

Ifremer
Service de la documentation et des bibliothèques
Bibliothèque La Pérouse (<http://www.ifremer.fr/blp/>)

Auteur : [Fred Merceur](#)
Référence : R.INT.BLP/FM/2005-4
Version : version 1.00
Mise à jour : 23 novembre 2005

Archimer,

ou la mise en place d'une Archive Institutionnelle à l'Ifremer

Table des matières

1. Introduction	4
2. Les archives institutionnelles	4
2.1. Contexte national et international	4
2.2. Intérêt pour l'Ifremer.....	5
2.2.1. Soutenir le mouvement Open Access	5
2.2.2. Valoriser la production scientifique.....	5
2.2.3. Créer une nouvelle base de données dédiée aux sciences de la mer	6
2.2.4. Renouer des liens entre les équipes de recherche et les bibliothèques.....	6
2.2.5. Améliorer la visibilité du site Internet de l'Ifremer.....	6
3. Principes généraux	7
3.1. Choix de la plateforme de développement.....	7
3.2. Types de documents enregistrés.....	7
3.3. Format de diffusion	7
3.4. Conservation des documents.....	7
3.5. Modalités d'enregistrement des documents.....	8
3.5.1. Enregistrement des thèses, des actes de congrès ou des rapports internes	8
3.5.2. Enregistrement des publications récentes	8
3.5.3. Enregistrement des articles publiés entre 2000 et 2005.....	9
4. Description fonctionnelle	9
4.1. La base de données « Bibliométrie »	9
4.1.1. Considérations générales.....	9
4.1.2. Description technique du projet.....	9
4.2. Archimail	11
4.3. Site WEB d'administration.....	13
4.4. Site WEB de consultation.....	18
4.4.1. Page d'accueil	18
4.4.2. Mode de recherche simplifiée	19
4.4.3. Mode de recherche avancée.....	21
4.4.4. Feuilletage	22
4.4.5. Panier	22
4.4.6. Export, impression et email des données bibliographiques.....	22
4.5. Serveur OAI-PMH	22
4.6. Statistiques d'utilisation.....	24
4.6.1. Méthodologie du suivi statistique de l'utilisation du site WEB d'Archimer	24
4.6.2. Méthodologie du suivi statistique du nombre de téléchargements de fichiers.....	25
4.6.3. Illustration : présentation des résultats synthétiques	25
4.6.4. Limite de fiabilités des résultats	26

5. Description de l'Architecture technique	27
6. Lancement d'Archimer : état des lieux	28
6.1.1. Enregistrement des publications récentes	28
6.1.2. Enregistrement des articles publiés entre 2000 et 2005	29
7. Objectifs à court et moyen termes	29
7.1. Court terme	29
7.1.1. Poursuite de l'enregistrement de l'arriéré des publications (années 2000-2005)	29
7.1.2. Développement d'un prototype de moissonneur spécialisé en sciences de la mer	29
7.2. Moyen terme	30
7.2.1. Collecte des « versions auteurs » de publications dès l'acceptation par l'éditeur	30
7.2.2. Elargissement du système à d'autres types de documents	30
7.2.3. Numérisation et enregistrement de documents sur support papier	30
Références bibliographiques	32

1. Introduction

En Août 2005, l'Ifremer lance Archimer (<http://www.ifremer.fr/docelec/>), son Archive institutionnelle, une base de texte intégraux proposant un ensemble de publications, de thèses, de rapports internes et d'actes de congrès. Ces documents scientifiques sont ainsi accessibles à tous, gratuitement, via internet.

Près de quatre mois après son ouverture, ce document présente les principes généraux de développement et de fonctionnement de cette nouvelle base ainsi que les fonctionnalités de ses différents modules.

La description des fonctionnalités rend compte de l'articulation de ce système avec les autres ressources électroniques proposées par le service de documentation de l'Ifremer. Ainsi la phase de repérage et d'enregistrement des documents est associée à sa base Bibliométrie, la phase de récupération des textes intégraux est liée à son portail de revues électroniques. Schémas et vues d'écran explicitent ainsi toutes les phases : repérage des références bibliographiques, vérification des droits de copyright, chargement des textes intégraux, indexation des références.

L'outil statistique associé y est décrit et souligne la place désormais accordée à la mesure de l'audience dans la gestion des ressources documentaires.

Ce document souligne également l'opportunité offerte au personnel du service de documentation de nouer de nouveaux liens avec les équipes de recherche de l'Ifremer et de participer activement à la valorisation de leurs travaux.

2. Les archives institutionnelles

2.1. Contexte national et international

Depuis le début des années 90, afin de contrer les politiques commerciales abusives de certains éditeurs scientifiques, des communautés scientifiques ont créé des serveurs de pré-publications pour offrir un accès gratuit et immédiat à leurs travaux (ex: ArXiv, en physique et RePec, en économie).

En 2001, l'organisation OAI (Open Archive Initiative) a formalisé un protocole d'interrogation de ces archives. Cette organisation a, non seulement défini un modèle commun pour les métadonnées (date, auteur, titre, nom du journal, ...), pour décrire les documents en texte intégral (qui eux, peuvent être stockés sous différents formats), mais, aussi, un protocole de collecte de ces métadonnées. Les archives, quand elles respectent ce protocole, deviennent donc "interopérables". Leurs métadonnées peuvent être récoltées automatiquement pour créer des bases de données référençant des documents pouvant être conjointement parcourus et localisés, comme s'ils appartenaient à une collection universelle, accessible à tous.

Ces premières initiatives sont aujourd'hui confortées par le mouvement « Open Access ». En effet, ce mouvement, qui cherche à rendre accessible à tous la documentation savante, en la diffusant gratuitement sur le WEB, s'est construit autour de trois conférences (Budapest, Bethesda et Berlin), qui ont donné lieu à la rédaction de manifestes en faveur du libre accès : manifestes signés, par exemple, en France, par le CNRS, l'INRIA, l'INRA et l'INSERM.

Pour atteindre l'objectif fixé, ce mouvement propose donc deux solutions :

- la création de revues en libre accès,
- la création d'archives ouvertes.

Ces dernières années, nous avons donc assisté à de nombreux engagements en faveur de ce mouvement de la part d'universités, d'organismes internationaux ou de gouvernements de certains pays. Cela a conduit à une multiplication de projets d'archives ouvertes. Ces nouvelles archives ont principalement été créées par des organismes de recherche ou par des universités (on parle alors d'Archives Institutionnelles), qui diffusent, par ce canal, une partie de leur production scientifique.

En France, par exemple, le CNRS est l'organisme de recherche le plus actif au sein du mouvement Open Access. Il est le premier organisme à avoir lancé ses archives ouvertes : HAL pour les pré-publications et

TEL pour les thèses. L'INRIA collabore depuis cette année au développement de ces deux archives et utilise ces deux serveurs pour y déposer ses propres publications. L'INRA a également annoncé le lancement prochain de ProdInra, son Archive Institutionnelle.

Au plan International, près de 350 archives compatibles OAI-PMH, sont référencées par l'organisation OAI. Dans le secteur des sciences de la mer, les communications faites à Rome, début octobre, dans le cadre de la 32^{ème} conférence de l'Association IAMSLIC (International Association of Marine Science Libraries and Information Centers), par nos homologues, montrent que plusieurs projets sont, soit déjà lancés, soit prévus. C'est notamment le cas des projets suivants :

Projets existants :

- Celui de la Woods Hole Oceanographic Institution (WHOI), dont l'archive institutionnelle, intitulée « Woods Hole Open Service », a été lancée cet été,
- Celui de la Scripps Institution of Oceanography (cf. California University Archive), qui serait également lancé.

Projets devant être lancés dans les prochains mois :

- Celui du Flanders Marine Institute (VLIZ), qui vise à rassembler et valoriser, dans le cadre d'une archive nationale, la production scientifique belge relative à la mer,
- Celui de l'Alfred Wegener Institute for Polar and Marine Research (l'AWI), en cours de conception.

Projet à caractère très prospectif : le projet « aquatic commons », présenté conjointement par nos collègues du Centre National d'Océanographie de Southampton et de l'Université de Floride. Il vise à la mise en place, au sein de IAMSLIC, d'une archive ouverte dédiée aux sciences aquatiques.

2.2. Intérêt pour l'Ifremer

2.2.1. Soutenir le mouvement Open Access

L'ouverture d'une archive institutionnelle est un soutien concret au mouvement « Open Access », dont l'Ifremer pourrait, à long terme, profiter des avancées. En effet, depuis plusieurs années l'Ifremer subit, de la part des plus grand éditeurs scientifiques, et comme toutes les autres grandes bibliothèques scientifiques, des augmentations de coûts d'abonnements aux revues scientifiques sans aucun rapport avec l'inflation. Ces augmentations l'obligent à consacrer une partie toujours plus importante de son budget aux contrats d'abonnement aux revues, et ce, au détriment d'autres sources d'information.

Si la majorité des publications de l'ensemble de la communauté scientifique internationale devenait accessible gratuitement sur le WEB, à travers un réseau d'Archives Ouvertes, elles pourraient constituer **une réelle alternative aux abonnements proposés par les éditeurs scientifiques**. Même sans imaginer pouvoir un jour nous passer de ces abonnements, **nous pouvons envisager être à terme mieux armés pour négocier, du fait de cette nouvelle donne, nos contrats avec les grands éditeurs scientifiques**.

2.2.2. Valoriser la production scientifique

Si l'accès gratuit à l'ensemble de la documentation scientifique internationale est un objectif à long terme, la mise en place d'une archive institutionnelle à l'Ifremer devrait pouvoir avoir un effet immédiat sur la visibilité de ses travaux. En effet, plusieurs études démontrent que les articles en libre accès sont plus cités que les articles uniquement accessibles à partir des sites WEB des éditeurs scientifiques (voir réf. 1, 2, 3, 4). La diffusion gratuite des publications de l'Ifremer, via Archimer, pourrait donc améliorer sensiblement leur impact scientifique.

2.2.3. Créer une nouvelle base de données dédiée aux sciences de la mer

Lorsque le nombre de documents disponibles dans Archimer aura atteint une masse critique, nous espérons que le personnel de l'Ifremer considérera cette base, non seulement comme un moyen de valoriser ses travaux à l'extérieur de l'Institut, mais aussi comme une **base de travail utile à ses recherches. Cette base devrait, en effet, agréger un ensemble de documents aujourd'hui disséminés sur plusieurs serveurs.** Elle devrait également donner accès à des documents, et, notamment, à des thèses, auxquelles Archimer est actuellement le seul moyen d'accès.

2.2.4. Renouer des liens entre les équipes de recherche et les bibliothèques

Les équipes de recherche utilisent aujourd'hui massivement les ressources électroniques (bases de données bibliographiques, journaux électroniques...) mises à leur disposition par les bibliothèques. **Le personnel de l'institut ne se déplace donc plus, ou peu, dans les salles de lecture des bibliothèques. Il a, en effet, accès à toutes ces ressources directement à partir de son poste de travail.**

Cette situation constitue, bien sûr, un réel progrès. Elle permet, par exemple, à tout le personnel de l'institut, quelle que soit sa localisation, d'avoir un accès à une très grande partie de la documentation mise à disposition par les bibliothèques et de bénéficier d'outils performants de recherche et de veille documentaires (ex. : bases de données bibliographiques, alertes de recherches automatiques...).

Cette situation a, par contre, tendance à isoler le personnel des bibliothèques des équipes de recherche. Ces dernières peuvent ainsi être amenées à mésestimer le travail effectué par les bibliothèques, dont l'élaboration et la mise en œuvre d'une politique d'acquisition des droits d'accès à des sources d'information sélectionnées). **Nous rencontrons, par exemple, régulièrement des chercheurs qui pensent que les articles des éditeurs scientifiques sont accessibles gratuitement sur le WEB.** En effet, comme les accès aux ressources de ses éditeurs sont protégés par contrôle de l'adresse IP, les chercheurs y accèdent de manière transparente, sans imaginer, par exemple, l'importance du travail que nécessite la négociation d'un contrat d'abonnement avec un éditeur comme Elsevier.

La mise en place d'une archive offre l'occasion aux personnels des bibliothèques de renforcer ses contacts avec les chercheurs, par l'intermédiaire, par exemple, de la collecte personnalisée des publications à enregistrer dans Archimer.

2.2.5. Améliorer la visibilité du site Internet de l'Ifremer

Les documents enregistrés dans l'Archive de l'Ifremer sont, non seulement, accessibles par l'intermédiaire du site WEB de consultation d'Archimer, mais aussi par l'intermédiaire de moteurs de recherche et par celui de moissonneurs OAI-PMH.

L'étude des statistiques d'utilisation d'Archimer démontre que les moteurs de recherche, dont, principalement Google, sont les principaux points d'accès à nos documents. Elle met en évidence le fait qu'une partie des utilisateurs, qui accède à nos documents directement par un moteur de recherche, continue ensuite sa visite vers la page d'accueil de notre site WEB. A partir de cette page, certains d'entre eux consultent d'autres documents disponibles via ce site. D'autres utilisateurs, quant à eux, continuent leur visite vers le site WEB institutionnel de l'Ifremer et y découvrent d'autres informations relatives à l'Institut.

Les documents enregistrés dans Archimer, sont, par conséquent, autant de nouveaux points d'entrée au site WEB institutionnel de l'Ifremer. Ils contribuent donc à améliorer l'audience du site WEB de l'Ifremer : site qui est un outil de communication essentiel pour l'Institut.

3. Principes généraux

3.1. Choix de la plateforme de développement

Pour réaliser le système Archimer, **nous avons fait le choix interne** à l'aide des technologies Java, JSP et Oracle, car nous avons, à l'origine du projet, l'objectif de réutiliser une partie des modules d'Archimer dans le cadre d'autres projets de la bibliothèque et, notamment, celui d'un projet de rénovation des sites WEB de consultation de nos catalogues. Nous souhaitons, de plus, lier ce nouveau système à d'autres modules informatiques déjà existants comme notre base « Bibliométrie » ou notre portail de revues électroniques. Un développement spécifique nous semblait alors la meilleure solution pour atteindre nos objectifs.

Les sites WEB de ce projet sont développés à l'aide de pages JSP et sont exécutés par le serveur Apache/Tomcat central de l'Ifremer. **Ces technologies ont été sélectionnées pour leur conformité avec la politique générale du département Informatique de l'Ifremer.**

Les données bibliographiques des documents enregistrées dans Archimer sont stockées dans une base de données Oracle : base hébergée sur un serveur mutualisé entre tous les départements de l'Ifremer. L'utilisation d'Oracle est particulièrement intéressante dans le cadre d'Archimer car elle permet l'intégration de fonctionnalités de recherche documentaire avancées.

3.2. Types de documents enregistrés

Actuellement Archimer a été conçue pour permettre l'enregistrement et la diffusion de thèses cofinancées par l'Ifremer, de rapports internes, d'actes de congrès et d'articles publiés dans des journaux scientifiques.

Pour faciliter l'acceptation de ce projet par le personnel de l'Ifremer, nous avons souhaité limiter, dans un premier temps, l'enregistrement des articles aux seules post-publications. Leur diffusion gratuite sur le WEB fait aujourd'hui l'unanimité, contrairement à la diffusion des pré-publications qui est parfois critiquées par certains auteurs par peur de plagiat, ou car la qualité de la publication n'est pas assurée par une validation par les pairs.

3.3. Format de diffusion

Nous avons choisi le format PDF comme format unique de diffusion. Tous les documents enregistrés dans Archimer sont donc convertis en PDF, et ce quelque soit l'outil utilisé pour sa rédaction (Word, Latex...). Nous avons sélectionné ce format de diffusion pour les raisons listées ci-dessous :

- L'assurance de la pérennité du format PDF, du fait de sa très large utilisation et de la publication de ses spécifications,
- Sa mise en œuvre simple, qui permet de réduire le temps de traitement et d'enregistrement des documents dans Archimer,
- Sa bonne adaptation à la diffusion électronique de documents volumineux, comme les publications ou les thèses.

3.4. Conservation des documents

La conservation des documents à très long terme n'a pas, jusqu'à présent, été l'une de nos préoccupations majeures dans le cadre de la définition de ce projet. D'emblée, nous avons, par exemple, exclu de convertir les documents en XML/SGML pour s'assurer de leur pérennité. Le temps de traitement d'une telle conversion nous semblait, en effet, incompatible avec les ressources humaines mises à disposition du projet.

Toutefois, considérant la masse de documents aujourd'hui stockés en PDF, nous espérons que, si ce format devenait un jour obsolète, il existerait alors des outils de conversion de nos fichiers PDF qui nous permettraient de convertir facilement ces fichiers PDF.

3.5. Modalités d'enregistrement des documents

Les documents sont enregistrés dans Archimer par le personnel des Bibliothèques de l'Institut, qui assure :

- la saisie des métadonnées,
- le classement des documents par domaines scientifiques (ex : biologie, aquaculture, pêche, ...),
- l'ajout de mots-clés, si utile,
- la remise en forme du texte intégral et sa conversion en PDF si nécessaire,
- le transfert du texte intégral vers le serveur d'Archimer.

3.5.1. Enregistrement des thèses, des actes de congrès ou des rapports internes

Pour les thèses, les actes de congrès ou les rapports internes, ce sont les auteurs eux-mêmes qui nous proposent l'enregistrement de leurs documents.

Pour diffuser un document de ce type via Archimer, **les auteurs doivent donc nous fournir, par email, les données bibliographiques nécessaires au référencement de leur document. Ils doivent également envoyer le texte intégral de leur document**, par email, CDROM ou au format Word ou au format PDF, en fonction de la taille du fichier.

Si les auteurs fournissent le texte intégral de leur document sous forme d'un ou de plusieurs fichiers Word, nous les convertissons et les fusionnons en un fichier PDF unique, à l'aide du logiciel Acrobat, avant transfert sur le serveur d'Archimer.

3.5.2. Enregistrement des publications récentes

Quelques auteurs nous signalent eux-mêmes les publications qu'ils souhaitent voir diffuser à partir d'Archimer. Dans ce cas, nous vérifions quelles sont les règles fixées par l'éditeur de la publication en matière d'auto-archivage. Si l'éditeur autorise cet auto-archivage, nous indiquons aux auteurs les éléments nécessaires à l'enregistrement de ces publications.

Toutefois, pour permettre l'enregistrement et la diffusion d'un plus grand nombre de publications, nous ne comptons pas uniquement sur des dépôts spontanés de la part des auteurs Ifremer, **mais nous procédons nous-mêmes aux veille et collectes suivantes :**

- **Toutes les semaines, nous repérons les publications rédigées par le personnel de l'Ifremer.** Toutes ces publications sont enregistrées dans la base « Bibliométrie » de l'Ifremer, dont l'alimentation est décrite dans l'un des prochains paragraphes,
- Nous étudions ensuite la politique de chacun des éditeurs des publications Ifremer repérées, et ce à l'aide, notamment, du site WEB Sherpa/Romeo. Si la politique de l'éditeur n'est pas déclarée, ni sur le site Sherpa/Romeo, ni sur son propre site WEB, nous essayons systématiquement alors de contacter l'éditeur pour lui demander l'autorisation d'enregistrer ces articles dans Archimer.
- **Si l'éditeur autorise l'auto-archivage de ses propres fichiers PDF** (ex : EDP Sciences, The Company of Biologists...), nous déchargeons nous même le fichier PDF correspondant à l'article considéré à partir du site de l'éditeur et nous l'enregistrons dans Archimer. **La majorité des données bibliographiques sont automatiquement transférées de la base « Bibliométrie » à la base Archimer.** Les données bibliographiques manquantes sont copiées manuellement à partir du site de l'éditeur. Dans ce dernier cas de figure, l'enregistrement est donc effectué sans qu'il soit nécessaire de contacter les auteurs.
- **Si l'éditeur autorise l'auto-archivage, mais limite cette dérogation à son droit de copyright au dernier « draft » de l'auteur** (c'est à dire la version envoyée par l'auteur à l'éditeur : version qui contient toutes les corrections demandées par les pairs lors du processus de relecture, mais n'a pas été mise en page par l'éditeur), nous contactons les auteurs « Ifremer » de la publication pour leur demander cette version. S'ils sont en mesure de nous la fournir, nous produisons nous mêmes, à partir des fichiers envoyés par les auteurs, le fichier PDF correspondant à nos critères de diffusion, et ce avant de l'enregistrer dans Archimer. Deux cas se présentent :

- **Soit l'auteur nous transmet sa publication sous la forme d'un ou de plusieurs fichiers Word** (un pour le texte, un autre pour les tableaux et les figures par exemple), nous fusionnons ces fichiers avant de les convertir en un fichier PDF unique. Nous reconstruisons également la première page de la publication, non seulement pour uniformiser la présentation de nos publications, mais surtout pour répondre aux règles fixées par les éditeurs (ajout de la citation complète et normalisée de la citation de la publication, ajout d'un lien vers le site de l'éditeur, ajout d'un texte explicatif spécifique à chaque éditeur...)
- **Soit l'auteur nous transmet sa publication en format PDF**, nous reconstruisons la première page avant de l'enregistrer dans Archimer.

3.5.3. Enregistrement des articles publiés entre 2000 et 2005

En parallèle de l'enregistrement des publications récentes que nous traitons régulièrement, nous avons entrepris de récolter les publications Ifremer publiées depuis l'année 2000. Pour conduire cette recherche, nous avons, à l'aide de la base « Bibliométrie », édité une liste des revues dans lesquelles le personnel de l'Ifremer a publié des articles au cours de ces cinq dernières années. Pour chacune de ces revues, nous étudions la politique de copyright de l'éditeur en matière d'auto-archivage. Si l'éditeur autorise l'auto-archivage de sa propre version PDF, nous l'enregistrons sans avoir besoin de contacter les auteurs. Si l'éditeur limite cette autorisation au dernier « draft » des auteurs, nous contactons, par une procédure automatisée, tous les auteurs Ifremer qui ont publié des articles dans cette revue. Ces messages ne sont bien sûr pas systématiquement fructueux. En effet, plusieurs cas d'insuccès peuvent se présenter :

- l'auteur qui s'est chargé de la rédaction de la publication peut avoir quitté l'institut,
- le dernier « draft » de la publication peut avoir été perdu ou supprimé,
- plusieurs auteurs ne prennent pas le temps de répondre à nos messages,
- quelques auteurs nous ont répondu que la publication que nous recherchions ne présentait plus d'intérêt scientifique,
- ...

4. Description fonctionnelle

4.1. La base de données « Bibliométrie »

4.1.1. Considérations générales

Nous décrivons, ci-après, l'architecture de la base Bibliométrie, car elle occupe, comme nous l'avons mentionné dans le paragraphe précédent, une place centrale dans les processus de collecte et d'enregistrement des publications dans Archimer.

La base « Bibliométrie » de l'Ifremer, est une base de données développée et alimentée par la Bibliothèque La Pérouse, dont l'objectif est de répertorier les articles publiés, par le personnel de l'Ifremer, dans des revues à comité de lecture. Cette base de données a été développée, à la demande de la Direction Scientifique, dans le contexte de la mise en place d'indicateurs nationaux d'évaluation de la production scientifique des organismes de recherche. Les références des publications enregistrées dans cette base de données, sont d'ailleurs, depuis l'année 2004, utilisées pour consolider les résultats obtenus par l'Observatoire des sciences et des techniques (OST), dont la mission est de concevoir et de produire des indicateurs quantitatifs sur la R&D. Ces indicateurs contribuent à la détermination du positionnement stratégique de la France en Europe et dans le monde.

4.1.2. Description technique du projet

Suite à une série d'analyses, la base de données Current Contents a été retenue comme source d'information nous permettant de repérer les publications d'agents de l'Ifremer parues dans des revues à comité de lecture. Ce repérage se fait sur la base des notices dans lesquelles l'Ifremer est cité (cf. notice suivante) :

ARTICLE TITLE: Structural and functional properties of hemocyanin from ...
ARTICLE AUTHOR: Chausson-F; Bridges-CR; **Sarradin-PM**; Green-BN; Riso-R; **Caprais-JC**; **Lallier-FH**
REPRINT AUTHOR: Chausson, F; CNRS; Biol Stn; BP 74; F-29682 Roscoff; France
RESEARCH INSTITUTION: UPMC, Observ Oceanol Roscoff, CNRS, UPR 9042, Roscoff, France; **IFREMER**, DRO, EP, Plouzane, France; Univ Dusseldorf, Inst Zoophysiol, D-4000 Dusseldorf, Germany; UBO, Lab Chim Marine, Plouzane, France; IUEM, CNRS, UPR 9042, Plouzane, France
SOURCE: PROTEINS-STRUCTURE-FUNCTION-AND-GENETICS. DEC 1 2001; 45 (4) : 351-359
PUBLICATION YEAR: 2001
ISSN: 0887-3585

Toutes les semaines, le responsable de l'application soumet un profil de recherche à la base réactualisée des Current Contents, via le serveur ERL de la société Ovid. Une partie des informations disponibles (cf. Titre de l'article, titre de la revue, année de publication, volumes, pages, noms, initiales des prénoms des auteurs, mots-clés, résumé et lien vers le texte intégral...) est extraite de ces références dans un fichier au format texte (voir fig. 1/2). Ce fichier texte est ensuite importé dans la base de données (voir fig. 1/1), via le site WEB d'administration (voir fig. 1/3).

Les Current Contents ne permettent pas de déterminer quel est l'organisme d'appartenance des différents auteurs (ex : Mr. Sarradin, Caprais et Lallier dans l'exemple précédent). Pour détecter tous les auteurs de l'Ifremer, le système se base sur l'annuaire LDAP de l'Institut (voir fig. 1/4).

Cependant, l'annuaire LDAP représente une image de la population à un moment donné. Une interrogation directe de la base de données LDAP ne permettrait pas, par exemple, de détecter un article d'un thésard de l'Ifremer qui aurait quitté l'Ifremer depuis la parution de cet article dans les CC. Pour améliorer la détection des chercheurs Ifremer dans une publication, le système analyse également les logs de l'annuaire LDAP (voir fig. 1/6). Ces fichiers contiennent la trace de toutes les actions intervenues sur les données de l'annuaire (ex : création, modification ou suppression d'un compte Ifremer).

Toutes les nuits, un outil (voir fig. 1/5) est lancé pour analyser ces fichiers logs. Cet outil permet de constituer une base de données de tous les agents actuels ou passés de l'Ifremer.

Quand un nouveau fichier d'articles est enregistré dans la base (voir fig. 1/6), le système commence par vérifier s'il est possible d'identifier chacun des auteurs directement dans l'annuaire LDAP. Pour les auteurs qu'il n'est pas possible d'identifier dans l'annuaire LDAP, il vérifie, à partir de la base constituée (à partir des logs de l'annuaire), s'il s'agit d'agents ayant quitté l'Ifremer dans les 3 dernières années.

A la fin de l'enregistrement d'un fichier d'articles, le système propose à l'administrateur de traiter, à l'aide d'interfaces WEB spécifiques, les ambiguïtés détectées lors du traitement (voir fig. 1/7). Le système pointe, par exemple, les publications, dans lesquelles la référence d'un auteur peut correspondre à plusieurs personnes à l'Ifremer (homonymie). Le système pointe également, par exemple, les publications dans lesquelles aucun des auteurs n'a pu être identifié comme étant Ifremer.

Pour finaliser l'enregistrement, un email est envoyé automatiquement aux auteurs repérés dans les publications enregistrées pour leur demander de contrôler les informations relatives à leurs articles. Enfin, pour apprécier la qualité des articles, les facteurs d'impact de plus de 5000 revues sont extraits du « Science Citation Index » (voir fig. 1/8) et enregistrés dans la base de données. Le système est donc capable d'associer un facteur d'impact à un article à partir du ISSN de la revue, présent à la fois dans la notice des Current Contents et dans le « Science Citation Index ».

Toutes les informations stockées dans la base de données sont accessibles à partir d'un site WEB (voir fig. 1/9) ouvert sur l'Intranet de l'Ifremer. Différents critères de recherche et différents formats d'édition sont disponibles pour limiter l'affichage des publications.

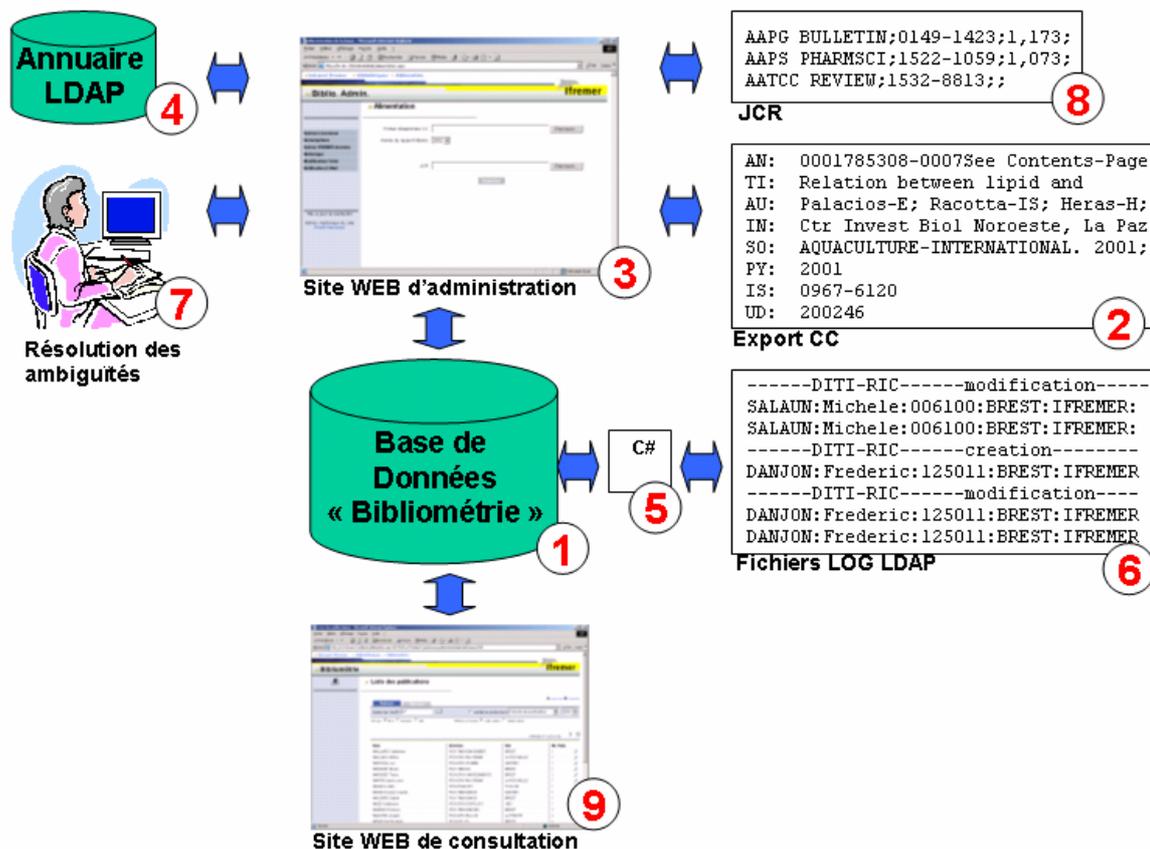


Figure n°1 : architecture du système « Bibliométrie »

4.2. Archimail

Toutes les semaines, nous étudions les « rapports » de la base « Bibliométrie ». Ces « rapports » présentent la liste des dernières publications Ifremer enregistrées dans cette base de données (voir fig. 2/1). Pour chacune des publications référencées dans ces « rapports », nous vérifions si l'éditeur de la revue (voir fig. 2/2) autorise l'auto-archivage. Ci c'est le cas, mais s'il limite cette autorisation au dernier « draft » de l'auteur ou s'il autorise l'archivage de ses propres fichiers PDF (sans que nous y ayons accès, si la bibliothèque n'a pas souscrit d'abonnement à cette revue par exemple), nous contactons tous les auteurs de cette publication et leur demandons les éléments qui nous manquent pour procéder à l'enregistrement de leur article.

Pour simplifier les contacts avec les auteurs de ces publications, nous avons développé le module « Archimail ». Il permet la génération de messages pré-rédigés et personnalisés en fonction de la publication à traiter. Quand, dans un des « rapports » publiés par la base « Bibliométrie » (voir fig. 2/1), nous détectons un article publié dans une revue, dont l'éditeur autorise l'auto-archivage (voir fig. 2/2), il suffit en effet de copier l'identifiant de l'article considéré dans le rapport (voir fig. 2/3) et de la copier dans le formulaire d'accueil d'ArchiMail (voir fig. 2/1). A l'aide de cet identifiant, Archimail récupère, dans la base « Bibliométrie », les données nécessaires à la composition du message, et, notamment, le titre de la publication et l'adresse email de tous les auteurs Ifremer repérés dans la publication considérée. A l'aide des ces informations, Archimail compose ensuite un message (voir fig. 2/5), qu'il est possible de personnaliser, avant de l'envoyer automatiquement à tous les auteurs Ifremer repérés dans cet article.

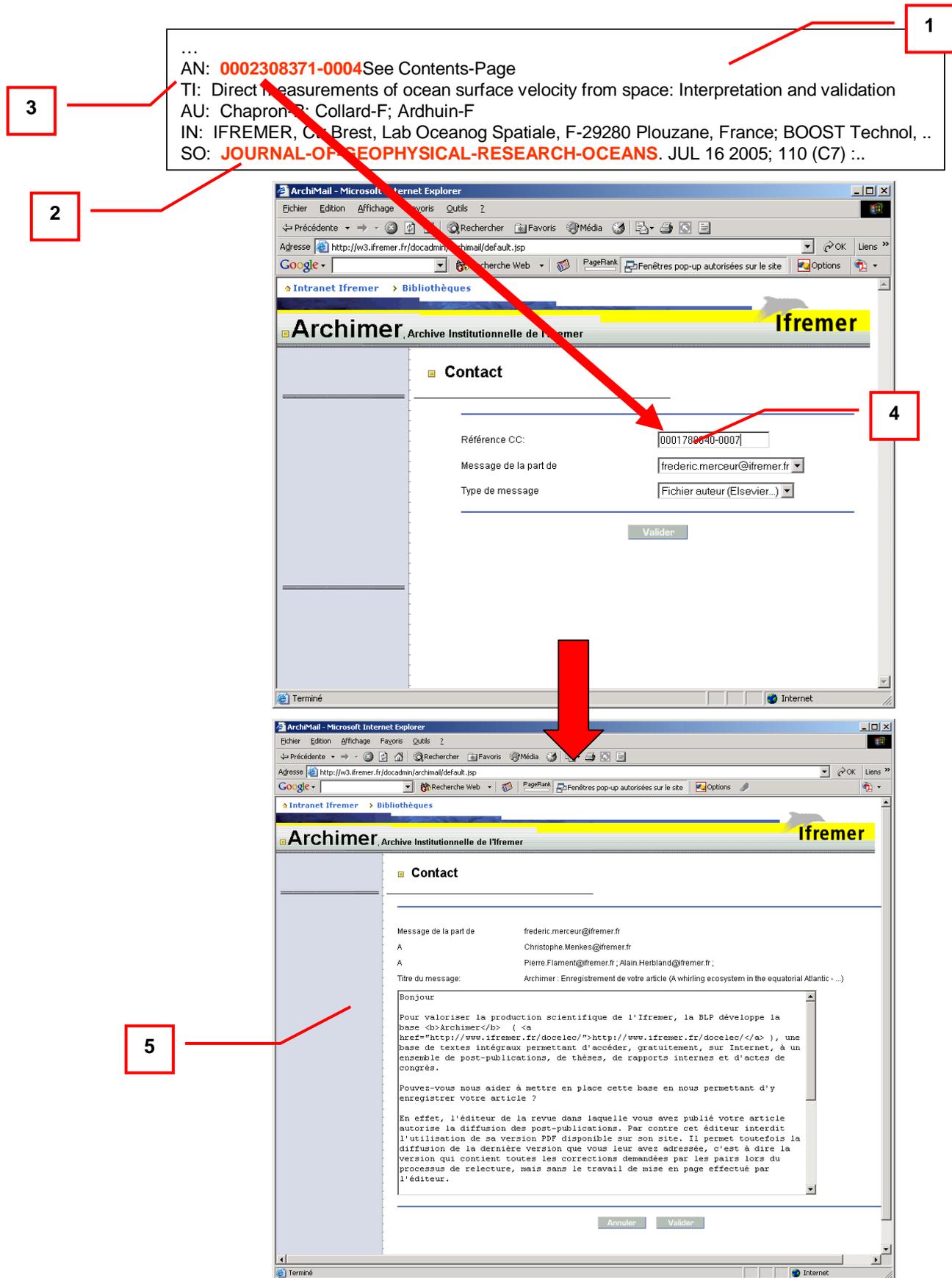


Figure 2 : Archimail

4.3. Site WEB d'administration

Quand toutes les données sont réunies pour référencer un document, le personnel de la BLP dispose d'un module d'administration, basé sur une interface WEB, pour enregistrer le texte intégral et ses données bibliographiques.

La page d'accueil du site d'administration d'Archimer permet de sélectionner le type de document à enregistrer (voir fig. 3/1). En effet, les données bibliographiques à saisir seront différentes en fonction de chaque document (voir tableau 1). Pour enregistrer une publication repérée dans la base Bibliométrie, il est également possible de saisir sa clé (voir fig. 2/3, 3/2). Dans ce cas, les métadonnées disponibles dans cette base « Bibliométrie » sont automatiquement transférées dans les masques de saisie du module d'administration d'Archimer.

Tableau 1 : liste des données bibliographiques enregistrées et indexées pour chaque type de document

	Thèse	Acte de congrès	Publications	Rapport interne	Indexé en recherche
Titre français	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	✓
Titre anglais	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	✓
Résumé français	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	✓
Résumé anglais	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	✓
Mots clés français	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	✓
Mots clés anglais	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	✓
Liste des auteurs (Nom, prénoms, organismes)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	✓
Classement dans une ou plusieurs des catégories	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	✓
Université	<input checked="" type="checkbox"/>				
Discipline	<input checked="" type="checkbox"/>				
Directeur de thèse (Nom, Prénom, organisme)	<input checked="" type="checkbox"/>				
Co-Directeur de thèse (Nom, Prénom, organisme)	<input type="checkbox"/>				
Financement de la thèse	<input checked="" type="checkbox"/>				
Date de publication	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	✓
Date d'enregistrement	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Langue du texte intégral	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Référence rapport interne				<input type="checkbox"/>	✓
URL du texte intégral	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
Taille du texte intégral (en Ko)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
URL de l'annexe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Taille de l'annexe (en Ko)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Nom de la revue			<input checked="" type="checkbox"/>		
URL de la revue			<input checked="" type="checkbox"/>		
Nom de l'éditeur			<input checked="" type="checkbox"/>		
URL de l'éditeur			<input checked="" type="checkbox"/>		
Volume			<input checked="" type="checkbox"/>		
Numéro			<input checked="" type="checkbox"/>		
Pages			<input checked="" type="checkbox"/>		
DOI			<input type="checkbox"/>		✓
Publication Copyright			<input checked="" type="checkbox"/>		
Acte Congrés		<input checked="" type="checkbox"/>			
Source		<input checked="" type="checkbox"/>			

: obligatoire

: optionnel

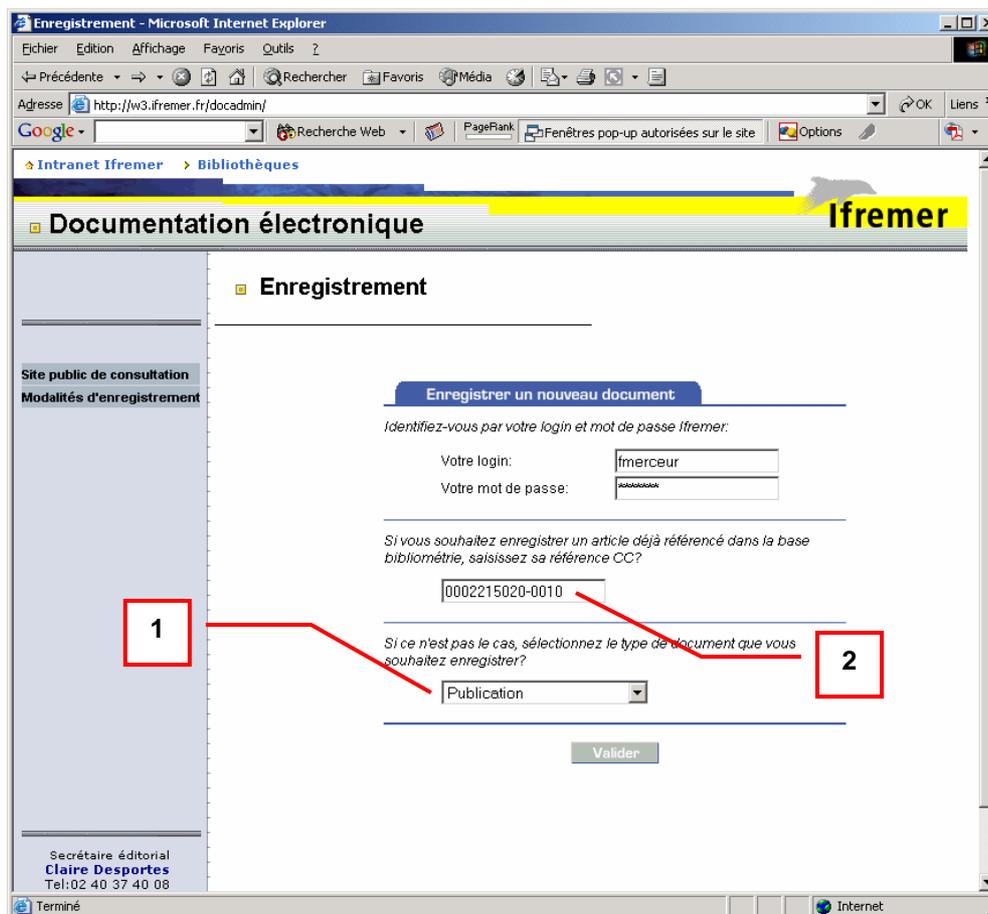


Figure 3 : Page d'accueil du site d'administration

A titre d'exemple, la figure 4 présente le formulaire de saisie des principales données bibliographiques d'une publication : après que l'administrateur ait saisi la clé d'une publication référencée dans la base « Bibliométrie ». Les champs « titre de l'article, date de publication, volume, numéro, résumé, mots-clés, auteurs » sont automatiquement pré-remplis. **Si la bibliothèque a souscrit un abonnement à la revue électronique, un lien vers le texte intégral de l'article disponible sur le site de l'éditeur est également transféré de la base « Bibliométrie » vers Archimer.** Ce lien (voir fig. 4/1) permet à l'administrateur d'accéder rapidement au site de l'éditeur (voir fig. 4/2) et d'y copier les données qui n'ont pas été obtenues automatiquement à partir de la base « Bibliométrie » (ex. : le DOI ou le copyright par exemple).

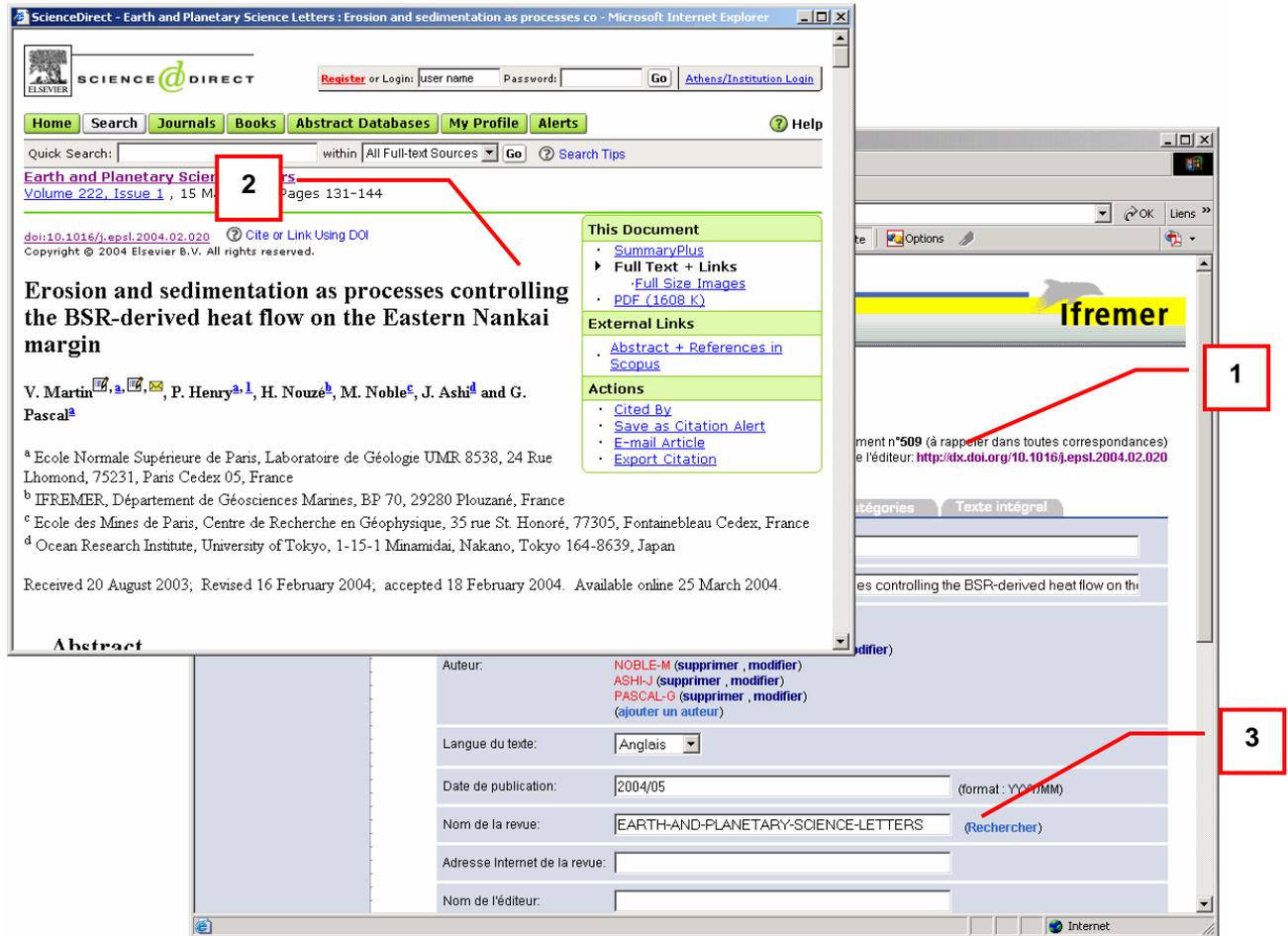


Figure 4 : Page de données générales du site d'administration

Pour enregistrer un article, il est nécessaire d'enregistrer le nom de la revue, son adresse Internet, le nom de son éditeur et l'adresse Internet de ce dernier. **Pour obtenir automatiquement ces données, l'administrateur a la possibilité de lancer une recherche dans la base de données « Revues électroniques ».** Cette base de données contient les références de tous les titres de revues électroniques auxquelles la BLP a souscrit un abonnement. En trois clics, il est donc possible de lancer une recherche dans cette base et de récupérer automatiquement les données relatives à cette revue (voir fig. 5).

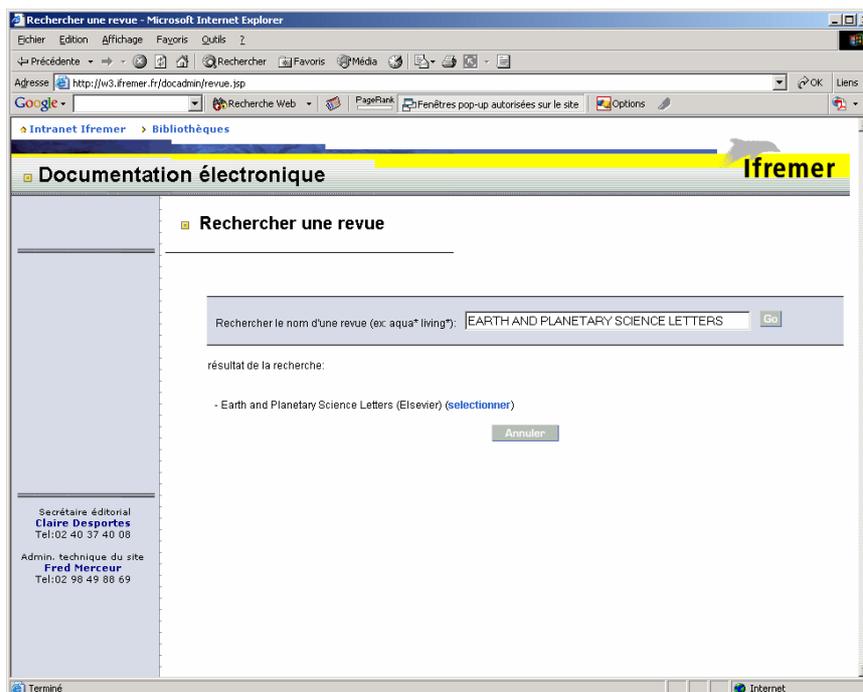


Figure 5 : Recherche d'une revue dans la base « Revues électroniques »

Quand toutes les données de l'onglet général ont été saisies, pour compléter l'enregistrement, il est ensuite nécessaire de :

- compléter les données présentées dans les onglets mot-clés (voir fig. 6) et résumés (voir fig. 7),
- classer le document dans une ou plusieurs catégories (voir fig. 8),
- enregistrer le fichier PDF correspondant au texte intégral du document.

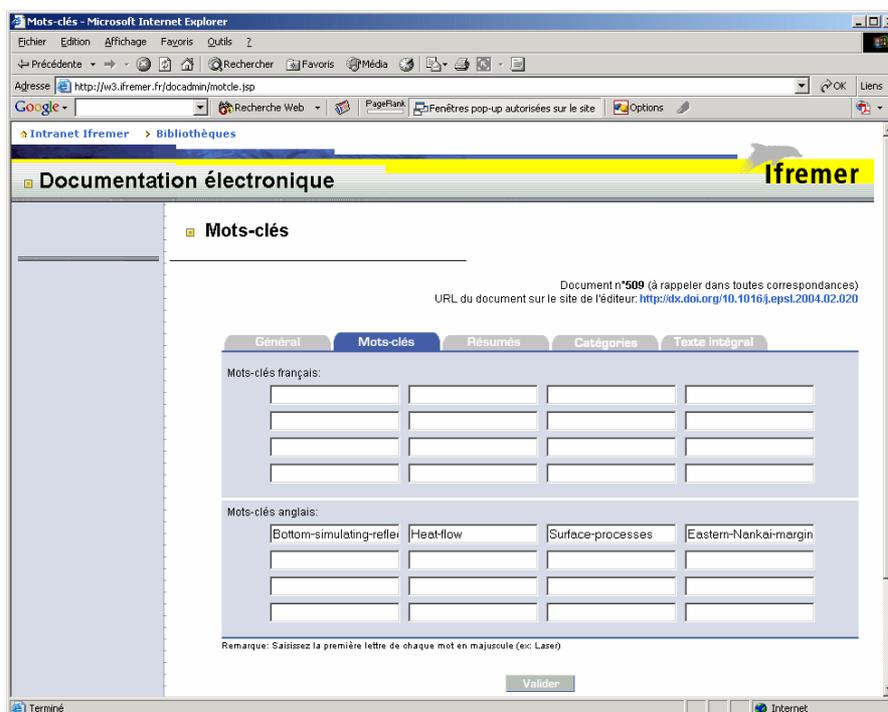


Figure 6 : Enregistrement des mots clés

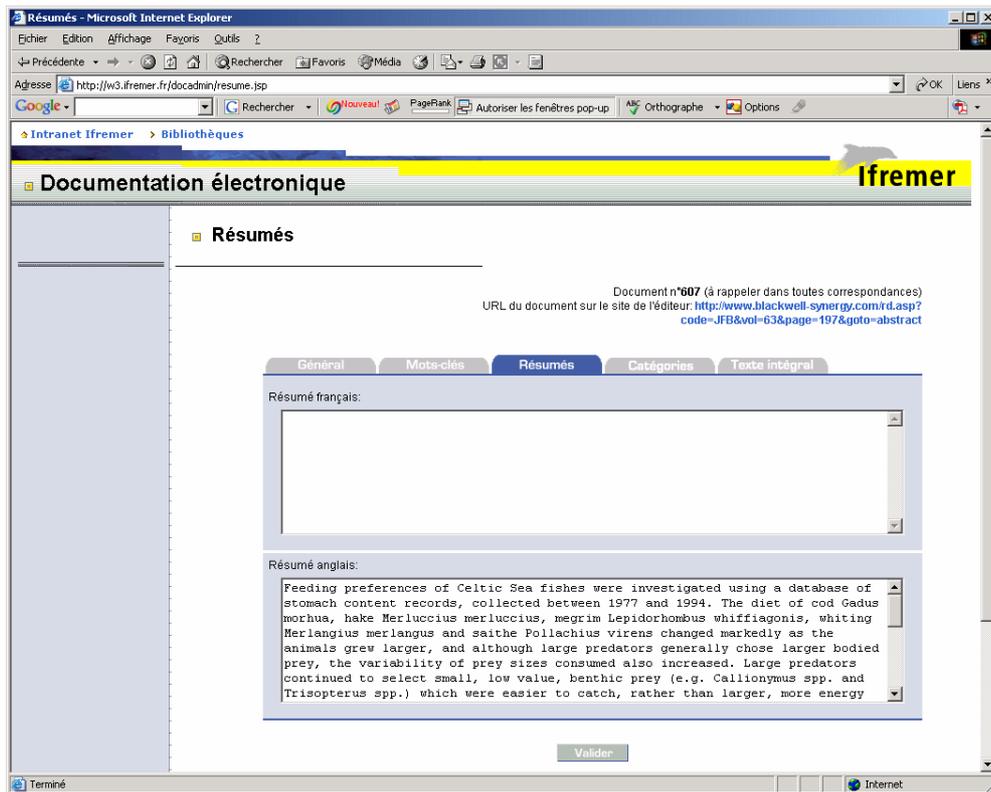


Figure 7 : Enregistrement des résumés

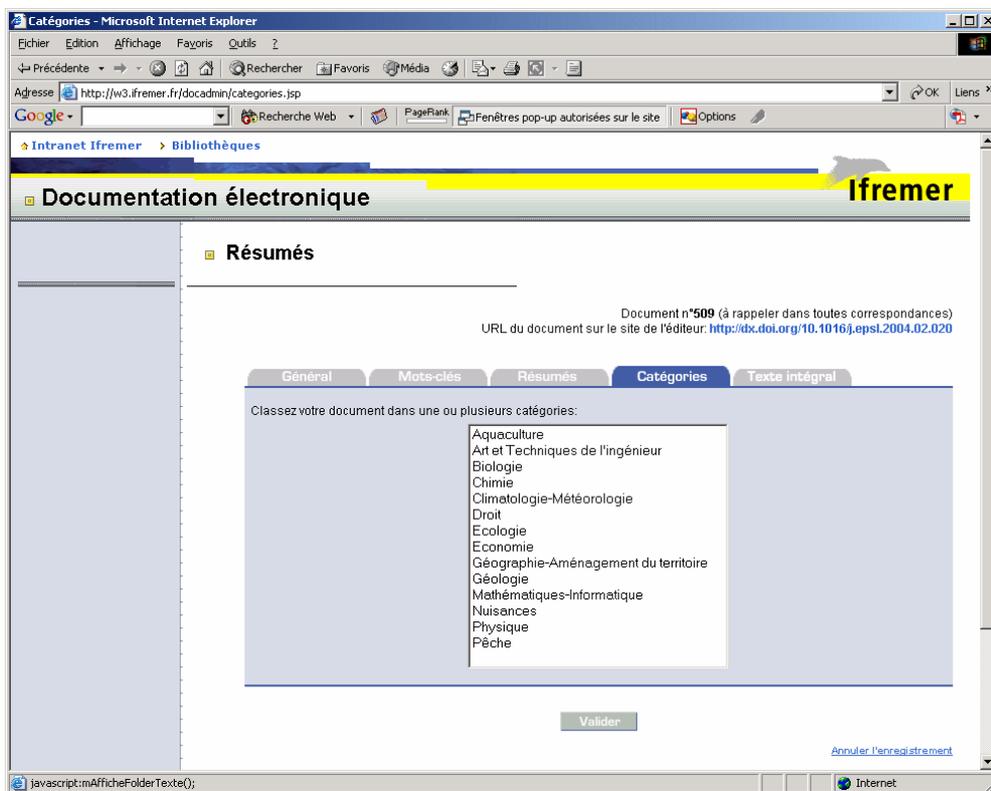


Figure 8 : Classement du document

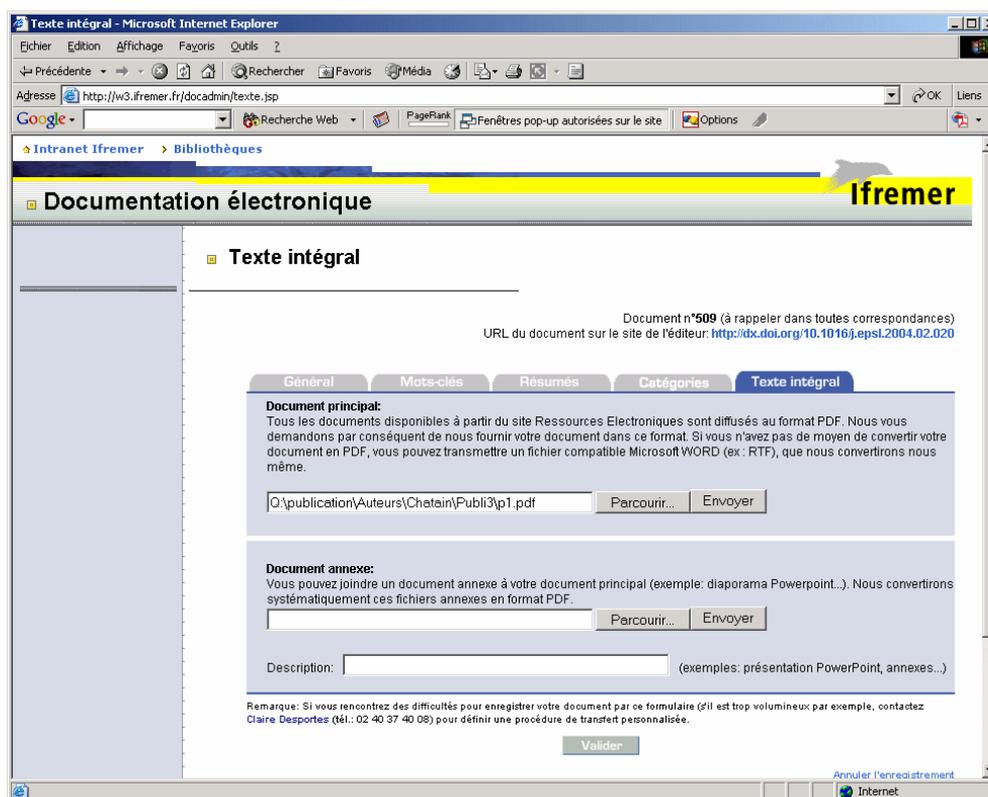


Figure 9 : Enregistrement du texte intégral

4.4. Site WEB de consultation

4.4.1. Page d'accueil

Le site WEB de consultation du système Archiver est accessible librement à partir de l'Internet. La page d'accueil du site WEB de consultation présente, non seulement, les objectifs de ce projet (voir fig. 10/1), mais donne, également, accès aux différents modules d'accès aux documents (voir fig. 10/2). De plus, elle propose un ensemble de liens hypertextes (voir fig. 10/3) vers :

- les modalités d'enregistrement des différents documents,
- des informations techniques concernant notre serveur OAI-PMH,
- un lexique des expressions utilisées dans le mouvement « Open Access »,
- un ensemble de liens vers d'autres Archives Ouvertes.

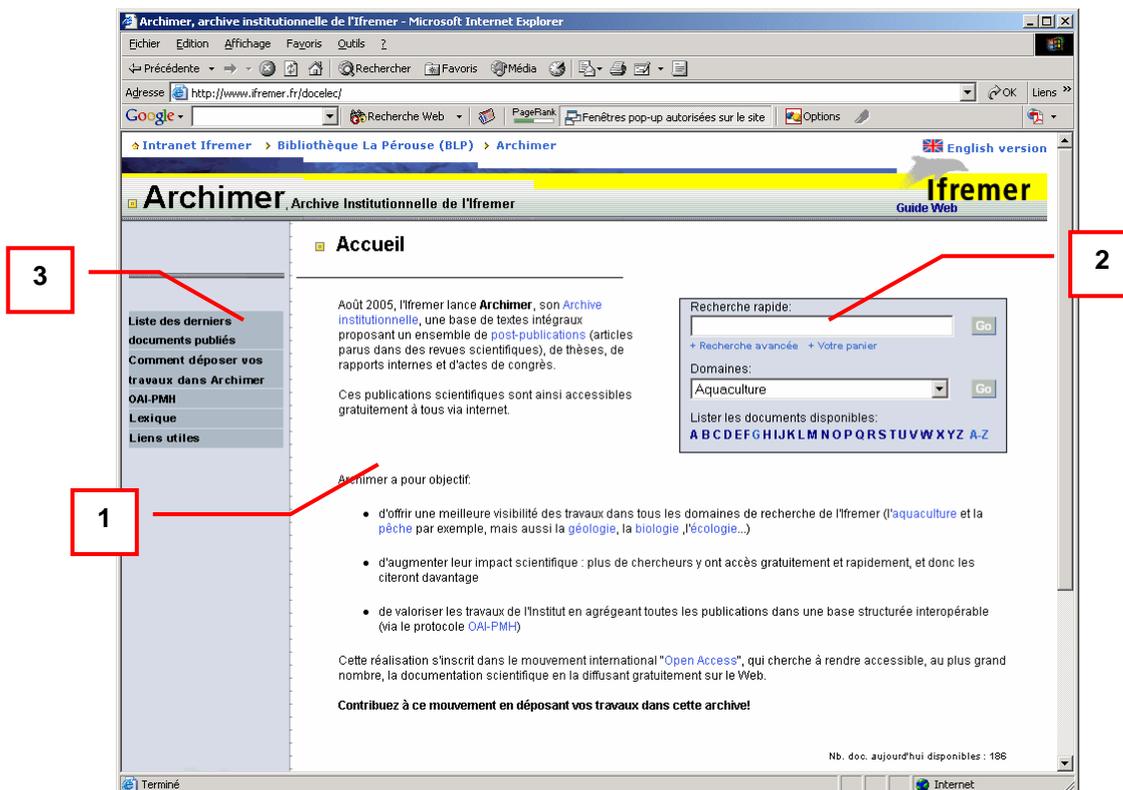


Figure n°10 : page d'accueil du site de consultation d'Archimer

4.4.2. Mode de recherche simplifiée

Le mode de recherche simplifiée est le principal point d'accès aux documents. Il permet de lancer des recherches conjointes (voir fig. 11/1) sur les données bibliographiques les plus importantes (titres, mots-clés, résumés, nom d'auteur, références, DOI...). Pour offrir une syntaxe de recherche simple et homogène par rapport aux autres systèmes proposés par la bibliothèque, nous avons défini notre propre langage de recherche. **Archimer intercepte les requêtes lancées par les utilisateurs pour les analyser et les traduire dans le langage de la base de données actuellement utilisée (Oracle).** Nous sommes ainsi plus indépendants par rapport à la base de données sélectionnée.

Archimer offre par exemple la possibilité d'utiliser des connecteurs booléens, des caractères génériques (ex : troncature), plusieurs niveaux de parenthésages, des opérateurs de proximités... La syntaxe supportée par Archimer est décrite en détail dans l'aide en ligne accessible à partir du site de consultation.

Quand un utilisateur lance une requête (voir fig. 11/1), Archimer affiche son résultat sous forme de liste (voir fig. 11/2). Pour chacun des documents de cette liste, Archimer affiche le titre, la date de publication, les auteurs et le type de documents. A partir de cette liste de résultats, l'utilisateur a la possibilité d'afficher la notice d'un document en cliquant sur son titre (voir fig. 11/3). Archimer affiche alors l'ensemble des données bibliographiques disponibles pour le document sélectionné (voir fig 12). Cette notice propose systématiquement un lien vers le texte intégral du document (voir fig. 12/1). Si l'utilisateur sélectionne ce lien, le texte intégral du document est ouvert dans une fenêtre annexe (voir fig. 13). Dans le cas des publications, la notice propose également, par l'intermédiaire du DOI, un lien vers le site de l'éditeur (voir fig. 12/2).

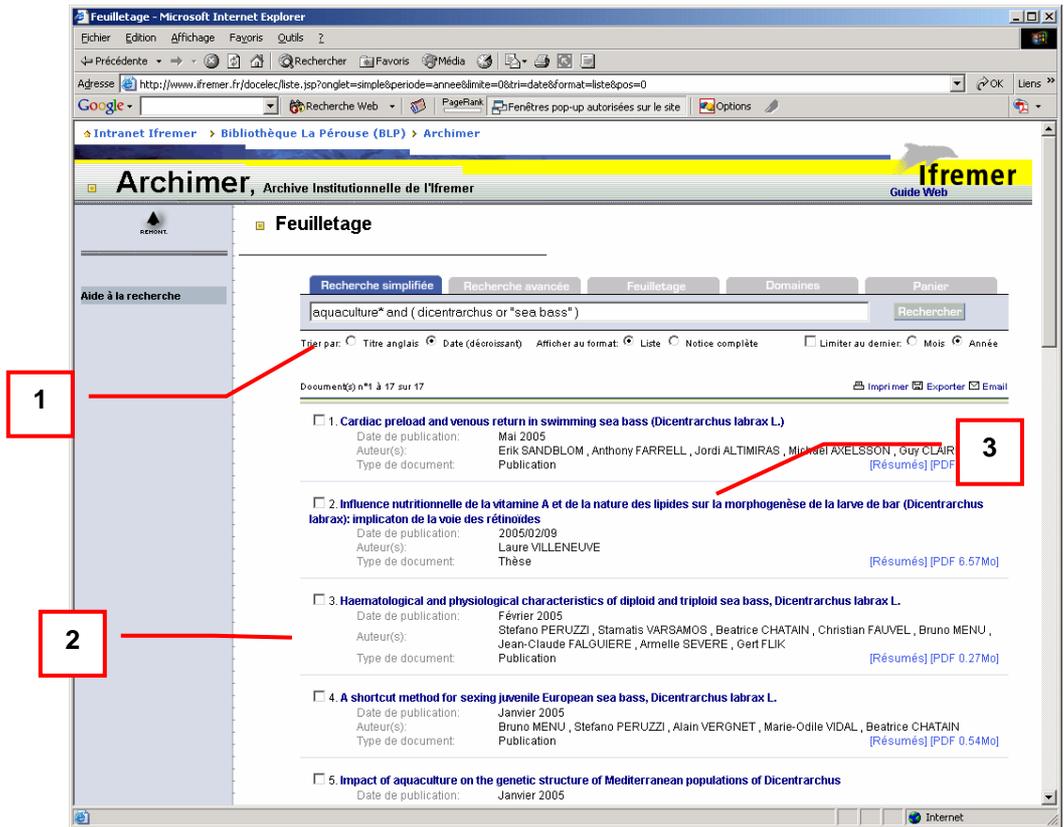


Figure n°11 : Mode de recherche simplifié

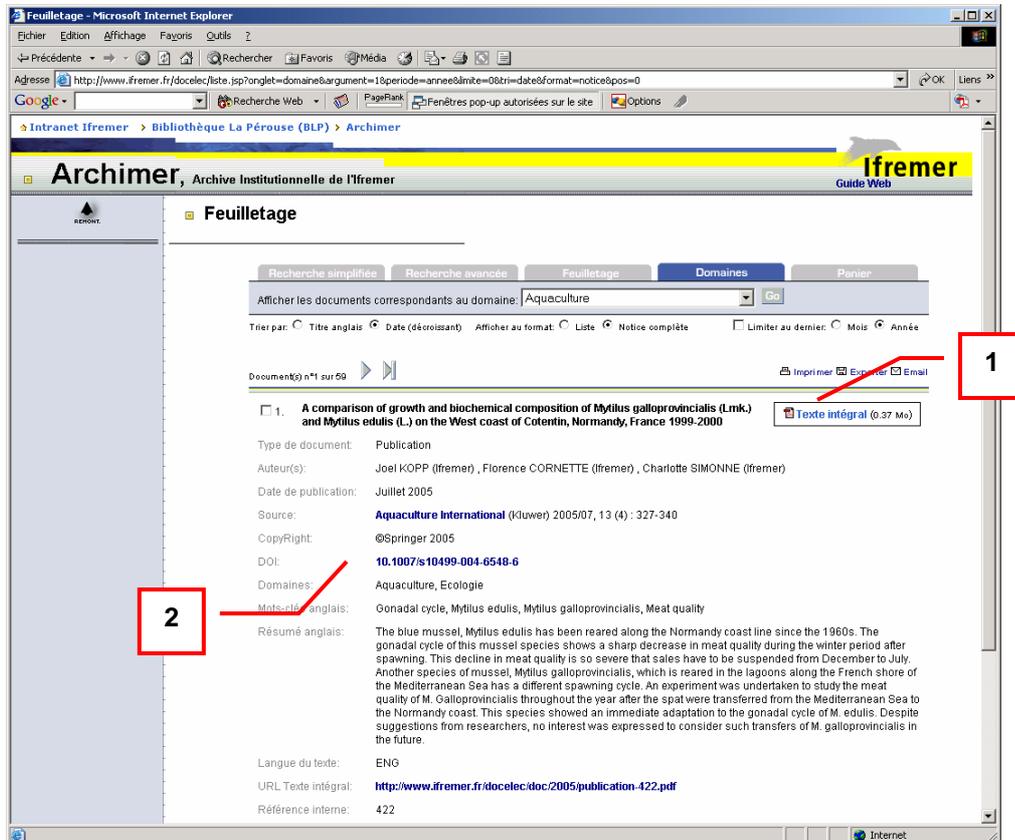


Figure n°12 : Affichage de la notice

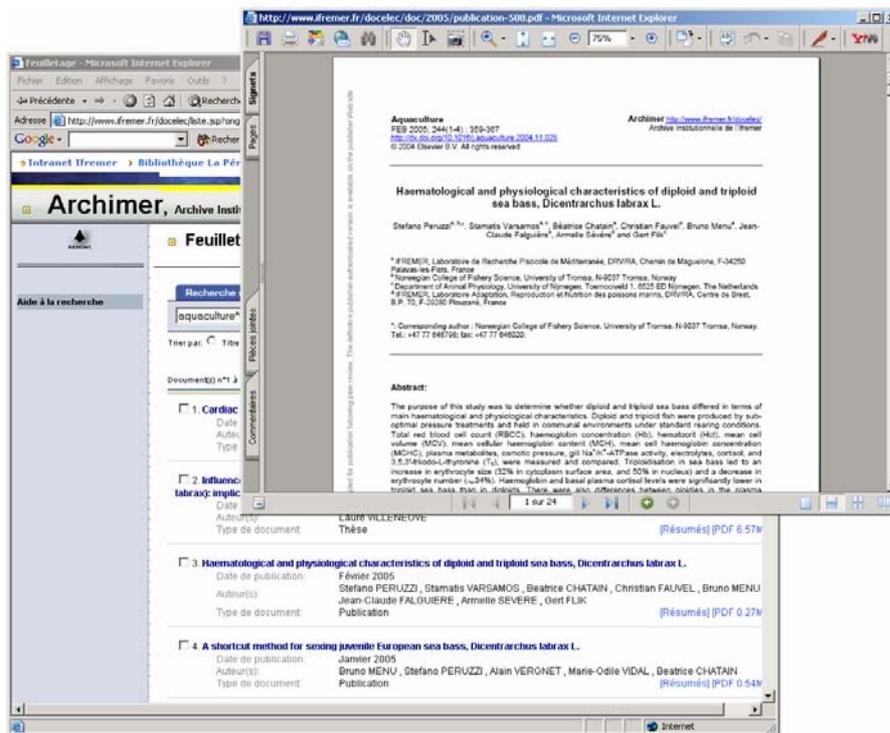


Figure n°13 : Affichage du texte intégral

4.4.3. Mode de recherche avancée

Le mode de recherche avancée permet de poser une requête sur un ensemble d'index, et notamment sur le texte intégral. L'utilisateur a la possibilité de spécifier des recherches dans plusieurs indexes et de combiner ces recherches à l'aide d'un connecteur booléen (voir fig. 14/1). Le système affiche les résultats intermédiaires (voir fig. 14/2), ainsi que le résultat global de la requête (voir fig. 14/3) :

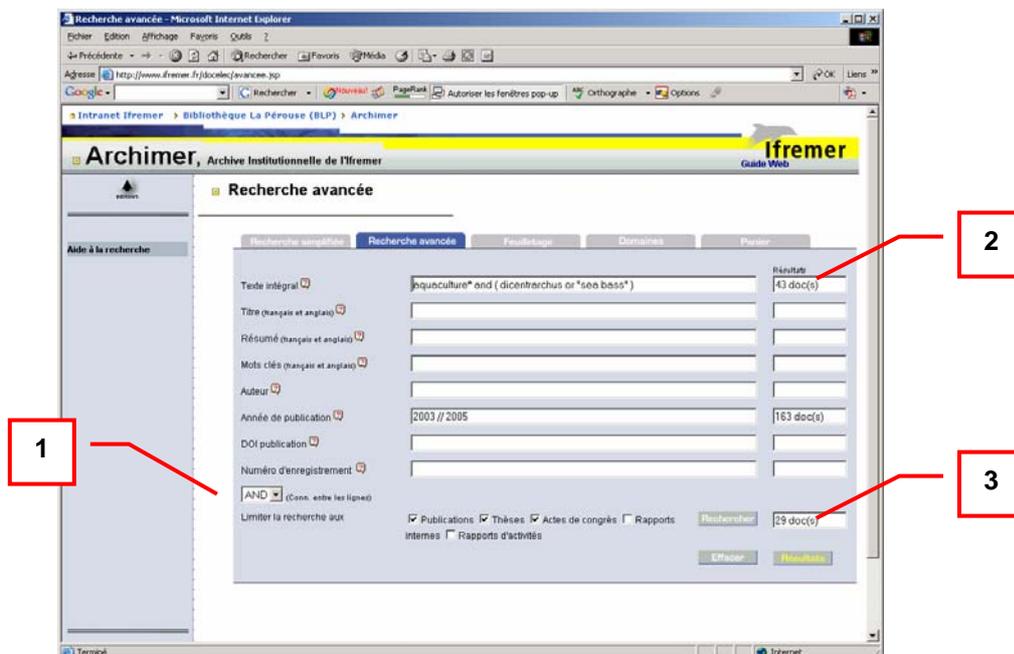


Figure n°14 : Mode de recherche avancé

4.4.4. Feuilletage

En supplément des modes de recherches simplifiées et avancées, les utilisateurs disposent de deux modes de « feuilletage » pour accéder aux documents :

- un mode de feuilletage par titre de documents,
- un mode de feuilletage par domaines : tous les documents sont classés dans un ou plusieurs des 15 domaines que nous avons définis. Les utilisateurs ont donc la possibilité de lister les documents classés dans le domaine qui les intéresse.

4.4.5. Panier

La case à cocher associée à chaque document de la liste de résultats permet à l'utilisateur de sauvegarder une sélection de références. Chaque sélection dans la liste de résultats provoque son ajout systématique dans la page « panier ». Cette sélection est cumulative, elle n'est pas réinitialisée à chaque recherche. Elle peut être composée de références sélectionnées à partir du résultat de plusieurs recherches différentes. La fonction « Panier » peut donc, par exemple, offrir les 2 services suivants :

- Elle peut servir à conserver une liste de documents, consultables fréquemment, sans devoir les rechercher à nouveau,
- Elle peut servir à garder une liste des résultats de recherches intermédiaires de documents, jugés intéressants, pendant que l'on effectue d'autres recherches.

4.4.6. Export, impression et email des données bibliographiques

Les utilisateurs ont également la possibilité d'exporter, d'imprimer ou d'envoyer par email les données bibliographiques des documents présents dans Archimer. L'utilisateur peut paramétrer cette fonction en spécifiant la plage de documents à exporter (tous les documents, les documents du panier ou le document courant) et en sélectionnant les champs bibliographiques (titres, résumés, auteurs...).

4.5. **Serveur OAI-PMH**

Le serveur Archimer est également compatible OAI-PMH. L'« *Open Archive Initiative Protocol for Metadata Harvesting* » (OAI-PMH) ou « *Protocole de Collecte de Métadonnées de l'Initiative Archives Ouvertes* », définit un standard pour transférer des collections de métadonnées. Il a été conçu dans le cadre de l'Initiative Archives Ouvertes pour implanter des bases inter opérables de pré-publications scientifiques soumises par leurs auteurs.

Ce protocole d'échange permet de créer, d'alimenter et de tenir à jour, par des procédures automatisées, des réservoirs d'enregistrements qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer, ni modifier leur localisation d'origine.

Grâce au protocole OAI, une bibliothèque agissant en tant que fournisseur de données a la possibilité d'offrir une visibilité accrue à ses documents, notamment à ses publications électroniques ou à ses fonds spécialisés.

Réciproquement, en tant que fournisseur de service, une bibliothèque peut réaliser une base de données ou un portail documentaire dans son domaine de spécialité ou sur un thème quelconque, en collectant les données descriptives de ressources et documents de tous types, accessibles sur l'Internet dans des entrepôts OAI.

*Dans ce contexte, le protocole définit le langage par lequel communiquent le fournisseur de données (**entrepôt**) et le fournisseur de services (**agrégateur**), qui rassemble des données collectées par un moissonneur. Pour alimenter l'agrégateur, le moissonneur visite plusieurs entrepôts, qu'il doit interroger massivement une seule fois ou en plusieurs étapes, pour extraire les enregistrements des items qui l'intéressent. Après la collecte, le moissonneur dépose les données dans une base que le fournisseur de services rend accessible à ses clients. L'interrogation de cette base est directe et ne sollicite pas les entrepôts d'origine. En effet, l'utilisateur final interroge uniquement le réservoir de notices, constitué par moisson, du fournisseur de service, qui lui retourne en réponse la liste des notices pertinentes.*

Ces notices proposent notamment un lien hypertexte vers le document primaire, seulement accessible sur le serveur du fournisseur de données : en activant ce lien, l'utilisateur arrive sur le site et dans

l'environnement graphique de l'institution productrice, dont le serveur n'est finalement sollicité que pour la fourniture de ce seul document.

Texte extrait du document « Le protocole OAI et ses usages en bibliothèque » (voir Réf.5)

En pratique, un moissonneur d'entrepôt (voir fig. 15/7) peut récolter les données bibliographiques des documents enregistrés dans Archimer en interrogeant notre base de données à l'aide de requêtes HTTP du type : <http://www.ifremer.fr/docelec/oai/OAIHandler?verb=Identify>.

Le moteur OAI-PMH d'Archimer, qui se présente sous la forme d'une Servlet (voir fig. 15/6), répond à ces requêtes HTTP en renvoyant la description, dans un format XML, de nos documents, qu'il récupère en interrogeant la base de données d'Archimer (voir fig. 15/1). Ces flux XML contiennent la description de nos documents (titre, résumés, auteur...) ainsi que l'adresse URL du texte intégral du document hébergé sur le site Ifremer (voir fig. 15/2).

Le moissonneur OAI-PMH enregistre ensuite ces données dans sa propre base (voir fig. 15/8). Les références des documents d'Archimer peuvent ainsi être agrégées avec les références de documents récoltées à partir des archives institutionnelles d'autres organismes de recherches (voir fig. 15/9). Le moissonneur peut ainsi offrir à ses utilisateurs une interface HTML (voir fig. 15/10) qui permet l'interrogation d'une base composée des références de plusieurs entrepôts.

En conclusion, les documents enregistrés dans Archimer sont notamment accessibles par l'intermédiaire:

- du site WEB de consultation d'Archimer (voir fig. 15/3, 15/4),
- de moteurs de recherche comme Google (voir fig. 15/5),
- de moissonneurs OAI-PMH comme Oaister (voir fig. 15/11).

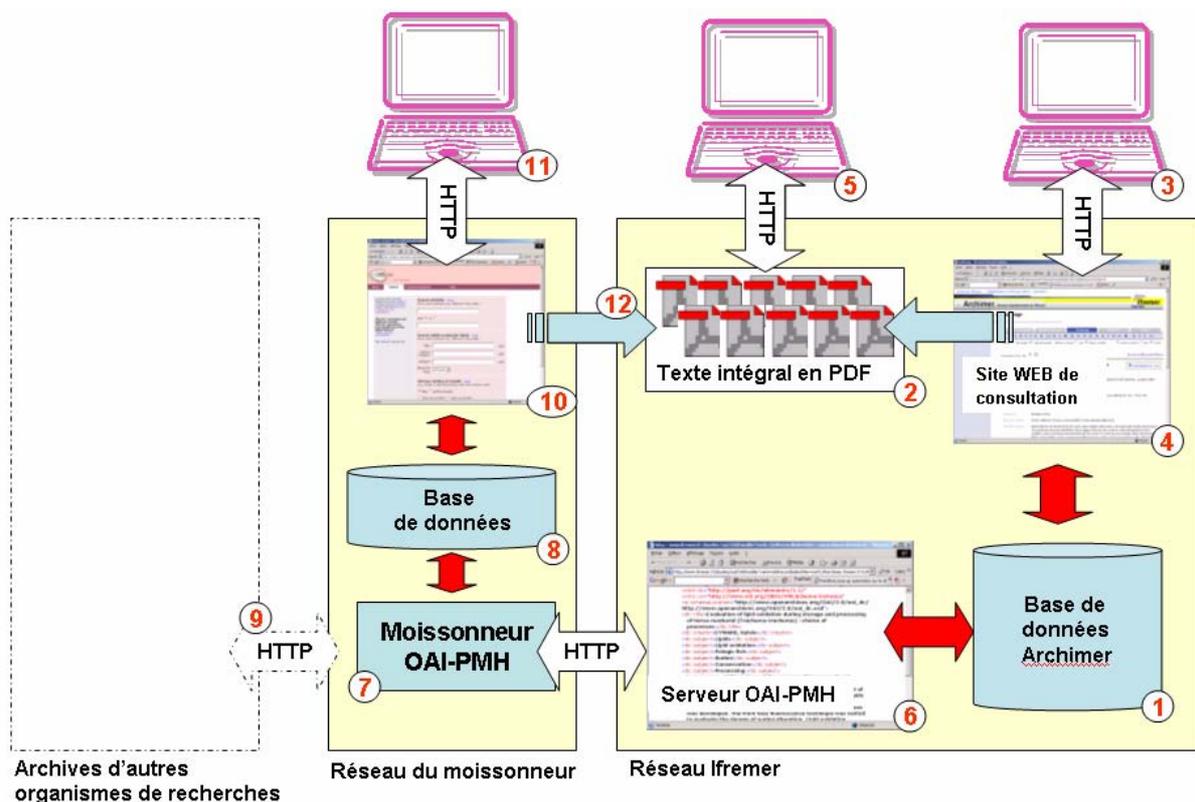


Figure n°15 : Structures des échanges OAI-PMH

4.6. Statistiques d'utilisation

4.6.1. Méthodologie du suivi statistique de l'utilisation du site WEB d'Archimer

Pour suivre l'utilisation du site WEB d'Archimer, un mécanisme a été implémenté dans la page d'accueil du site. Il enregistre, pour chaque session de travail, les informations listées ci-dessous dans une base de données. Ces informations sont disponibles en temps réel à partir d'un site WEB, accessible à tout le personnel du service de documentation :

- **L'horaire de passage** : Date et heure de démarrage de chaque session de travail, enregistré dans la base. Cette information permet de lancer l'édition de statistiques sur une période fixée par le personnel du service de documentation .
- **Le nom du navigateur** : Archimer stocke le nom du navigateur Internet de l'utilisateur. **Cette information nous sert à filtrer les connexions des robots des moteurs de recherche (ex : Google, MSN) et des aspirateurs de sites. En effet, les connexions de ces robots peuvent représenter, en fonction des périodes, de 60 à 90% du trafic.** Des résultats statistiques qui n'auraient pas été nettoyés de ces connexions, n'auraient donc pas de sens. Voici quelques exemples de noms de navigateur :
 - MnoGoSearch/3.2.31
 - Mozilla/4.0 (compatible; MSIE 5.5; Windows NT 5.0)
 - Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)
 - Pompos/1.3 <http://dir.com/pompos.html>

Dans ses résultats statistiques, Archimer prend uniquement en compte les connexions de navigateurs dont le nom contient le mot Mozilla (voir ex. 2 et 3). **Ce premier filtre permet d'éliminer un nombre important de passages de robots.** Dans la liste de connexions répondant à cette première condition, **Archimer élimine ensuite toutes les connexions de navigateurs dont le nom contient une liste de mots considérés comme correspondants à des robots** : Google, Crawl, Pompos, Mirago, Msnbot... (voir ex. 3).

- **L'adresse IP** : cette adresse IP nous permet d'obtenir une idée approximative de l'identité des utilisateurs, en identifiant dans un premier temps, les connexions du personnel de l'Ifremer, puis les connexions provenant de l'extérieur. S'agissant des connexions extérieures, Archimer détermine le pays et l'organisme (ou à défaut leur fournisseur d'accès à Internet) des utilisateurs considérés. Cette détermination est faite à l'aide d'une base de données éditée par la société MaxMind (<http://www.maxmind.com/app/account>), dont la BLP a fait l'acquisition.
- **L'URL de provenance (Referring URL)** : Cette URL est stockée quand un utilisateur arrive sur le site Archimer après avoir suivi un lien Internet sur un autre site WEB. L'information ainsi stockée nous permet de savoir où est référencé notre site WEB et de savoir quels sont les chemins les plus utilisés pour parvenir jusqu'à Archimer.
- **Le pourcentage d'utilisation des différents modes de recherche** : plusieurs modes d'accès aux documents sont proposés par Archimer : des modes de recherche simplifiée, ou avancée, et des modes de feuilletage, par titre et par domaine. Archimer stocke des données relatives au taux d'utilisation de ces différents modules. Elles nous donnent des indications sur le comportement des utilisateurs face à l'ergonomie de nos produits, et nous permettent, également, d'affiner nos résultats statistiques. En effet, les robots des moteurs de recherche ne sont pas capables de soumettre un formulaire dans une page WEB. En analysant le comportement des utilisateurs, il est donc possible de détecter le passage de robots, ou d'aspirateurs, qui auraient échappé aux contrôles réalisés à l'aide du nom du navigateur.
- **Les notices dynamiques visualisées** : Il s'agit de notices générées dynamiquement par Archimer lors de son interrogation par un utilisateur : interrogation qui induit l'enregistrement, par le système, de la référence du document affiché. Ces données viennent compléter les résultats obtenus en analysant les logs de notre serveur WEB (voir paragraphe suivant). Elles permettent l'identification des documents les plus consultés.

- **Les requêtes** : exécutées par l'intermédiaire des modes de recherches simplifiés et avancés, les requêtes sont sauvegardées dans une base de données. L'étude, manuelle, de ces requêtes permet, par exemple de déterminer si la syntaxe du langage de recherche est correctement utilisée. Si ce n'est pas le cas, ces informations permettent d'améliorer l'aide en ligne et les messages d'information affichés à la suite de mauvaises manipulations de l'utilisateur.

4.6.2. Méthodologie du suivi statistique du nombre de téléchargements de fichiers

S'il est possible d'accéder aux notices et au texte intégral des documents par l'intermédiaire de l'interface d'Archimer, il est également possible d'y accéder par l'intermédiaire de moteurs de recherche (ex : Google) ou par l'intermédiaire de moissonneurs OAI (ex : Oaister). L'analyse statistique de l'utilisation de l'interface WEB d'Archimer ne permet donc pas d'obtenir une vision complète du nombre de téléchargements des notices et des fichiers de texte intégral des documents.

Pour compléter les données statistiques récoltées par l'interface WEB d'Archimer, nous analysons les logs du serveur WEB de l'Ifremer. En effet, tous les accès HTTP de tous les sites WEB de l'Ifremer passent par ce serveur Apache et chaque appel à un fichier (HTML, PDF, Image...) est enregistré dans des fichiers de logs. L'analyse de ces fichiers nous permet de détecter tous les accès aux notices statiques et aux fichiers PDF correspondants aux documents enregistrés dans Archimer.

Pour ce faire, un programme JAVA analyse, toutes les nuits, l'activité du serveur Apache et isole les accès aux documents lfremer. Ce programme filtre également :

- les accès des robots des moteurs de recherche et les accès des aspirateurs de sites,
- les téléchargements, par blocs, des fichiers PDF. En effet, depuis les dernières versions d'Acrobat reader, un document PDF, quand il est optimisé pour une lecture sur le WEB, peut-être téléchargé en tâche de fond (la première page du document PDF s'affiche dans le navigateur de l'utilisateur avant que la totalité du document ne lui parvienne). Pendant qu'il lit la première page, le document continue ainsi à lui parvenir, en tâche de fond. Alors que, normalement, une ligne dans un fichier de log du serveur Apache correspond à la visualisation d'un document, ce transfert de fichiers PDF, en plusieurs blocs, peut laisser des centaines de lignes pour une seule lecture.

Le programme JAVA filtre donc, pour les fichiers PDF, ce type de lecture pour ne retenir qu'une seule référence par lecture. Les données de tous les accès aux documents « Archimer » sont ensuite sauvegardées dans une base de données.

4.6.3. Illustration : présentation des résultats synthétiques

En supplément du site WEB qui permet au personnel de la BLP de consulter, en temps réel, le détail des données statistiques non seulement pour Archimer, mais aussi pour les autres ressources électroniques mises à disposition des usagers de la bibliothèques (bases de données bibliographiques, revues électroniques...), des résultats synthétiques sont adressés au personnel de l'ensemble des bibliothèques de l'Ifremer, sous forme d'un bulletin, édité tous les deux mois. Pour Archimer, par exemple, ce bulletin présente l'évolution du nombre de connexions au site WEB d'Archimer (voir fig. 16) ainsi que l'évolution du nombre de téléchargement de documents en texte intégral (voir fig. 17).

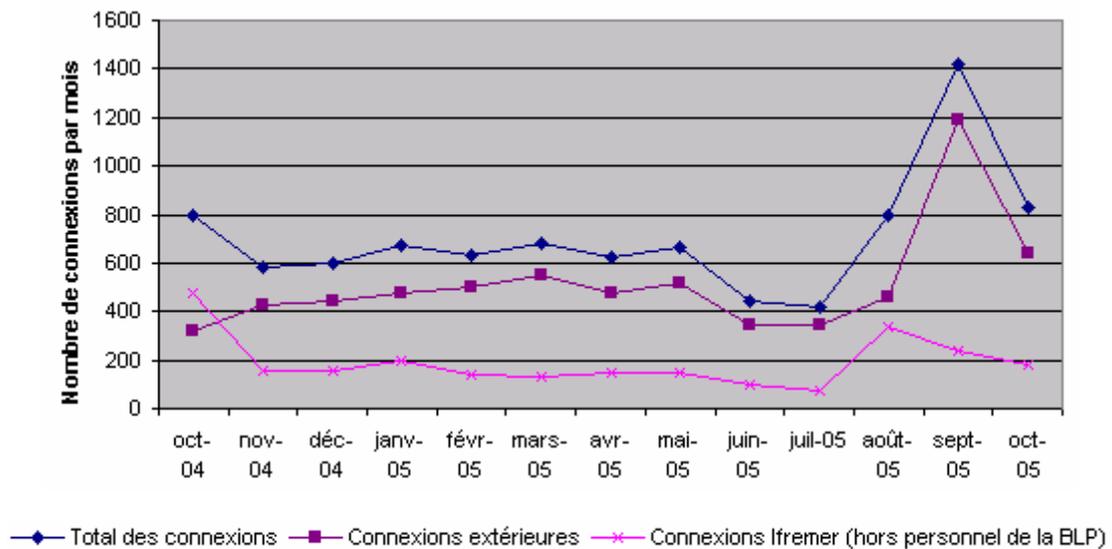


Figure 16 : Evolution, au cours de l'année passée, du nombre de connexions à Archimer.

Remarque : le pic de connexion en septembre 2005, est lié à un message d'information sur Archimer postée sur la liste d'information biblio.fr

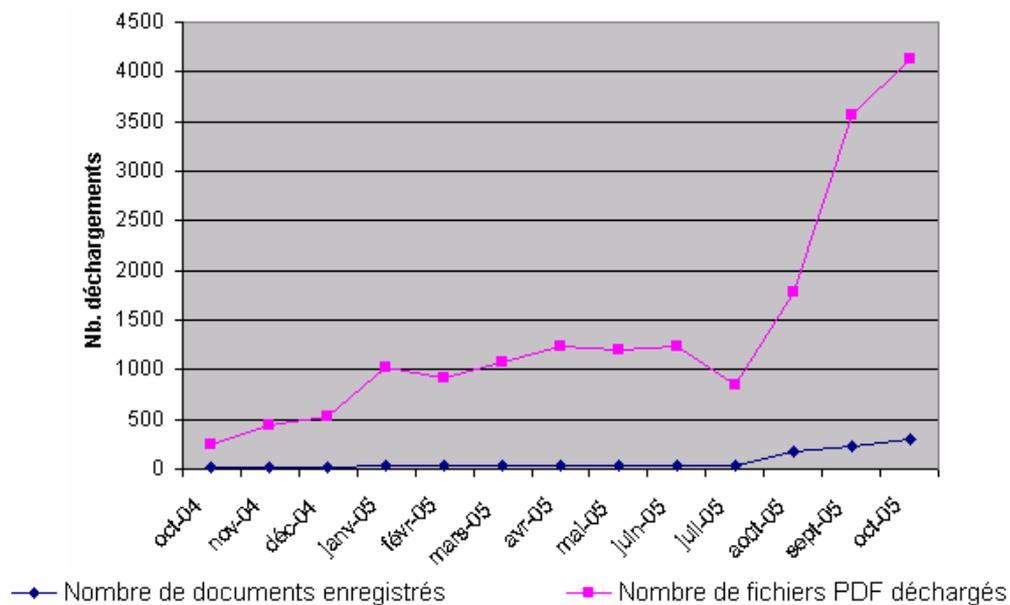


Figure 17 : Evolution, au cours de la même période du nombre de fichiers PDF téléchargés, ainsi que du nombre de documents enregistrés dans Archimer.

4.6.4. Limite de fiabilités des résultats

La principale difficulté rencontrée pour mesurer l'audience d'un site WEB ouvert sur l'Internet réside dans le filtrage des accès des robots de moteur de recherche et des aspirateurs de site. En effet, le passage des robots et des aspirateurs peut représenter, dans le cas d'Archimer, par exemple, de 60 à 90% des connexions au site WEB.

Nous avons donc mis en place un ensemble de mesures pour repérer et supprimer les connexions de ces robots et de ces aspirateurs de nos résultats statistiques, mais il est possible que certaines de ces connexions soient comptabilisés parmi dans nos résultats statistiques.

5. Description de l'Architecture technique

La figure n°18 présente l'organisation technique des principaux modules du système Archimer. Certains modules annexes, comme le module de statistiques ou le module Archimail ne sont toutefois pas représentés sur ce schéma.

L'architecture du système Archimer s'organise autour d'une base de données Oracle (voir fig. 18/1), qui contient les données bibliographiques, et d'un espace disque où le texte intégral des documents est stocké sous forme de fichiers au format PDF (voir fig. 18/2).

Pour enregistrer un nouveau document dans Archimer, le personnel chargé de l'enregistrement (voir fig. 18/3) se connecte à un site WEB (voir fig. 18/4) accessible à partir de l'Intranet de l'Ifremer. Ce site WEB propose un ensemble de formulaires WEB spécifiques aux types de documents à enregistrer (Thèses, rapports internes, publications...).

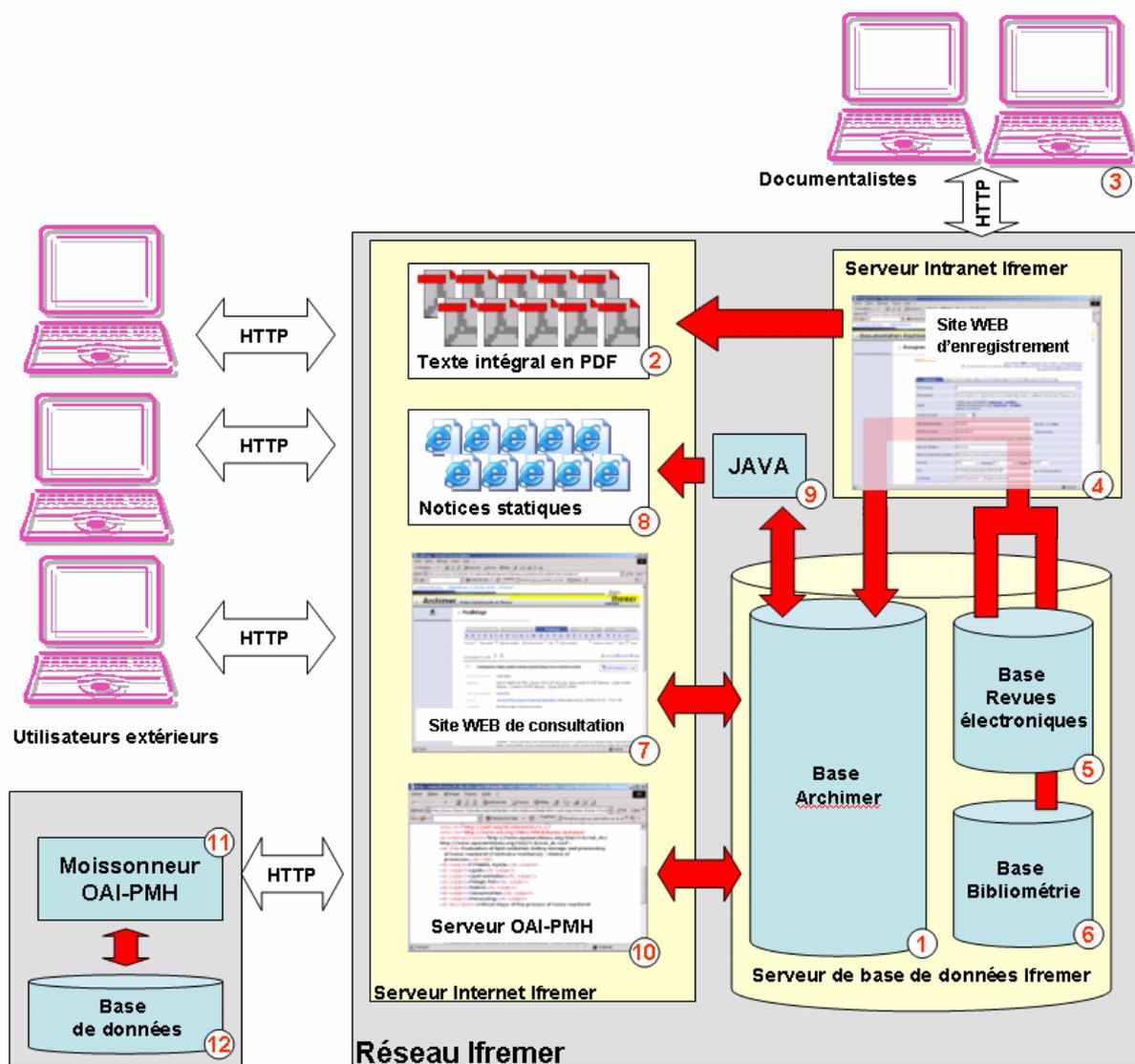


Figure n°18 : architecture du système Archimer

Ces formulaires permettent l'enregistrement des données bibliographiques du document (titre, résumé, auteur...) qui seront sauvegardées dans une base de données (voir fig. 18/3). Ces formulaires permettent

également l'enregistrement du texte intégral, sous forme d'un fichier PDF, stocké sur un espace disque du serveur Internet de l'Ifremer (voir fig. 18/2).

Pour enregistrer une publication déjà référencée dans la base Bibliométrie (voir fig. 18/6), le personnel précité peut saisir l'identifiant associé à ce document dans cette base de données. Cette option permet de transférer automatiquement les données bibliographiques disponibles dans cette base de données vers les formulaires de saisie. Dans ce cas, pour le personnel considéré, il ne reste plus, pour compléter l'enregistrement, qu'à saisir les informations manquantes (DOI, copyright, texte intégral).

Dans le cas de l'enregistrement d'une publication, pour obtenir automatiquement les données relatives à la revue dans laquelle l'article est publié, le personnel de la bibliothèque peut également faire une recherche dans la base « revues électroniques » (voir fig. 18/5). Cette base de données contient la liste cumulative de tous les titres auxquels Ifremer a souscrit un abonnement. Le personnel peut ainsi saisir quelques mots du titre (ex : aqua* liv*), pour retrouver le titre correspondant, et transférer l'ensemble de ces données dans les formulaires de saisie (titre complet de la revue, URL de la revue, nom de l'éditeur, URL de l'éditeur).

Les utilisateurs extérieurs peuvent consulter les documents disponibles via le site WEB d'Archimer (voir fig. 18.7). Dans ce cas, les notices des documents enregistrés sont construites dynamiquement à l'aide de pages JSP en fonction des requêtes exécutées par les utilisateurs. Ces notices proposent des liens vers le texte intégral des documents (voir fig. 18.2).

Toutes les nuits, un programme JAVA (voir fig. 18.9) construit un fichier HTML statique pour chaque nouveau document enregistré (voir fig. 18.5). Ce fichier statique propose la notice du document ainsi qu'un lien vers le texte intégral (voir fig. 18.2). Ces fichiers statiques sont construits à l'intention des robots des moteurs de recherches Internet (ex : Google, MSN). De cette façon, ces notices et le texte intégral des documents sont directement accessibles à partir de ces moteurs de recherche.

Archimer est également compatible OAI-PMH. Un moissonneur (voir fig. 18.11) peut donc récolter les données bibliographiques des données enregistrées dans Archimer en interrogeant son serveur OAI-PMH (voir fig. 18.10). Ce moissonneur pourra ainsi alimenter sa propre base de données bibliographique (voir fig. 18.12) à partir des références récoltées dans plusieurs archives. Ces moissonneurs (ex : Oaister), offriront ainsi, à partir de leur propre interface de consultation, un accès aux notices statiques d'Archimer (voir fig. 18.8).

6. Lancement d'Archimer : état des lieux

Le 24 octobre 2005, 324 documents étaient disponibles, en texte intégral, à partir d'Archimer. Ces 282 documents correspondaient à :

- 35 thèses,
- 279 publications,
- 6 actes de congrès,
- 4 rapports internes.

6.1.1. Enregistrement des publications récentes

Entre le 17 août et le 24 octobre, nous avons repéré 63 publications dans la base Current Contents. Parmi ces 63 publications :

- 5 articles étaient publiés chez des éditeurs qui autorisaient l'auto-archivage de leurs propres fichiers PDF auxquels nous avons accès. Nous avons donc pu les enregistrer sans l'aide des auteurs,
- 13 articles étaient publiés chez des éditeurs qui interdisaient l'auto-archivage. Il s'agissait principalement des éditeurs AMS et Inter-research. Nous n'avons donc pas pu les enregistrer,
- 1 article était publié chez un éditeur qui imposait un embargo de deux ans avant l'archivage souhaité. Nous n'avons donc pas pu enregistrer l'article considéré,

- 44 articles étaient publiés chez des éditeurs qui limitaient le droit d'auto-archivage aux derniers drafts de l'auteur ou chez des éditeurs qui autorisaient l'auto-archivage de leurs propres fichiers PDF auxquels nous n'avions pas accès. Dans tous ces cas, nous avons contacté les auteurs pour leur demander les données qui nous manquaient pour compléter l'enregistrement. A noter que les auteurs de 16 de ces 44 articles nous ont fourni les données que nous recherchions.

Nous avons donc enregistré 21 des 63 publications répertoriées dans la base de données Current-Contents. Cela correspond à un pourcentage de 33% des publications récentes.

6.1.2. Enregistrement des articles publiés entre 2000 et 2005

L'exploitation de la base de données Current Contents nous a permis de repérer qu'environ 275 publications sont rédigées ou co-rédigées par le personnel de l'Ifremer chaque année. Ce repérage a été confirmé par le contenu de notre base de données « Bibliométrie », à savoir près de 1500 références d'articles collectées pour la période 2000-2005. Ce sont ces 1500 références que nous avons commencées de traiter. Nous en avons, pour l'instant, analysées plus de 850. Parmi ces 850 publications :

- **près de 150 articles** ont été publiés chez des éditeurs qui n'autorisent pas l'auto-archivage,
- **nous avons pu enregistrer près de 50 articles** sans contact avec les auteurs **car l'éditeur autorisait l'archivage de ses propres fichiers PDF,**
- **650 articles étaient publiés chez des éditeurs qui autorisent uniquement l'archivage du dernier « draft » de l'auteur.** Nous avons posté des messages à tous les auteurs Ifremer de ces 650 publications. Ces messages nous ont permis de récolter environ 100 articles supplémentaires.

7. Objectifs à court et moyen termes

7.1. Court terme

7.1.1. Poursuite de l'enregistrement de l'arriéré des publications (années 2000-2005)

Par rapport aux 1500 publications Ifremer publiées entre les années 2000 et 2005, **il nous en reste près de 650 à traiter.** Nous espérons terminer ce traitement courant 2006.

7.1.2. Développement d'un prototype de moissonneur spécialisé en sciences de la mer

Nous avons également le projet de développer un moissonneur OAI-PMH spécialisé en sciences de la mer. Ce moissonneur interrogera, au plan international, un ensemble de serveurs compatibles OAI-PMH (dont Archimer), pour en extraire les références de documents relatives aux sciences de l'océan. Ces références seront agrégées dans une base de données cumulative, dont l'interrogation sera possible par l'intermédiaire d'un nouveau site WEB. Ce projet serait l'occasion :

- d'offrir au personnel de l'Ifremer une nouvelle source d'informations, en agrégeant des documents dont l'accès est aujourd'hui difficile du fait de leur dissémination dans un ensemble de serveurs au contenu non indexé dans les bases de données bibliographiques traditionnelles,
- d'offrir une visibilité supplémentaire aux publications de l'Ifremer enregistrés dans Archimer : en effet, en agrégeant ces publications avec un ensemble de documents relatifs aux sciences de l'océan, nous pourrions espérer atteindre une masse critique de documents qui pourrait faire de cette base de données une source d'informations de référence dans le domaine des sciences de la mer. Nous pourrions ainsi conduire un public spécialisé à consulter cette base de données dans laquelle les documents Ifremer auraient une place privilégiée.

7.2. Moyen terme

7.2.1. Collecte des « versions auteurs » de publications dès l'acceptation par l'éditeur

Pour l'instant, très peu d'auteurs déposent leurs documents spontanément dans Archimer. Nous avons obtenu la majorité des documents actuellement enregistrés dans Archimer en contactant nommément leurs auteurs. Cependant, cette méthode présente plusieurs limites :

- Lorsque nous cherchons à contacter les auteurs d'une publication, ils ont parfois déjà quitté l'Ifremer. Cela s'explique par le fait qu'il peut s'écouler plus d'une année entre la soumission d'un article à une revue et son apparition dans les Current Contents. Quand un thésard publie un article pour présenter ses travaux à la fin de sa thèse, il a donc souvent quitté l'Ifremer au moment où son article est publié et qu'il apparaît dans les Current Contents,
- Si un éditeur autorise uniquement la diffusion du dernier « Draft » d'une publication, au moment où nous contactons les auteurs pour obtenir cette version, il est souvent trop tard, car ces fichiers ont été perdus ou supprimés.

Nous envisageons donc mettre en place une collecte systématique des « versions auteurs » dès leur acceptation par une revue. En effet, si les auteurs nous déposaient systématiquement leurs fichiers dès l'acceptation de leur publication par une revue, nous pourrions, non seulement améliorer le pourcentage de collecte des publications Ifremer, mais surtout, sous réserve d'une compatibilité avec les règles de copyright fixées par l'éditeur, diffuser les publications « In-Press ». Nous pourrions ainsi participer à l'accélération du processus de diffusion des résultats des recherches de l'Ifremer, en diffusant les publications plusieurs mois avant qu'elles n'apparaissent sur le site Internet de l'éditeur.

7.2.2. Élargissement du système à d'autres types de documents

Actuellement, Archimer permet l'enregistrement et la diffusion de thèses, de post-publications, de rapports internes, de rapports d'activités et d'actes de congrès. Nous pourrions, ultérieurement, envisager l'intégration d'autres types de documents :

- Publications « In-Press »,
- Pré-publications,
- Brevets,
- Posters,
- Ouvrages
- ...

7.2.3. Numérisation et enregistrement de documents sur support papier

Début 2006, nous installerons une chaîne de numérisation, pour permettre l'enregistrement et la diffusion, sur Internet, de documents anciens dont nous ne possédons pas de version électronique. Cette chaîne de numérisation pourrait, par exemple, être appliquée à quelques thèses en taxonomie. Elle sera composée d'un scanner professionnel, pour la numérisation des pages, et d'un logiciel d'OCR, pour générer un fichier PDF de type texte, dont le contenu pourrait être indexé par les robots des moteurs de recherche.

De plus, cette chaîne de numérisation pourrait également être appliquée à certains fichiers PDF, que nous recevons et qui posent problèmes. Ces fichiers, préparés avec le logiciel Latex, ressemblent à des fichiers PDF texte, mais n'en sont pas. Ils permettent de sélectionner des portions de texte à l'aide du logiciel Acrobat reader, mais non de « copier/coller » ou de « Rechercher ». De la même façon, les robots des moteurs de recherche ne peuvent pas extraire le texte de ces fichiers et ne peuvent donc pas les indexer. Ces documents sont les moins visibles sur Internet, comme le démontre l'analyse de nos résultats statistiques.

D'après les informations que nous avons obtenues sur ce problème, si nous n'avons pas accès aux sources Latex, la seule façon de corriger ces documents serait de les importer dans un logiciel d'OCR.

Archimer, ou la mise en place d'une Archive Institutionnelle à l'Ifremer

La chaîne de numérisation et le logiciel précités devraient donc aussi nous permettre de restaurer, en vue de leur intégration dans Archimer, de tels documents.

Références bibliographiques

- (1) Lawrence, S. (2001) Free online availability substantially increases a paper's impact. *Nature* 411, (6837) : 531 (Web Debates)
- (2) Brody, T., Stamerjohanns, H., Harnad, S. Gingras, Y. & Oppenheim, C. (2004) The effect of Open Access on Citation Impact. Presented at: National Policies on Open Access (OA) Provision for University Research Output: an International meeting, Southampton, 19 February 2004.
<http://opcit.eprints.org/feb19prog.html>
<http://www.ecs.soton.ac.uk/~harnad/Temp/OA-TAadvantage.pdf>
- (3) Kurtz, Michael J.; Eichhorn, Guenther; Accomazzi, Alberto; Grant, Carolyn S.; Demleitner, Markus; Murray, Stephen S.; Martimbeau, Nathalie; Elwell, Barbara. (2003) The NASA Astrophysics Data System: Sociology, Bibliometrics, and Impact. *Journal of the American Society for Information Science and Technology* <http://cfa-www.harvard.edu/~kurtz/jasis-abstract.html>
- (4) Kurtz, M.J. (2004) Restrictive access policies cut readership of electronic research journal articles by a factor of two, Harvard-Smithsonian Centre for Astrophysics, Cambridge, MA
<http://opcit.eprints.org/feb19oa/kurtz.pdf>
- (5) Texte extrait du document « Le protocole OAI et ses usages en bibliothèque »
François NAWROCKI - Ministère de la culture et de la communication
28 janvier 2005 - <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>