école doctorale, de les y former dès le début de la rédaction et de les assister tout au long de la mise en forme du texte.

Quel intérêt pour les auteurs ?

L'usage de feuilles de style est d'ores et déjà requis par nombre d'éditeurs et d'organismes de colloques, il est bon que les jeunes chercheurs y soient formés au plus tôt. De plus, la structuration de la thèse fournit une version de base de bonne qualité qui permettra une lecture aisée en ligne, ainsi que l'archivage dans un format durable. Elle facilite l'identification de métadonnées et l'exploitation ultérieure de parties du texte.

Collaborer avec tous les acteurs

Simultanément à notre démarche, d'autres projets de diffusion de littérature grise savante sont menés, souvent par les chercheurs eux-mêmes. Il demeure que la gestion de collections importantes de contenus électroniques correctement indexés et donc se prêtant à la recherche documentaire suppose une coopération des différents acteurs.

Avant tout, les producteurs de documents doivent prendre conscience de leur appartenance à une communauté dépassant les limites de leur établissement : c'est par l'échange de textes et/ou de métadonnées qu'ils pourront constituer des bibliothèques de taille critique ou des entrepôts de métadonnées. Tous les acteurs du secteur public et non commercial doivent y trouver avantage : bibliothèques, composantes de l'université, chercheurs, sociétés savantes, éditeurs institutionnels, presses universitaires, services d'enseignement à distance, ...

³ Détails sur le site de la sous-direction des bibliothèques
[Http://www.sup.adc.education.fr/bib/H](http://www.sup.adc.education.fr/bib/H)

Pour le bibliothécaire, rien de plus familier que cette démarche dans la droite ligne de pratique fort anciennes (échanges nationaux et internationaux, catalogues collectifs). C'est pourquoi dans une prochaine étape Grisemine sera adapté pour se conformer au protocole standardisé des archives ouvertes, OAI-PMH⁴.

L'archivage institutionnel constitue dans certains établissements une première approche intéressante permettant la diversification des services d'une bibliothèque : beaucoup d'universitaires ressentent le besoin de rattachement collectif de l'information produite dans leurs laboratoires et concluent à la nécessité de se doter d'un entrepôt numérique pour archiver et diffuser leurs productions. La bibliothèque intervient ici pour assurer non seulement la qualité et la normalisation de l'archivage mais également le prolongement de la démarche à l'échelle d'un réseau national et international. Son rôle de fournisseur d'information se trouve ainsi enrichi.

Une intervention plus active en amont (production), en cours (traitement, métadonnées), en aval (diffusion) nous assigne une nouvelle place dans le cycle de l'information, celui d'un prestataire de service à valeur ajoutée.

De quoi être optimiste...

⁴ The Open Archives Initiative Protocol for Metadata Harvesting
<http://www.openarchives.org/OAI/openarchivesprotocol.html>