



HAL
open science

La valorisation des mémoires d'étudiants sur une archive ouverte institutionnelle : un fonds et un contexte spécifiques

Vivien Mann

► To cite this version:

Vivien Mann. La valorisation des mémoires d'étudiants sur une archive ouverte institutionnelle : un fonds et un contexte spécifiques. domain_shs.info.docu. 2010. mem_00502624

HAL Id: mem_00502624

https://memic.ccsd.cnrs.fr/mem_00502624v1

Submitted on 15 Jul 2010

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.



Mann Vivien

Master 1, mention ICD
(Information communication Documentation)

Mémoire de stage

Mission effectuée du 19 avril au 31 mai 2010
à l'UFR IDIST, Université Lille 3 Charles-de-Gaulle

La valorisation des mémoires d'étudiants sur une archive ouverte institutionnelle : un fonds et un contexte spécifiques

Mission :

Dépôt des mémoires des anciens étudiants sur MemSIC et analyse des besoins pour la mise en place d'une archive ouverte propre à l'UFR IDIST

Sous la direction de :

Joachim Schopfel (tuteur universitaire)
Damien Besegher (tuteur professionnel)

Soutenu le 21 juin 2010 à l'UFR IDIST
Université Lille 3 Charles de Gaulle (campus Pont de Bois)
BP 60 149,59653 Villeneuve d'Ascq Cedex

Remerciements :

Je tiens à remercier Damien Besegher et Joachim Schöpfel pour leur accueil et l'aide qu'ils m'ont apporté durant ce stage. Je remercie également toute l'équipe du secrétariat de l'UFR IDIST pour sa disponibilité.

Préambule

Notre mission a consisté en le dépôt rétrospectif des mémoires des anciens étudiants de l'IDIST sur MemSIC : une plateforme nationale dédiée à l'auto-archivage des étudiants au niveau Master en Sciences de l'Information et de la Communication.

Par ailleurs, nous avons du réaliser une analyse des besoins de l'Université, de l'UFR IDIST et des étudiants pour la mise en place d'une archive ouverte propre à l'UFR. Celle-ci comprend une description formalisée des objectifs et publics visés par le projet, une analyse de l'existant, et une série de recommandations techniques. A ce jour, nous n'avons pas fini sa rédaction mais la joindronss en annexe à ce mémoire. Vous pourrez vous y reporter pour obtenir toute information complémentaire sur le détail du projet, ou sur la description des dispositifs abordés.

Ce mémoire propose une réflexion théorique destinée à mettre en valeur la spécificité du fonds qu'entend diffuser l'UFR : des mémoires de stage, et la spécificité d'un projet de mise en place d'archive ouverte en contexte institutionnel.

Afin de le montrer, nous chercherons dans une première partie à caractériser les mémoires de Master au regard de la « littérature grise », en montrant notamment qu'ils répondent à des besoins informationnels spécifiques. Nous nous intéresserons ensuite aux archives ouvertes comme instrument pour la valorisation d'un tel patrimoine, en insistant sur la dimension institutionnelle d'une telle entreprise. Le projet de l'UFR étant encore au stade d'expérimentation, ce sera l'occasion de présenter un certain nombre d'options possibles tenant aux procédures de dépôt comme au développement technique.

Table des matières

| | |
|---|----|
| Introduction..... | 7 |
| Présentation de la structure..... | 7 |
| Le projet de l'UFR IDIST..... | 8 |
| Description de la mission..... | 9 |
| Tâches effectuées..... | 9 |
| Première partie : Les mémoires d'étudiants, une littérature grise à valoriser..... | 11 |
| 1 - La littérature grise : définitions, problématiques, enjeux..... | 12 |
| 1.1 - Définitions..... | 12 |
| 1.2 - Une masse informationnelle hétérogène..... | 13 |
| 1.2.1 - Hétérogénéité des types de documents..... | 13 |
| 1.2.2 - Hétérogénéité des sources productrices..... | 14 |
| 1.3 - Quelques chiffres..... | 14 |
| 1.3.1 - Volumétrie..... | 14 |
| 1.3.2 - Impact..... | 15 |
| 1.4 - Un contrôle bibliographique en cours d'élaboration mais encore insuffisant..... | 15 |
| 1.4.1 - Le contrôle bibliographique..... | 15 |
| 1.4.2 - Les enjeux de la normalisation..... | 16 |
| 1.4.3 - La normalisation : un processus lent et difficile..... | 16 |
| 1.4.5 - TEI, TEF, « Nancy Style », OAI-PMH et Dublin Core..... | 17 |
| 1.5 - De la nécessité d'une démarche de valorisation structurée..... | 18 |
| 2 - Les mémoires d'étudiants : une littérature grise spécifique..... | 19 |
| 2.1 - Statuts académiques et juridiques..... | 19 |
| 2.1.1 - Au niveau académique : une hésitation terminologique..... | 19 |
| 2.1.1 - Des œuvres protégées par le droit d'auteur..... | 19 |
| 2.1.3 - Des œuvres collectives ?..... | 20 |
| 2.2 - Une dissémination insuffisante des exemplaires imprimés..... | 21 |
| 2.3 - Une production marginale dans les archives ouvertes..... | 22 |
| 2.4 - Particularités par rapport à la production universitaire « grise » de 3ème cycle..... | 23 |
| 2.5 - Des ressources répondant à des besoins informationnels particuliers..... | 25 |
| 2.5.1 - Un questionnaire pour collecter des données..... | 25 |
| 2.5.2 - Hypothèses de départ..... | 25 |
| 2.5.3 - Synthèse des résultats..... | 26 |
| 3 – L'expérimentation de l'UFR IDIST..... | 28 |
| 3.1 - Des ressources documentaires inexploitées..... | 28 |
| 3.2 - Une politique de sélection qualitative..... | 29 |
| 3.2.1 - Qualité ou « ranking » ?..... | 29 |
| 3.2.2 - Le rôle des procès verbaux..... | 30 |
| 3.3 - Numérisation et dépôt rétrospectifs..... | 30 |
| 3.4 - Les métadonnées : signalement et amplification..... | 33 |
| 3.5 - Conclusion..... | 33 |

| | |
|---|----|
| Deuxième partie : les archives ouvertes, un instrument pour la mise en accès et en visibilité des mémoires de Master..... | 35 |
| 1 - Les archives ouvertes : état des lieux..... | 36 |
| 1.1 - « Green Road » et « Open Access » : principes fondateurs..... | 36 |
| 1.2 - Le modèle de la Communication Scientifique Directe..... | 37 |
| 1.3 - Définition et Typologie des archives ouvertes | 37 |
| 1.3.1 - Un terme ambiguë | 37 |
| 1.3.2 - Typologie des archives ouvertes..... | 38 |
| 1.4 - Chiffres et tendances..... | 39 |
| 1.4.1 - Un succès à pondérer..... | 39 |
| 1.4.2 - L'essor des archives institutionnelles..... | 40 |
| 1.5 - La centralisation : spécificité du modèle français (rôle du CCSD)..... | 41 |
| 2 - Le projet de l'IDIST : des procédures à formaliser..... | 43 |
| 2.1 - Auto-archivage ou dépôt institutionnel : un choix organisationnel..... | 43 |
| 2.2 - Quel circuit pour la modération ?..... | 44 |
| 3 - Le projet de l'UFR : perspectives de développement..... | 45 |
| 3.1 - Les contraintes de l'environnement : sécurité et continuité institutionnelle | 45 |
| 3.2 - Arguments en faveur du dépôt sur une plateforme nationale (MemSIC)..... | 47 |
| 3.3 - Temps court / temps long : deux démarches opposées mais complémentaires..... | 49 |
| 3.5 - L'intégration partielle au S.I : un compromis pragmatique ?..... | 51 |
| 3.6 - Quelle marge de manœuvre pour la personnalisation ?..... | 52 |
| Bibliographie..... | 55 |
| Annexes..... | 56 |
| Synthèse de l'inventaire..... | 57 |
| Méthodologie du questionnaire | 59 |
| Le questionnaire : « pratiques de consultation et de diffusion des mémoires des étudiants M1 et M2 »..... | 61 |
| Les résultats..... | 66 |

Introduction

Présentation de la structure

L'UFR IDIST (Information Documentation Information Scientifique et Technique) forme des professionnels de la documentation depuis plus de 20 ans. Son offre de formation va de Bac +2 (DEUST) à Bac +8 (Doctorat). Dans le cadre de notre mission, les parcours dont nous avons diffusé les mémoires sont :

1. Master 1 ICD (Information Communication Documentation) ;
2. Master 2 ICD, parcours GIDE (Gestion de l'Information et du Document en Entreprise), IDEMM (Ingenierie Documentaire, Edition et Médiation Multimédia), et PRISME (Produits de l'Information Spécialisée et Médiation Electronique) ;
3. Master 2 rSIC (Recherche en Sciences de l'Information et de la Communication) ;
4. DESS SID (ancien Master ICD – jusqu'en 2003-2004).

L'UFR a acquis une solide légitimité, fondée, entre autre, sur les arguments suivants :

1. Des parcours professionnels en phase avec l'émergent ;
2. Un fort taux de réussite aux examens ;
3. Un fort taux d'insertion à la sortie des formations ;
4. De multiples partenariats avec des organismes professionnels régionaux, nationaux et internationaux qui offrent aux étudiants des perspectives de stage ou d'embauche.

Tous les Masters de l'IDIST imposent aux étudiants la rédaction d'un livrable de fin d'année : principalement des mémoires de stages. Ceux-ci sont le reflet de la qualité des enseignements, ainsi que des connaissances théoriques et des compétences techniques acquises par les étudiants pendant l'année. Aussi, leur diffusion la plus large possible est susceptible d'intéresser de nombreux publics, au premier rang desquels les auteurs eux-mêmes.

Le projet de l'UFR IDIST

Aujourd'hui, les mémoires de Master de l'IDIST sont difficilement accessibles. Plus de 1000 exemplaires imprimés sont stockés en salle B5-113 parmi d'autres documents sans faire l'objet d'aucune valorisation spécifique.

Afin de rendre cette production accessible, l'UFR souhaite diffuser les mémoires sous forme électronique au sein d'une archive ouverte. En effet, de nombreuses études ont montré l'intérêt de telles plateformes pour accroître, si ce n'est décupler, la visibilité et l'impact de la « littérature grise ». A terme, le projet mise également sur l'auto-archivage des étudiants : les auteurs devront eux-mêmes déposer leurs productions.

Par ce projet, l'UFR poursuit les objectifs suivants¹ :

1. Valoriser des ressources intéressantes ou utiles, aujourd'hui invisibles et inaccessibles ;
2. Valoriser le parcours universitaire des auteurs, en misant sur l'impact positif d'une telle démarche à l'insertion professionnelle ;
3. Accroître la visibilité de son offre de formation auprès des étudiants potentiels ;
4. Renforcer le label-qualité de l'établissement auprès des partenaires institutionnels et professionnels.

Le projet est encore en phase expérimentale. Le dépôt rétrospectif des mémoires des anciens étudiants sur MemSIC doit être conçu comme un test. Celui-ci permettra d'effectuer les observations nécessaires sur les procédures de dépôt et les usages de consultation des fichiers afin de concevoir une archive propre à l'IDIST. Rien n'a encore été décidé quant à la nature et à la forme de cette archive : HaL, MemSIC ou DUMAS, simple « guichet » vers une plateforme nationale ou archive locale.

L'expérience menée par l'IDIST intéresse néanmoins la Présidence de l'Université de Lille 3 qui souhaite également développer une archive ouverte étendue à l'ensemble des composantes de formation. Mais, là encore, rien n'a été décidé quant aux aspects techniques de la solution, ni même aux procédures de dépôts. A terme, l'IDIST pourrait donc jouer le rôle de « tête de pont » sur ce projet.

¹ Pour toute précisions sur les objectifs du projet, se référer à l'analyse des besoins

Description de la mission

Notre mission s'est articulée autour de trois axes :

1. Dépôt rétrospectif des mémoires sur MemSIC ;
2. Rédaction d'une analyse des besoins préparatoire au projet de l'UFR ;
3. Participation à un wiki collaboratif animé par Joachim Schopfel, Sophie Chauvin et Gabriel Gallezot à l'occasion d'une communication universitaire sur le thème des mémoires de Master dans les archives ouvertes².

Tâches effectuées

Le dépôt des fichiers a nécessité l'inventaire - sélection des mémoires susceptibles d'être diffusés, selon 4 critères:

1. Une note supérieure ou égale à 14/20 ;
2. Autorisation à diffusion sur le web de l'étudiant auteur ;
3. Autorisation à diffusion sur le web du Jury ;
4. Autorisation à diffusion sur le web du lieu de stage.

Sur 298 mémoires répondants à ces critères, nous en avons retenu 168, pour lesquels nous possédions les adresses mails des auteurs³. Nous avons ensuite procédé à la collecte par mail des exemplaires au format « .pdf ». Cette étape a nécessité un travail d'identification des sources : différents « annuaires des anciens » constitués par des étudiants ou des enseignants dans le cadre de projet. Afin de faciliter la gestion du projet, nous avons également réalisé une petite base de données relationnelle avec l'outil OpenBase.

La collecte par mail ayant obtenue un très faible taux de réponses, nous avons procédé à la numérisation rétrospective d'une cinquantaine de mémoires. Comme nous le verrons, cette étape a posé des problèmes qu'il nous faut encore corriger en réalisant différents traitements sur les fichiers. A ce jour, une trentaine de mémoires ont été déposés mais nous n'avons pas encore terminé notre

² Préparation au colloque du CIUEN 2010, Strasbourg.

³ Synthèse de l'inventaire disponible en annexe

mission. A terme, entre 50 et 70 mémoires devraient être déposés. La « masse critique » nécessaire pour le test sera atteinte.

Dans le cadre de la contribution au Wiki universitaire et pour la rédaction de l'analyse des besoins, nous avons procédé à un travail de formalisation des objectifs et publics visés par le projet. De même, nous avons effectué une analyse de l'existant menée sur 10 sites français et internationaux contenant des mémoires de Master, en portant notre attention sur les procédures de dépôt, les métadonnées, les fonctionnalités de recherche et les statistiques d'usages⁴

⁴ Grille d'analyse recommandée par Joachim Schopfel, notre tuteur universitaire

*Première partie : Les mémoires d'étudiants,
une littérature grise à valoriser*

1 - La littérature grise : définitions, problématiques, enjeux

1.1 - Définitions

On trouve de nombreuses définitions de la « littérature grise ». L'AFNOR nous propose la suivante :

« [Tout] document dactylographié ou imprimé, produit à l'intention d'un public restreint, en dehors des circuits commerciaux de l'édition et de la diffusion, et en marge des dispositifs de contrôle bibliographique ». Ainsi des thèses, des rapports de recherche, des brevets, des actes de colloques, et des mémoires d'étudiants.

Selon cette définition, la littérature grise présente donc les caractéristiques suivantes :

1. Un manque de « contrôle commercial »⁵, qui pose la question de sa visibilité.

La « littérature grise », ou « non-conventionnelle », par opposition à la « littérature blanche » n'est pas distribuée dans les circuits habituels de l'édition scientifique. Elle fait donc l'objet d'une diffusion alternative, au sein d'archives ouvertes, de portails ou de bases de données spécialisées.

Notons cependant que la frontière entre ces deux mondes, « la littérature grise » et la « littérature blanche », est de moins en moins tranchée. Certaines thèses ou conférences sont publiées par des éditeurs commerciaux en tant que monographies ou dans des revues. Inversement, les éditeurs commerciaux proposent de plus en plus des services tournés vers ou fondés sur le Libre Accès, complémentaires à leur offre payante.

2. Un manque de « contrôle bibliographique » qui pose la question de son accès.

Mal signalée dans les catalogues, la littérature grise est difficilement identifiable et localisable lors d'une recherche d'information. Ainsi peut on dire, à la suite de Jacques Keriguy, que « la production et la diffusion de la littérature grise entraîne des difficultés presque insurmontables à toutes les

5 Schöpfel. Voir bibliographie entrée [12]

étapes de la chaîne documentaire »⁶.

3. Une vocation non-marchande (bien que certains services proposent un accès payant).
4. Une préservation au long terme non assurée : la littérature grise n'est pas soumise à l'obligation de dépôt légal⁷.
5. Une diffusion à l'intention d'un « public restreint », ce qui ne signifie pas qu'elle ne puisse pas être utile à des publics élargis (le rapport de l'expert sert d'abord d'aide à la décision pour le décideur, mais aussi à l'étudiant pour son travail).

La définition « de Luxembourg » (3ème Conférence internationale sur la littérature grise en 1997), souligne quant-à-elle les principales sources de productions de cette production :

« [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 ».

1.2 - Une masse informationnelle hétérogène

Par son hétérogénéité, la littérature grise résiste à une typologie exhaustive.

1.2.1 - Hétérogénéité des types de documents

Du point de vue des types de documents, thèses, conférences, rapports et travaux de recherche universitaires (« *pre-print* ») sont les plus souvent cités pour la caractériser. Néanmoins, une catégorie générale recouvre souvent des types de documents très diverses. Ainsi, pour les « rapports » : rapports institutionnels, rapports techniques, rapports de recherche, rapports d'activités, rapports de projet, etc.

En outre, la « littérature grise » englobe des documents plus volatiles et éphémères que l'on rassemble parfois sous l'appellation de « matériaux gris » (« *grey content* ») : documents de travail

⁶ Voir bibliographie entrée [10]

⁷ Du moins, en France. Voir par exemple le projet hollandais DARE (p43)

(« *working papers* »), supports de cours (« *learning objects* »), résultats de recherche « bruts » (« *data sets* », « *raw datas* »)⁸.

Plusieurs éléments laissent penser que la littérature grise sera de plus en plus hétérogène, en particulier le développement du web de données et des e-sciences, marqué par l'émergence de cyber-infrastructures destinées à stocker et diffuser les résultats de recherche brutes (visualisations, données satellites, résultats de computations et autres « *raw datas* »).

1.2.2 - Hétérogénéité des sources productrices

Du point de vue des sources productrices, la diversité prévaut aussi. La littérature grise est facilement assimilée à l'ensemble de la production non publiée émanant des établissements d'enseignements supérieurs et de recherche (actes de colloques, thèses, « *pre-print s* », etc.). En réalité, c'est l'arbre qui cache la forêt. L'ensemble du secteur public (administrations, ministères, instituts, organismes publiques et para-publiques) ainsi que le secteur industriel sont les plus gros producteurs de « matériaux gris », sous la forme privilégiée de rapports ou de « *working papers* »⁹.

Derrière chaque document se trouvent donc des modes de financement, de validation, de diffusion et des finalités très différentes. Aussi, le traitement documentaire de la littérature grise résiste à une approche globale. Dans l'idéal, chaque type de document nécessiterait des procédures de signalement, de cataloguage et diffusion spécifique.

1.3 - Quelques chiffres

1.3.1 - Volumétrie

La diversité et le mauvais signalement de la « littérature grise » compliquent l'analyse de son impact et sa volumétrie. Néanmoins, tout porte à croire qu'elle tient une importance considérable dans la production scientifique.

Ainsi, une étude quantitative menée en 2009 par l'Université Lille 3 et l'INIST sur 150 archives

⁸ Typologie utilisée par l'annuaire international d'archives ouvertes *OpenDOAR*.

⁹ Schöpfel. Voir bibliographie entrée [12]

ouvertes a montré que 74 % des dispositifs interrogés contiennent au moins un type de littérature grise¹⁰. En outre, les auteurs évaluent à 17 % la part de littérature grise sur l'ensemble des dépôts réalisés au sein de ces plateformes. Quant aux types, l'enquête révèle que les documents les plus nombreux sont les conférences (157, 257 articles déposés), puis les thèses (70,488 articles), suivies des rapports (36,186 articles). Enfin, une part importante des dépôts n'est pas « typée » et difficilement identifiable (43,308 dépôts). Il s'agit souvent de documents souffrant d'un contrôle bibliographique insuffisant.

1.3.2 - Impact

On mesure l'impact d'un article scientifique à son nombre de citations. La littérature grise n'étant pas publiée, un moyen pour y parvenir consiste à analyser les statistiques de consultation – téléchargement des ressources grises au sein d'archives ouvertes contenant aussi des ressources « conventionnelles ». Celles-ci montrent que les ressources grises, tout types confondus, y sont davantage consultées que les ressources blanches.

Ainsi, toujours selon la même enquête, les statistiques d'usages de l'archive IFREMER pour l'année 2009 montrent que les articles sont en moyenne téléchargés 7 fois par mois, contre 33 fois pour les thèses. De même, l'archive de l'école d'ingénieur de Toulouse OATAO (Open Archive Toulouse Archive Ouverte) montre qu'un article est en moyenne consulté 49 fois, contre 109 fois pour une thèse. Enfin, les statistiques de RePEc (Research Papers in Economics) montrent que les « *working papers* » sont 2 à 7 fois plus consultés que les articles de revues.

1.4 - Un contrôle bibliographique en cours d'élaboration mais encore insuffisant

1.4.1 - Le contrôle bibliographique

Selon Michèle Hudon, le « contrôle bibliographique » est « l'ensemble des opérations qui visent la création et la maintenance d'un système d'enregistrement adéquat de tous les types de documents publiés ou non, imprimés, audiovisuels ou numériques, qui enrichissent la somme des connaissances humaines et de l'information. L'inventaire ainsi réalisé permet de distinguer ces documents les uns des autres et d'en faciliter le repérage. »¹¹

10 Schopfel-Prost. Voir bibliographie entrée [11]

11 Voir bibliographie entrée [09]

En contexte numérique, le contrôle bibliographique soulève la question des métadonnées employées pour la description des œuvres, et leur normalisation.

1.4.2 - Les enjeux de la normalisation

Au support papier, les données bibliographiques assurent l'identification et la localisation des œuvres depuis un catalogue. Dans les fiches signalétiques, l'accent est donc surtout porté sur les « métadonnées descriptives » (bibliographiques). Les technologies numériques permettent une reproduction quasi immédiate et potentiellement infinie des œuvres. Aussi, la description juridique des œuvres (« métadonnées de gestion ») compte désormais autant que leur description bibliographique. Celles-ci garantissent la traçabilité des œuvres et expriment les licences d'exploitation. Au final, les métadonnées de gestion doivent répondre aux questions suivantes : qui est l'auteur ? Qui sont les ayants droits ? Quels sont les termes de la licence ? Qui héberge la ressource ?

Or, il n'existe pas à ce jour de format bibliographique standard pour la description de la littérature grise. Aussi, « le repérage et l'identification des documents gris, tout comme leur accès, sont les fonctions qui pâtissent le plus d'une absence de normalisation »¹².

1.4.3 - La normalisation : un processus lent et difficile

Différents éléments compliquent les processus de normalisation des formats bibliographiques.

En premier lieu, la grande diversité des types de documents et des contextes d'usages. Face aux cas particuliers, les normes internationales s'avèrent souvent incomplètes ou inadaptées. Ainsi, certaines spécifications sont des variantes de normes plus générales adaptées à des types de documents et des contextes spécifiques (ex : la TEF, déclinaison de la TEI). D'autres sont des créations destinées à un type de document particulier (ex : la « Nancy Style »)¹³.

Ensuite, le caractère « [universel et ubiqué] »¹⁴ de la littérature grise rend difficile l'harmonisation

12 Comberousse. Voir bibliographie, entrée [03]

13 Voir *infra*. 1.4.5 - TEI, TEF, « Nancy Style », OAI-PMH et Dublin Core...

14 Schöpfel, Voir bibliographie entrée [12]

des recommandations à l'échelle internationale. Instances de normalisation et initiatives nationales peinent à se coordonner. Preuve en est l'abandon de la norme ISO pour le catalogage des rapports en 2005 (ISRN - International Standard Report Number - ISO 10444:1994)¹⁵.

Enfin, les organismes producteurs de littérature grise sont réticents à appliquer ces recommandations dans la chaîne de signalement car leur mise en œuvre est souvent fastidieuse, sinon coûteuse, alors que la littérature grise n'a aucune rentabilité commerciale directe. A l'arrivée, « 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.) »¹⁶

1.4.5 - TEI, TEF, « Nancy Style », OAI-PMH et Dublin Core...

Dans le domaine de la normalisation des métadonnées bibliographique sur le web, il existe malgré tout quelques initiatives qui ont fait date. Elles sont surtout le fruit de groupements de type « consortiums » dont les modes de fonctionnement et les processus de décision sont plus souples que ceux des instances de normalisation (ISO, AFNOR, ANSI, etc.).

Citons d'abord la TEI (« Text Encoding Initiative ») du Consortium TEI fondé en 1987. Sa mission est la rédaction de règles (« guidelines ») pour la description de textes littéraires ou de SHS, et la conception de jeux de métadonnées standards adaptés aux spécificités des corpus. Dans le domaine de la littérature grise, l'apport de la TEI consiste principalement en une tentative d'adaptation du DUBLIN CORE aux rapports et aux thèses.

En France, une variante de la TEI existe pour la description des thèses : la TEF (« *Thèses Électroniques Françaises* »). Elle repose sur l'utilisation d'un schéma XML unique pour la description des thèses soutenues en France¹⁷. Depuis 2008, la version 1.2 de la TEF comprend un volet de métadonnées de gestion (droits d'auteurs, administration, conservation).

Citons ensuite la « Nancy Style », une recommandation émanant du GLISC (Grey Literature International Steering Committee) élaborée en décembre 2005 lors de la 7ème Conférence

15 Schöpfel, Voir bibliographie entrée [12]

16 ibid

17 ibid

Internationale sur la Littérature Grise à Nancy. Cette recommandation concerne la présentation et la structuration des rapports I.S.T.

Enfin, dans le domaine des archives ouvertes, le protocole d'échange OAI-PMH intègre les éléments du Dublin Core, ce qui assure la visibilité des notices des productions au sein d'autres systèmes que les archives de dépôt¹⁸.

1.5 - De la nécessité d'une démarche de valorisation structurée

Le processus de mise en accès de la littérature grise est amorcé depuis longtemps. Les perspectives qu'offrent aujourd'hui les environnements numériques, archives ouvertes en tête, pour la dissémination de ce patrimoine sont inédites. En effet, la multiplicité des services, la baisse des coûts de mise à disposition et les « effets de longue-traine » inhérents à l'économie des réseaux constituent d'indéniables atouts pour accroître la visibilité des productions « non-conventionnelles ».

Toutefois, une démarche non-structurée est susceptible de retourner ces atouts contre les objectifs poursuivis. En effet, sans un contrôle bibliographique minimum, les productions mises en ligne risquent de tomber dans les profondeurs du « web invisible ».

Or, celle-ci constitue une manne, non seulement pour les chercheurs, mais aussi pour tous les usagers potentiels. Ainsi peut on penser que « la littérature grise est certes une littérature éphémère, une littérature de l'éphémère, mais elle représente sans conteste une source toujours renouvelée de connaissances scientifiques et technologiques trop longtemps ignorée de bon nombre d'utilisateurs

18 Voir 3.4 – Métadonnées : signalement et amplifications pour plus de précision (p34)

potentiels »¹⁹.

2 - Les mémoires d'étudiants : une littérature grise spécifique

2.1 - Statuts académiques et juridiques

On peut envisager le statut des mémoires de Master selon deux aspects : académique (des examens de fin d'années), juridique (des œuvres protégées par la propriété intellectuelle).

2.1.1 - Au niveau académique : une hésitation terminologique

La définition académique d'un mémoire de Master n'est pas normalisée. S'il existe autant de définitions qu'il y a de formations²⁰, toutes s'accordent à dire qu'un mémoire de Master est le résultat individuel, écrit, et évalué lors d'une soutenance d'un projet de recherche construit selon une méthodologie choisie, destinée à mettre en valeur les savoirs et savoir-faire acquis pendant sa formation par l'étudiant.

Pour les mémoires de stage, formations, UFR et universités hésitent entre deux termes : « mémoires de stage » et « rapports de stage ». Traditionnellement, on distingue les mémoires par leur aspect réflexif : ceux-ci doivent être le lieu d'une prise de recul critique de l'étudiant sur sa mission. Les rapports désignent davantage une relation factuelle des tâches effectuées et une description de l'environnement. Néanmoins, les termes sont encore employés l'un pour l'autre au sein des universités.

2.1.1 - Des œuvres protégées par le droit d'auteur

Selon le Code de la propriété Intellectuelle (CPI), sont protégées par le droit d'auteur « toutes les œuvres de l'esprit, quels qu'en soient le genre, la forme d'expression, le mérite ou la destination » (art. L. 112-1). A priori, le statut d'œuvre intellectuelle accordé aux mémoires de Master semble

¹⁹ Keriguy. Voir bibliographie entrée [10]

²⁰ Source : Wiki

donc évident, et se fonde sur les raisons suivantes :

1. Un auteur identifié : l'étudiant divulgue un mémoire en son nom propre ;
2. Une « créations de l'esprit » présentant un caractère de « d'originalité ».

Par leur aspect réflexif et rédigé, les mémoires sont bien des créations de l'esprit « originales », c'est à dire : « empruntes de la personnalité de leur auteur »²¹. Ils sont donc protégés par les deux volets de la propriété intellectuelle : le droit moral, relatif à la responsabilité et à la personne de l'auteur, et le droit patrimonial, relatif aux ayants droits sur l'exploitation de l'œuvre.

Ainsi donc, les mémoires de Master ne peuvent être ni diffusés, ni corrigés, ni modifiés sans l'autorisation préalable des auteurs. En cas de réutilisation, la source doit-être citée : interdiction du plagiat, pillage intellectuel. Par ailleurs, toute utilisation qui nuirait à la personne de l'auteur est interdite (diffamation, atteinte à l'image, divulgation de données personnelles, etc.). Enfin, en cas de libre diffusion comme sur une archive ouverte, la réutilisation des mémoires entre dans le régime des « usages raisonnables » (« fair-use » du copyright) : la reproduction est autorisée à des fins non-marchands.

2.1.3 - Des œuvres collectives ?

Dans le cas des mémoires de Master, différents arguments pourraient jouer en faveur d'une propriété intellectuelle partagée. Tout d'abord, ceux-ci engagent l'image et la responsabilité de l'établissement de rattachement, comme celles du directeur de mémoire, parce qu'ils sont le reflet de la qualité de la formation et de l'encadrement. D'autre part, les mémoires de stage peuvent contenir des informations relatives au fonctionnement et aux choix stratégiques des organismes d'accueil : notes d'intention, études prospectives, chiffres clés, détail des « *process* », etc.

Mais alors, si ni l'université, ni le lieu de stage n'ont le droit de diffuser un mémoire sans l'accord de l'auteur, pourquoi celui-ci le pourrait ? Une solution « protectionniste » consisterait à conférer aux mémoires de Master le statut « d'œuvre collective », dont la définition est la suivante :

« [tout œuvre] créée sur l'initiative d'une personne physique ou morale qui l'édite, la publie et la

21 Dorstter. Voir bibliographie entrée [04]

divulgue sous sa direction et son nom et dans laquelle la contribution personnelle des divers auteurs participant à son élaboration se fond dans l'ensemble en vue duquel elle est conçue, sans qu'il soit possible d'attribuer à chacun d'eux un droit distinct sur l'ensemble réalisé » (art. L.113-2 – CPI).

Cependant, comme le montre cette définition, la « contribution personnelle » de plusieurs auteurs est nécessaire pour accorder à une création un tel statut. Or, un mémoire de Master se définit précisément comme un travail individuel. Pour être dans la légalité, l'auteur doit être le seul contributeur.

Comme nous le verrons plus tard, la seule procédure légitime et légale qui puisse concilier ces parties est la demande systématique d'une autorisation officielle à diffusion au moment de la soutenance²².

2.2 - Une dissémination insuffisante des exemplaires imprimés

Selon la définition donnée par l'AFNOR, les mémoires d'étudiants relèvent bien de la « littérature grise ». Leur contenu est non-marchand. Leur contrôle bibliographique est la plupart du temps insuffisant. Ils sont rarement accessibles depuis les guichets « conventionnels » (portails d'éditeurs scientifiques ou catalogue des bibliothèques numériques). Leur diffusion se limite le plus souvent à quelques exemplaires imprimés stockés dans les archives administratives des composantes de formation ou dans les bureaux des enseignants-chercheurs.

Un cadre légal et réglementaire définit assez précisément le statut et le circuit des thèses de 3ème cycle au niveau national²³. En tant que documents administratifs, elles sont communicables dans leur version finale après soutenance à toute personne extérieure aux établissements de rattachement (avis émis par la CADA lors de la séance du 03/03/2005, référence 20051020)²⁴, sauf si elles présentent un caractère confidentiel établi par le Président de l'Université. Les établissements ont donc obligation légale de conserver un exemplaire des thèses pour la consultation au dans les murs.

Le dépôt et la collecte des thèses sont quant-à eux est régis par l'arrêté du 6 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction et de diffusion des thèses ou des travaux présentés en soutenance en vue du doctorat. Trois semaines avant la soutenance, l'étudiant remet 3

22 Voir 3.2.2 – Le rôle des procès verbaux (p31)

23 http://scd.u-bourgogne.fr/services/Circuittheses/Circuit_theses.htm

24 <http://www.cada.fr/fr/conseils/frame.htm>

exemplaires à la B.U de l'université de rattachement : un pour la consultation sur place, un pour le PEB (si autorisation de diffusion par le Bureau des Doctorants), un pour l'ANRT (Atelier National de Reproduction des Thèses) de Grenoble ou Lille 3. La B.U peut alors cataloguer la thèse dans le SUDOC ou procéder au dépôt de l'exemplaire sous forme numérique sur la plateforme de son choix.

En comparaison, le circuit de collecte, de signalement et de diffusion des mémoires de Master est nettement moins complet et contrôlé. Des similarités existent : en tant que documents administratifs, les secrétariats des UFR ont obligation légale de les conserver. En principe, les B.U centrale des universités doivent les collecter. Cependant, les mémoires rassemblées ne font l'objet d'aucun signalement dans le SUDOC ou tout autre catalogue de ressources universitaires. Les mise en place d'une politique de valorisation des exemplaires rassemblées dans les B.U ou dans les bibliothèques d'application des UFR reste à la discrétion des établissements.

Dans ces conditions, l'auto-archivage dans une archive ouverte semble une solution toute trouvée pour la mise en accès et la diffusion élargie des mémoires de Master, à l'instar des thèses. En effet, « alors qu'une thèse imprimée, conservée en bibliothèque universitaire, est consultée en moyenne une fois tous les dix ans, un site de thèses électroniques est visité chaque jour par plus de 300 personnes (statistiques de l'Université Lyon 2) »²⁵. En outre, l'auto-archivage des étudiants permet d'exploiter le support natif des mémoires : numérique. Comme nous le verrons, la numérisation rétrospective par une institution pose de nombreux problèmes.

2.3 - Une production marginale dans les archives ouvertes

Malgré les avantages de la diffusion électronique, les mémoires d'étudiants sont sous représentés dans les archives ouvertes. Seul 10 archives contenant ce type de documents sont signalées dans le répertoire international OpenDOAR²⁶.

En France, les mémoires d'étudiants au niveau Master ne représenteraient aujourd'hui que 0,1% du contenu des archives ouvertes nationales²⁷. Deux archives ouvertes seulement sont spécifiquement consacrées à la diffusion des mémoires de Master²⁸ au niveau national, et sont sous-utilisées :

25 <http://www.cybertheses.org/?q=fr/node/19>

26 Source : Wiki

27 Ibid

28 Il existe aussi un site à caractère commercial : <http://www.memoireonline.com/> , et une plateforme associative en marge de toute institution : <http://librapport.org/>.

1. [DUMAS](#) : une archive ouverte multidisciplinaire rassemblant 246 documents en texte intégral (01/06/2010) ;
2. [MemSIC](#) : une archive ouverte consacrée aux mémoires de Master en Science de l'Information et de la Communication rassemblant 199 documents en texte intégral (01/06/2010).

Pourtant, une étude menée par le consortium COUPERIN en 2007 auprès de 74 établissements ayant initié des projets d'archive ouverte établissait que 27 % (20 organismes) estimaient la valorisation des travaux d'étudiants (mémoires, thèses d'exercice et rapport un objectif « très important »²⁹.

Différentes raisons peuvent expliquer cette sous-utilisation :

1. La relative jeunesse de ces dispositifs : MemSIC a été ouvert aux publics en 2003 (référencement dans ROAR en novembre 2003), et DUMAS au printemps 2008 (référencement dans ROAR au printemps 2008) ;
2. Une communication insuffisante à destination des étudiants ;
3. Une communication insuffisante à destination des communautés d'enseignants-chercheurs les plus réticentes au dépôt électronique³⁰. Ceux-ci archivant eux mêmes peu leurs productions n'incitent pas leurs étudiants à le faire.

A ce sujet, l'enquête COUPERIN 2007 révèle que 19,4 % des établissements ont rencontré pour principale difficulté un manque d'implication des chercheurs. Quelles qu'en soient les raisons, l'auto-archivage des étudiant n'est pas rentré dans les usages. Aujourd'hui, les projets de diffusion de mémoires de Master se limitent surtout à des initiatives personnelles ou confinées aux ENT. Ce type d'initiative ne donne pas aux mémoires d'étudiants la visibilité nécessaire pour toucher l'ensemble des usagers potentiels. Or, nous allons voir qu'ils existe une demande et des besoins informationnels spécifiques en la matière.

29 GTA0. Voir bibliographie entrée [08]

30 Par exemple, les communautés SHS traditionnellement attachées à l'ouvrage est peu familières des pratiques d'auto-archivage : lettres, philosophie ou histoire. Voir ...

2.4 - Particularités par rapport à la production universitaire « grise » de 3ème cycle

Mais avant, il convient s'arrêter sur certaines caractéristiques des mémoires de Master qui les distinguent des productions universitaires « grises » de 3ème cycle : thèses, rapports de recherche, « *pre-prints* », etc.

1. Une validation normée par l'institution universitaire.

Comme les thèses de troisième cycle, les mémoires de Master n'ont pas d'autre validité qu'institutionnelle (soutenance devant un jury). Une évaluation alternative de type *peer-review* ne s'applique pas à ces productions. Ainsi, MemSIC et DUMAS ne diffusent aucun *pre-print* mais seulement des versions finales et validées des mémoires. De ce fait, les mémoires d'étudiants n'entrent pas totalement dans le modèle de la communication scientifique directe³¹.

2. Une littérature exclue du marché de l'I.S.T

Le propre de la littérature grise est sa vocation non-marchande. Nous entendons par là que l'auteur d'un rapport de recherche ne tire aucun profit direct de la publication de son œuvre, contrairement aux monographies. Néanmoins, certaines productions universitaires « grises » peuvent faire, *in fine*, l'objet d'une publication et entrer dans le cadre de l'économie marchande. Ainsi des thèses publiées et vendues comme ouvrages, ou des conférences comme actes de colloques par les éditeurs scientifiques.

Si certains sites à caractère commerciaux font payer l'accès aux mémoires d'étudiants en texte intégraux, ces derniers n'ont jusqu'à présent jamais fait l'objet de publication par des éditeurs spécialisés et sont totalement exclus du marché de l'I.S.T.

3. Une littérature qui ne s'insère pas dans l'économie de la citation

On mesure l'impact d'une production scientifique, « grise » ou « blanche », à son nombre de

31 Voir 1.2 – Le modèle de la communication scientifique directe (p38).

citations. Si les chercheurs touchent rarement de revenus directs sur l'exploitation de leurs œuvres, la citation représente leur gratification symbolique. Aussi doivent-ils en principe donner à leurs travaux une visibilité maximale, en les soumettant aux comités scientifiques de revues jouissant d'un fort facteur d'impact³². Les productions « grises » n'étant par définition pas publiées, un tel positionnement stratégique doit passer par le dépôt dans des bibliothèques numériques spécialisées ou dans des archives ouvertes, complété par un travail de signalement dans les annuaires et moteurs de recherche spécialisés.

Un tel système n'existe pas pour les mémoires d'étudiants. D'une part, ceux-ci ne sont pas produits dans l'intention d'être cités (ce qui ne veut pas dire qu'ils ne puissent pas l'être). Surtout, la pratique de la citation n'est pas, à ce niveau, institutionnalisée. Aussi, les mémoires ne s'insèrent pas dans le cadre de l'économie de la citation qui régit l'information scientifique.

2.5 - Des ressources répondant à des besoins informationnels particuliers

2.5.1 - Un questionnaire pour collecter des données

Afin de mieux cerner les besoins informationnels auxquels les mémoires de stage peuvent répondre, nous avons réalisé un petit questionnaire destiné à collecter des données sur les pratiques de consultation et d'auto-archivage des mémoires d'étudiants par les étudiants³³.

Les résultats détaillés de ce questionnaire, ainsi qu'une présentation de l'échantillon et de la méthodologie adoptée pour sa réalisation sont disponibles en annexe.

Ce questionnaire ne faisait pas parti de notre mission. Aussi ne constitue-il en aucun cas une quelconque étude quantitative ou d'usages, tout au plus un petit sondage. Une telle entreprise devra être menée une fois le projet de l'UFR porté à son terme et ouvert aux publics.

Nous ne l'avons pas conçu selon le protocole qui préside à la conception des enquêtes sociologiques : état de l'art, problématique, hypothèses de réponses. Nous n'avons pas non plus construit notre échantillon scientifiquement, mais empiriquement, selon les personnes que nous

32 IF (*impact factor*) : rapport entre le nombre de citations reçues par un périodique pendant 3 ans et le nombre d'articles publiés sur la même période.

33 Questionnaire disponible en annexe p 61.

connaissances et susceptibles d'y répondre. Aussi n'est-il ni représentatif, ni neutre : une majorité d'étudiants de M1 ICD, familiers des problématiques liées à l'auto-archivage et aux bibliothèques numériques.

Toutefois, nous avons malgré tout quelques hypothèses de départ concernant les types d'informations recherchés par les étudiants dans les mémoires.

2.5.2 - Hypothèses de départ

Selon nous, la différence essentielle entre les mémoires de Master et les productions universitaires de 3^{ème} cycle tient à la relativité de leur valeur scientifique pour les usagers. D'une part, un travail de 2nd cycle ne peut pas prétendre au degré de maturité scientifique d'une thèse, ni même d'un article de recherche. En outre, les mémoires de Master répondent à un projet moins scientifique (faire avancer la discipline) que pédagogique (valider des acquis). Aussi, le contrat de lecture est différent et les usages sont spécifiques.

C'est a fortiori le cas des mémoires de stage, principalement diffusés par l'UFR IDIST. Parce qu'ils s'adosent à l'expérience concrète d'une mission, ceux-ci contiennent des types d'informations dont certains les distinguent des productions de recherche :

1. États de l'art ;
2. Références bibliographiques ,
3. Études de cas ;
4. Retours d'expériences ;
5. Informations méthodologiques.

Ces types d'informations sont susceptibles de répondre à des besoins informationnels autres que scientifiques :

1. Aide à la rédaction des mémoires ou travaux en cours d'années (illustrations méthodologiques) ;
2. Aide à la recherche de stage (exemples de missions et de lieu de stage) ;
3. Aide à l'orientation (cas pratiques de situations professionnelles).

2.5.3 - Synthèse des résultats

Pratiques de consultation³⁴

Indéniablement, les mémoires de Master, stage ou recherche, constituent une source d'information exploitée par les étudiants. Sur 27 étudiants, un seul déclare ne jamais avoir consulté de mémoire. Cependant, il ne s'agit pas d'une ressource quotidienne, mais ponctuelle : une majorité d'étudiants (18) en consultent « de temps en temps », seulement 2 « régulièrement ».

Les étudiants consultent d'abord dans le cadre de la rédaction de leurs propres mémoires (24 étudiants) et recherchent en premier lieu des informations de type méthodologique (17 étudiants). Les « contenus scientifiques » ont bien une valeur relative : 13 étudiants déclarent en rechercher « principalement », 9 de manière « secondaire », et 3 ne s'y intéressent « pas du tout ». Les références bibliographiques sont plus importantes qu'on ne pourrait le penser : 10 étudiants déclarent en rechercher « principalement », 9 les estiment « secondaires », et 6 ne s'y intéressent pas. Quant-aux exemples de sujets – problématiques de recherche et missions – lieux de stages : 8 et 6 étudiants déclarent en rechercher « principalement », 10 et 8 de manière « secondaire », 7 et 11 « pas du tout ».

Les supports de consultation les plus utilisés se partagent à parts égales entre papier et numérique : 13 étudiants pour chaque. Sachant que les étudiants consomment de plus en plus de ressources électroniques, la persistance du support papier témoigne de la marginalité des mémoires de Master sur les réseaux. Ceci semble confirmé par les réponses à la question portant sur les modalités d'accès les plus fréquentes : la moitié des étudiants (14) accèdent encore aux mémoires par la bibliothèque de leur UFR ou par un enseignant – membre du personnel administratif. 13 étudiants déclarent y accéder par le web.

Pratiques de diffusion sous forme électronique³⁵

Deux étudiants seulement déclarent avoir déjà diffusé leur mémoire sur le web, à leur propre initiative. Un seul a précisé le dispositif : un site à caractère commercial et un site à caractère

³⁴ Présentation et méthodologie du questionnaire disponible en annexe, voir « dimensions » abordées par le questionnaire (p59)

³⁵ idem

personnel (« Blog, site web, page personnelle »). Ceci confirme ce que nous savions : l'auto-archivage des étudiants, comme le dépôt des mémoires par l'institution ne sont pas encore rentrés dans les usages.

En revanche, et bien qu'on se situe là plus que jamais sur le terrain du conditionnel, les étudiants interrogés semblent prêts pour la mise en place de telles procédures, à un niveau institutionnel. Une majorité d'étudiants estiment que la mise en ligne des mémoires « réussis » peut valoriser leur parcours universitaire lors d'une candidature à une offre de stage ou d'emploi (26 étudiants sur 27). De même, une majorité se déclare favorable à la mise en ligne des mémoires par leur UFR (26 étudiants sur 27 également).

Cependant, ceux-ci sont favorables à une mise en ligne contrôlée, à des degrés divers. En effet, seul 4 étudiants se disent favorables à une mise en ligne sans « aucune condition particulière ». En revanche, plus de la moitié des étudiants (15 sur 26) sont pour l'instauration d'une politique de sélection « qualitative » (une note pallier supérieure ou égale à 14/20). 13 étudiants sont pour un accès restreint par des mesures techniques de protection des fichiers (copier-coller, impression désactivée, etc.). Seulement 5 étudiants souhaiteraient un accès limité de type intranet (« accès limité aux étudiants et enseignant-chercheurs de votre formation - discipline »).

La seule personne s'étant déclarée contre une telle initiative invoque les raisons suivantes : la protection des données personnelles - vie privée, la peur d'être plagié, et le caractère immature des mémoires qui les rendrait inutiles ou non-communicables.

Aussi précaires que soient les résultats de ce sondage, ils révèlent malgré tout une demande des étudiants pour la mise en ligne des mémoires de Master en texte intégral. Le terrain est donc favorable pour le projet de l'IDIST. Les réticences semblent minoritaires, au moins sur le papier. Les étudiants interrogés sont malgré tout sensibles à la question de la qualité des mémoires diffusés, et à celle de leurs droits d'auteurs. Apparemment, certains restent à convaincre des avantages du Libre Accès pour la visibilité des productions et le partage des connaissances. En effet, la moitié des étudiants souhaitent voir leur fichier protégé en écriture, ce qui est contraire à la philosophie des archives ouvertes, et au projet de l'IDIST.

3 – L'expérimentation de l'UFR IDIST

3.1 - Des ressources documentaires inexploitées

A l'UFR IDIST, la situation est un peu particulière et joue contre la visibilité des travaux d'étudiants. En effet, l'UFR a transféré sa bibliothèque d'application à la Bibliothèque des bibliothèques du SCD de Lille 3 il y a une dizaine d'années. Les mémoires de Master antérieurs à l'année universitaire 1996-1997 sont stockés à la B.U, mais ne font l'objet d'aucune politique de signalement - mise en accès. Les mémoires postérieurs à cette date sont stockés en salle B5-113, sans faire l'objet d'aucune mise en valeur.

L'inventaire réalisé par Aurore Plichon, stagiaire en M1 ICD l'an passé, a fait état de plus de 1000 travaux d'étudiants consignés, dont une majorité de mémoires de Master et d'anciens DESS – DEA. L'IDIST n'ayant plus de bibliothèque, les mémoires restent invisibles et inaccessibles : la salle B5-133 n'est pas en accès libre car des documents administratifs confidentiels y sont entreposés. Par ailleurs, une enquête réalisée auprès des étudiants de M1 ICD en 2008-2009 a révélé que l'ensemble des étudiants ignoraient l'existence de ces ressources et comment se les procurer.

Pourtant, comme le rappelle notre questionnaire, il existe une demande pour leur mise en ligne. Le fonds est essentiellement constitué de mémoires de stage de M1 ICD, M2 GIDE, IDEMM et, dans une moindre mesure, PRISME. En effet, très peu d'étudiants de niveau M2 s'inscrivent en parcours recherche (Master rSIC). En outre, contrairement aux mémoires de stage, il manque à ce jour une procédure susceptible de garantir la traçabilité de ces productions, comme leur conformité au Code de la Propriété Intellectuelle. Aucun document officiel (procès verbal) n'existe pour consigner l'autorisation des auteurs de mémoires rSIC. Aussi avons nous renoncé à leur dépôt rétrospectif sur MemSIC.

3.2 - Une politique de sélection qualitative

3.2.1 - *Qualité ou « ranking » ?*

MemSIC s'est imposé comme l'environnement naturel pour la diffusion des mémoires : une archive thématique consacrée aux mémoires de Master en Sciences de l'Information et de la Communication, gratuite d'utilisation.

Or, MemSIC ne pratique aucune politique de sélection éditoriale, comme son équivalent multi-disciplinaire DUMAS. Le service laisse aux établissements de rattachement le soin de mettre en place une politique éditoriale.

Partant, deux approches sont envisageables. La première, fondée sur la quantité, consiste à imposer une sélection minimale (10/20) tout en laissant aux usagers le soin d'évaluer la qualité de l'information par eux-mêmes. Trois arguments justifient cette politique : la neutralité et la transparence de la démarche, l'habitude des usagers à évaluer l'information depuis les moteurs sur le web, un meilleur « ranking » de l'institution dans les moissonneurs d'archives ouvertes ou dans les moteurs de recherche orientés IST. En effet, plus il y a de dépôts, plus un site génère de trafic, meilleur est le référencement des fichiers. Cependant, la mise en ligne de mémoires défailants risque de desservir les auteurs, nuire à la réputation de l'établissement, dévaluer la qualité des bons mémoires et l'ensemble du dispositif.

La seconde, fondée sur la qualité, consiste à imposer une note pallier, résolument celle de l'UFR IDIST. En effet, cette démarche est plus à même d'atteindre les objectifs initiaux du projet : valoriser le cursus universitaire des étudiants, en escomptant un effet positif sur l'insertion professionnelle, et renforcer le label de l'institution. Les détracteurs de cette politique invoqueront sa dimension subjective, comme les effets pervers qu'elle peut introduire : sur-notations et concurrences entre UFR. Toutefois, la note de 14/20 (mention Bien) est généralement retenue comme critère d'admission dans divers parcours universitaires (entrée en Thèse et dans certains Masters). Elle est donc susceptible de constituer une référence commune à tous les établissements. Enfin, si le dépôt des mémoires se systématisait, les Jury pourront noter en conséquence.

3.2.2 - Le rôle des procès verbaux

Les procès verbaux sont les documents officiels établis par les Jury au moment des soutenances. Aussi constituent ils des preuves de la conformité des mémoires au droit d'auteur et au règlement d'utilisation de MemSIC. Ceux-ci attestent que le mémoire à été soutenu et consignent sa note. Ils contiennent également l'autorisation à diffusion des auteurs et du Jury : consultation sur place (exemplaire imprimé), diffusion sur un intranet, diffusion sur internet. Aussi, leur renseignement par le Jury et leur conservation par l'administration est obligatoire, et indispensable pour le dépôt rétrospectif comme pour l'auto-archivage. Sur les 298 mémoires ayant obtenu une note supérieure ou égale à 14/20, 129 avaient un procès verbal incomplet (volet « autorisation » manquant ou non renseigné).

En outre, dans une optique de numérisation et dépôt rétrospectif, le volet autorisation présente une lacune : les modalités des éventuelles clauses de confidentialité ne sont pas précisées et rarement signalée par les Jury (parties à « caviarder » ou durée de l'embargo). C'est renoncer à leur signalement dans les catalogues, à défaut d'une diffusion en texte-intégral.

3.3 - Numérisation et dépôt rétrospectifs

Sur l'ensemble des mémoires de Master, 168 mémoires remplissaient les conditions de qualité et de légalité pour la diffusion. Restait à collecter les exemplaires sous forme électronique par mail. Cette entreprise n'ayant pas porté ses fruits, il a fallu numériser rétrospectivement. Le schéma ci dessous décrit une GED idéale. Nous verrons après comment les tâches se répartissent avec MemSIC.

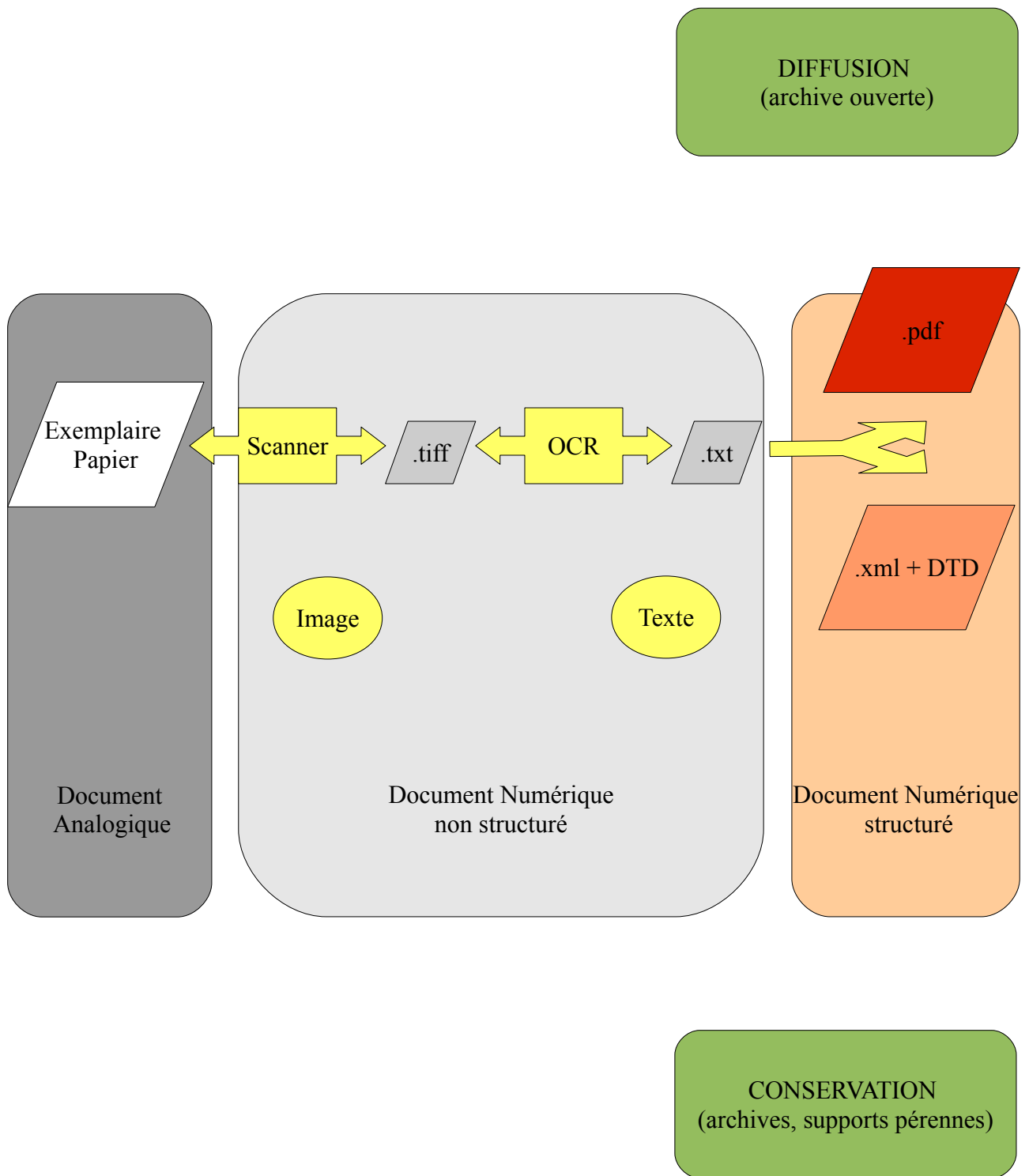


Schéma d'une GED idéale

La chaîne de numérisation idéale converge ainsi vers la production de deux documents de sortie : un exemplaire au format .pdf (Portable Document Format), ainsi qu'une version structurée et décrite du document au format .xml (eXtensible Markup Language), assortie de sa DTD (Doctype Definition). Ces deux formats assurent un accès le plus ouvert possible au document, ainsi que sa sauvegarde à moyen – long terme. En effet, le format .pdf est utilisé pour la diffusion car ouvert pour la lecture et compatible avec la plupart des systèmes. Le format .xml garantit quant à lui la portabilité du document vers un autre système non compatible en cas d'évolution des formats.

L'obtention de ces deux états de sortie implique une série de traitements sur l'exemplaire imprimé (conversion – génération – compilation), et la production de formats intermédiaires :

1. **Scanner** : production d'un document en mode image, poids important, possibilités de manipulation limitées ;
2. **Océrisation** : production d'un document en mode texte, possibilités de manipulations importantes (correcteur orthographique, mise en forme via une feuille de style) ;
3. **Structuration – Description** : production d'un schéma de document et d'une version .xml automatiquement générée via une feuille de style) ;
4. **Compilation** : production d'un fichier .pdf, taux de compression important, lisible sur tous les systèmes d'exploitation.

Sur ArchiveSIC (modèle de MemSIC pour les travaux de 3ème cycle), le CCSD prend en charge tous ces traitements, hormis le passage au scanner³⁶. En principe, MemSIC les assure aussi. Néanmoins, ceci vaut dans le cadre des activités normales de la plateforme : dépôts ponctuels des étudiants. Dans le cas d'un dépôt rétrospectif « massif », l'équipe du CCSD n'a pas nécessairement le temps de réaliser ces traitements, d'autant plus qu'elle doit administrer toutes les plateformes du portail HaL en parallèle. Aussi, dans le cadre de cette expérience, nous devons encore océriser les fichiers. C'est également une condition indispensable à leur référencement dans les moteurs de recherche généralistes.

Une telle expérience souligne l'avantage de l'auto-archivage pour un UFR. Celui-ci permet « d'externaliser » l'ensemble de ces tâches : les traitements à MemSIC, la collecte et le dépôt aux étudiants. Cela représente surtout une économie de temps et de ressources humaines pour l'IDIST qui est équipé d'un scanner multi-fonction et d'un logiciel d'océrisation efficace³⁷.

36 Gallezot. Voir bibliographie entrée [06]

37 Voir 2.& - « Autoarchivage ou dépôt institutionnel » (p44).

3.4 - Les métadonnées : signalement et amplification

La saisie des métadonnées est l'étape la plus importante du dépôt sur MemSIC.³⁸ Relativement complètes, elles garantissent un contrôle bibliographique élémentaire et la traçabilité des mémoires :

1. description juridique : auteur, lieu de stage, embargo si confidentialité ;
2. Identification institutionnelle : date de soutenance, laboratoire - établissement de rattachement, responsable de formation, directeur de mémoire, parcours – mention universitaire, domaine scientifique (nomenclature du CNRS) ;
3. Identification thématique : abstract, mots-clé.

Ces métadonnées constituent des « chemins d'accès » cohérents aux mémoires. Surtout, elles sont compatibles au format OAI-PMH (Open Archive initiative – Protocol for Metadata Harvesting), un protocole d'échange entre « data providers » divers (archives ouvertes, bibliothèques numériques, portails d'éditeurs, OPAC) et « services providers » : des moissonneurs d'archives ouvertes comme OAIster, ou WorldCat, et des méta-moteurs orientés références - citations comme Scopus, CiteBase ou Citeseer³⁹. Les métadonnées de MemSIC assurent donc l'amplification du signalement des mémoires déposées en rendant les notices visibles sur les principaux canaux d'accès à l'information scientifique grise.

Ajoutons pour finir qu'il existe d'autres lieux de visibilité moins « conventionnels » pour les mémoires d'étudiants sur le web : les moteurs de recherche thématiques (Google Scholar), et les moteurs de recherche .pdf (PDF Search Engine, BePDF). Plusieurs facteurs influent sur le référencement des fichiers .pdf⁴⁰. Néanmoins, les critères déterminants sont le mode-texte et le poids : les fichiers doivent être composés de chaînes de caractère pour être indexés par les moteurs de recherche, et suffisamment légers pour que l'ensemble de leur contenus soit interprétés.

3.5 - Conclusion

Si MemSIC s'est imposé comme l'environnement naturel pour mener l'expérimentation de l'UFR, la

38 Pour le détail de la procédure et des métadonnées, se référer à l'analyse des besoins

39 Aubry. Duploux. Voir bibliographie entrée [01]

40 Pour le détail, se référer à l'analyse des besoins

plateforme ne sera pas nécessairement la solution adoptée à l'arrivée. En effet, telle quelle, il lui manque des fonctionnalités et services de personnalisation. MemSIC est une archive thématique nationale dont le but est de valoriser une discipline, pas les productions d'une institution particulière. Aussi, l'auto-archivage des étudiants sur MemSIC valorisera surtout le travail des étudiants.

Néanmoins, les choses ne sont pas figées. D'une part, MemSIC devrait évoluer en une version 2 sous la forme d'une archive HaL destinée à l'ensemble des productions « non-conventionnelles» du portail. D'autre part, le CCSD est habitué à proposer des services aux institutions soucieuse de valoriser leurs productions sur la base de solutions existantes « adaptées ». C'est cette possibilité que nous allons creuser en deuxième partie.

Deuxième partie :

*les archives ouvertes, un instrument pour la
mise en accès et en visibilité des mémoires de
Master*

1 - Les archives ouvertes : état des lieux

1.1 - « *Green Road* » et « *Open Access* » : principes fondateurs

La « *Green Road* », diffusion dans les archives ouvertes, comme la « *Gold Road* », diffusion dans les revues en accès libre, constituent deux modes de diffusion de l'information scientifique alternatifs à l'offre des éditeurs commerciaux.

La « *Green Road* », qui nous intéresse ici, repose sur les principes suivants :

1. Le Libre Accès (*Open Access*) ;
2. La Gratuité ;
3. Un modèle de communication scientifique directe.

Le Libre Accès à la recherche scientifique peut s'aborder sous deux angles : économique et technique.

D'un point de vue économique, le Libre Accès implique la gratuité de consultation pour les lecteurs, et la gratuité de publication pour les auteurs. Ainsi, l'utilisation des archives ouvertes comme MemSIC ou DUMAS est gratuite en amont du dépôt, comme en aval. Pour les établissements, la gratuité permet d'effectuer une double économie : baisse des budgets alloués à l'acquisition de ressources, baisse des budgets alloués à la publication des chercheurs (la soumission d'un article au comité scientifique d'une revue commerciale est payante).

Ainsi, le coût des archives ouvertes pour les universités française est estimé 500 à 1000 fois inférieur aux dépenses en abonnements pour des revues scientifiques⁴¹. Il se limite aux frais de personnel engagés pour le développement, la maintenance technique et l'administration des archives. Si le projet repose sur l'utilisation d'une solution existante (plateforme externalisée au CNRS) et sur un service gratuit (maintenance assurée par le CCSD), la dépense en propre est insignifiante pour l'institution⁴². Ce modèle économique repose sur le financement en amont : la dépense a été intégrée à la dotation globale de la recherche publique et de l'enseignement supérieur français.

41 Baruch . Voir bibliographie entrée [02]

42 GTA0. Voir bibliographie entrée [08]

D'un point de vue technique, le principe corolaire du Libre Accès est l'interopérabilité, au cœur de la philosophie des archives ouvertes. Ainsi, pour assurer le Libre Accès à la recherche scientifique, notamment aux « ressources grises », les archives ouvertes doivent respecter les normes de compatibilité entre les systèmes, notamment le protocole d'échange de métadonnées OAI-PMH. En principe, elles ne doivent pas non plus restreindre leur accès par un identifiant, mot de passe, et surtout par des mesures de protection techniques des fichiers.

1.2 - Le modèle de la Communication Scientifique Directe

Les archives ouvertes visent à la mise en place d'un modèle de communication scientifique directe, tel que l'a pensé Paul Ginsparg, le fondateur d'ArXive en 1991. A l'origine, il s'agissait de répondre à un besoin de la communauté des chercheurs en Physique pour l'accélération de la publication des résultats de recherche.

Un tel modèle est marqué par la suppression des intermédiaires entre les chercheurs-auteurs et leur communauté. Il s'oppose au circuit de l'édition scientifique dont les processus de validation éditoriale étaient vécus comme autant de pesanteurs contraignantes.

Ce modèle a fait l'objet de nombreuses critiques de la part des éditeurs spécialisés. Celles-ci portaient notamment sur la qualité et la fiabilité des contenus diffusés. Au cœur du problème : l'évaluation de l'information grise. En effet, les archives ouvertes sont généralement composées de documents non évalués au statut hétérogène (« pre-prints », « semi-published » ou « working papers »). Les communautés scientifiques ont progressivement réussi à répondre à ces critiques en organisant un mode de validation collaboratif : le « peer-review » (évaluation par les pairs), alternatif à la soutenance devant un jury (thèses, mémoires) ou à la soumission au comité scientifique d'une revue (articles).

1.3 - Définition et Typologie des archives ouvertes

1.3.1 - Un terme ambigu

Le monde des archives et celui de la documentation se différencient par leurs approches. Le premier

est centré sur la question de la conservation, des supports, et manifeste un souci de "faire collection". Le second est centrée sur les questions liées à l'accès et à la diffusion de l'information. Or, les "archives ouvertes", telles qu'elles, ont rarement une dimension patrimoniale. Elles sont avant tout des entreprises documentaires. Leur nom fait référence à ArXive.

D'autre part, l'adjectif "ouvert" se confond facilement avec gratuité. Or, comme nous l'avons vu, les archives ouvertes ont un coût. Il n'est simplement pas assumé par l'utilisateur. L'épithète désigne ici surtout une ouverture technique, par référence à l'Open Access.

Ainsu, Gabriel Gallezot définit les archives ouvertes comme : « des réservoirs d'articles (pré-publication ou post-publication) auto-archivés, en accès libre, sans barrières économiques ni juridiques et intégrant le protocole d'interopérabilité OAI-PMH »⁴³.

Néanmoins, cette définition générique, aussi précise soit-elle, ne rend pas compte de la diversité des dispositifs.

1.3.2 - Typologie des archives ouvertes

Pour caractériser les archives ouvertes, Gabriel Gallezot propose la typologie suivante ⁴⁴ :

1. Les archives institutionnelles : « projets relevant d'universités, de grandes écoles, d'organismes de recherche, d'associations professionnelles ».

Ex : Archimer, l'archive institutionnelle de l'Ifremer⁴⁵

2. Les archives disciplinaires : « réalisations de communautés scientifiques généralement regroupées en domaine de recherche assez large ».

Ex : ArchiveSIC, le modèle de MemSIC dédié aux travaux de recherche de 3ème cycle en SIC⁴⁶.

3. Les archives thématiques : « entrepôts qui se focalisent sur un sujet ou un type de document

43 Gallezot. Voir bibliographie entrée [07]

44 Gallezot. Voir bibliographie entrée [05]

45 <http://archimer.ifremer.fr/>

46 <http://archivesic.ccsd.cnrs.fr/>

singulier ».

Ex : Cyberthèses, le portail de l'organisation internationale de la francophonie dédié au dépôt des thèses électroniques⁴⁷.

4. Les méta-archives : « une collation de notices au protocole OAI qui renvoie, pour la consultation du texte intégral, à l'archive d'origine » (ndlr : les moissonneurs).

Ex : moteur d'archive ouverte OAIster, aujourd'hui intégré au Worldcat⁴⁸.

A quelques variantes près, cette typologie est généralement celle employée pour analyser les plateformes dédiées à la diffusion de ressources grises en libre accès. Toutefois, cet outil théorique doit être employé avec souplesse. Il n'existe pas de dispositif «chimiquement pur ». Comme nous le rappelle l'auteur : « une même interface [peut] mélanger les genres, les types de documents, et impliquer différents types d'acteurs ».

1.4 - Chiffres et tendances

1.4.1 - Un succès à pondérer

Les archives ouvertes sont en constante augmentation depuis 10 ans. Au niveau international, le répertoire OpenDOAR signalait 349 archives en 2006 et 1681 en juin 2009, ce qui témoigne d'un taux de croissance de plus de 25 % par an⁴⁹. Il convient toutefois nuancer cette hausse. Ouvert en 2005, OpenDOAR n'avait pas assez de moyens et de visibilité pour dresser un inventaire exhaustif à l'époque.

47 <http://www.cybertheses.org/>

48 Le Très Grand Equipement de l'OCLC, le catalogue mondiale des bibliothèques interconnectées : <http://www.oclc.org/ca/fr/worldcat/default.htm>

49 <http://www.opendoar.org/onechart.php?clID=&ctID=&rtID=&clID=&lID=&potID=&rSoftWareName=&search=&groupby=r.rDateAdded&orderby=&charttype=growth&width=600&height=350&caption=Growth%20of%20the%20OpenDOAR%20Database%20-%20Worldwide>

Au niveau national, l'enquête de 2009 menée par l'Université de Lille 3 et l'INIST⁵⁰ confirme malgré tout cette hausse exponentielle. Le nombre d'archives identifiées est passé de 56 en 2008 à 150 en 2009, ce qui implique un taux de croissance de plus de 168 %.

Le nombre total de dépôt a lui aussi augmenté : 703 178 dépôts en 2008 contre 1 878 520 dépôts en 2009. Mais là encore, ce chiffre est à relativiser en raison du morcellement de l'offre. En effet, les dépôts se répartissent sur un grand nombre de petites et moyennes archives, contenant peu de dépôts (entre 100 et 10 000), et une minorité de grosse structures au niveau nationale, concentrant la majorité des dépôts.

En outre, les archives ouvertes sont de plus en plus victimes d'un dévoiement de leurs objectifs initiaux. Démarche à l'origine communautaire, les archives ouvertes intéressent désormais les institutions qui les intègrent dans leurs stratégies d'établissement pour accroître leur visibilité, sans pour autant offrir un accès en texte intégral. Ainsi l'enquête démontre que 16 % des dépôts réalisés dans les archives françaises en 2009 ne proposaient qu'un accès aux notices bibliographiques. Les sites dédiés aux mémoires de Master (MemSIC et DUMAS) n'échappent pas à la règle, en proposant aux auteurs de mémoires confidentiels des embargos pour retarder la mise en ligne. En outre, 12 % des archives interrogées restreignent leur accès via des identifiants et des mots de passe.

Ces éléments tendent à pondérer le succès remporté par les archives ouvertes. Toutefois, ces pratiques émergentes restent relativement récentes et le processus de libération de l'accès à la recherche scientifique est en pleine transition. Le paysage est donc marqué par d'importants décalages : certaines communautés scientifiques sont en avance, et d'autres encore réticentes au dépôt parce qu'attachées à l'ouvrage et aux circuits de publications normalisés.

1.4.2 - L'essor des archives institutionnelles

En France comme ailleurs, la tendance est à l'éclosion des archives ouvertes institutionnelles.

La répertoire international OpenDOAR signale 1681 site, dont 81 % d'archives institutionnelles (01/06/2010).

En France, l'enquête COUPERIN 2007 montre que 74 % des organismes avaient initié des projets d'archives institutionnelles à l'échelle locale. 54,5% des établissements avaient inscrit leur projet

⁵⁰ Schöpfel – Prost. Gallezot. Voir bibliographie entrée [12]

dans leur contrat d'établissement. 53 % des établissements étaient encore au « niveau 1 » de développement : phase d'expérimentation, conception, réalisation. 47 % avaient atteint le niveau 2 : installation, test, ouverture aux publics.

Selon Gabriel Gallezot, l'essor des archives institutionnelles s'explique d'abord par l'accélération de la recherche et la concurrence rampante entre les laboratoires qui en découle. Dans un tel contexte, les institutions seraient poussées à rechercher les moyens d'une visibilité maximale, d'où l'initiation de projets d'archives ouvertes. Par ailleurs, la tendance s'expliquerait aussi par les relations souvent conflictuelles qu'entretient le monde des archives ouvertes avec les éditeurs scientifiques : les institutions occuperaient une position plus légitime aux yeux des éditeurs pour diffuser gratuitement les productions dont elles sont la source.

L'enquête COUPERIN souligne enfin la tendance à l'éparpillement, au cloisonnement, et au manque de coordination des initiatives. Seul 16 organismes avaient initié des projets d'envergure nationale ou inter-établissement. Ce modèle de développement pose donc la question de l'accès unifié aux ressources déposées. Celui-ci suppose l'interconnexion des bases de données (protocole OAI -PMH) et la construction de cyber-infrastructure (Très Grand Équipement) capables d'intégrer l'ensemble des ressources accessibles depuis des guichets éparpillés⁵¹.

1.5 - La centralisation : spécificité du modèle français (rôle du CCSD)

La particularité française tient à la forte centralisation de la structure nationale d'archivages ouvertes autour d'une entité administrative : le CCSD, et d'une solution technique : le logiciel d'archives ouvertes HaL⁵².

Le CCSD (Centre pour la Communication Scientifique Directe) est une unité du CNRS créée en 1998, avec pour mission principale le développement d'outils adaptés pour une base de documents centrale, sur le modèle d'ArXive. L'objectif : créer une plateforme nationale pour le dépôt des productions scientifiques, dans un but de communication directe.

51 Par exemple, le TGE – ADONIS du CNRS (Accès Unifié au Documents Numériques des Sciences Humaines et Sociales), accessible à l'adresse suivante : <http://www.tge-adonis.fr/>

52 Baruch. Voir bibliographie entrée [02]

Le logiciel d'archive ouverte HaL (Hyper Articles en Ligne) est développé pendant deux ans, et le service ouvre aux publics fin 2000⁵³. Il s'agit d'une solution Open Source compatible au protocole OAI-PMH dont l'interface est depuis proposée aux institutions pour créer des guichets d'accès personnalisés aux différentes bases HaL.

Si HaL s'inspire d'ArXive, il couvre un spectre de disciplines beaucoup plus large, en rassemblant des archives aussi bien institutionnelles (portail HaL INSERM) que thématiques (Tel – serveur de thèses multidisciplinaires) ou disciplinaires (portail HaL – SHS).

En 2006, un protocole national est signé pour 2 ans, associant l'ensemble des établissements d'enseignement supérieur et de recherche, les Grandes Écoles et le CNRS pour le dépôt systématique des publications sur HaL. Aujourd'hui, HaL regroupe 41 portails institutionnels⁵⁴. L'accord a pris fin en 2009, mais fait l'objet d'un renouvellement tacite. Cette démarche appuie la tendance à la création d'archives institutionnelles dans le paysage français. Tout comme le consortium COUPERIN, elle en est la cause et l'effet.

Si la politique de diffusion reste à charge des établissements, le développement technique, la maintenance et l'administration des plateformes sont assurés par le CCSD. Les plateformes nationales dédiées aux mémoires d'étudiants au niveau Master (MemSIC et DUMAS) sont elles aussi hébergées par HaL, et administrées par le CCSD.

Par sa forte centralisation et sa logique « descendante », le modèle français s'oppose aux modèles anglo-saxons fondés sur une logique mutualisée de type « fédération » où l'on part de la base pour s'organiser au niveau national dans un second temps⁵⁵. Citons par exemple le projet DEPOT aux Royaumes Unis, le projet DIVA en Suède, ou le projet DARE aux Pays Bas.

Le projet DARE⁵⁶, lancé en 2004, prévoit la construction d'un réseau d'archives ouvertes institutionnelles interconnectées. Les universités membres s'engagent à construire une archive avec le solution de leur choix, pourvu qu'elle soit compatible avec le protocole d'échange OAI – PMH. Elles assurent la diffusion des ressources et sont libres de déterminer les procédures de dépôt qui les intéressent. La Bibliothèque Nationale des pays Bas prend en charge le dépôt légal des documents et assure leur préservation au long terme au sein de bases de données adaptées.

53 <http://hal.archives-ouvertes.fr/>

54 Liste des différents portails HaL accessible à l'adresse suivante : http://hal.archives-ouvertes.fr/index.php?halsid=j5m0f74hpsk8nuo0ltbbqk023&action_todo=list_portail

55 Baruch. Voir bibliographie entrée [02]

56 <http://dare.uva.nl/en>

Comme l'explique Pierre Baruch, le contrôle du système par le CCSD faisait au départ l'objet de vives critiques de la part des acteurs français des archives ouvertes. En effet, la logique de centralisation, avec ses processus décisionnels de type « top-bottom » s'opposent en principe à la culture communautaire des archives ouvertes, et ses processus « bottom-up »⁵⁷. Néanmoins, un tel appareil, marqué par le relai systématique du pouvoir central, à le mérite d'harmoniser les initiatives, dans un contexte marqué par l'éparpillement.

2 - Le projet de l'IDIST : des procédures à formaliser

2.1 - Auto-archivage ou dépôt institutionnel : un choix organisationnel

Notre analyse de l'existant a fait ressortir deux types de procédures de dépôt impliquant des choix organisationnels différents pour l'IDIST :

1. L'auto-archivage : les déposants sont les auteurs des mémoires (MemSIC) ;
2. Le dépôt institutionnel : les déposants sont des référents universitaires autorisés par l'archive (DUMAS).

Les deux procédures possèdent leurs avantages et leurs inconvénients. L'auto-archivage permet aux établissements « d'externaliser » les tâches de collecte et le dépôt en tant que tel. Ce dernier est relativement rapide (10 minutes en moyenne) et ne nécessite aucune compétences particulières. Il consiste essentiellement en l'enregistrement de l'auteur dans la base (premier dépôt), le renseignement des métadonnées, et le chargement du fichier⁵⁸. Toutefois, systématisé à l'ensemble des mémoires « candidats à diffusion », cette étape représente une charge de travail potentiellement importante pour les UFR qui ne jouissent pas forcément des ressources nécessaires pour y procéder. L'auto-archivage mise également sur la bonne volonté des auteurs, et risque de diminuer la visibilité potentielle des formations.

57 Piotr. Voir bibliographie entrée [11]

58 Pour le détail des procédures de dépôt et des métadonnées à renseigner sur MemSIC et DUMAS, se référer à l'analyse des besoins.

Le dépôt institutionnel permet quant à lui de normaliser la procédure, ce qui, pour les établissements, offre un meilleur contrôle sur la qualité des mémoires diffusés, ainsi qu'une « gestion homogène et standardisée »⁵⁹ des fichiers (cohérence des métadonnées, harmonisation de la présentation via une feuille de style standard, etc.). Cette procédure peut donc servir de base à la mise en place d'une proto « charte éditoriale », et faciliter la création de « collections » pour les UFR.

2.2 – Quel circuit pour la modération ?

Toutefois, sur MemSIC comme sur DUMAS, la procédure est mixte : la mise en ligne des fichiers est soumise à la modération d'un administrateur, qu'ils aient été déposés par l'auteur ou par l'établissement de rattachement. La modération ne constitue en aucun cas une ingérence éditoriale de l'administrateur. Les archives du CCSD obéissent à un principe de neutralité et laissent aux établissements le soin de déterminer une politique de diffusion particulière. Ainsi par exemple, aucune note pallier n'est fixée. Sur ArchiveSIC, le système est même conçu selon un principe de « modération partagé » : les domaines scientifiques sont répartis entre plusieurs modérateurs qui peuvent faire opposition aux choix des autres pour juguler toute « velléité éditoriale »⁶⁰.

Cette modération est d'abord technique et concerne la lisibilité des fichiers⁶¹. Les seuls aspects thématiques concernent l'adéquation entre le contenu et le thème de l'archive (les sciences de l'information), et l'adéquation entre les métadonnées renseignées et le contenu du fichier : titre, domaine scientifique, mots-clés. La modération concerne enfin certains aspects juridiques et réglementaires :

1. Le mémoire doit avoir été soutenu (conformité au règlement d'utilisation de l'archivage),
2. L'exemplaire déposé doit être conforme à la version transmise au Jury pour la soutenance ;
3. Conformité du mémoire avec le droit d'auteur et la propriété intellectuelle (pas de plagiat).

MemSIC étant nettement moins utilisé qu'ArchiveSIC, la procédure n'obéit pas au principe de « modération partagée ». Après soumission du mémoire par l'auteur, le modérateur prend contact avec le directeur de stage ou le responsable de formation (métadonnées obligatoires à renseigner lors du dépôt) pour obtenir confirmation sur les points mentionnés précédemment. Ceux-ci sont

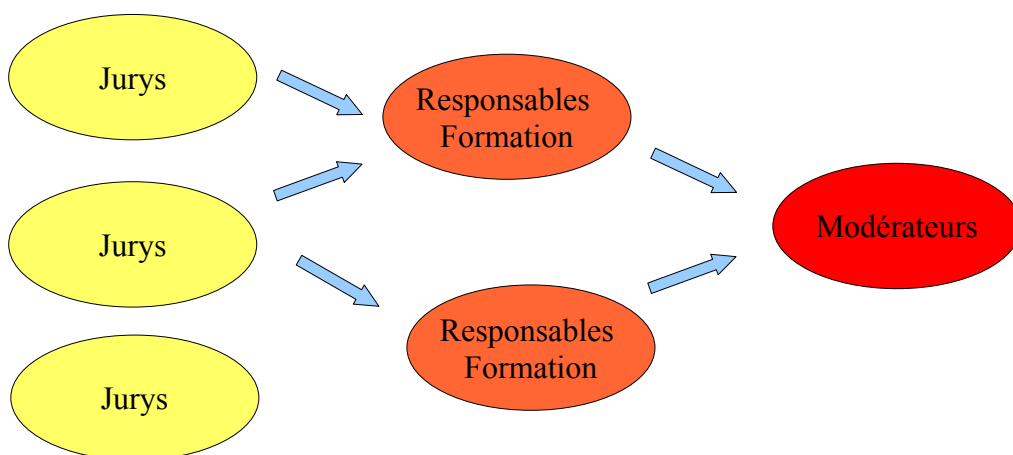
59 <http://www.couperin.org/archivesouvertes/spip.php?rubrique6>

60 Gallezot. Voir bibliographie entrée [06]

61 Pour le détail de la modération, se référer à l'analyse des besoins

chargés de faire l'interface entre le modérateur et l'auteur si des corrections sont nécessaires (les refus sont très rares).

Aussi, il semble important de rationaliser la procédure pour éviter les échanges inutiles. La solution la plus simple consisterait à transmettre au modérateur la liste des titres des mémoires autorisés, les noms d'auteurs et leur adresses mails par le biais d'un circuit officiel que l'on peut formaliser par le schéma suivant :



Ainsi, l'administrateur pourrait valider directement les mémoires soumis en vérifiant la liste préalablement contrôlée.

3 - Le projet de l'UFR : perspectives de développement

3.1 – Les contraintes de l'environnement : sécurité et continuité institutionnelle

Le projet d'archive ouverte de l'IDIST s'insère dans un environnement technique et politique restrictif. En effet, celui-ci s'intègre dans la « stratégie d'établissement » de Lille 3. Par conséquent, le projet doit recevoir l'aval de la Présidence et du CRI (Service Informatique de l'Université de Lille 3). Or, la priorité du CRI n'est pas aujourd'hui de proposer des services à valeur ajoutée mais de développer un système d'information protégé, stable et cohérent. Ainsi, pour des raisons de sécurité réseau et de continuité institutionnelle, l'ensemble des projets web initiés par les composantes de Lille 3 ont vocation à être absorbés par le système d'information. Ceci vaut aussi pour les ressources numériques portant le label des UFR qui doivent être hébergées sur les serveurs de Lille 3.

Ces contraintes posent les questions suivantes :

1. Intégration ou non de l'archive au système d'information ?
2. Développement d'une « application maison » (via un SIGB) ou recours à un logiciel d'archive existant (HaL, Eprints, Dspace, Digital Commons, etc.) ?
3. Hébergement en local ou à l'extérieur sur un service existant (MemSIC ou DUMAS) ?

Ces questions, formulées en termes techniques, se résument en une : stratégiquement, faut-il déposer les ressources sur un serveur national ou local ?

En France comme ailleurs, la tendance actuelle est à la constitution d'archives institutionnelles locales. Lille 3 n'échappe pas à la règle : un projet de constitution d'une archive en local est en cours de discussion. A terme, cette archive pourrait d'ailleurs être étendue à l'ensemble des universités de Communauté Urbaine de Lille (Lille 1 et à Lille 2).

Cependant, ce type de projet nécessite un fort investissement et s'inscrit dans le temps long : nombreuses contraintes de développements à surmonter, processus décisionnels soumis à la chaîne de validation institutionnelle, poids du facteur politique, etc. Aussi, L'UFR IDIST ne souhaite pas initier un tel projet. En effet, son objectif est d'impulser une dynamique d'auto-archivage suffisante pour atteindre rapidement, facilement et à moindre frais une « masse critique » raisonnable.

De plus, un projet aussi lourd doit faire l'objet d'un positionnement adéquate pour ne pas être un échec (sous-utilisation du dispositif). Pour ce faire, la conception doit se fonder sur des « retours d'usages ». Or, les étudiants ne déposant aujourd'hui pas leurs travaux, il est impossible d'en obtenir. Dans un premier temps, il semble donc plus stratégique de maintenir le projet à un stade expérimental : une archive de l'IDIST à destination des étudiants de l'IDIST afin d'observer les usages.

3.2 - Arguments en faveur du dépôt sur une plateforme nationale (MemSIC)

1. Un investissement quasi nul

Les services du CCSD sont gratuits, et la solution existe déjà. En d'autres termes, les coûts de développement seraient minimes (tout dépend du degrés d'intégration au système d'information de l'université)⁶². Par ailleurs, le CCSD se charge de la maintenance technique et de l'administration du site. Enfin, si le projet s'oriente vers l'auto-archivage des étudiants, les tâches de collecte et le dépôt en tant que tel seront délégués aux étudiants. En d'autres termes, ce projet nécessiterait un investissement financier et humain minime.

2. Un meilleur « ranking » institutionnel

Le CNRS est un des organismes de recherche les plus important au niveau international. Selon un récent classement webometrics des 400 archives ouvertes les plus visibles, 3 portails HaL arrivent dans les 10 premiers résultats au classement général et sur l'indice de visibilité : HaL – CNRS, (2nde position), immédiatement suivi d'HaL – INRIA (3ème position), et HaL-SHS (8ème position).

62 Voir *infra* 3.5 - L'intégration partielle au S.I : un compromis pragmatique ? (p53)

| | taille | visibilité | richesse | Classement général) |
|-------------|--------|------------|----------|---------------------|
| HaL - CNRS | 8 | 3 | 3 | 2 |
| HaL - INRIA | 17 | 4 | 19 | 3 |
| HaL - SHS | 80 | 5 | 109 | 8 |

« Ranking » mondial de HaL (janvier 2010, Cybermetrics Lab, CSIC, Espagne)⁶³

En cas d'hébergement sur les serveurs du CCSD, les fichiers profiteraient de cette très bonne identification institutionnelle.

3. Trafic et Référencement

L'économie des réseaux procèdent selon un cercle vertueux : plus un site génère de trafic, plus le nombre de requêtes sur les fichiers est important, meilleur est leur référencement. Contrairement aux idées reçues, un petit nombre de dépôts ne garantit pas une meilleure visibilité – distinction des ressources. D'une certaine manière, le contraire se produit : plus elles se fondent dans un dispositif important, plus elles ont de chance de remonter dans les pages de résultats des moteurs généralistes. De ce fait, la version 2 de MemSIC, une archive HaL dédiée à l'ensemble des productions non « conventionnelles » ou « typées » devrait favoriser le référencement des fichiers. En effet, MemSIC est aujourd'hui sous-utilisé et présente un trop petit nombre de dépôts pour que les mémoires soient correctement référencés par les moteurs.

Enfin, MemSIC est compatible au protocole d'échange OAI-PMH, ses métadonnées bibliographiques sont donc moissonnées par les méta-moteurs d'archives ouvertes.

4. Un hébergement réactif

Le CCSD utilise une puissante infrastructure réseau. En effet, les archives HaL, comme DUMAS et MemSIC, sont hébergées, entre autre, sur les serveurs de l'IN2P3 (Institut de Physique Nucléaire et de Physique des Particules – CNRS)⁶⁴. Ces serveurs ont une capacité de calcul importante et

⁶³ http://repositories.webometrics.info/top400_rep_inst.asp

⁶⁴ <http://www.in2p3.fr/>

peuvent gérer de gros volumes de transactions. Aussi, utiliser cette infrastructure *via* MemSIC peut induire un confort d'utilisation aux usagers, en offrant un temps de consultation quasi instantanée. Ce critère ne jouera sans doute pas tout de suite dans le projet de l'UFR. Néanmoins, si le projet évolue en une archive locale pour Lille 3, Lille 2 et Lille 1 intégrée au système d'information, un hébergement sur le serveur du CNRS pourrait permettre de répartir la charge.

5. Sauvegarde au long terme des fichiers

Jusqu'à peu, la plateforme commune HaL n'intégrait pas de dimension patrimoniale, contrairement à ses homologues européens comme DARE ou DIVA. Les seules procédures mises en place pour assurer la conservation des fichiers consistaient à générer une version XML et une DTD des documents pour assurer leur portage vers d'autres systèmes. Ceci posait la question de leur sauvegarde au long terme (30 ans et plus) sur des supports adaptés.

Cependant, la signature récente d'un accord entre le CINES (Centre Informatique National d'Enseignement Supérieur) et le CCSD dans le cadre du projet TGE-ADONIS devrait à l'avenir assurer la sauvegarde à long terme des fichiers selon des procédures spécifiques. L'opération a été lancée le 3 avril 2010, 45 000 documents sur les 150 000 documents en texte intégral totalisé par HaL ont déjà été traités.

3.3 - Temps court / temps long : deux démarches opposées mais complémentaires

En analysant les stratégies de développement des établissements, l'enquête COUPERIN 2007 fait émerger deux types de démarche - projet :

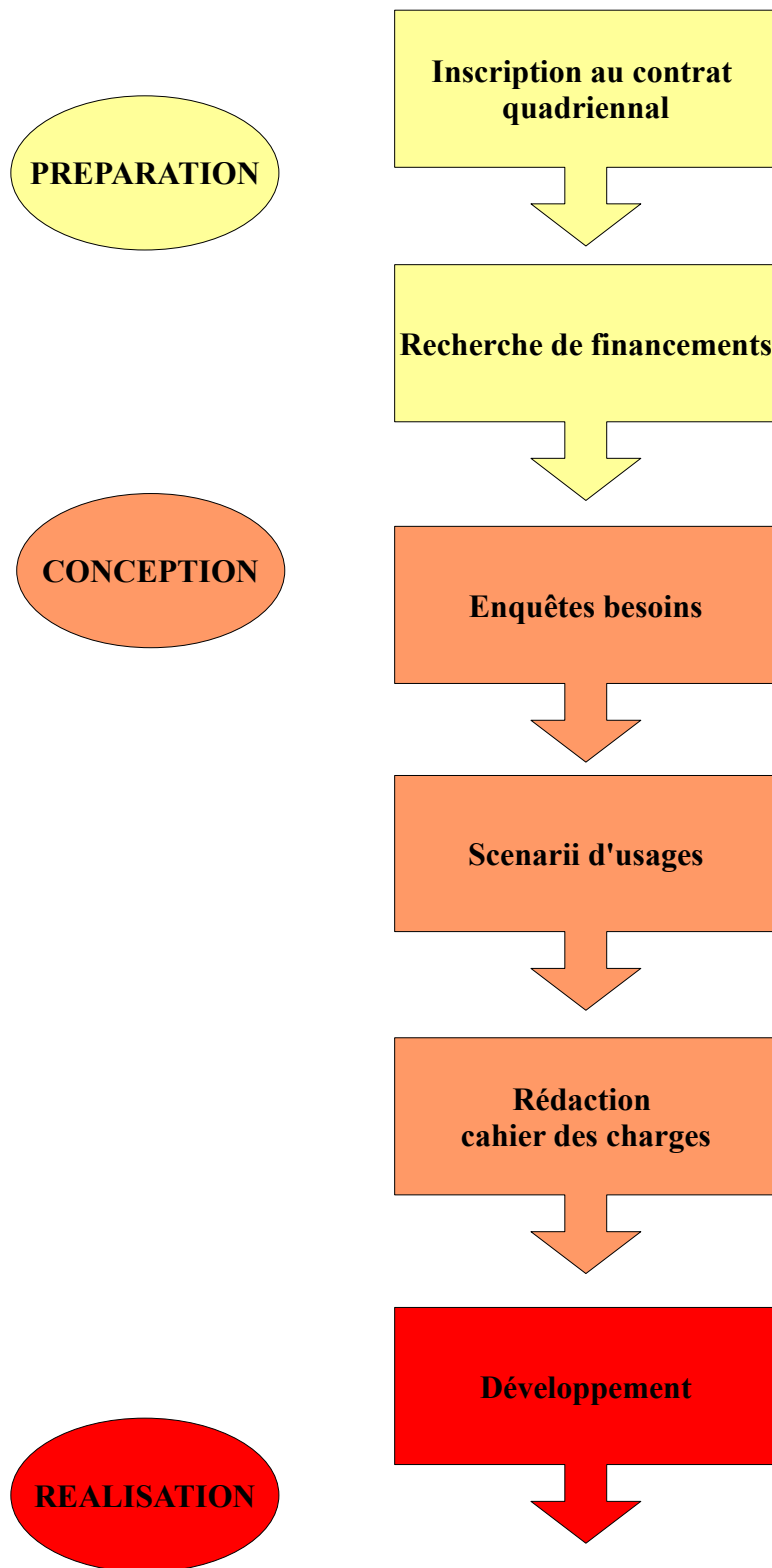
1. Une démarche empirique, s'inscrivant dans le temps court (9 mois maximum).

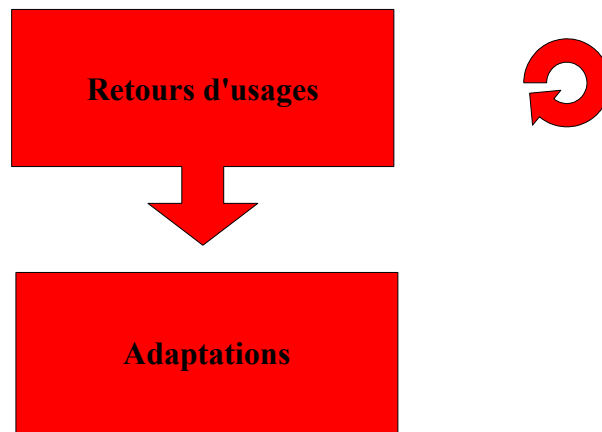
Cette démarche repose sur l'utilisation d'un logiciel clé en main type HAL, avec parfois quelques adaptations : francisation, personnalisation de l'interface. Cette démarche suit le processus suivant :



2. Une démarche structurée, s'inscrivant dans le temps long (2 ans ou plus).

Cette démarche implique généralement l'intégration des archives ouvertes aux ENT des établissements. Elle suit le processus suivant :





Cependant, ces deux démarches ne sont pas incompatibles : « certains établissements démarrent avec un projet simple, adoptent une technologie parce qu'existante et disponible (par exemple HAL) et poursuivent par ailleurs une démarche projet s'inscrivant dans un temps plus long (intégration dans l'ENT). » Ainsi, une complémentarité peut être trouvée entre le projet de l'IDIST et celui de l'université.

Les acteurs de l'université l'ont bien compris : Patrice de La Broise, chargé de mission à la « professionnalisation » pour le service de la Présidence suit le projet de l'IDIST et le considère comme un test reproductible à plus grande échelle. Quelque soit la formule adoptée, deux types d'archives peuvent cohabiter de manière complémentaire et non concurrentielle, pour peu qu'on délimite clairement leur territoire :

1. Plusieurs plateformes (une par UFR) hébergées sur un serveur national (temps court) : auto-archivage des étudiants ;
2. Une plateforme centrale propre à Lille 3 hébergée en locale (temps long) : dépôt par l'institution ou auto-archivage des étudiants, préservation au long terme, mise en « collection », etc.

En outre, si le protocole d'échange OAI-PMH est respecté et que les bases sont interconnectées, les notices peuvent être automatiquement moissonnées. Les mémoires déposés dans l'une sont visibles dans l'autre.

3.5 - L'intégration partielle au S.I : un compromis pragmatique ?

L'enquête COUPERIN 2007 révèle également qu'il existe différents degrés d'intégration des archives aux systèmes d'information. A l'arrivée, seulement 20 % des établissements ont procédé à une intégration totale, et 53 % à une intégration partielle. Parmi ces derniers, 20 % estiment leur archive « bien intégrée », 22 % « moyennement intégrée », et 11 % « peu intégrée ».

Pour le projet de l'UFR, une intégration partielle offrirait un compromis pragmatique entre les prérogatives de l'environnement institutionnel, et les avantages du dépôt sur un serveur national.

Cette intégration partielle pourrait prendre la forme d'un « interfaçage » en local. Les ressources seraient ainsi accessibles depuis le site de l'UFR IDIST, mais hébergées sur les serveurs du CCSD. Là encore, le développement peut être plus ou moins poussé :

1. Interface légère : seul les modules de requêtes et consultation sont accessibles.

Il s'agit a priori de la solution la plus simple et rapide à développer. Si les modalités techniques restent à déterminer, l'opération consisterait normalement en l'importation d'une brique du code de MemSIC dans le site de l'UFR.

2. Interface complète : développement d'une interface personnalisée pour l'interrogation, la consultation, la navigation et le dépôt dans l'archive.

Quoi qu'il en soit, l'UFR devra opérer une clarification juridique et technique en partenariat avec le CCSD et le Service Informatique de l'université pour savoir ce qu'il est possible et légal de faire. En principe, HaL offre aux institutions la possibilité d'utiliser la solution pour créer des « guichets » personnalisés vers des ressources hébergées dans d'autres bases HaL. De plus, les serveurs du CCSD peuvent prendre en charge de gros volumes de transaction. Enfin, si la seconde solution est adoptée (interface complète), il faudra déterminer la marge de manœuvre pour sa personnalisation.

3.6 - Quelle marge de manœuvre pour la personnalisation ?

MemSIC n'est pas une archive institutionnelle mais thématique. Son objectif n'est pas la valorisation d'une institution mais d'une discipline. Contrairement à DUMAS ou à HaL, MemSIC

ne propose pas de service de personnalisation. Les institutions ne peuvent pas rassembler leurs productions sous la forme de « collections » tamponnées sur un sous-site paramétrable : charte graphique, logos, animations, texte de présentation, etc. L'auto-archivage des étudiants sur MemSIC servira d'abord à valoriser la production des étudiants. L'UFR y gagnera en visibilité, mais son label ne sera pas mis en valeur.

Si une intégration partielle au système d'information via un logiciel clé en main comme HaL est retenu, il pourrait être intéressant de personnaliser l'interface, ceci sur deux niveaux :

1. Aspects formels : charte graphique en cohérence avec la communication de l'IDIST, disposition générale des contenus ;
2. Fonctionnalités de recherche, d'indexation, d'import-export des données.

Nous nous arrêtons ici sur les fonctionnalités. Tout d'abord, MemSIC comme DUMAS présentent généralement des fonctionnalités de recherche classiques (les applications sont très similaires), dans la lignée des OPAC et des bibliothèques numériques. Ainsi, trouve-on une première entrée de type annuaire (« browse »)⁶⁵, et une seconde de type automatique (« search »)⁶⁶.

En revanche, on ne trouve aucun critères pour spécifier le type de mémoire (stage ou recherche), ou de lieu de stage selon une typologie par secteur d'activité ou par raison sociale. Or, nous savons que le fonds de l'IDIST consiste principalement en des mémoires de stages. Dans cette optique, il pourrait être intéressant d'adapter les critères de recherche à la spécificité du fonds pour répondre aux besoins informationnels de certains usagers en recherche de stage par exemple. Reconnaissons que ce n'est toutefois pas le rôle d'une archive ouverte que de proposer un annuaire de lieux de stages détourné.

En outre, dans le cadre de la préparation à la communication pour le CIUEN 2010, les observations de Sophie Chauvin ont souligné différentes lacunes des archives ouvertes sur le sujet.

On ne trouve aucune fonctionnalité offrant la possibilité de parcours transversaux fondés sur les usages émergents. Citons notamment l'exploration par « tags » ou « nuages de tags », et les outils de

65 Consultation par auteur, institutions, date de publication, etc. Pour le détail des fonctionnalités et critères de recherche sur MemSIC et DUMAS, se référer à l'analyse des besoins.

66 Recherche simple et avancée (multi-critère. Idem.

visualisation de l'information (ex : moteurs de recherche ou thésauri cartographiques).

Par ailleurs, on ne trouve aucune aide à l'indexation lors de saisie des métadonnées (mots-clés). L'indexation libre soulève, à terme, la question de la pertinence du système d'information. Pour l'instant, le fonds déposé n'est pas assez important pour faire de MemSIC un système d'information « bruyant ». Si l'auto-archivage des étudiants de l'IDIST, de Lille 3 ou d'ailleurs venait à se développer, l'accès aux mémoires s'en trouverait compliqué. Dans cette optique, il pourrait être intéressant d'établir une aide à l'indexation sur deux niveaux :

1. Institutionnel : indexation contrôlée par un thésaurus adapté aux spécificités du fonds ;
2. Communautaire : indexation libre sous forme de « tags », avec suggestion des « tags » les plus populaires.

Notons que les folksonomies aideraient à amplifier le signalement des mémoires en s'appuyant sur le référencement par les communautés : sites de « *micro-blogging* » (Twitter), sites de « *social bookmarking* » (Delicious, Diigo), systèmes de recommandation divers et variés.

Enfin, MemSIC, et c'est une caractéristique que la plateforme partage avec beaucoup d'archives institutionnelles, ne sont pas des espaces de travail, mais de simples lieux de visibilité et de consultation des documents. Plus particulièrement, on ne trouve aucune fonctionnalité permettant d'exporter les métadonnées bibliographiques, citations et références vers un système de gestion (CiteYouLike ou Zotero). On ne trouve pas non plus d'espaces contributif permettant d'enrichir les documents pour rédiger des commentaires, ou des annotations à titre personnel ou public.

Bibliographie

- [01]AUBRY, Sara. DUPLOUY, Laurent. *Open Archive Initiative*. Groupe PIN. Département de la bibliothèque numérique, BNF (2004). Présentation disponible à l'adresse suivante : http://www.aristote.asso.fr/PIN/presentations/2004/oai_bnf.pdf
- [02]BARUCH, Pierre. *La diffusion libre du savoir Accès libre et Archives ouvertes*. In: *L'Archicube*, no. 3, dossier "L'édition et le Numérique", Association des anciens élèves, élèves et amis de l'Ecole normale supérieure, Paris; no. 3, p. 77-95 (2007)
- [03]COMBEROUSSE, Martine, « La littérature grise ». In : *BBF*, 1993, n° 5, p. 60-62 (1993). Consulté le 01 juin 2010. Accessible en ligne à l'adresse suivante : <http://bbf.enssib.fr/consulter/bbf-1993-05-0060-016.pdf>
- [04]DORSTTER, Laura. *Le concept d'originalité dans la législation française du droit d'auteur et dans celle du copyright anglais*. Article publié sur le blog juridique M2BDE. Université de Paris Ouest. posté par Jean-Sylvestre Bergé le 8/01/2009. Consulté le 01/06/2010. Disponible en ligne à l'adresse suivante : <http://m2bde.u-paris10.fr/blogs/dpi/index.php/post/2009/01/08/Le-concept-doriginalite-dans-la-legislation-francaise-du-droit-dauteur-et-dans-celle-du-copyright-anglais-par-Laura-DORSTTER>
- [05]GALLEZOT, Gabriel. *Le Libre Accès (Open Access) : partager les résultats de la recherche*. In : *Colloque international : L'information numérique et les enjeux de la société de l'Information*. Tunis, 14-16 Avril 2005. ISD, Tunis : France (2005)
- [06]GALLEZOT, Gabriel. *ArchiveSIC, Archive Ouverte en Sciences de l'Information et de la Communication : rôle, fonctionnement et usage*. In : *Guide Pratique, Les publications électroniques*. Archimag. (2003)
- [07]GALLEZOT, Gabriel. *Archives Ouvertes : définition et constat français*. Schedae, 2008, prépublication n° 3, (fascicule n° 1, p. 23-34). Disponible en ligne à l'adresse suivante : <http://www.unicaen.fr/services/puc/ecriture/preprints/preprint0032008.pdf>
- [08]GTAO, Groupe de Travail Archive Ouverte. *Enquête 2007 : panorama des réalisations et des projets*. Consortium Couperin. 2007. Disponible en ligne à l'adresse suivante : <http://www.couperin.org/archivesouvertes/spip.php?article118>
- [09]HUDON, Michèle. *Le contrôle bibliographique*. In : *Terminologie de Base des Sciences de l'Information*. Université de Montréal. [Consultée le 31/05/2010] Disponible en ligne à l'adresse suivante : <http://www.ebsi.umontreal.ca/termino/00000062.htm>
- [10]KERIGUY, Jacques. *L'accès à la littérature grise*. *BBF*, 1984, n° 2, p. 138-143. Article consulté le 22 mai 2010, disponible en ligne à l'adresse suivante : <http://bbf.enssib.fr/consulter/bbf-1984-02-0138-001>
- [11]PIOTR. *La « French Touch » des archives ouvertes*. Article publié sur blogo-numericus. Posté le 14/10/2006. Consulté le 20/05/2010. Disponible en ligne à l'adresse suivante : http://www.ccsd.cnrs.fr/IMG/pdf/RevueDePresse/Blogo_Numericus_La_french_touc

[h_des_archives_ouvertes.pdf](#)

- [12]SCHOPFEL, Joachim. *Le devenir de la littérature grise. Quelques observations*. Article publié dans la revue *Perspectives Documentaires en Éducation*, 2006, n°62. Disponible en ligne à l'adresse suivante :http://archivesic.ccsd.cnrs.fr/index.php?halsid=ku39o76955rbl8gr1462uki585&view_this_doc=sic_00136829&version=1
- [13]SCHOPFEL, Joachim. PROST, Hélène. *Usage of grey literature in open archives : state of the art and empirical results*. In : *GL11 Eleventh International Conference on Grey Literature: The Grey Mosaic - Piecing It All Together*. Washington :14-15 December 2009. Disponible en ligne à l'adresse suivante :
http://archivesic.ccsd.cnrs.fr/index.php?halsid=ku39o76955rbl8gr1462uki585&view_this_doc=sic_00480308&version=1

Annexes

Synthèse de l'inventaire

Document de référence : procès verbaux des soutenances (P.V)

- Mémoires « sélectionnés » : note > ou = à 14/20 ;
- Mémoires « autorisés » : note > ou = à 14/20, autorisation de l'étudiant et de l'organisme lieu de stage pour diffusion sur le web ;
- Mémoires « candidats à diffusion » : note > ou = à 14/20, PV incomplets (champ « autorisation » absent ou non renseigné) ;
- Mémoires « bons pour diffusion » : note > ou = à 14/20, mémoire autorisé, exemplaire électronique acquis.

Chiffres généraux

- Sur les 298 mémoires sélectionnés :
- 168 mémoires « autorisés »
→ Reste à collecter les exemplaires électroniques ;
- 129 mémoires « candidats » :
→ Reste à contacter les auteurs pour demander autorisation et exemplaire électronique.

Représentation des formations

Sur l'ensemble des mémoires sélectionnés (« autorisés » et « candidats » compris) :

- 171 M1 ;
- 127 M2 / DESS (année 2003-2004).

Parmi les M2 – DESS, on trouve :

- 49 GIDE : 20 autorisés, 29 candidats ;
- 36 IDEMM : 25 autorisés, 11 candidats ;
- 6 PRISME : 5 autorisés, 1 candidat ;

- 14 DESS GPIO (2003-2004) : 6 autorisés, 8 candidats ;
- 9 DESS GPSI : 1 autorisé, 8 candidats ;
- 12 DESS MSMM : 2 autorisés, 10 candidats.

14 mémoires « bons pour diffusion » :

- 10 mémoires de M1 (année 2008-2009) ;
- 4 mémoires de M2 GIDE (2008-2009).

Méthodologie du questionnaire

Dimensions principales abordées par le questionnaire

1. Identité – cursus (questions 1 et 2)

2. Pratiques de consultation (questions 3 à 10)
 - Type de mémoire le plus souvent consulté ;
 - Support le plus souvent consulté ;
 - Fréquence de consultation ;
 - Type d'information recherché, par ordre d'importance ;
 - Circonstances de consultation ;
 - Modalités d'accès ;
 - Critères d'évaluation pour la qualité – fiabilité des mémoires, par ordre d'importance.

3. Pratiques de diffusion sous forme électronique (questions 11 à 17)
 - Dispositifs utilisés pour dépôt ;
 - Circonstances – personne à l'initiative du dépôt ;
 - Conditions souhaitées pour la diffusion des mémoires au format électronique ;
 - Objections pour la diffusion des mémoires ;

Seuls deux étudiants ayant déjà diffusé leurs mémoires sur le web, les réponses de cette partie relèvent surtout du déclaratif.

Échantillon

Au final, 27 étudiants ont répondu au questionnaire : 20 M1 et 7 M2. Plus particulièrement :

- 17 M1 ICD ;
- 1 M1 en biochimie moléculaire ;
- 6 M2 orientés « urbanisme » ;
- 2 M2 orientés « gestion du patrimoine » ;

- 1 M2 orienté « management de projet ».

Modalités de collecte

Questionnaire en ligne réalisé au moyen de l'outil de sondage Google Documents, du 10 mai au 1er juin 2010. Formulaire envoyé par mail. Enregistrement automatique des réponses dans une feuille de calcul Excell.

Le questionnaire : « pratiques de consultation et de diffusion des mémoires des étudiants M1 et M2 »

1 - En quelle année êtes vous inscrit actuellement ?

2 - Quelle est votre formation actuellement ?

Donnez l'intitulé approximatif de votre formation - université de rattachement.

3 - Avez-vous déjà consulté un mémoire d'étudiant au niveau Master 1 ou 2 ?

Si non, veuillez cocher et passer directement à la question 11.

Oui

Non

4 - Quel(s) type(s) de mémoire(s) avez vous consulté ou consultez vous généralement ?

Vous pouvez cocher plusieurs réponses.

Recherche

Stage

Autre :

5 - Sous quelle forme avez vous consulté ou consultez vous le plus souvent des mémoires d'étudiants ?

- Papier
- Numérique

6 - A quelle fréquence avez-vous consulté ou consultez vous des mémoires d'étudiants ?

- Une fois seulement
- De temps en temps
- Régulièrement

7 - Quelles informations recherchez ou recherchez-vous, par ordre d'importance?

| | Principalement | Secondaires | Pas du tout |
|---|-----------------------|-----------------------|-----------------------|
| Contenus scientifiques | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Références bibliographiques | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Informations méthodologiques | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Exemple de sujet de recherche | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Exemples d'organismes - missions de stage | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

8 - Dans quelles circonstances avez-vous consulté ou consultez-vous le plus souvent ces mémoires ?

Vous pouvez cocher plusieurs réponses.

- Pour la rédaction de votre propre mémoire

- Pour un travail à rendre en cours d'année
- Pour vous aider à trouver un sujet de recherche
- Pour vous aider à trouver un stage
- Dans le cadre d'une démarche d'orientation universitaire ou professionnelle
- Autre :

9 - Comment vous êtes vous procuré ce ou ces mémoires le plus souvent ?

Vous pouvez cocher plusieurs réponses.

- Par l'auteur du mémoire lui même
- Par un étudiant connaissant l'auteur et/ou possédant un exemplaire de son mémoire
- Par un professeur, enseignant-chercheur ou membre du personnel administratif de votre composante de formation
- A la bibliothèque de votre UFR - Université
- Sur un site web
- Autre :

10 - Comment estimez vous l'importance des critères suivants pour évaluer la qualité - fiabilité d'un mémoire ?

| | Critère Important | Critère Secondaire | Pas important |
|--|-----------------------|-----------------------|-----------------------|
| Note obtenue | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Vous connaissez l'auteur | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Vous connaissez le directeur de mémoire | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Précision - exhaustivité des références bibliographiques | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

11 - Avez-vous déjà diffusé votre ou vos mémoires d'étudiant sur le web (stage ou recherche)?

Si non, veuillez cocher et passer directement à la question 14.

12 - Qui était à l'origine de cette décision ?

Vous pouvez cocher plusieurs réponses.

- Vous même
- Votre directeur de mémoire ou de stage
- Votre responsable de formation
- La bibliothèque de votre UFR - Université
- Autre :

13 - Sur quel(s) dispositif(s) ? Vous pouvez cocher plusieurs réponses.

- Intranet de votre UFR - Université
- Site web de votre UFR - Université
- Blog, site internet d'un enseignant-chercheur
- Votre Blog, site web, page personnelle
- Site à caractère commercial
- Autre :

14 - Pensez-vous que la diffusion sur le web des mémoires d'étudiants réussis en texte intégral puisse valoriser leur parcours universitaire ?

Par exemple, dans le cadre d'une candidature à une offre de stage, d'emploi ou à une formation sélective.

15 -Si votre UFR décidait de diffuser sur le web les mémoires de ses étudiants en texte intégral, l'autoriseriez-vous?

Si non, veuillez cocher et aller directement à la question 17.

16 - Si oui, à quelle(s) condition(s)?

Vous pouvez cocher plusieurs réponses. Une fois terminé, vous pouvez envoyer le questionnaire sans répondre à la question 17.

- Une note pallier (ex : sup. ou égale à 14/20)
- Accès restreint aux étudiants et enseignants-chercheurs de votre formation - discipline
- Protection des fichiers en écriture (copier-coller, impression impossible, etc.)
- Aucune condition particulière
- Autre :

17 - Si non, pour quelle(s) raison(s)?

Vous pouvez cocher plusieurs réponses. Une fois terminé, vous pouvez envoyer le questionnaire.

- Protection de votre vie privée - données personnelles ou confidentielles
- Vous pensez qu'en général, un mémoire d'étudiant n'est pas d'assez bonne qualité pour être

diffusé

Vous ne voulez pas être cité par d'autres étudiants ou enseignants-chercheurs dans leurs travaux

Vous avez peur d'être plagié

Autre :

Les résultats

Pratiques de consultation (questions 3 à 10)

Sur 27 étudiants, 26 ont déjà consulté un mémoire de Master 1 ou 2, quelque soit son type (stage ou recherche).

Types de mémoires les plus souvent consultés

| Stage seulement | Recherche seulement | Stage ou recherche |
|-----------------|---------------------|--------------------|
| 2 | 12 | 12 |

(/26 étudiants)

Supports de consultation les plus souvent utilisé

| Papier | Numérique |
|--------|-----------|
| 13 | 13 |

(26 étudiants)

Fréquence de consultation

| Une fois seulement | De temps en temps | Régulièrement |
|--------------------|-------------------|---------------|
| 6 | 18 | 2 |

(/26 étudiants)

Types d'informations recherchés, par ordre d'importance

| | Principalement | Secondaire | Pas du tout |
|------------------------------|----------------|------------|-------------|
| Contenus Scientifiques | 13 | 9 | 3 |
| Références bibliographiques | 10 | 9 | 6 |
| Informations méthodologiques | 17 | 5 | 6 |
| Exemples de sujet de | 8 | 10 | 7 |

| | | | |
|---|---|---|----|
| recherche | | | |
| Exemples de missions – organismes de stage | 6 | 8 | 11 |

(/25 étudiants)

Circonstances de consultation les plus fréquentes

| | | | | | |
|---|---|---|------------------------------------|---|-------|
| Pour la rédaction de votre propre mémoire | Pour un travail à rendre en cours d'année | Pour vous aider à trouver un sujet de recherche | Pour vous aider à trouver un stage | Dans le cadre d'une démarche d'orientation universitaire ou professionnelle | Autre |
| 24 | 10 | 4 | 0 | 3 | 1 |

Autre : « par curiosité »!

(/26 étudiants)

Modalités d'accès les plus fréquente

| | | | | | |
|----------------------------------|---|---|---|-----------------|-------|
| Par l'auteur du mémoire lui même | Par un étudiant connaissant l'auteur et/ou possédant un exemplaire de son mémoire | Par un professeur, enseignant-chercheur ou membre du personnel administratif de votre composante de formation | A la bibliothèque de votre UFR - Université | Sur un site web | Autre |
| 4 | 0 | 6 | 8 | 13 | 3 |

Autre : centre de documentation ou personne membre du lieu de stage

(/26 étudiants)

Critères d'évaluation de l'information, par ordre d'importance

| | | | |
|--------------------------|-------------------|--------------------|---------------|
| | Critère important | Critère secondaire | Pas important |
| Note obtenue | 12 | 9 | 4 |
| Vous connaissez l'auteur | 3 | 9 | 14 |
| Vous connaissez le | 7 | 8 | 9 |

| | | | |
|--|----|---|---|
| directeur de mémoire | | | |
| Précision - exhaustivité des références bibliographiques | 12 | 9 | 4 |

(/26 étudiants)

Pratiques de diffusion sous forme électronique (questions 11 à 17)

2 étudiants sur 27 déclarent avoir déjà diffusé leur mémoire sur le web, tout deux à leur initiative personnelle. Un seul a précisé le type de dispositif : à la fois sur un « site à caractère commercial » et sur un « blog, site web, page personnel ».

26 étudiants sur 27 pensent que la diffusion sur le web des mémoires d'étudiants réussis en texte intégral peut valoriser le parcours universitaire des auteurs.

1 étudiant sur 27 pense que la diffusion sur le web des mémoires d'étudiants réussis en texte intégral peut valoriser le parcours universitaire des auteurs.

26 étudiants sur 27 seraient prêts à autoriser la diffusion de leur mémoire sur le web en texte intégral si celui-ci était « réussi ».

1 étudiants sur 27 n'autoriserait pas la diffusion de son mémoire sur le web en texte intégral si celui-ci était « réussi ».

Conditions souhaitées pour la diffusion sur le web

| Une note pallier (ex : sup. ou égale à 14/20) | Accès restreint aux étudiants et enseignants-chercheurs de votre formation - discipline | Protection des fichiers en écriture (copier-coller, impression impossible, etc.) | Aucune condition particulière | Autre |
|---|---|--|-------------------------------|-------|
| 15 | 5 | 13 | 4 | 4 |

Autre : accord préalable des auteurs (2), anonymisation des données personnelles (1), attention aux

clauses de confidentialité (1).

(/26 étudiants)

Raisons invoquées pour l'objection à diffusion

| Protection de votre vie privée - données personnelles ou confidentielles | Vous pensez qu'en général, un étudiant n'est pas d'assez bonne qualité pour être diffusé | Vous ne voulez pas être cité par d'autres étudiants ou enseignants-chercheurs dans leurs travaux | Vous avez peur d'être plagié | Autre |
|--|--|--|------------------------------|-------|
| 1 | 1 | 0 | 1 | 0 |

(/1 étudiant)

Résumé

Les mémoires de Master sont des « ressources grises » produites en quantité. Or, leur potentiel documentaire est rarement exploité par les universités. L'accès aux exemplaires imprimés est limité. Leur diffusion au sein des archives ouvertes est insuffisante. Pourtant, les modèles du libre accès et du dépôt électronique semblent aujourd'hui les seuls capables d'offrir à la « littérature grise » la visibilité et l'impact qui lui manquent. En s'appuyant sur l'expérience menée à l'UFR IDIST, ce mémoire propose d'analyser la particularité des mémoires de Master par rapport aux autres « ressources grises ». Il se penche aussi sur certains aspects techniques et procéduraux spécifiques à la mise en place d'une archive ouverte en contexte institutionnel.

Mots-clés

littérature grise ; ressources grises ; matériaux gris ; mémoires d'étudiants ; mémoires de Master ; Archives ouvertes : archives ouvertes institutionnelles ; MemSIC ; CCSD ; OAI