

Master en Sciences de l'Information
et des Bibliothèques
Option : « Réseaux d'information et document
électronique »

**Les publications scientifiques
publiées par les praticiens et les
chercheurs de l'Institut Claudius
Regaud :**

**réalisation d'une base de données
des publications.**

Natacha BIED-CHARRETON

Sous la direction de Catherine Delamaire
Coordinateur Information-Documentation
Institut Claudius Regaud
20-24 rue du Pont Saint-Pierre
31052 Toulouse Cedex



Remerciements

Nombreuses sont les personnes auxquelles je tiens à remercier dans le cadre de ce stage d'application.

Je remercie tout d'abord Catherine Delamaire, responsable de la médiathèque médicale de l'Institut Claudius Regaud, pour m'avoir confié l'étude de la problématique des publications scientifiques et pour m'avoir suivi tout au long de la mise en œuvre du projet.

Je remercie également Clarys Lheureux et Laurent Mazoyer, documentalistes à la médiathèque médicale pour leur accueil et pour m'avoir fait part de leurs réalisations documentaires.

Je remercie les personnes comme Pierre Canal, responsable du Département d'Information Hospitalière et Fañçoise Foglieni, Urbaniste des données, pour m'avoir conseillé lors de l'étude des solutions envisageables.

Je remercie les praticiens et les chercheurs pour avoir accepté de répondre à mon enquête.

Enfin, je remercie les quatre correcteurs du présent rapport, Catherine, Clarys , Bertrand, Geneviève pour leur disponibilité et leurs précieux conseils.

Résumé :

Les publications scientifiques ont une place importante dans la pratique médicale à titre de capitalisation des connaissances issue d'un processus cumulatif d'accroissement des connaissances et dans le cadre de la pratique du consensus médical. Plus particulièrement, les enjeux propres à un établissement spécialisé dans les traitements contre le cancer comme l'Institut Claudius Regaud, sont de pouvoir réaliser les rapports d'activité scientifiques et de réaliser la capitalisation des connaissances en oncologie.

Ce rapport a pour objectif de présenter les étapes qui ont conduit à la réalisation d'une base de données des publications scientifiques produites par les praticiens et les chercheurs de l'Institut Claudius Régaud. Une analyse de l'existant évaluant les points forts et les points faibles du système de gestion des publications scientifiques, nous conduit au choix d'une solution parmi les différents scénarios proposés pendant le stage. La dernière partie présente la mise en œuvre du nouveau système et les perspectives d'évolution.

Descripteurs : Publication scientifique, base de données, publication électronique, édition scientifique, logiciel de références bibliographiques, oncologie, santé, Archives Ouvertes, Système d' Information, capitalisation des connaissances.

Abstract :

Scientific publications have an important role in practical medicine : it is useful for knowledge management and for practicing medical consensus. More specifically, in a cancer Institute dedicated to comprehensive medical care, the objectives are realizing scientific report and constituting scientific knowledge created by practitioners and researchers.

This report aims at describing steps to creating an online database of medical publications published by practitioners and researchers. A study of the actual system points at analyzing the Information flow and the management of scientific publications. A solution to the new information system is chosen after having studying several possibilities. The last part describes the implementation and new perspectives for a future time.

Keywords : scientific publications, databases, electronic publication, scientific edition, bibliographic software, oncology, medicine, Open Archives Initiative, Information System, knowledge management.

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée

Sommaire

INTRODUCTION	6
ANALYSE PREALABLE	8
1. ELABORATION D'UN PLAN DE TRAVAIL	8
2. ETUDE DU SYSTÈME D'INFORMATION DE L'ICR	8
2.1. <i>Rôle et mission du Département d'Information Hospitalière (DIH)</i>	<i>8</i>
2.2. <i>Rôle et mission du Système d'Information Hospitalière (SIH)</i>	<i>10</i>
2.3. <i>Le dossier informatisé du patient, un exemple d'application du DIH</i>	<i>11</i>
3. LA GESTION DOCUMENTAIRE, UN EXEMPLE D'APPLICATION TRANSVERSALE DU SYSTÈME D'INFORMATION HOSPITALIER.....	12
3.1. <i>GEIDE.....</i>	<i>13</i>
3.2. <i>Intranet.....</i>	<i>15</i>
3.3. <i>Worklow et travail collaboratif.....</i>	<i>16</i>
3.4. <i>Solution technique employée pour la gestion documentaire</i>	<i>17</i>
4. ANALYSE DE LA GESTION DES DOCUMENTS SCIENTIFIQUES DANS LE DOMAINE MÉDICAL	19
4.1. <i>L'existant à l'Institut Claudius Regaud.....</i>	<i>19</i>
4.2. <i>Analyse de l'existant en externe : expériences dans le domaine de l'oncologie et en sciences</i>	<i>21</i>
5. ETUDE DES PRATIQUES ET DES BESOINS	24
5.1. <i>Enquête par entretiens.....</i>	<i>24</i>
5.2. <i>Objectifs du questionnaire</i>	<i>24</i>
5.3. <i>Synthèse des entretiens</i>	<i>24</i>
5.4. <i>Bilan du fonctionnement actuel.....</i>	<i>30</i>
5.5. <i>Analyse des besoins exprimés par les utilisateurs (Médecins, chercheurs, Département du Système d'Information Hospitalier)</i>	<i>38</i>
PROPOSITION D'UN NOUVEAU DISPOSITIF.....	42
1. ENJEUX DU NOUVEAU SYSTÈME	42
1.1. <i>Enjeux du nouveau dispositif</i>	<i>42</i>

1.2. <i>Public visé</i>	42
2. ELABORATION D’UN CAHIER DES CHARGES FONCTIONNEL ET TECHNIQUE.....	42
2.1. <i>Contraintes du système</i>	43
2.2. <i>Architecture fonctionnelle du nouveau système d’information</i>	44
2.3. <i>Caractéristiques techniques du nouveau système</i>	47
3. ANALYSE CONCEPTUELLE DES DONNÉES À INFORMATISER	48
3.1. <i>Nouveaux flux d’information</i>	48
3.2. <i>Modèle conceptuel des données à informatiser (MCD)</i>	51
3.3. <i>Modèle Physique des données à informatiser (MPD)</i>	52
MISE EN ŒUVRE DU SYSTEME	54
1. SCÉNARIOS POSSIBLES	54
1.1. <i>Module web Reference Manager version 11 (module web publisher)</i> .	54
1.2. <i>Solution Open Source web dynamique : PHP3, Mysql, Apache</i>	56
1.3. <i>Solution mixte : PHP, Oracle</i>	56
1.4. <i>Entrepôt OAI (Open Archives Initiatives, Auto archivage)</i>	58
1.5. <i>Structuration de documents basée sur le langage XML</i>	59
2. EVALUATION DES SCÉNARIOS	60
3. PROJET RETENU	61
4. MISE EN ŒUVRE DU SYSTÈME	62
4.1. <i>Etapas</i>	62
4.2. <i>Collecte des données</i>	63
4.3. <i>Mise en ligne de la base</i>	68
4.4. <i>Bilan de la réalisation de la base de données</i>	70
CONCLUSION : ÉVOLUTION ET PERSPECTIVES DU SYSTÈME	71
BIBLIOGRAPHIE	73
TABLE DES ANNEXES	78

Introduction

L’Institut Claudius Regaud (ICR) ¹ est le Centre Régional de lutte contre le cancer en Région Midi-Pyrénées. C’est un établissement privé à but non lucratif, participant au Service public. Il fait partie de la Fédération Nationale des Centres de Lutte Contre le Cancer² (FNCLCC) qui regroupe les 20 centres français.

L’établissement a pour but de dépister et de diagnostiquer les tumeurs, d’effectuer les soins auprès des patients, de participer à la mission de recherche pour développer de nouveaux médicaments et de réaliser des activités de formation continue auprès des professionnels de la santé.

A notre arrivée, le projet de stage est défini avec la responsable de la médiathèque médicale sous forme d’objectifs généraux à atteindre : réaliser une base de données des publications scientifiques des praticiens et des chercheurs de l’Institut Claudius Regaud.

Deux raisons principales rendent urgent d’effectuer ce projet au sein de l’Organisme :

- Pouvoir réaliser la mémoire de l’entreprise et la capitalisation des connaissances diffusées sur internet
- Pouvoir rendre compte de l’activité scientifique de l’Organisme auprès des décideurs au travers de documents spécifiques comme l’élaboration du rapport d’activités annuel (évaluation de la production des documents scientifiques, calcul du facteur d’impact des articles dans les revues) ; rapport INCA (Institut National du Cancer), DOMES, ICARES...

Avant le démarrage du projet, il convient d’en cerner les enjeux en s’interrogeant sur le sens de la publication scientifique et sur sa mise en contexte professionnel.

¹ Voir la fiche récapitulative de l’Organisme en annexe 1 page 78

² <http://www.fnclcc.fr/>

La question est de savoir quels sont le rôle et la place d'une publication scientifique dans la pratique clinique et dans l'activité de recherche. En effet, « la publication scientifique n'est pas un processus homogène, elle ne peut être dissociée de son contexte de production et de diffusion, en particulier des contraintes du travail scientifique, de son environnement politique et économique. Les modalités d'inscription et de diffusion qu'elle privilégie se diversifient en conséquence »³

Les publications scientifiques prennent une place importante dans l'exercice de la pratique médicale. En effet, dans le milieu hospitalier, une enquête a montré que la collecte de données scientifiques publiées joue un rôle majeur :

- à titre de capitalisation des connaissances issue d'un processus cumulatif d'accroissement des connaissances
- pour travailler en réseau dans le cadre du consensus médical
- pour la formation médicale (la quantité de données à connaître et à maîtriser ne cesse de s'accroître, il est nécessaire d'opérer une veille scientifique)

L'Institut Claudius Regaud souhaite pouvoir élaborer un dispositif pour mettre en valeur les documents. La démarche qui sera adoptée consistera à analyser les besoins des usagers puis à élaborer un nouveau système qui tiendra compte de deux impératifs :

- la mise en adéquation du nouveau dispositif avec le Système d'Information Hospitalier existant (GEIDE, Intranet, procédures de travail collaboratif, Internet)
- le choix d'un système informatique pérenne et évolutif

Plusieurs initiatives ont vu le jour avec l'essor des Nouvelles Technologies de l'Information et de la Communication comme le mouvement des Archives Ouvertes avec l'essor des logiciels libres : diverses solutions techniques sont possibles. Il s'agira donc de les étudier pour opérer un choix.

³ Chartron ,G. Elément pour une approche comparée de la publication scientifique, 2003 (voir bibliographie).

ANALYSE PREALABLE

1. Elaboration d'un plan de travail

Le projet démarre par l'élaboration du plan de travail. Les principales étapes du projet sont l'analyse préalable au projet, l'étude d'une solution et la mise en œuvre du système. La première étape consiste en une étude approfondie des pratiques des usagers en matière de publications scientifiques et une analyse des besoins. Un diagramme de GANT a été utilisé pour planifier la démarche : les différentes étapes du projet sont récapitulées dans le document figurant dans l'annexe 3 page 83.

2. Etude du Système d'Information de l'ICR

Le Système d'Information de l'ICR, appelé Système d'Information Hospitalier (SIH) est géré par le Département d'Information Hospitalière (DIH). L'objectif fixé de cette première étape est de comprendre les missions du DIH et d'analyser la circulation des informations gérées par le SIH. Le projet pourra ainsi s'intégrer dans le Système existant.

2.1. Rôle et mission du Département d'Information Hospitalière (DIH)

Un entretien avec l'Urbaniste des Données du Département ⁴ du système d'Information Hospitalier (DIH) ainsi que la consultation des documents de

⁴ L'urbaniste des données a en charge l'évolution et la conception du Système d'Information comme l'assistance aux projets relatifs aux architectures fonctionnelles et techniques ainsi qu'aux règles d'urbanisme ; il gère aussi l'administration et la proposition de règles de cohérence du référentiel de données.

présentation du DIH élaborés dans le cadre de la démarche Qualité implantée dans l'organisme ont permis de comprendre le rôle de ce service.

Le DIH est directement rattaché à la Direction : l'organigramme figurant dans l'annexe 2, page 81 présente le département. Ce dernier pilote l'informatisation des services de l'ICR sur les plans technique, fonctionnel et stratégique. Il met en adéquation les besoins des patients de l'ICR, la politique de l'établissement et les technologies offertes sur le marché. L'usage des nouvelles technologies est bénéfique pour le patient. En effet, la circulation de l'information est plus rapide sur :

- les résultats d'examen
- les traitements ou les consultations et les législations concernant les droits du patients imposent la traçabilité des actes médicaux.
- la diffusion de l'information entre services et à l'extérieur de l'hôpital (notamment pour participer aux réseaux de soins).
- le contrôle des processus et des coûts notamment pour les essais cliniques s'effectue par la tarification à l'activité (T2A).

En conséquence, il est essentiel pour le DIH de veiller à la priorité d'informatisation et à la préparation aux changements nécessaires. Face à la multiplicité des systèmes d'information, la mission majeure est d'intégrer les différentes sources et bases de données existantes dans l'établissement.

Le DIH à vocation à devenir à terme, la structure fédérant l'ensemble des contributeurs au recueil, à la validation et à l'exploitation des informations médicales et non médicales.

Il s'appuie sur trois pôles d'activités :

- le pôle infrastructure technique : réseaux, sécurité des réseaux, administration des systèmes et de la convergence images-données,
- le pôle Etude et Développement du système d'Information : urbanisation des données (modélisation des données, choix des applications et implantation des systèmes), projet d'informatisation et accompagnement au changement.

- Le pôle Production d'informations médicales (assistanat médical) notamment pour le dossier du patient informatisé (DPI) et contrôle de l'information par le Département d'Information Médicale (DIM).

2.2. Rôle et mission du Système d'Information Hospitalière (SIH)

Deux domaines constituent le SIH :

- les applications médicales ou liées à l'hospitalisation
- les applications administratives ou sectorielles

Les applications médicales sont gérées par un Progiciel de Gestion Intégré (PGI) appelé « Quadrat » comprenant des déclinaisons pour les différents départements de soins et pour la délégation « Recherche médicale », cette dernière se composant en plusieurs spécialités :

- Radiothérapie
- Médecine nucléaire
- Biologie,
- Diététique,
- Chimiothérapie,
- Imagerie (PACS),
- Anapathologie,
- Stérilisation, PMSI

Les applications administratives ou sectorielles sont gérées par différents logiciels en fonction des activités administratives de l'Organisme Hospitalier :

- Gestion des risques (logiciel Erisk)
- Médecine du Travail
- Gestion du patrimoine/ Bâtiment
- Gestion documentaire (logiciel Sharepoint)
- Gestion du Parc informatique
- Gestion des Ressources Humaines (logiciel Hypervision)
- Trésorerie

- Enterprise Ressource Planning (ERP)⁵ administratif

L'application GAP permet de relier les applications comme la transmission des données administratives du patient, la facturation hospitalière et la facturation des prestations.

2.3. Le dossier informatisé du patient, un exemple d'application du DIH

Depuis 1996, un projet de création de réseau d'archivage et de communication d'images médicales pour le dossier du patient médicalisé a été réalisé entre le DIH et le personnel de la communauté médicale : le réseau s'appelle PACS.

Les objectifs sont :

- de stocker pendant deux ans et de rendre accessible par l'application PACS les images provenant des différentes sources sur un serveur dédié ainsi que sur CD-Rom pour le stockage à long terme,
- de faire circuler les images au sein du département des radiations,
- d'intégrer les images dans le SIH,
- de reconstituer l'exhaustivité des images médicales concernant un patient en y intégrant des vidéos réalisées aux blocs opératoires et en diffusant les images aux cliniciens dans le cadre de la création du dossier du patient hospitalisé.

La création d'un réseau d'images permet ainsi la prise en charge globale du patient par l'intégration d'un dossier électronique. Le dossier est mis à disposition du

⁵ ERP : Les ERP (en anglais *Enterprise Resource Planning*), aussi appelés *Progiciels de Gestion Intégrés (PGI)*, sont des applications dont le but est de coordonner l'ensemble des activités d'une entreprise (activités dites verticales telles que la production, l'approvisionnement ou bien horizontales comme le marketing, les forces de vente, la gestion des ressources humaines, etc.) autour d'un même système d'information. *Comment ça marche*, <http://www.commentcamarche.net/entreprise/erp.php3>, consulté le 10/10/2005

réseau de praticiens au moyen d'une expertise collective ce qui vise à supprimer les cloisonnements entre disciplines.

La particularité de ce réseau est :

- d'intégrer le dossier administratif (SIR, dossier du patient comprenant les demandes d'examens d'imagerie) au dossier médical (PACS) tout au long du traitement médical,
- de pouvoir archiver le dossier dans son intégralité tout au long du traitement.

Sur le plan de l'architecture du système, les données sont enregistrées séparément sur des stations de travail en fonction des différents stades du traitement médical (radiodiagnostic, sénologie, médecine nucléaire, radiothérapie, bloc opératoire.) Ces données sont ensuite rassemblées sur un même serveur informatique puis répliquées par routage sur un serveur web (web1000). L'objectif étant de rendre les informations accessibles aux cliniciens, celles-ci étant exportées sur un serveur de résultats au fur et à mesure des demandes (Mediweb). Des sauvegardes sont effectuées sur des serveurs de grande capacité des stockage (2, 5 Teraoctets) et des baies de DVD.

L'activité de production de dossiers PACS est importante : 5404 dossiers sont stockés sur quatre mois, principalement des dossiers de radiologie.

3. La Gestion Documentaire, un exemple d'application transversale du Système d'Information Hospitalier

La médiathèque médicale joue un rôle transversal entre les services et les différents métiers. Elle a en charge la communication de documents scientifiques pour les praticiens et les chercheurs (en interne et en externe). Elle a également la mission d'alimenter la page d'accueil de l'Intranet, une partie de l'Internet, et la

GEIDE (Gestion Électronique d'Informations et de Documents pour l'Entreprise⁶). Elle participe enfin à la mise en place d'un système de travail collaboratif à l'aide du Workflow.

3.1. GEIDE

Un système de GEIDE est mis en place depuis cinq ans. Il est complété depuis deux ans par un dispositif de Workflow.

Au moment de la création du document, l'utilisateur peut décider si son document est versé sur Intranet ou sur le serveur GEIDE :

- si le document concerne l'ensemble de la communauté, il sera versé sur Intranet (l'accès est ouvert à tous),
- si le document est réservé à un groupe déterminé, à un sous-groupe, ou pour archivage, il sera versé sur le serveur GEIDE (l'accès est limité).

3.1.1. Définition des objectifs

L'objectif principal est de permettre un accès sécurisé en fonction des groupes d'utilisateurs déclarés au préalable. Il offre en outre l'avantage de visualiser le suivi du cycle de vie du document : versionning (création, publication, diffusion, conservation). Enfin, il permet de générer des documents multi-formats dans un environnement microsoft.

3.1.2. Sécurisation des accès

Des droits d'accès aux documents sont attribués pour les lecteurs, les auteurs, les contributeurs et les administrateurs. Un lecteur peut lire le document pour lequel il est autorisé. Les auteurs peuvent créer des procédures mais ils peuvent modifier

⁶ Selon la définition disponible sur le site EDUCNET <http://www.educnet.education.fr/ecogest/veille/communication/com02.htm>

uniquement les documents qu'ils ont créés. Les contributeurs approuvent le document, ils peuvent créer les groupes de personnes habilitées à valider la procédure y compris exclure l'administrateur. Les deux administrateurs gèrent l'ensemble du circuit des documents : ils peuvent examiner tous les documents, sauf ceux exclus par les contributeurs.

3.1.3. Typologie des interfaces disponibles pour les « geideurs »

Deux interfaces sont possible pour les « geideurs » :

- L'interface en client léger s'adresse au « petits contributeurs » pour les personnes a faible activité de publication : elle permet une navigation et une ergonomie conviviales, grâce au rendu HTML. La majeure partie des accès est fournie en client léger (25 accès actuellement). Cette interface offre l'historique des publications du document et de vérifier si le document est bien publié.
- L'interface en client lourd pour les « grands contributeurs » à forte activité de publication comme les responsables de services : elle offre la possibilité de verser plus de documents à la fois dans la GEIDE .

3.1.4. Arborescence des fichiers de données

L'arborescence et la structure des fichiers ont été élaborées d'une part en respectant la classification de la Haute Autorité de Santé (ANAES) et d'autre part en tenant compte de la structure organisationnelle de l'Institut. Après une utilisation de cinq années, cette arborescence va être amenée à évoluer pour mieux tenir compte de l'évolution structurelle de l'Institut.

3.1.5. Elaboration de documents modèles et de procédures

Des méthodes de publication sont élaborées par la médiathèque pour normaliser la mise en forme des documents et pour expliquer les procédures de versement sur le serveur dédié à la publication des documents GEIDE. Une feuille de style est disponible sur l’Intranet ainsi que dans un onglet spécifique installé dans le logiciel word.

3.2. Intranet

Véritable système de gestion de circulation de l’information, le portail Intranet a pour objectif d’offrir l’accès à distance d’un certain nombre de documents à l’ensemble de la communauté de l’ICR. La page d’accueil donne l’accès aux informations d’actualité concernant la vie de l’ICR (sous forme de brèves comme l’annonce de projets ou de conférences.) et offre l’accès aux documents produits par les différents services. La navigation vers les documents s’opère grâce à l’arborescence des fichiers et à l’interrogation par un moteur de recherche. Les documents sont publiés en version HTML grâce à une couche applicative développée par un prestataire de service.

Il est possible à partir de la page d’accueil d’accéder à d’autres applications comme le logiciel de déclaration des risques ou la gestion des emplois du temps des salariés (Chronos) .

Les ressources de la médiathèques comme la liste des ouvrages, la liste des revues électroniques, les photos et la liste des publications scientifiques de l’ICR sont accessibles à l’ensemble de la communauté par cet outil (Rubrique médiathèque).

La recherche d’information est possible grâce au moteur de recherche intégré au portail Sharepoint : il offre une possibilité de rechercher à travers tout le site et en plein texte les données de type texte issus de documents de l’environnement Microsoft office. Un moteur de recherche spécifique pour trouver un document image a donc été rajouté pour la rubrique « Photothèque ».

3.3. Workflow ⁷ et travail collaboratif

3.3.1. Workflow de demandes de documents scientifiques

Des procédures facilitant le travail collaboratif ont été créées pour signaler l'état d'avancement d'un document.

Par exemple, pour faire une demande de document scientifique (article, thèse, ouvrage) non disponible à la médiathèque, un formulaire de demande d'article est créé. Le suivi de la demande est alors disponible pour tous les usagers dans la rubrique médiathèque, puis la rubrique workflow demande d'articles. En se connectant avec un login, chaque demandeur visualise ainsi l'état de sa demande sur l'Intranet (il peut s'agir d'une nouvelle demande, d'une demande en cours et d'un avis de réception de l'article).

3.3.2. Workflow Qualité

La démarche qualité fait partie intégrante du fonctionnement de l'Institut : elle permet d'apporter une offre de services tournés vers le patient. L'accréditation de l'Institut par la Haute Autorité de Santé garantit cette qualité ⁸.

Un grand nombre de modes opératoires et de documents modèles ont été élaborés à tous les niveaux des services par les Correspondants Qualité ⁹ et par la Délégation Qualité. Le document standard comprend un cartouche donnant la date de création et de mise à jour, la diffusion, le code du document, la validation, l'état de la procédure, le titre du document.

L'avantage de cette démarche est de pouvoir vérifier et valider le document par les approbateurs concernés et par la Délégation Qualité. Le document peut-être ré-

⁷ Le Workflow est système permettant l'automatisation des procédures du flux de travail où les documents et les tâches sont transmises d'un participant à un autre selon des règles (source : <http://www.commentcamarche.net/>)

⁸ Pour en savoir plus sur la démarche d'accréditation : <http://www.anaes.fr>

⁹ Les correspondants qualité sont des médiateurs entre les services et les qualitiens

adressé à l'auteur pour modification par téléchargement sous forme de lien ou en enregistrant la pièce jointe.

Le tableau de bord accessible sur Intranet permet de vérifier à tout moment la progression du document.

Un serveur dédié permet de conserver le document en cours d'élaboration.

L'ensemble des documents publiables (procédures, référentiels, cartographie des procédés) sont accessibles sur Intranet.

3.4. Solution technique employée pour la gestion documentaire

L'ICR a choisi, après une étude de marché, la solution Sharepoint Portal Serveur pour gérer Intranet et la GEIDE. Sharepoint n'offrant pas de page web d'accueil, une page d'accueil est créée avec le logiciel Dreamweaver offrant une interface web conviviale. Elle est alimentée quotidiennement par la médiathèque.

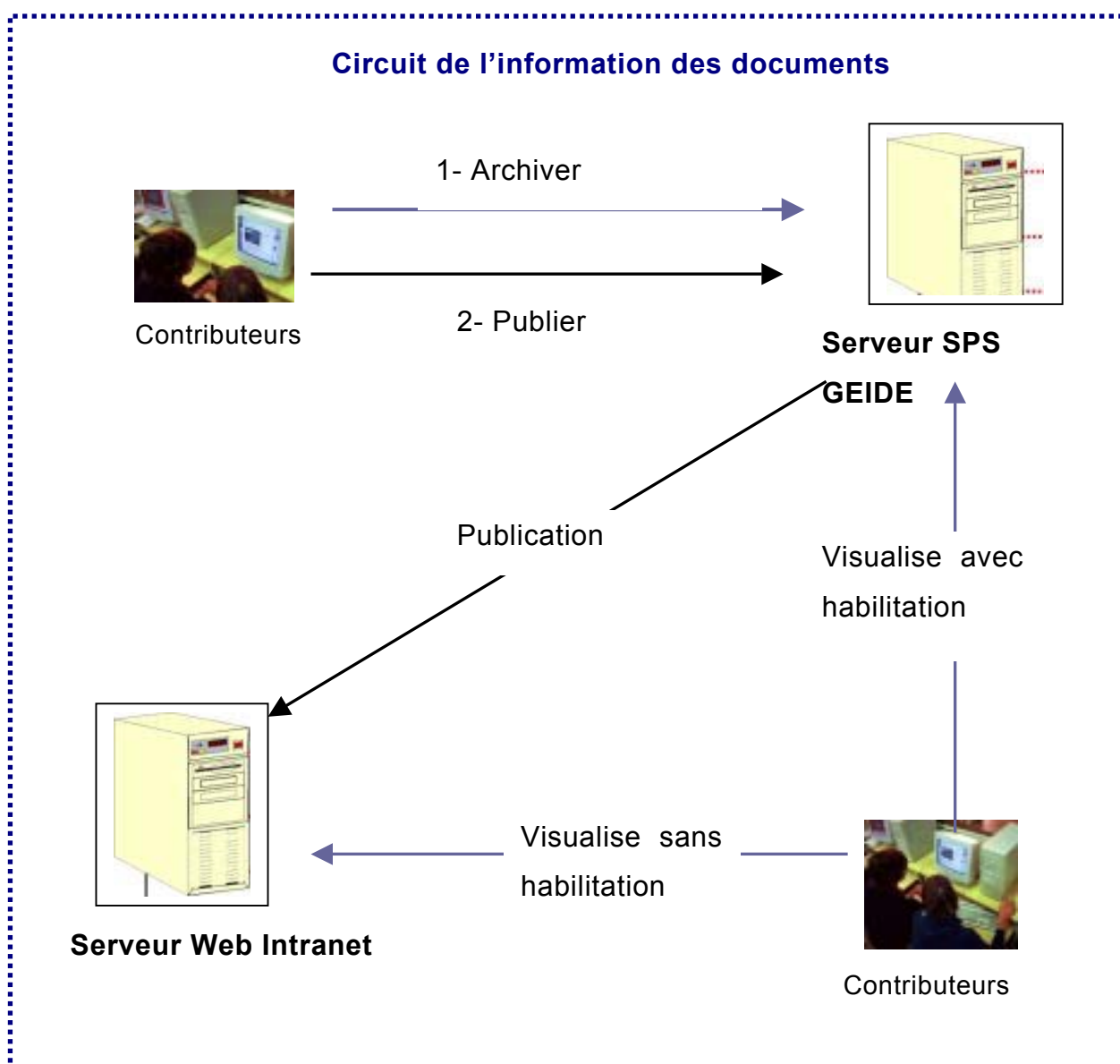
La couche informatique permettant de transformer les documents natifs en format HTML a été développée par une société extérieure.

Les données sont gérées ainsi selon les profils de documents pour l'Intranet et la GEIDE :

- Un serveur GEIDE conserve les données propres à la gestion de contenu des documents,
- Un serveur Intranet reçoit les données spécifiques à Intranet,
- Un espace disque partagé pour gérer les informations propres à chaque service (création des pages web, enregistrement de données documentaires pour la médiathèque) servant de base pour l'alimentation de la GEIDE et de l'Intranet.

Les données générées par la médiathèque sont envoyées à distance sur le serveur correspondant grâce à la technologie du Push¹⁰ avec le logiciel Dreamweaver : les informations sont mises en ligne instantanément.

Chaque application (Intranet, GEIDE) est installée sur un serveur dédié. Des liens sont prévus pour envoyer des documents entre les serveurs



¹⁰ La technologie du Push (littéralement « pousser ») permet d'obtenir une information ciblée selon une périodicité adaptée aux besoins de l'utilisateur (CHU Rouen, glossaire internet. [EN LIGNE], 200. Disponible sur :< <http://www.chu-rouen.fr/dsii/html/glossaire.html>> (consulté le 07/09/2005)

4. Analyse de la gestion des documents scientifiques dans le domaine médical

4.1. L'existant à l'Institut Claudius Regaud

4.1.1. Type de documents

Actuellement, les documents répertoriés sont les articles scientifiques ainsi que les communications diffusées en interne ou à l'occasion de congrès. Des bases de données de références bibliographiques internationales gratuites comme Pubmed¹¹ ou payantes comme Embase¹², permettent de répertorier les documents publiés par l'ensemble de la communauté scientifique.

Deux gestions des documents publiés par les praticiens et les chercheurs (provenant d'une Unité INSERM et chercheurs affiliés à l'ICR) sont opérées à l'Institut Claudius Regaud : une gestion pour les références bibliographiques et une gestion pour les documents en texte intégral.

4.1.2. Gestion des références bibliographiques

La médiathèque a en charge de répertorier la production scientifique de l'Institut : elle effectue une veille automatisée sur la base de données Pubmed à partir des auteurs des articles et du nom de l'Institut. Le service ne souscrit pas à d'autres bases de données médicales comme Embase pour des raisons financières.

¹¹ **PubMed** est une base de données bibliographiques centrée à l'origine sur la documentation en sciences biologiques et développée par le National Center for Biotechnology Information (NCBI) de la National Library of Medicine (NLM), (adresse du site : <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>)

¹² **Embase** est une base de données bibliographique produite par l'éditeur Elsevier, recensant la documentation mondiale en sciences médicales fondamentales, médecine, pharmacie, médecine dentaire et sciences infirmières. En plus de ses propres références bibliographiques, Embase couvre également l'ensemble du contenu de Medline ; sa particularité est de couvrir essentiellement des revues européennes.

Les références récupérées sont stockées dans des logiciels de références bibliographiques comme Reference Manager version réseau, logiciel préconisé par l'Institut pour l'ensemble des praticiens.

Les praticiens et les chercheurs stockent également un certain nombre de références avec ou sans logiciel de gestion bibliographique.

4.1.3. Documents en version texte intégral

La médiathèque a entrepris depuis 2003 le référencement de la version électronique des articles à partir des sites des éditeurs de revues. En tant que bibliothèque associée elle a un poste à l'ICR dédié aux universitaires de la bibliothèque de Santé de Toulouse. Par ailleurs les chercheurs accèdent au portail BiblioInserm de l'Inserm. Les chercheurs signalent par mail leurs productions de manière plus ou moins exhaustive. Les articles sont enregistrés sur le serveur de GEIDE pour une durée de trois ans. Ils sont présentés par ordre chronologique de publication sur Sharepoint. Le fichier est nommé selon les éléments suivants : auteur appartenant à l'Institut : titre du périodique, année de publication, volume, pagination. Un signalement des dernières parutions se fait en parallèle sur la page d'accueil de l'Intranet (rubrique Publications).

4.1.4. Récapitulation des données informatiques existantes

Type	Taille	Stockage	Diffusion
Références bibliographiques Export possible (TXT et XML)	162 références stockage années 2003-2004	Base de données de références bibliographiques Reference Manager	Références bibliographiques (3 par semaine)
Articles (PDF	149 références depuis 2003	Fichier Serveur GEIDE/ Intranet	Fichiers Intranet
Communications internes/ Congrès (Power Point)	25 documents depuis 2002	Dossiers Serveur GEIDE Intranet	Fichiers Intranet

4.2. Analyse de l'existant en externe : expériences dans le domaine de l'oncologie¹³ et en sciences

Une prospection a été effectuée dans le domaine de l'oncologie médicale pour connaître l'existence éventuelle de base de données de publications scientifiques. Le recours au réseau des Instituts en cancérologie français s'est avéré efficace : tout d'abord, avant de démarrer le stage, une visite a été faite au centre de documentation de l'IARC (Centre International de la Recherche sur le Cancer) pour prendre connaissance de la gestion électronique des publications scientifiques de l'Organisme. Ensuite, au démarrage du stage, une prospection sur les sites internet de la Fédération nationale des Centres de Lutte Contre le Cancer (FNLCC) a été effectuée. Un entretien téléphonique a permis de contacter les personnes pour obtenir des informations techniques sur les projets réalisés.

Les résultats ont montré que deux expériences similaires au projet ont été réalisées dans le domaine de l'Oncologie médicale :

- L'Institut Paoli Calmettes est le seul Centre de Lutte Contre le cancer qui a mis en ligne une base de données sur Internet (les références uniquement pour des raisons de droits d'auteur) et sur Intranet (les références et les articles).
- L'IARC (Organisme dépendant de l'Organisation Mondiale de la Santé) a aussi mis en ligne les données sur Intranet .

4.2.1. Institut Paoli Calmettes- Marseilles ¹⁴

Les données sont présentes sur le site Internet et sur le site Intranet

La médiathèque alimente la base de données Reference Manager ; les informations sont ensuite exportées sous Access (une seule table est créée sans index) par le service informatique. Pour ne pas mobiliser un serveur, la base est stockée sur un

¹³ Oncologie : le terme désigne l'étude des tumeurs. Le terme *oncologie* est de plus en plus souvent utilisé pour désigner l'étude des cancers malgré l'existence du terme *cancérologie*. http://w3.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp

¹⁴ <http://www.institutpaolicalmettes.fr/>

poste client. L'interrogation des données se fait en asp (scripts, requêtage, envoi de la page de résultats sur le serveur). Un lien ODBC fait le lien avec ACCESS.

Au préalable, le service informatique a dû bien réfléchir sur le mode d'alimentation de la base et sur l'uniformité des champs et du contenu.

Les responsables du projet sont satisfaits du fonctionnement et n'envisagent pas d'autres évolutions.

4.2.2. IARC ¹⁵

Cet organisme est un Centre International de la Recherche sur le Cancer. Les publications qui en émanent sont importantes : des monographies thématiques utiles pour l'évaluation des risques cancérigènes et pour la prise d'une décision scientifique, nécessitent de rassembler toute la littérature produite par les chercheurs de l'Organisme (sur 300 chercheurs, 50 publient régulièrement.)

Il s'est avéré nécessaire de centraliser les données pour effectuer la traçabilité des informations et pour pouvoir les consulter rapidement.

Chaque Département alimente sa propre banque de données avec le logiciel Reference Manager. Une personne centralise les références bibliographiques avec le logiciel Reference Manager (depuis 2003-2005) : les auteurs signalent la parution de leurs articles par mail. Les références sont ensuite hébergées sur le site Intranet sous forme de page HTML statiques, comportant un lien sur la base de données Pubmed (par l'identifiant « Pubmed ID »). L'accès à l'article dans sa version intégrale est fourni : la médiathèque a en charge de récupérer les articles depuis le site des éditeurs et par scannérisation des documents.

La difficulté majeure est d'obtenir les informations auprès des chercheurs : la Direction a mis en place une réunion mensuelle qui permet de signaler les publications In-Press (sachant que ce qui compte est le document dans sa phase

¹⁵ <http://www.iarc.fr/ENG/General/index.php>

finale (en Post-Publication). Ceci s'avère être efficace pour avoir une liste complète.

Le projet de la médiathèque sera d'étudier la possibilité de verser les données sur le logiciel documentaire Sydneyplace déjà utilisé pour la gestion des données documentaires.

4.2.3. Autre exemple de réalisation en sciences

L'usage de la liste de diffusion du logiciel de gestion de références bibliographiques Reference Manager a été utilisée pour avoir des expériences de création de bases de données en science. Grâce à cette liste de diffusion, il a été possible de prendre connaissance d'un projet de publications scientifiques mené à la London's Global University (UCL), Université de médecine-santé : une base de données ¹⁶ a été créée rassemblant toutes les publications de tous les Départements (une cinquantaine) depuis 1997. La base a été développée en PHP4/Mysql/ avec un serveur Apache et l'application a été créée avec Myopia (solution Open Source avec contrat de licence Institution) : la solution est téléchargeable sur SourceForge ¹⁷.

Plusieurs niveaux d'accès ont été prévus : l'accès grand public, l'accès administrateur (utile pour la gestion des publications et des authentifications), l'accès des auteurs (utile pour gérer la mise en ligne de leurs publications).

Les données peuvent être insérées manuellement ou automatiquement avec les logiciels de gestion de références bibliographiques ou avec les bases de données comme Pubmed.

¹⁶ <http://www.ucl.ac.uk/research/publications/new/>

¹⁷ <http://myopia.sourceforge.net>

5. Etude des pratiques et des besoins

5.1. Enquête par entretiens

La méthode par entretiens est choisie pour recueillir des informations d'ordre qualitatif sous forme de discussion.

Les entretiens se sont déroulés auprès d'un échantillonnage représentatif d'acteurs reflétant chaque secteur d'activité : quatre personnes de la délégation recherche puis une à deux personnes du service médecine, chirurgie, radiothérapie ont été interrogées. Les praticiens et les chercheurs sont interrogés au titre de producteurs et de consommateurs d'informations.

Un questionnaire a suivi de fil conducteur pour diriger l'entretien.

A l'issue du questionnaire, une synthèse des résultats de l'enquête en dégagera les idées principales. Le questionnaire servira également comme point de repère pour effectuer une évaluation du projet.

5.2. Objectifs du questionnaire

Le questionnaire ¹⁸ vise deux objectifs : le premier est de prendre connaissance des pratiques des praticiens et des chercheurs en matière de gestion des publications scientifiques. Il doit répondre à la problématique du degré d'intégration de l'activité scientifique dans l'activité de recherche

Le deuxième vise le recueil des besoins éventuels pour améliorer le fonctionnement actuel et pour vérifier la pertinence du projet.

5.3. Synthèse des entretiens

5.3.1. Analyse des usages

Les personnes interrogées ont recours à l'utilisation de leurs propres publications sous forme de références bibliographiques pour :

¹⁸ Le contenu du questionnaire figure dans l'annexe 4 page 85

- à titre individuel effectuer leur épreuve de titres
- à titre collectif rédiger le rapport d'activité du département une fois tous les deux ans et pour le Standard Option Recommandation (SOR) ¹⁹
- faire les demandes de subventions pour la Recherche auprès des Organismes habilités comme l'Institut National du Cancer, l' Association pour la Recherche sur le Cancer.

Les références bibliographiques stockées servent pour :

- préparer une future publication,
- faire de la veille scientifique sur un thème de recherche ou sur un cas clinique à traiter,
- rédiger un CV.

5.3.2. Analyse du corpus de documents utilisés

5.3.2.1. *Articles publiés et références bibliographiques*

L'ensemble des personnes interrogées conserve les articles publiés (en texte intégral) sous forme de tiré-à-part en papier et les trois-quarts les conservent en format électronique (PDF), certains depuis quinze ans.

Cinq personnes sur neuf conservent des références bibliographiques et utilisent un logiciel de références bibliographiques pour conserver les références des articles publiés. Il s'agit du logiciel Reference Manager, logiciel préconisé par la FNLCC. Quatre personnes sur sept ont une gestion structurée des informations sous forme de banques de données thématiques. De manière générale, la gestion des références se fait au coup par coup. Une banque de données contient en moyenne 120 références.

Dans l'ensemble, l'utilisation des références se fait de manière individuelle, il n'y a pas d'échanges entre les praticiens. Le département de radiothérapie vient

¹⁹ Le SOR correspond à la recommandation pour la pratique clinique

d'émettre le besoin de mettre en ligne une banque de données pour l'ensemble de la communauté des praticiens.

La Délégation Recherche éprouve le besoin de mettre en commun les références bibliographiques pour mesurer son activité scientifique : les publications sont rassemblées par une personne, mais d'après les personnes interrogées, ce système ne reste pas optimal (la base de données n'est pas suffisamment actualisée).

5.3.2.2. *Nature des publications*

La délégation Recherche publie des documents essentiellement en recherche fondamentale du fait de son appartenance à une unité INSERM.

Pour les départements ayant une activité clinique, les documents sont de nature clinique, et dans le domaine de la recherche en transfert et sur des essais précoces.

5.3.2.3. *Type de publications référencées*

Toutes les personnes interrogées référencent principalement les articles scientifiques issus de la base de données Pubmed.

Les communications à des congrès, les thèses, et autres documents appartenant à la littérature grise, sont très peu répertoriées.

La difficulté est que ces documents ne font pas l'objet d'un signalement systématique et que les auteurs ne conservent pas toujours dans leurs dossiers le document en version électronique. Ce constat est paradoxal avec le fait que le responsable du département de radiothérapie signale que l'activité de publication sur un thème est souvent collective : une question donnée mobilise un ensemble de type de publications (cas cliniques, essais précoces, recherche en transfert.)

5.3.2.4. *Fréquence moyenne de publication*

La fréquence de publication indique le volume de données qu'il sera nécessaire d'intégrer dans le nouveau système. D'après les réponses, la fréquence de publication à titre d'auteur principal est équivalente entre l'activité de la délégation recherche et les activités de clinique (entre un à trois articles par an par individu).

Les personnes interrogées ayant une activité clinique ont souligné le fait que leur fréquence de publication est plus limitée que le secteur de la recherche.

5.3.2.5. *Lieu de publication*

Pour l'ensemble des départements et la délégation recherche, l'essentiel des publications porte dans les revues à fort facteur d'impact, comme les revues américaines. Les exigences des éditeurs sont très précises : il s'agit de faire part d'un sujet innovant basé sur beaucoup de résultats expérimentaux.

5.3.3. Les pratiques en recherche documentaire

L'ensemble des personnes interrogées a une pratique régulière de la recherche documentaire (pour la médecine, une fois par mois.) La moitié fait des recherches ciblées dans sa spécialité, les recherches sont peu faites sur des thèmes transversaux (sauf pour le département de radiothérapie).

La principale source utilisée est la base de données Pubmed (elle est gratuite) bien que les praticiens et les chercheurs soient conscients des limites de cette source : elle ne répertorie pas toutes les publications de l'ICR et l'interrogation de la base demande une certaine pratique malgré sa bonne ergonomie (notamment pour le choix des mots-clés).

Une personne interrogée signale le recours aux sites de revues électroniques régulière (sept titres de revues).

La Délégation Recherche utilise toutes les sources disponibles : réseau INSERM, réseau des bibliothèques Universitaires toulousaines.

La consultation du site Intranet de l'ICR est faible. Quatre personnes consultent fréquemment la page d'accueil Intranet pour consulter les dernières publications des confrères mais sur les quatre, la moitié passe la page d'accueil pour accéder aux documents dans leur version intégrale.

La recherche documentaire sur Internet n'est pas optimisée : l'usage de profils de veille pour effectuer un profil d'alerte sur un sujet donné est rare (deux personnes sur les sept ont créé un profil de veille documentaire profitant à l'ensemble d'une collectivité). Peu de personnes ont créé un système d'alerte basé sur la table des matières (Table of Content, TOC) pour faire de la veille sur un titre de revue, faute de temps.

L'alternative aux modèles économiques traditionnels n'est pas connue par les praticiens et les chercheurs : les personnes interrogées ignorent l'existence du protocole des Archives Ouvertes (Open Archives Initiatives).

5.3.4. Recueil des besoins et des suggestions

Les personnes interrogées sont satisfaites des services proposés par la médiathèque pour la mise en valeur des publications scientifiques (signalement des articles parus, Workflow de demandes d'articles).

Il est à noter que par rapport à l'idée de mettre en ligne des publications de l'ICR la majorité des personnes interrogées n'a pas vu d'intérêt majeur. Il a donc été nécessaire de présenter les avantages du projet lors de l'entretien. A l'issue de l'entretien les personnes ont paru convaincus de l'intérêt : les informations accessibles rapidement à distance favoriseront le travail collaboratif plus systématique en Internet et en externe et pourront s'ajouter au système de GEIDE existant.

5.3.4.1. Accès à la base de données

A l'unanimité, les personnes interrogées souhaitent que la base de données soit accessible à la fois en interne et en externe. L'intérêt majeur est de pouvoir accéder au texte intégral sur l'Intranet.

5.3.4.2. Consultation des données

Il a été suggéré de créer une rubrique située sur la page d'accueil du site Intranet qui permettrait d'accéder directement à un formulaire d'interrogation.

Il a été suggéré de créer une rubrique sur la page d'accueil du site internet, par exemple dans la partie accès professionnel.

Le formulaire d'interrogation s'inspirerait du site de l'Institut Paoli Calmettes : il contiendrait une possibilité d'interrogation sur les Auteurs et par date de publication.

Plusieurs accès thématiques ont été suggérés :

- Un accès thématique en fonction de l'activité de l'Institut : médecine, radiothérapie, recherche en transfert, chirurgie, autre,.
- Un accès par genre : soins, recherche pour le site Internet,

Une remarque a été faite pour bien distinguer les accès en fonction des thèmes transversaux et des thèmes plus spécifiques à une activité.

5.3.4.3. Type de documents référencés

Quatre personnes ont émis le souhait d'étendre le référencement des articles pour alimenter la base à plusieurs bases de données.

Il a été suggéré d'effectuer le signalement sur plusieurs types de documents : thèses, congrès...

Le département de Radiologie fait part d'une volonté de publier dans les revues françaises : le responsable du département explique que l'objectif est de faire une mise au point scientifique et de faire savoir que l'ICR travaille sur un sujet donné.

5.3.4.4. Mode d'alimentation de la base

La majorité des personnes interrogées ont souligné le fait que la méthode d'alimentation de la base doit s'effectuer le plus simplement possible (ils ne peuvent pas accorder beaucoup de temps à cela) : le signalement par mail au service médiathèque d'une publication à venir est la plus efficace.

Le département de Radiologie constate que ce système n'est pas fiable et a suggéré de déléguer une personne du service médiathèque pour effectuer l'intégralité du suivi des demandes de publications auprès des éditeurs. Cela centraliserait le circuit informationnel et permettrait la constitution d'une base de données fiable.

Une personne pense que le projet ne doit pas être redondant avec Pubmed puisque toutes les publications ne sont pas référencées dans la base Pubmed. La réflexion doit donc s'attacher à offrir un autre service qui serait un complément indispensable de cette source d'information.

5.4. Bilan du fonctionnement actuel

A l'issue de l'enquête et de l'analyse de l'existant (flux d'information et utilisation des données), un certain nombre de points ont pu être soulevés. L'analyse a été présentée auprès de la responsable de la médiathèque afin de définir des objectifs précis pour la réalisation de la base de données.

5.4.1. Pratiques des médecins et des chercheurs

Les pratiques des médecins et des chercheurs en matière de gestion des publications scientifiques sont inégales : le logiciel Reference Manager préconisé n'est pas utilisé de façon systématique, surtout par les praticiens et la pratique de l'outil informatique est encore limitée.

L'analyse du fonctionnement des praticiens et des chercheurs révèle un paradoxe : le travail de recherche documentaire et de gestion des publications scientifiques a tendance à être individualiste alors que les publications sur un sujet de recherche sont collectives et donnent lieu à une série d'écrits. Par exemple les articles rédigés pour les Standards Options Recommandations (SOR) donnent lieu à une publication collective.

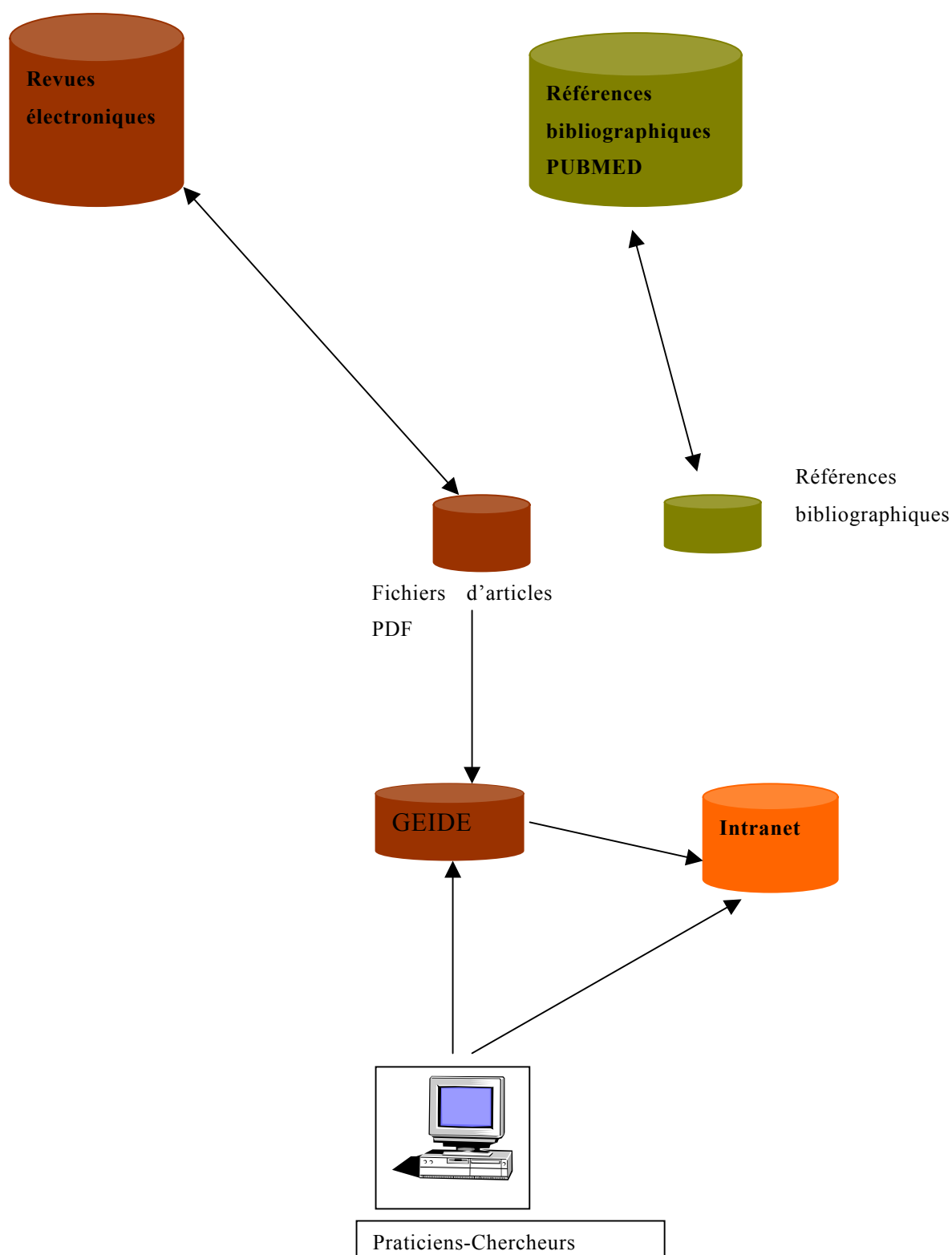
La méthode de recherche documentaire se base principalement sur une seule source de données, Pubmed, alors que les publications ne sont pas toutes répertoriées dans cette base.

Les solutions proposées pour optimiser et systématiser la recherche documentaire sont faiblement utilisées, faute de temps pour créer le profil et pour faire une formation (système d'alerte comme Biomail ou My NCBI directement intégré à Pubmed).

5.4.2. Analyse du flux d'informations entre les praticiens, chercheurs et documentalistes

La circulation de l'information entre l'obtention de l'information et la communication des données représentée dans le schéma ci-dessus montre que l'accès à l'information n'est pas centralisée pour l'utilisateur (source d'accès : GEIDE, Intranet).

Flux d'information gestion des publications scientifiques



5.4.3. Modélisation de processus métiers pour la gestion de l'information

5.4.3.1. Identification des processus

Deux tableaux présentent les acteurs et l'identification des processus pour la gestion de l'information : recherche et accès à l'information publiée par les praticiens et chercheurs de l'ICR.

Identification des processus pour les praticiens et les chercheurs

Processus	Où	Comment
Faire des recherches bibliographiques sur un sujet de recherche	Base de données Pubmed	Requête sur les champs auteurs et mots-clés
Rédiger une publication	Envoi de l'article à l'éditeur	Logiciel traitement de texte Rédaction une à deux fois /an à titre d'auteur principal
Informar de la parution de de l'article auprès de la médiathèque	Service médiathèque Secrétaire du Département Recherche Unité Inserm	Communication formelle (courrier électronique) ou informelle
Récupérer le texte intégral des documents publiés	Demande de tiré –à-part par mail	Stockage dans word ou dans un logiciel de références bibliographiques (Endnote ou Ref Manager)

Identification des processus pour les documentalistes

Processus	Où	Comment
Collecter et stocker les références bibliographiques	Pubmed Google Scholar	Profil de veille : biomail Cubby Logiciel Reference Manager
Récupérer les documents : articles Congrès	Site web éditeurs	Abonnement aux revues électroniques Demande de tiré- à-part
Stocker Diffuser l'ensemble des données	Serveur GEIDE Intranet	

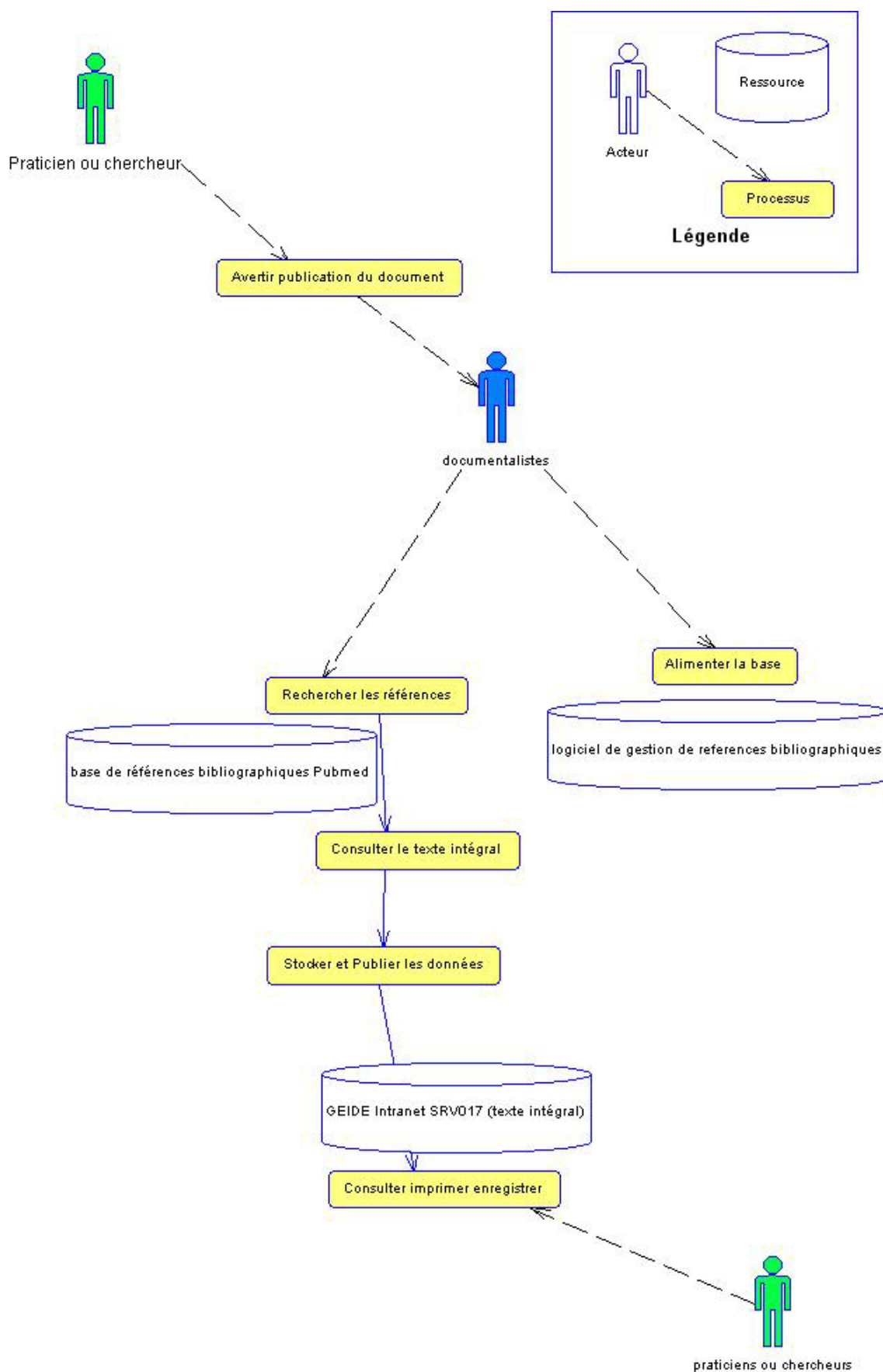
5.4.3.2. Modélisation processus des métiers

En s'inspirant de la Démarche Qualité utilisée par l'ensemble du personnel de l'Institut, la représentation des processus des métiers a été jugée utile pour modéliser les opérations réalisées par les acteurs identifiés (praticiens et documentalistes.)

La méthode choisie est la modélisation des données avec le logiciel Power AMC²⁰ utilisé par le Département d'Information Hospitalier : le propre des premiers est de produire l'information scientifique et pas seulement d'y avoir recours. La médiathèque joue le rôle d'intermédiaire pour faciliter le recueil et la circulation des données.

Le schéma ci-dessous présente les différents processus opérés par les acteurs pour accéder à un document.

²⁰ Power AMC, logiciel édité par Sybase <http://www.sybase.com/home>



5.4.4. Bilan du fonctionnement actuel

Un certain nombre d'atouts en matière de gestion des données peuvent être relevés. La médiathèque répertorie les données : un recensement exhaustif effectué pour les deux années antérieures est effectué (références bibliographiques, articles, communication et posters à un congrès).

Des moyens pour transmettre l'information sont utilisés au moyen de l'Intranet.

A partir des schémas et des résultats de l'enquête, il est constaté que la principale difficulté est de pouvoir recenser les publications : le signalement des documents en cours de publication n'est pas effectué de façon systématique. De ce fait, la médiathèque a un rôle moteur pour répertorier les informations et les diffuser selon la méthode Push (l'information est diffusée collectivement sur l'Intranet). Plusieurs méthodes sont utilisées pour obtenir l'information : l'examen des sources externes (bases de données Pubmed et sites de revues électroniques liées à Pubmed) et sollicitation directe auprès des auteurs (par courrier électronique ou de manière informelle (par voie orale).

En matière de collecte de l'information, le fait d'avoir une seule source n'est pas suffisant pour atteindre l'exhaustivité.

Conjointement au service de la médiathèque, la Délégation Recherche a une activité de recensement des publications pour la réalisation du rapport d'activité. Depuis 2003, les chercheurs sont autonomes dans leur recherche, de ce fait il n'y a plus de personne responsable du recensement des publications.

Toutefois, la personne anciennement chargée de cette activité, récupère les avis de publication signalés par la médiathèque et les insère dans un fichier word à partir des données signalées par la médiathèque pour garder une trace de l'activité dans son Unité de recherche. A l'issue des entretiens, il s'est avéré que les chercheurs sont conscients de cet état de fait et mesurent l'importance de faire une gestion exhaustive des références bibliographiques et des articles en texte intégral.

Un certain nombre de problèmes peuvent être soulevés, notamment le fait que la gestion des informations recueillies est éclatée pour les activités de collecte, de stockage et de mise à disposition des données. Par exemple, les communications aux congrès des praticiens de l'ICR ainsi que les posters sont l'objet d'une gestion à part dans le dossier « photothèque » de la rubrique médiathèque et les références bibliographiques ne sont pas publiées dans leur intégralité auprès de l'ensemble de la communauté scientifique .

Le stockage des informations n'est pas prévu pour qu'il soit pérenne. Cela a des conséquences sur l'accès aux données. Pour des raisons de stockage entre les références et les articles, les données sont stockées dans deux lieux différents (serveur médiathèque pour les articles, disque dur local pour les références contenues dans Reference Manager) ce qui ne donne pas la possibilité d'accéder au texte intégral .

En terme de Système d'Information Documentaire, la médiathèque, n'ayant pas de base de données pour gérer les documents, effectue une étude pour installer un logiciel documentaire et le choix n'est pas encore opéré à ce jour.

En conclusion, il est constaté qu'il n'y a pas de réelle pratiques de gestion des publications : les données font l'objet d'un stockage individuel pour les praticiens et les chercheurs. Un certain nombre font l'objet d'une diffusion sur Intranet mais il n'existe pas de moteur de recherche pour interroger à distance l'ensemble des données publiées lors de l'année en cours et lors des années antérieures.

5.5. Analyse des besoins exprimés par les utilisateurs (Médecins, chercheurs, Département du Système d'Information Hospitalier)

5.5.1. Médecins-chercheurs

A titre collectif, les enjeux répertoriés sont d'effectuer un historique de l'ensemble des publications pour l'édition de rapports d'activités de l'ICR, pour la rédaction des recommandations pour la pratique clinique comme le Standard Option Recommandation (SOR.). A moyen terme, il serait envisagé de faire part à la communauté des chercheurs non ICR et aux patients de l'activité scientifique innovante.

A titre individuel, il s'agit de pouvoir effectuer un recensement exhaustif et systématique de ses propres publications pour préparer d'autres publications et rédiger une épreuve sur titres.

Une autre suggestion a été formulée : pouvoir mutualiser les banques de données déjà créées. Le Département de Radiothérapie souhaiterait mettre en commun sur le site Intranet les banques de données existantes créées par les praticiens : les références répertoriées par la communauté médicale de l'ICR sur un thème précis ont fait l'objet d'un tri à partir de la base Pubmed. Les informations seraient mises à disposition des différents départements.

L'objectif attendu du nouveau système est de pouvoir accéder rapidement à la totalité des publications et à leur version intégrale à savoir :

- consulter l'intégralité des parutions du mois dans toutes les disciplines,
- consulter l'ensemble des données dans tous les domaines.

Le suivi des dernières publications scientifiques est nécessaire pour la rédaction des documents-recommandations pour la pratique clinique exigé par les Standards Options Recommandation.

Les moyens suggérés sont de centraliser les informations et les rendre accessibles sur Internet et sur Intranet.

En terme de gestion, il serait souhaitable qu'une personne dont la mission serait d'effectuer le secrétariat scientifique notamment de réaliser le suivi intégral des publications de l'Institut (par exemple de la pré-publication à la post-publication).

Les documents jugés utiles seraient les articles, les communications aux congrès, les brevets et en dernier lieu les thèse.

5.5.2. Médiathèque

Les besoins sont :

- Pouvoir recenser les documents de manière exhaustive et pérenne dans un même lieu de stockage,
- Pouvoir offrir un accès distant à ces données (de la référence bibliographique au texte intégral),
- Répertorier les informations pour l'année en cours dans une même base et réfléchir pour la reprise de l'antériorité,
- Signaler le Facteur d'Impact dans la base de données lié à la référence bibliographique pour réaliser les rapports d'activités scientifiques en prenant en compte la mise à jour annuel du Facteur d'Impact,
- Augmenter les actions de formation pour rendre autonomes les praticiens et les chercheurs et pour favoriser une meilleure collaboration avec la médiathèque en vue de l'alimentation de la base de données par les auteurs d'articles. Par exemple en effectuant des formations au logiciel Reference Manager, à l'utilisation de Pubmed et aux fonctions de veille automatisée (Biomail, Cubby),
- Diversifier l'accès aux bases de données de Références Bibliographiques pour la communauté des praticiens et des chercheurs :

offrir par exemple l'accès à la base Embase ou à la Cochrane Libray pour la médecine fondée sur les preuves (EBM).²¹.

A propos du choix des bases de données en médecine, une étude menée par Stéphane Ratté ²²montre que dans le domaine médical la principale base utilisée est Pubmed (celle-ci est très complète en terme de contenu : le volume de périodique indexé dans le biomédical est très important, le délai d'indexation est faible et le nombre d'articles potentiellement disponibles en plein texte et conséquent). L'auteur de l'étude constate que le volume important d'information disponible est une force mais aussi une faiblesse (le temps requis pour chercher et analyser les résultats des études demeure un obstacle à l'utilisation de la littérature scientifique par les médecins).

De ce fait, il est suggéré d'utiliser en deuxième position les sources axées sur la médecine fondée sur les preuves plutôt que la base Embase, ceci dans un but qualitatif plutôt que quantitatif.

La médiathèque a émis le besoin que le nouveau système puisse intégrer, à terme d'autres documents d'intérêt général pour l'entreprise comme la revue de presse concernant l'Institut Claudius Regaud.

5.5.3. Département informatique

Les besoins sont :

- Faire en sorte que la base de données puisse s'interfacer avec les applications existantes : serveurs, environnement, logiciel documentaire,
- Prendre une solution pérenne et fiable pour le stockage et l'accès aux articles en texte intégral,
- Respecter l'intégrité des données,

²¹ Consulter par exemple l'article de G. Chartron à ce sujet : *Eléments pour une approche comparée de la publication scientifique*, Archives nationales du Québec, 2003

²² RATTE,S, Une analyse descriptive et comparative d'une sélection de bases de données bibliographiques du domaine biomédical. [EN LIGNE], 2004. Disponible en ligne <http://bibliothecaire.com/stephane_ratte.html>, visité le 05/05/2005

- Prévoir la structuration de la base envisageant le maximum de champs pour alimenter plusieurs types de documents.

Tableau récapitulatif des besoins recensés à l'ICR

<i>Public visé</i>	<i>Accès</i>	<i>Rubrique</i>	<i>Contenu</i>
<i>L'ensemble de l'Organisme</i>	Internet/Intranet	Rubrique médiathèque	Références et articles
<i>Patients</i>	Internet	Lien direct vers la page de recherche	Références
<i>Etudiants en recherche de stage</i>	Internet	Rubrique médiathèque	Références
<i>Communauté scientifique et médicale</i>	Internet	Rubrique médiathèque	Références

PROPOSITION D'UN NOUVEAU DISPOSITIF

1. Enjeux du nouveau système

1.1. Enjeux du nouveau dispositif

D'après l'analyse des besoins des utilisateurs, un certain nombre d'enjeux se dégagent : il est pertinent de fédérer l'ensemble des documents publiés en les rendant accessibles à tout moment sur Internet et Intranet au moyen d'une base de données. Un dispositif de gestion est envisagé pour optimiser le flux des informations en entrées (variétés des sources) et en sortie (édition de documents).

1.2. Public visé

En interne : il s'agit de la communauté scientifique (praticiens, chercheurs de l'Unité INSERM travaillant pour l'Institut ou chercheurs directement affiliés à l'Institut). En externe, il s'agit des praticiens, des chercheurs, des étudiants en médecine et des patients.

2. Elaboration d'un cahier des charges fonctionnel et technique

Un dispositif de gestion des publications scientifiques sera mis en place pour les fonctions de :

- Stockage et conservation des publications pour la mémoire de l'Institut,
- Diffusion des documents au moyen de l'Internet et de l'Intranet,
- Définition de règles pour faire évoluer le nouveau système d'information comme la mise à jour des données.

2.1. Contraintes du système

2.1.1. Contraintes techniques

Le système devra pouvoir prendre en compte un certain nombre de points :

- Pouvoir prendre en compte l'interfaçage et la compatibilité avec le système d'information actuel (projet d'un nouveau SIGB pour la médiathèque et Système d'Information Hospitalier, GEIDE),
- Pouvoir s'interfacer avec le logiciel Reference Manager
- Pouvoir effectuer la normalisation de la structuration des champs et l'harmonisation de la saisie des données.

2.1.2. Contrainte fonctionnelle

Le système devra favoriser une prise en main rapide du nouveau système pour permettre la mise à jour de la base et pour favoriser l'autonomie de la personne chargée du suivi de la base de données.

2.1.3. Contrainte organisationnelle

Un groupe restreint de personnes sera chargé de l'administration de la base pour veiller au choix des mots-clés et pour respecter une uniformité dans la saisie des données.

2.1.4. Contrainte de performance

Le système devra faciliter le rapatriement des données depuis différentes sources externes (bases de données).

Il devra intégrer les métadonnées de divers documents (métadonnées sur le titre, l'auteur, la source, la date de publication) et le document primaire téléchargeable sous différents formats.

2.2. **Architecture fonctionnelle du nouveau système d'information**

Une base de données commune et accessible à distance sera l'outil pour fédérer l'ensemble des données. Cette base de données s'intégrera dans l'architecture du système d'information existant : un accès aux données sera possible en interne (Intranet) et en externe (Internet) en deuxième temps.

2.2.1. Principales fonctionnalités attendues

2.2.1.1. *Mode de stockage*

Un système de gestion de base de données permettra le stockage des données émanant de tous les services producteurs d'information scientifique.

2.2.1.2. *Modes d'alimentation et de mise à jour de la base*

L'alimentation de la base de données se fera de manière centralisée par une seule personne afin d'éviter les doublons. Le nouveau fonctionnement permettra une mise à jour immédiate de la base de données sur les serveurs web.

Les producteurs de documents pourront signaler la parution de leurs écrits par courrier électronique au service producteur de la base de données.

2.2.1.3. *Consultation rapide des données pour le lecteur*

Le système offrira la fonctionnalité de consultation des données. Les caractéristiques de cette fonctionnalité sont :

- Offrir une simplicité du système (interrogation, rapatriement des données),
- Permettre l'interrogation à distance de la base (Internet/Intranet),
- Offrir la consultation des derniers articles parus depuis un mois, par exemple dans une rubrique «Actualité scientifique» sur Intranet ,
- Pouvoir choisir le nombre de résultats à afficher par page.

2.2.1.4. *Exploitation de données*

Les fonctionnalités suivantes devront être remplies pour l'exploitation des données :

- Pouvoir exporter les résultats des références bibliographiques en format RIS, ou fichier texte ou XML pour éditer un document spécifique (rapport d'activité par département),
- Pouvoir consulter et imprimer l'article en texte intégral au format PDF (fonctionnalité disponible sur l'Intranet pour tenir compte des droits d'auteurs),
- Pouvoir distinguer les publications au titre de l'ICR (recours à la mention d'affiliation) de celles publiées à titre individuel, pour le spécifier dans le rapport d'activités.

2.2.1.5. *Distinguer des niveaux d'accès*

Le dispositif devra pouvoir différencier les requêtes formulées en interne de celles formulées depuis l'extérieur (pour des raisons de droits d'auteurs, les documents publiés dans les revues scientifiques ne seront pas diffusés sur internet.)

2.2.1.6. *Description fonctionnelle du formulaire d'interrogation à distance des données*

CHAMPS
Auteurs (auteur principal)
Auteurs secondaires (le second de la liste des auteurs)
Année de publication
Thématique : médecine, radiothérapie, recherche en transfert, autre
Mots-clés définis par les auteurs (basés sur la classification du MESH de Pubmed)

2.2.1.7. Description fonctionnelle de l’affichage des données

CHAMPS
Auteur
Titre
Source de la publication
Mots-clés
Résumés
Facteur d’impact
Adresse de l’auteur principal
URL de l’article en HTML
URL de l’article en PDF

2.3. Caractéristiques techniques du nouveau système

2.3.1. Mise à jour des fichiers / signalement des nouvelles données

Le système prévoira la mise à jour automatique pour la mise en ligne Intranet ainsi qu'une réplication de base hebdomadaire sur l' Internet ou dans le cas contraire, une insertion de pages statiques pour les données concernant l'année en cours. Le système devra pouvoir signaler automatiquement sur l'Intranet les dernières données intégrées dans le système.

2.3.2. Archivage des données

Le système prévoira de disposer d'un espace sur les serveurs Intranet et Internet, pour stocker les références sous forme de tables, accompagnées des fichier PDF (texte intégral). Il offrira la possibilité de faire le lien entre les références bibliographiques et le texte intégral

2.3.3. Sauvegarde

Un système de sauvegarde sera mise en place par le service informatique pour effectuer une copie des données selon une périodicité définie au préalable.

2.3.4. Alimentation de la base de données

Une personne centralisera les données en utilisant le logiciel Reference Manager version 11 préconisé par l'Institut.

3. Analyse conceptuelle des données à informatiser

L'analyse conceptuelle des données complète la description du nouveau système opérée dans le Cahier des Charges.

3.1. Nouveaux flux d'information

Le nouveau flux de l'information met au cœur du nouveau dispositif la base de données.

3.1.1. Principaux flux en entrée

Il s'agit des documents produits par les médecins, chercheurs : articles de revues scientifiques, communication aux congrès, conférences en qualité de membres invités et brevets (document Word diffusés uniquement à l'éditeur).

3.1.2. Principaux flux en sortie

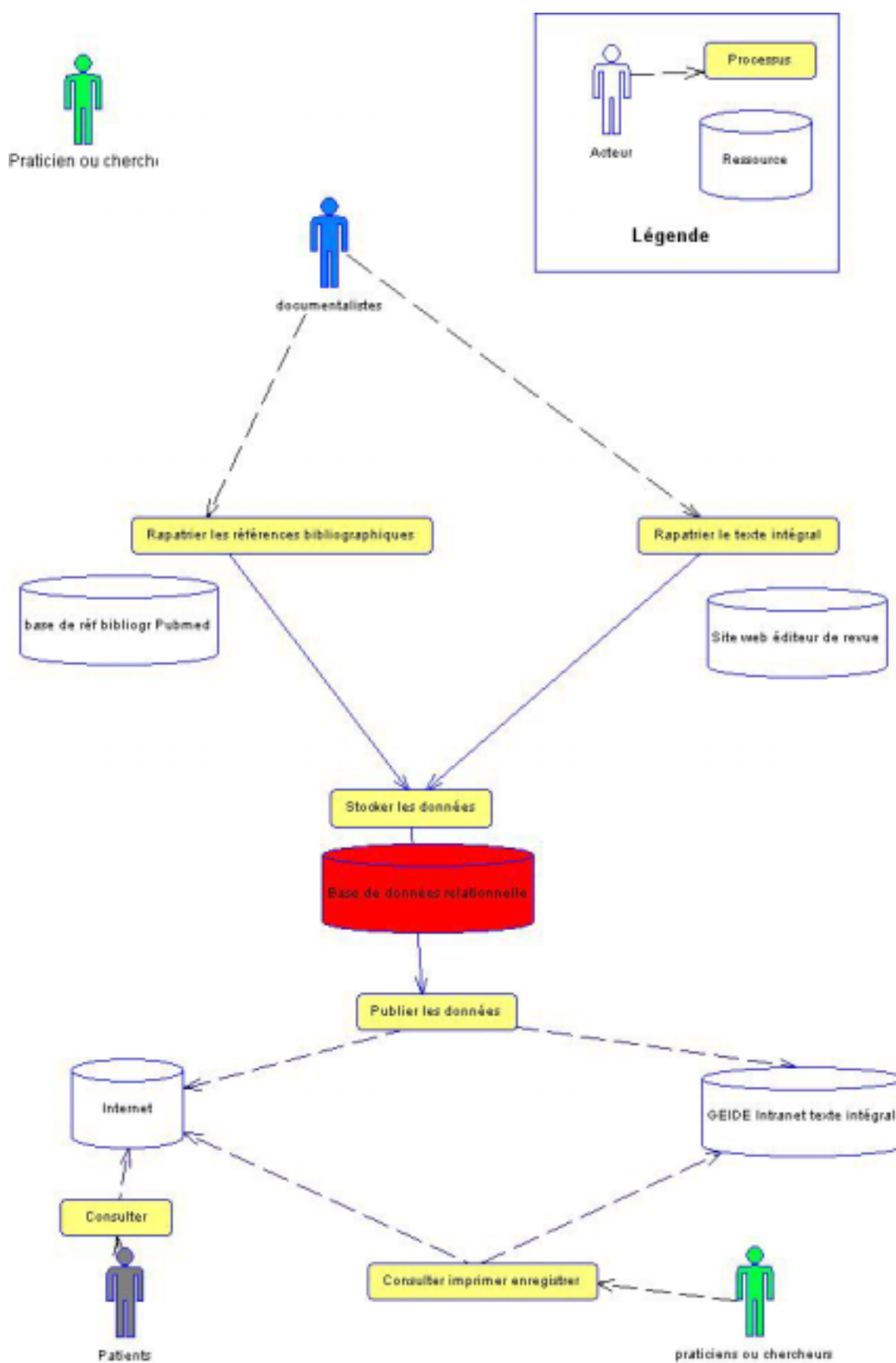
Les informations attendues en sortie sont la liste de références bibliographiques présentées sous des formats paramétrables, et la possibilité d'obtenir le document en format PDF (la version native n'étant pas conservée actuellement par l'auteur). Les résultats obtenus après interrogation de la base peuvent être exportés en formats TXT, XML pour exploitation de données.

3.1.3. Schéma général des flux : modélisation acteurs-processus

Le schéma présente le nouveau circuit de l'information en fonction des processus opérés par les acteurs.

Les acteurs de la chaîne sont :

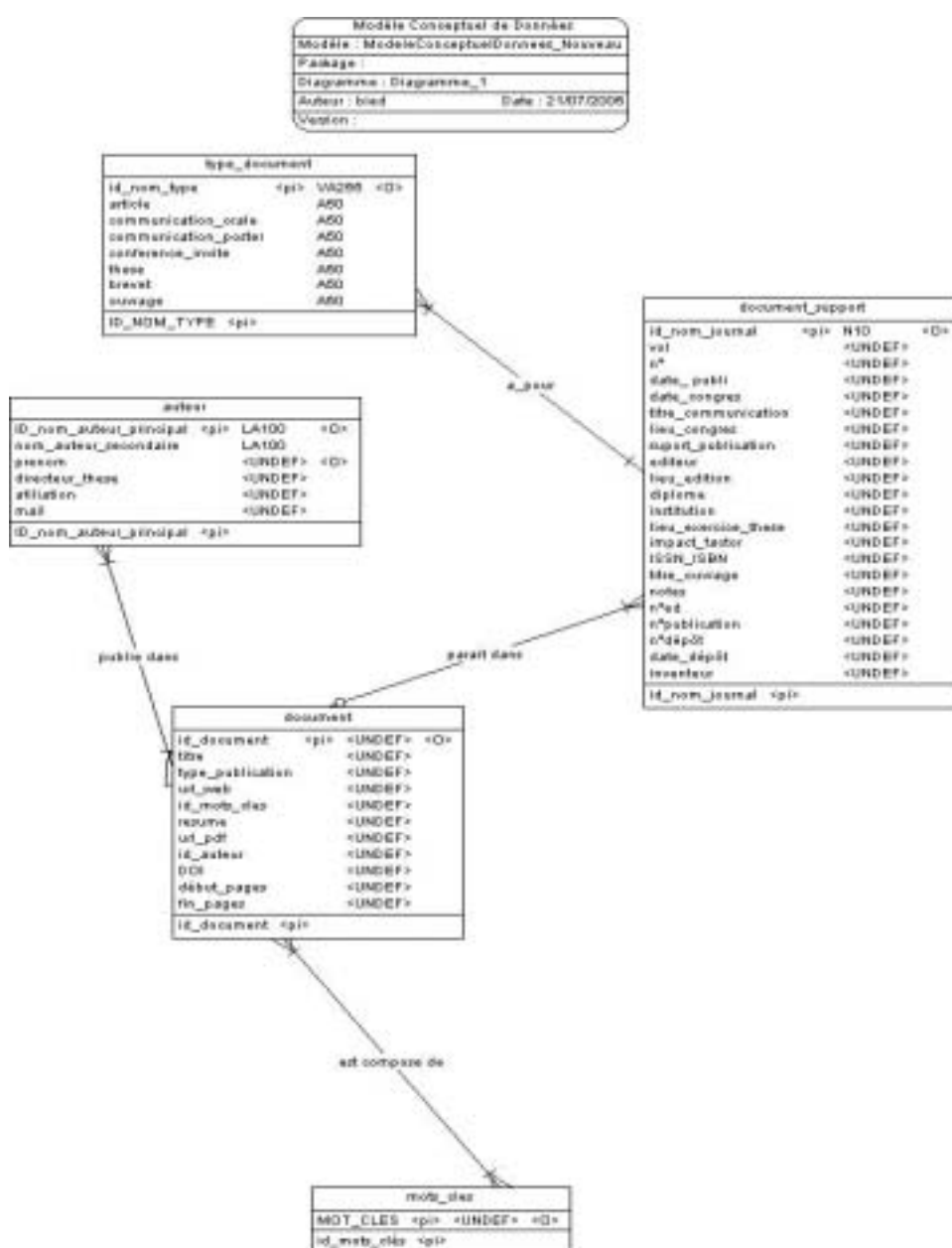
- les producteurs répartis dans différents services : praticiens et chercheurs, c'est à dire l'ensemble de la communauté scientifique de l'Institut,
- le service de la médiathèque qui fait office de médiateur entre les gisements d'information scientifique et le lecteur final au moyen de la nouvelle base de données en ligne,
- le lecteur final correspondant à la communauté scientifique interne à l'ICR et externe à l'ICR et in fine au patient.



3.2. Modèle conceptuel des données à informatiser (MCD)

La première étape dans la création d'un Système d'Information consiste à réaliser un MCD. Il s'agit de définir les entités-relations, de créer les tables en définissant les champs, les index éventuels et de mettre en relation les tables.

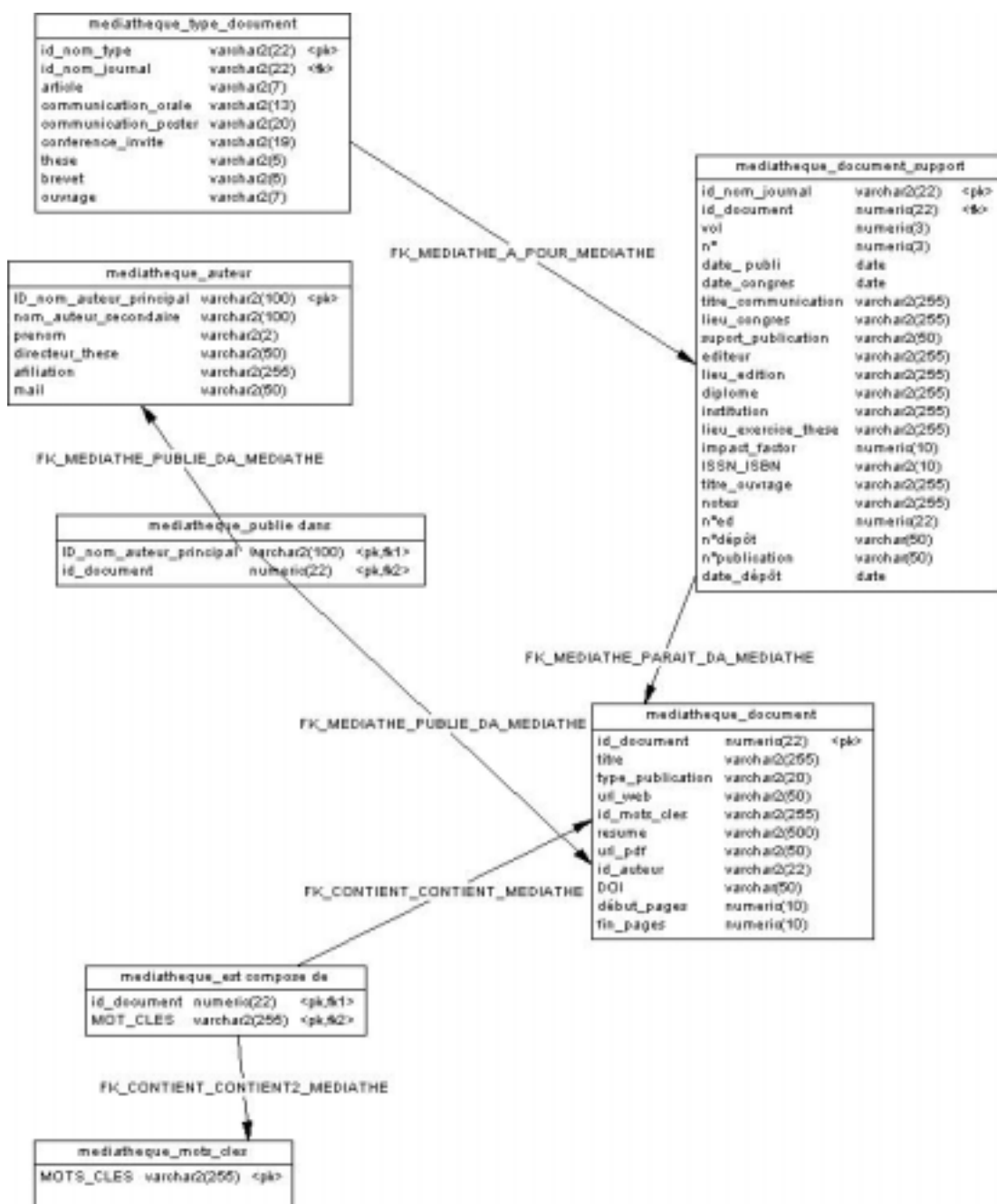
Les tables appelées « **type de document** », « **document** », « **document support** » (informations sur le document publiés dans une revue ou lors d'un congrès...), « **auteur** », « **mots-clés** » correspondent à la structure de la base de données..



3.3. Modèle Physique des données à informatiser (MPD)

La deuxième étape consiste en la réalisation du MPD, après l'élaboration du MCD. L'étape physique permet de préciser les systèmes de stockage employé et suggère un choix de matériel informatique pour le Système d'Information.

Concrètement, il permet d'implémenter le modèle dans le SGBD par la traduction dans un langage de définition de données. C'est dans ce graphique que les caractéristiques de chaque champ (type et taille) ont été mentionnées.



MISE EN ŒUVRE DU SYSTEME

1. Scénarios possibles

L'essor des logiciels libres et des langages de description de données (comme le XML) offre plusieurs scénarios pour la réalisation du projet. Ces possibilités ont été décrites et présentées au service Médiathèque avant de choisir une solution.

Parmi les possibilités, nous pouvons distinguer :

- Logiciel de gestion de références bibliographiques : Reference Manager module Web,
- Solution open source : PHP, Mysql, Apache,
- Solution mixte : PHP, Oracle,
- Solution entrepôt OAI (autoarchivage),
- Solution structuration de documents à partir du langage XML.

1.1. Module Web Reference Manager version 11 (module Web publisher)

Le logiciel de références bibliographiques Reference Manager version 11 élaboré par Thomson ISI ²³, contient un module Web publisher. L'Institut utilise ce logiciel et a acheté le module Web publisher.

Ce module Web est déjà paramétré : l'interface Web offre la possibilité de gérer 15 bases de données accessibles à distance et en réseau sur Internet et/ ou sur Intranet

1.1.1. Principe de fonctionnement

²³ <http://www.refman.com/>

Le module Web publisher s'installe sur un serveur Web (dans ce cas, l'installation d'un firewall est préconisée). Les bases de données peuvent être gérées de manière indépendantes et peuvent être aussi mises à jours à distance (si les bases sont autorisées en écriture lors de l'ouverture de la base).

Pour utiliser les fonctions d'import et d'export de références au format RIS et au format XML, il est nécessaire d'avoir installé en réseau la version 11 du logiciel, sur les postes souhaitant récupérer les données .

1.1.2. Fonctionnalités

Un module d'interrogation est intégré comprenant un mode d'interrogation simple et un mode d'interrogation avancé. Le dernier mode offre les fonctionnalités de recherche multi-critères sur 37 zones (cela correspond à la quasi-totalité des champs). Parmi les zones les plus utilisées, on distingue : auteur (différentes mentions de responsabilité), titre (différents niveaux de titres), mots-clés, date.

L'exploitation des références bibliographiques sélectionnées après une requête est possible par la fonction d'export en format RIS ou en format XML.

Plusieurs modes d'affichage des résultats sont possibles : format HTML ou format « bibliographique » utile pour générer une bibliographie (le format d'enregistrement possible est le HTML).

1.1.3. Paramétrage

Le paramétrage s'effectue à partir du menu « Web publisher », il est toutefois limité (une boîte de dialogue avec trois onglets est proposée pour l'ensemble du paramétrage).

Il est possible d'autoriser l'accès en écriture aux personnes ayant le logiciel Reference Manager : le bouton « create a new record » permet d'alimenter la centrale depuis le module Web publisher : un formulaire reprenant la structuration de la base du logiciel Reference Manager permet l'enregistrement automatique dans la base de données.

L'accès peut-être protégé en lecture et en écriture par un mot de passe (user name et password) et s'applique à toutes les personnes se connectant au site.

Le paramétrage du format d'affichage des données pour l'interface Web est possible (résumé, liens vers l'URL du document en texte intégral, vers les fichiers PDF...).

L'interface est en anglais. La personnalisation de l'interface est possible pour la page d'accueil et pour la page à « propos de ». Il est également possible de faire un lien pour créer un mode d'emploi approprié aux lecteurs.

1.2. Solution Open Source Web dynamique : PHP3, Mysql, Apache

Cette solution en logiciel libre permet de créer un système de gestion de base de données gratuit (SGBD) Mysql interfacée, au moyen de scripts d'interrogation et de mise à jour en langage PHP. Le SGBD est installé sur un serveur Apache gratuit, avec une machine Linux.

Le PHP3 est un outil idéal pour une SGBD de taille moyenne ou de grande taille, n'ayant pas à supporter un grand nombre de requêtes simultanées.

Le PHP3 est un langage embarqué dans les pages HTML et traité par le serveur. Il permet de construire dynamiquement des pages HTML contenant les résultats de calculs ou de requêtes SQL adressées à un SGDB.

Il est possible de reprendre les données existantes en format texte.

De plus en plus de projets utilisent cette solution peu coûteuse et très très fonctionnelle pour la création et la mise à jour d'une base de données.

1.3. Solution mixte : PHP, Oracle

L'Institut utilise majoritairement l'environnement Oracle pour créer et gérer les bases de données (l'environnement Oracle est inclus dans le schéma directeur). Il

est donc envisageable de récupérer les données en format texte pour créer une base de données Oracle interfacée avec le langage PHP. Il est possible de reprendre les données existantes en format texte.

La chaîne des traitements de l'information pour créer la base de données est automatique mais nécessite un certain nombre d'opérations ²⁴. Il faut d'abord créer la structure de la base avec le logiciel Power AMC pour la modélisation des données. Une fois que le modèle physique des données (MPD) est créé à partir du modèle conceptuel de données (MCD), il faut générer automatiquement un script pour créer une base de données SGBD. La base est stockée sous Oracle. L'application TOAD ²⁵ permet d'interroger les bases Oracle.

L'étape suivante consiste à exporter les données existantes en format TXT (en ayant bien défini le séparateur de champs) pour les stocker sur un serveur. Les données sont transformées automatiquement par l'application Sunopsis²⁶. Elles sont enfin stockées sous Oracle et peuvent être gérées au moyen de l'application TOAD.

Il est envisageable d'interfacer la base avec le progiciel de Business Intelligence (Datawarehouse) ²⁷, Business Project, pour élaborer des statistiques de consultation de la base de données depuis Intranet. Ce progiciel étant largement utilisé par le système d'information hospitalier à titre d'informatique décisionnel, il est possible de l'utiliser pour évaluer la qualité de cette nouvelle prestation.

L'avantage de cette solution est de pouvoir intégrer la base de données dans le Système d'Information Hospitalière. Toutefois, les traitements à faire, disproportionnés par rapport au volume de données prévu pour la base des publications, présentent un inconvénient majeur.

²⁴ Voir le schéma en annexe 5 page 88

²⁵ TOAD : outil d'administration et de développement rapide de base de données distribué par Quest Software

²⁶ Sunopsis : outil pour l'automatisation du dialogue entre les application d'une entreprise

²⁷ Datawarehouse et Business Object : progiciel de gestion des processus métiers et gestion des connaissances conçu pour représenter les processus et la connaissance d'un métier particulier.

Voir <http://phortail.org/webntic/Business-Objects-Presentation-19.html>

1.4. Entrepôt OAI (Open Archives Initiatives, Auto archivage)

Le projet serait de créer une base d'autoarchivage de pré-publications et de post-publications accessibles à distance sur Internet. Il serait envisagé de favoriser l'auto archivage en installant une base d'archives de publications électroniques compatibles OAI (Open Archives Initiatives).

L'objectif serait de permettre la conservation des documents dans ses différentes étapes depuis la première rédaction soumission auprès de l'éditeur, jusqu'à la publication dans les bases de données. Cela favoriserait la capitalisation et la conservation des productions scientifiques de l'Institut.

Techniquement, il est possible d'utiliser des logiciels pour stocker et exploiter les documents : par exemple le logiciel Arxiv²⁸ fournit une interface permettant aux auteurs de déposer sur la base du CCSD des manuscrits d'articles scientifiques. Il faut s'inscrire dans HAL (Hyper Article en Ligne) pour pouvoir déposer un document en texte intégral.

Trois types de dépôt sont possibles :

- Le document récent ou le document ancien (rétro-dépôt) dont on spécifie une date de rédaction antérieure,
- Le document dont le texte intégral est déjà disponible dans Arxiv,
- Une notice bibliographique avec si possible l'adresse DOI et les références de publication.

Une interface spécialisée permet de fournir des métadonnées : la présence de métadonnées structurées permet des recherches fines, comme des extractions automatiques des articles et des notices au niveau d'une Unité ou d'un laboratoire...

²⁸ Arxiv : <http://hal.ccsd.cnrs.fr>

1.5. Structuration de documents basée sur le langage XML

Ce choix serait une toute autre orientation technologique. Cette solution est utile pour une gestion de contenu orienté document et non orienté données.

Cela aurait un intérêt :

- Pour rassembler des données semi-structurées ou structurées de manière différentes,
- Pour normaliser un document (en particulier pour normaliser les formats d'échanges),
- A titre de stockage (pérennité du format),
- Pour la conversion dans l'optique d'intégrer les données dans d'autres systèmes d'information (extensibilité),
- Pour être facilement distribué par n'importe quel protocole, à même de transporter du texte (intégralité).

Deux solutions seraient envisageables pour créer une base de données :

- La première se baserait sur la création de notices (création de balises méta en XML : titre, auteur, date création du document...).
- La deuxième serait basée uniquement sur le XML (création de DTD, puis d'un schéma et d'une feuille de style XSLT) offrant ainsi des fonctionnalités d'interrogation en texte intégral et de déclinaison possibles en tout format. Une base de données native de documents serait alors constituée.

Pour la première solution, il peut-être possible d'associer les notices aux documents (dans ce cas, il convient d'opérer un stockage natif du document).

Pour cela, il conviendrait de définir un niveau de balisage normalisé des documents selon des normes de vocabulaire comme la TEI²⁹ ou comme Docbook³⁰ (vocabulaire de balisage XML adapté au ouvrages, aux articles notamment).

²⁹ TEI : Text Encoding Initiative permet le balisage de tout type de texte (prose, poésie, thèse, discours)

³⁰ Docbook : permet le balisage de documentation à caractère nettement plus technique (PHP, Linux, etc.) voir : <http://docbook.sourceforge.net/>

Cela serait d'autant plus envisageable que le balisage du document serait facilité par la structuration imposée par une publication scientifique. En effet, le document est soumis à des règles de rédaction quant à la forme et au contenu, règles très précises édictées par les éditeurs et diffusées dans les revues (rubrique instructions aux auteurs).

2. Evaluation des scénarios

	Avantages	Inconvénients
Reference Manager module Web Publisher	- Prise en main rapide : peu de paramétrage. - Autonomie pour la mise à jour - Logiciel utilisé actuellement par les praticiens, les chercheurs et la médiathèque médicale. - Interopérabilité du système : export en format RIS ou XML	- Paramétrage limité (interfaçage Web standard) - Interface en anglais fixée par défaut - Logiciel propriétaire - Usage essentiellement dédié à la gestion des références bibliographiques
PHP, Mysql, Apache	- Autonomie complète pour création de la base de données et pour la mise à jour - Prise en main rapide après formation - Système de plus en plus utilisé - logiciel libre	- Développement nécessaire : un mois de programmation à temps plein - Pas forcément adapté au faible volume d'information prévu - logiciel libre
PHP, Oracle	- Schéma directeur de l'Institut respecté - Temps de programmation : minimum un mois	- Compétences informatiques pour la base de données Oracle, Sunopsis - Système lourd (principe de fonctionnement lourd : trop d'étapes) - Compétences informatiques nécessaires - Pas forcément adapté par rapport au volume de données prévu
Base de données XML, XSLT document natif	- Fonctionne à partir d'un document natif - Utile pour une gestion de contenu orientée document	- Pas utile pour une gestion de contenu orientée données - Difficile d'obtenir la version originale des documents - Il n'est pas envisagé d'échanger les données

Archives Ouvertes	Un enjeu majeur pour la mémoire de l'entreprise : conservation du document et des métadonnées	- Nécessite la récupération de la version native du document - Le suivi d'un document de la pré-publication à la post-publication n'est pas indispensable à l'Institut (une seule Unité y aurait recours)
------------------------------	---	--

3. Projet retenu

Une réunion a eu lieu avec l'ensemble du personnel de la médiathèque médicale et le responsable du Département d'Information Hospitalière.

L'enjeu de cette réunion était de présenter le projet auprès du responsable (il n'avait pas pris connaissance du projet avant le début du stage) et ensuite de le solliciter pour une prise de décision sur le choix du système à implanter.

Deux solutions ont été étudiées : la solution avec le logiciel Reference Manager et la solution base de données sous Oracle (la dernière a été étudiée car Oracle est intégré au schéma Directeur de l'Organisme). Les autres ont été écartées car les solutions Open Source ne font pas encore partie du Schéma Directeur de l'Institut. La première a été retenue car elle semble adaptée au volume des données à intégrer. De plus, la dernière version du logiciel Ref Manager comprenant un module Web vient d'être achetée en version monoposte par l'Organisme et rentre dans les préconisations du cahier des charges.

A l'issue de la réunion, il a été décidé d'ouvrir l'accès à la base de données sur Intranet. Les dernières parutions des publications continueront à être diffusées. Une page statique présentant les publications de l'année sur Internet sera aussi réalisée.

A ce stade du projet, une réflexion personnelle s'est imposée concernant le choix qui s'est opéré : la solution sera respectée dans sa mise en oeuvre mais elle paraît être une solution transitoire car elle ne nous semble pas optimale. Elle impose des contraintes de stockage de documents et de données dans le même environnement informatique et elle implique l'achat d'une version réseau du logiciel. Une solution

orientée Open Archives Initiative avec le dépôt HAL par exemple permettrait de diffuser efficacement les publications sur le Web.

4. Mise en œuvre du système

La solution choisie par les acteurs est mise en place en procédant à des étapes définies au préalable.

4.1. Etapes

Il s'agit de récupérer les données et de les rassembler dans le logiciel de références bibliographiques Reference Manager version 11 avec le module Web Publisher.

Trois étapes sont nécessaires pour la constitution de la base :

1. Alimenter la base de données avec les références bibliographiques
2. Créer une banque de documents en texte intégral
3. Réaliser le lien entre les références bibliographiques constituées et les articles en version intégrale

La constitution de la base de références bibliographiques s'est opérée en deux temps. Dans un premier temps, la base a été alimentée à partir des articles scientifiques issus de la base Pubmed puis intégrés dans le logiciel Reference Manager.

Dans un second temps, d'autres documents ont été intégrés afin d'enrichir le contenu de la base de données avec des documents non référencés dans la base Pubmed :

- les communications aux congrès,
- les communications internes à l'Institut Claudius Regaud,
- les brevets,

- les thèses d'exercice en médecine et en pharmacie réalisées à l'ICR.

4.2. Collecte des données

4.2.1. Structure de l'application créée dans Reference Manager

Une base est créée pour chaque année dans le logiciel References Manager, (cela correspond aux six dernières années.) Une telle organisation permet de simplifier l'interrogation pour l'utilisateur (un index correspondant à chaque base facilite l'interrogation.)

Une base d'archives correspondant aux années antérieures à 2000 est également construite.

Au préalable, il est nécessaire d'adapter certains masques d'aide à la saisie : pour cette opération, la base doit être ouverte en mode d'écriture uniquement. Les champs inutiles dans les masques de saisies sont supprimés, d'autres champs jugés utiles sont rajoutés (comme le champ « **intitulé du congrès** ») et l'ordre d'apparition des champs est modifié pour une meilleure saisie.

Le masque de saisie correspondant au type de document « personal communication » est utilisé pour référencer les communications diffusées à l'Institut Claudius Regaud et non publiées : certains champs comme le nom et du lieu de la conférence sont rajoutés.

Au fur et à mesure de l'alimentation de la base, un travail de dédoublonnage (élimination des doublons) des références bibliographiques est effectué sur les champs auteurs, titre et mots-clés : le dédoublonnage est possible au moment de l'import des références ou une fois que la base est alimentée (Menu Tools, Fonction Batch Operations / Duplicate Search.)

4.2.1.1. Environnement informatique

Les moyens mis à disposition n'ont pas permis de mettre en œuvre de façon optimale la base de données. D'une part, un poste client est utilisé pour héberger le module Web du logiciel Ref Manager puisque la version réseau n'est pas acquise à

l'ICR. D'autre part, il n'a pas été possible durant le stage de disposer d'un serveur, ni d'un espace disque pour stocker de façon pérenne les fichiers PDF (les documents GEIDE ont une date de publication.) le travail de collecte des données a cependant pu être réalisé.

4.2.1.2. *Volumétrie*

En terme de volume des données, le travail porte sur l'alimentation de la base de données de manière exhaustive, soit en moyenne une centaine de références sur les six dernières années. La base des archives portant sur la période de 1980 à 1999, est moins exhaustive (neuf cent références au total) : en effet, cela nécessite la connaissance des noms des praticiens et des chercheurs à cette époque, ce qui ne nous a pas toujours été communiqué.

Après quatre mois de stage, la base constituée représente un échantillonnage pertinent de documents.

4.2.2. Traitement des références bibliographiques

4.2.2.1. *Recueil des articles à partir de Pubmed*

La National Library of Congress met gratuitement à disposition l'accès aux références bibliographiques.³¹

Deux méthodes sont possibles pour collecter les références d'articles (différents types de documents sont référencés dans cette base comme les articles, les articles de synthèse, les cas cliniques...) : depuis Pubmed ou directement à partir de Reference Manager après avoir installé les connecteurs appropriés (correspondant au bouton propriétés du serveur).

A partir des résultats obtenus depuis Pubmed, les fonctions d'affichage au format Medline permettent l'export au format TXT. Il suffit d'utiliser ensuite les fonctions d'import dans Reference Manager (un certain nombre de champs comme le lien URL vers Pubmed, les mots-clés sont automatiquement intégrés)

³¹ Pubmed : <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>

directement à partir de Reference Manager après avoir installé les filtres appropriés.

L'autre méthode se fait directement à partir de Reference manager après avoir installé les connecteurs ou filtres appropriés.

La recherche depuis Pubmed est utilisée car elle offre plus de fonctionnalités : les requêtes, l'historique des recherches, le panier conservant les résultats obtenus (Clipboard.)

Les requêtes les plus utilisées portent sur les auteurs (à titre d'auteur principal et d'auteur secondaire) : nom et l'initiale du prénom de l'auteur.

Exemple :

« Ferron G » [au]

Elles portent aussi sur le champs Adresse pour obtenir la mention d'affiliation :

Exemple :

« Institut Claudius Regaud » [ad]

Des limites sur le champs date sont utilisées (date de publication et non date de saisie de la référence dans Pubmed).

Exemple :

« Ferron G » [au] AND 2004 [dp]

Les filtres de Pubmed, utiles pour sélectionner les références en fonction des concepts (systematic reviews, evidence based medicine) ont été peu utilisés, seulement pour distinguer un article d'une review.

Les résultats obtenus sont conservés dans le panier (Clipboard) pour être ensuite être exportés au format TXT.

Les requêtes sont faites manuellement ; mais il est également possible de faire des recherches manuelles ou automatiques avec la création de profil de veille dans Pubmed ³² ou dans Biomail³³.

4.2.2.2. *Recueil des communications*

Les données sont fournies par le documentaliste-iconographe de l'ICR puisqu'il gère les fichiers Power Point des communications scientifiques de l'ICR. Les fichiers en texte intégral sont fournis en format PDF et Power Point. Les fichiers sont stockés dans la GEIDE et publiés sur l'Intranet.

Le travail consiste à rajouter les métadonnées (date, lieu de la conférence, mots-clés) nécessaires à la recherche d'un document.

Pour les mots-clés, un travail d'indexation est opéré à l'aide du thésaurus MESH de Pubmed et du thésaurus CISMEF³⁴ pour la traduction des termes français.

4.2.2.3. *Recueil des brevets*

Peu de brevets sont produits à l'ICR (5 à ce jour depuis 2002)

Les références bibliographiques sont rapatriées depuis le site internet de l'INPI ³⁵ et de l'organisme européen European Patent Office ³⁶.

4.2.2.4. *Recueil des thèses (à partir de 2000)*

Il s'agit des thèses d'exercice de médecine ou des thèses de doctorat. Les références bibliographiques sont collectées depuis le catalogue du SUDOC mais il n'est pas possible d'installer un filtre d'import. Les données sont enregistrées en TXT.

La recherche porte sur le champs auteur (ICR ou Université Paul Sabatier) croisée avec le mots-clés « cancer ».

³² Les fonctions de personnalisation de Pubmed comme la création de profil de recherche sont accessibles dans la rubrique MYNCBI

³³ <http://www.biomail.org/>

³⁴ cismef : <http://www.chu-rouen.fr/cismef/>

³⁵ http://www.inpi.fr/front/show_rub.php?rub_id=8

³⁶ <http://ep.espacenet.com>

Un tri portant sur la mention du directeur de thèse est effectué pour sélectionner les références.

4.2.3. Traitement du texte intégral

La médiathèque collationne les tirés-à- part des articles et formats PDF provenant des sites Internet des éditeurs de revues électroniques : Les fichiers PDF sont enregistrés dans la GEIDE (pour les trois dernières années) et sont publiés sur l’Intranet pour une durée limitée.

Afin de faire connaître le travail des cliniciens et des chercheurs à la communauté scientifique de l’ICR, nous avons rapatrié les documents selon trois méthodes :

- depuis Pubmed,
- depuis le portail de l’Université Paul Sabatier de Toulouse dont l’accès est ouvert pour le personnel universitaire de l’ICR
- depuis le portail BiblioInserm de l’INSERM.

Une numérisation du document peut être également envisagée si le document n’est pas disponible en ligne. Cela n’a pas été effectué au cours du stage.

Les autres documents récupérés sur le serveur GEIDE comme les communications lors des congrès et lors des conférences organisées par l’ICR sont publiées d’une part sur le serveur GEIDE et d’autre part sur l’Intranet pour une durée maximale d’un an.

Pour que les documents soient téléchargeables en ligne depuis leur référence, leur stockage doit être dans le même environnement informatique que la base de données. Or en terme de stockage la base ayant été mise en ligne depuis un poste client pour une période temporaire, cela implique de refaire les liens sur le lieu de stockage définitif.

L’autre solution envisagée est de faire un lien vers les articles déjà stockés dans la GEIDE, mais le temps de publication des documents étant limité et surtout l’accès à la GEIDE étant restreint, cette solution a été abandonnée.

Pour des raisons de droits d'auteurs, les éditeurs n'autorisant pas la publication des documents, les articles publiés feront l'objet d'un signalement uniquement sur l'Intranet. Un certain nombre d'éditeurs autorisent la publication de leurs documents dans le cadre des Archives Ouvertes (la liste des éditeurs est accessible sur le site Sherpa³⁷ : un résumé indique pour chaque éditeur la politique de transfert de Copyright.)

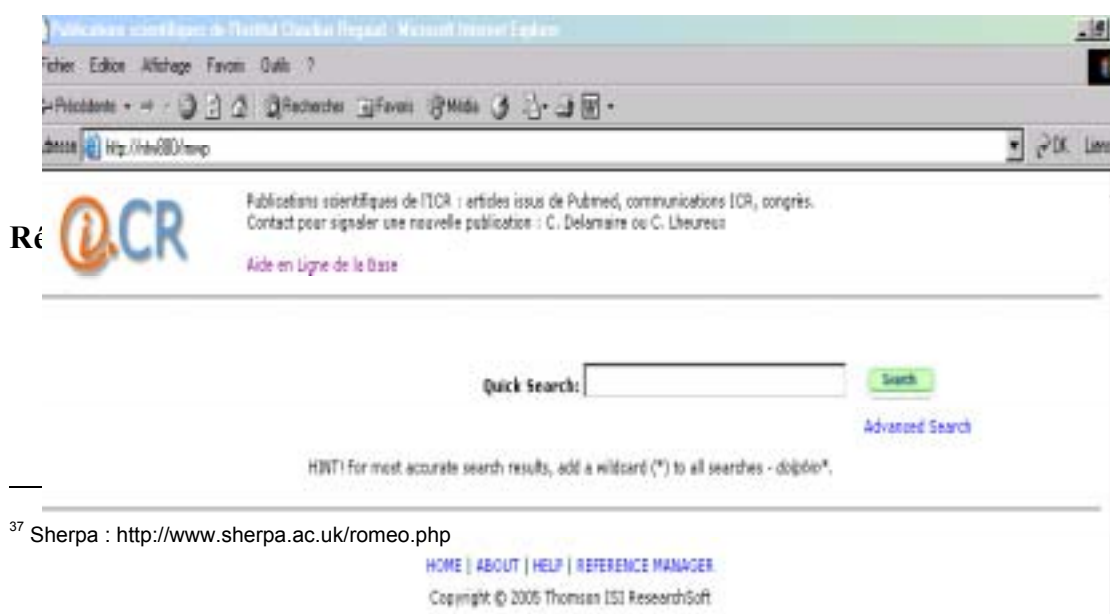
4.3. Mise en ligne de la base

4.3.1. Paramétrage de Web Publisher

La mise en ligne pour accéder à la base de donnée est prévue pour fonctionner en priorité sur l'Intranet. Le paramétrage de la page d'accueil est effectué : une brève présentation du contenu de la base a été réalisée ainsi qu'une aide en ligne pour une première prise en main des fonctionnalités d'interrogation et d'export.

Les fonctionnalités de personnalisation de l'interface sont réduites : les rubriques de la page d'accueil sont limitées en caractères, les modes d'affichages des résultats sont réduits et non modifiables (certains champs personnalisés lors de la création de la base n'apparaissent pas lors de l'affichage des résultats de recherche).

Page d'accueil de l'interface



³⁷ Sherpa : <http://www.sherpa.ac.uk/romeo.php>

Affichage des résultats de recherche



4.3.2. Publication de la base sur Intranet

Des liens vers la base et vers l'aide en ligne sont créés dans la rubrique des publications. Un signalement pourra être effectué à partir de la page d'accueil pour optimiser l'accès.



4.4. Bilan de la réalisation de la base de données

La publication de la base a été annoncée par mail auprès des praticiens et auprès du service du Département d' Information Hospitalière.

Le système répond-il aux besoins exprimés par les acteurs ?

Nous constatons que la réalisation de la base de données est bénéfique car elle offre l' accès à distance d'un échantillonnage représentatif des publications de l'ICR.

Cependant, le système a deux points faibles : il n'a pas été possible d'insérer les liens PDF dans la base de données, de ce fait le besoin exprimé par les usagers n'a pas pu être réalisé. Sur le plan technique, le poste client utilisé comme serveur Web du logiciel Reference Manager n'est pas une solution, surtout pour le stockage des fichiers PDF.

L'autre problème se situe dans le module Web notamment sur la possibilité d'importer les références. Cette fonctionnalité n'est offerte qu'avec la dernière version du logiciel. L'achat de la nouvelle version avec une licence réseau sera nécessaire pour l'ensemble de l'Institut. Le montant prévisionnel de la licence installée sur un serveur avec un accès simultané de cinq utilisateurs s'évalue à 2195 € (prix éducation valables jusqu'au 30 septembre 2005).

Conclusion : évolutions et perspectives du système

Le projet a pu se concrétiser par la réalisation d'une base de données en ligne offrant ainsi un nouveau service aux usagers. Les phases d'analyse préalable et d'étude de solutions ont pris le plus de temps, car cela est nécessaire pour proposer un système fiable et pérenne.

En terme d'évaluation du projet, nous avons pu constater que les moyens mis à disposition pendant le stage se sont avérés limités.

Nous préconisons une rapide évolution du système. En effet, la solution choisie est une transition vers d'autres dispositifs. La première évolution sera d'intégrer la base des publications dans le projet de management des connaissances (Knowledge Management) prévu à moyen terme par l'Institut Claudius Regaud. Un portail Intranet offrirait un point d'entrée commun à tous les différents services. Les connaissances « tangibles »³⁸ contenues dans les bases de données seraient accessibles au même niveau que les connaissances « intangibles », elles mêmes composées des savoirs-faire et des compétences de l'ensemble du personnel de l'Organisme.

Le projet de publication en ligne des documents scientifiques rentre dans ce cadre puisqu'il s'agit d'un projet de capitalisation et de partage des connaissances. Pour que cela soit intégré dans le projet, l'évolution possible serait de permettre la mise à jour de la base par les producteurs de documents eux-même, comme ce qui existe déjà avec les procédures de publication d'un document GEIDE produit par l'ICR : un dispositif de co-production pourra être étudié dans ce sens. Par exemple, l'objectif serait d'axer le dispositif en amont pour arriver à mieux collecter les

³⁸ La différence entre connaissances tangibles et intangible a des répercussions en terme de gestion des supports : voir Duizabo,S.,Guillaume, N. Les problématiques de gestion des connaissances dans les entreprises. 1997, Dauphine Marketing stratégie Prospective

informations (et non pas une fois qu'elles sont publiées) et optimiser la phase de collecte des données.

Le moyen serait d'établir une procédure pour une alimentation collective : une base commune serait installée sur un serveur, l'attribution de mot de passe pour les personnes habilitées à verser un document serait effectué.

Un formulaire d'alimentation de la base pourrait être réalisé prévoyant l'envoi des références avec le versement du document dans sa version native.

L'autre évolution sera de prévoir une récupération des données vers d'autres initiatives comme le mouvement des Archives Ouvertes, (OAI) favorisant ainsi le travail en réseau nécessaire pour la pratique médicale des chercheurs. Cette solution s'intégrerait à moyen terme avec les types de systèmes choisis et les logiciels Open Source prévus dans le futur Schéma Directeur Informatique de l'Organisme.

Ce projet devra prendre en compte le fait que les archives ouvertes sont peu connues à l'Institut Claudius Regaud : parmi les personnes interrogées lors de l'enquête, personne ne connaissait le sujet.

D'autre part, sur le plan technique et organisationnel, il est parfois difficile de récupérer les documents en version native puisque les auteurs ne les conservent pas mais les rapatrient une fois publiés en version PDF depuis les bases de données bibliographiques comme Pubmed.

La solution pourra consister à court terme à montrer l'intérêt des Archives Ouvertes à la communauté médicale et scientifique et à demander à sauvegarder la version native du document.

Bibliographie

PUBLICATIONS SCIENTIFIQUES ET INTERNET : UN OUTIL POUR LA CAPITALISATION DES CONNAISSANCES

BARTHELEMY, P., DOU, H. dir. de thèse. *Les réseaux pour la recherche, outil de diffusion de l'information scientifique.* Thèse de doctorat Sciences de l'Information et de la Communication. Aix-Marseille : Université Paul Cézanne. Faculté des sciences et techniques de Saint-Jérôme, 1993, 233 pp.

BUTTNER, A. *Comment rédiger un rapport ou une communication scientifique* [EN LIGNE], 2002. Disponible sur :
<<http://www.sante.univ-nantes.fr/cidmef/menu/CommentRediger.pdf>> (consulté le 05/06/2005).

CHARTRON, G. *Elément pour une approche comparée de la publication scientifique. Forum Universitaire : la communication scientifique en quatre dimensions.* 2003, Montréal. [EN LIGNE]. Disponible sur :
<http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/04/35/sic_00000435_00/sic_00000435.pdf> (consulté le 10/08/2005).

AGENCE UNIVERSITAIRE DE LA FRANCOPHONIE (AUF). *Infothèque francophone : ressources en ligne et actualités scientifiques francophone* [EN LIGNE], 2004. Disponible sur :
<http://www.infotheque.info/article.php?article=210> (consulté le 10/08/2005).

LEBEL, D., BUSSIÈRES, J.F.; LECOMPTE, L.L. La gestion des références dans une publication scientifique. *Pharmactuel*, 2000, 33, 3, 75-78 pp.

PIGNARD, N., PAILLARD, I. dir. de thèse. *Les nouvelles formes de publications scientifiques sur internet : la remise en cause du modèle éditorial traditionnel.* DEA Sciences de l'Information et de la Communication Option Communication et Médias. Grenoble: Institution de la communication et des médias- Université Stendhal, Grenoble 3, 2000, 112 p.

MAHE, A., SALAUN, J. dir. de thèse. *La communication scientifique en (r)évolution : l'intégration des revues électroniques dans les pratiques informationnelles de chercheurs en sciences de la nature comme révélateur des mutations du modèle traditionnel de la communication scientifique.* Thèse de doctorat. Lyon : Université Claude Bernard, 2002, 178 p.

MAROS, D. *L'édition électronique scientifique : la longue marche vers l'appropriation* [EN LIGNE], 2004. Disponible sur :
<<http://gaougalinr.chez.tiscali.fr/articles.php?lng=fr&pg=359>> (consulté le 10/09/2005).

ACADEMIE DES SCIENCES. *Les Publications scientifiques et techniques en langue française*. Paris : Edition Tech et Doc, 1998, 1-39 pp.

TIETSE, S., BOUCHE, R. dir. de thèse. *Internet et capitalisation des connaissances en médecine : construction de la valeur d'usage des outils de l'internet par les médecins hospitalo-universitaires*. Thèse pour l'obtention du titre de Docteur. Lyon : Université Claude Bernard Lyon 1- Ecole Doctorale Informatique et Information pour la Société, 2003, 195 p.

PRINCIPALES BASES DE DONNÉES EN MÉDECINE-SCIENCE, MODE D'EMPLOI

MONTREAL, U. D., MICHAUD, S. *Pubmed : aide en ligne* [EN LIGNE], 2005. Disponible sur :
<<http://www.bib.umontreal.ca/SA/caps35.htm>> (consulté le 02/08/2005).

RATTE, S. *Une analyse descriptive et comparative d'une sélection de bases de données bibliographiques du domaine biomédical* [EN LIGNE], 2004. Disponible sur :
<http://bibliothecaire.com/SRatte_BDB_20040604.pdf> (consulté le 02/08/2005).

THIRION, B. *Medline/Pubmed : mode d'emploi* [EN LIGNE], 2004. Disponible sur :
<<http://www.chu-rouen.fr/documed/pubmedhelp.html>> (consulté le 02/08/2005).

RÉALISATIONS DE BASE DE DONNÉES DE PUBLICATIONS SCIENTIFIQUES

(CENTRE PAUL PAPIN. *Publications des praticiens du Centre Paul Papin* [EN LIGNE], 2004. Disponible sur :
<http://www.centrepaulpapin.fr/esp_medecins/publications.htm> (consulté le 02/08/2005).

DALLOT-FEBIO, Y., PRADAL, T. dir. de thèse. *Etude de faisabilité d'une base de données des publications des chercheurs de l'INSA de Lyon*. DESS Informatique documentaire. Lyon: ENSSIB/Université Claude Bernard Lyon 1, 1997, 78 p.

POMPIDOU, C. *Base des articles scientifiques de l'IRCAM* [EN LIGNE], 2005. Disponible sur: <<http://articles.ircam.fr/>> (consulté le 02/08/2005)

CENTRE NATIONAL D'ETUDES SPATIALES. *Arsitote : pérennisation des informations numériques* [EN LIGNE], 2004. Disponible sur : <http://vds.cnes.fr/pin/pin_contacts.HTML> (consulté le 02/08/2005).

CENTRE CANADIEN DE TELEDETECTION. *Centre d'information sur la terre. Base de données GEOSCAN* [EN LIGNE], 2005. Disponible sur: <http://geoscan.ess.nrcan.gc.ca>. (consulté le 5/07/10/).

UNIVERSITY COLLEGE LONDON (UCL). *Information about research* [EN LIGNE], 2005. Disponible sur: <<http://www.ucl.ac.uk/research/publications/new/>> (consulté le 02/08/2005).

SYSTÈME D'INFORMATION ET MODÉLISATION DE DONNÉES : ÉLABORER UN CAHIER DES CHARGES

BENARD, C. *Le cahier des charges informatique : l'expression des besoins des utilisateurs*. 3ème tirage édition. Paris: Editions d'Organisation, 1993,190 p.

ACADEMIE DE GRENOBLE *Du modèle conceptuel de données au schéma physique de la base. BTS assistant de gestion PME/PMI* [EN LIGNE], 2000. Disponible sur : <<http://www.ac-grenoble.fr/ecogest/pedago/administration/prodpeda/mcd.htm>> (consulté le 02/08/2005).

Web-NTIC MOE MOA. *Web NTIC : cahier des charges fonctionnel général* [EN LIGNE], 2005. Disponible sur : <<http://phortail.org/webntic/>>(consulté le 20/08/2005).

ACCÈS LIBRE AUX PUBLICATIONS SCIENTIFIQUES : ARCHIVES OUVERTES

AUBRY, C., JANIK, J. *Les Archives Ouvertes : enjeux et pratiques. Guide à l'usage des professionnels de l'information*. PARIS: ENSSIB, 2005, 332 p. (Sciences et Techniques de l'Information)

BOSC, E. *Sélection thématique de sites Web : La communication scientifique en libre accès* [EN LIGNE], 2005. Disponible sur :
<http://www.tours.inra.fr/prc/internet/documentation/communication_scientifique/coms_cith.htm> (consulté le 02/08/2005).

BOSC, E. *L'Information au service du chercheur La communication scientifique revue et corrigée par internet* [EN LIGNE], 2005. Disponible sur :
<http://www.tours.inra.fr/prc/internet/documentation/communication_scientifique/coms_ci.htm> (consulté le 10/08/2005).

CCSD, CNRS. *Archives Sic. Liste par domaine : Communication et information scientifique* [EN LIGNE], 2002. Disponible sur :
<http://archivesic.ccsd.cnrs.fr/view_sic_comm_fr.HTML> (consulté le 02/08/2005).

CHANIER, T. *Archives ouvertes et publications scientifiques*. Paris: l'Harmattan, 2004, 186 p.

CNRS, CCSD. *Centre pour la communication directe CCSD* [EN LIGNE], 2002. Disponible sur :
<http://ccsd.cnrs.fr/rubrique.php3?id_rubrique=3> (consulté le 15/05/2005).

CNRS, CCSD. *Archives EDUC-TICE* [EN LIGNE], 2004. Disponible sur :
<<http://archive-edutice.ccsd.cnrs.fr/>> (consulté le 15/05/2005).

INIST, CNRS. *Libre accès à l'information scientifique et technique* [EN LIGNE], 2005. Disponible sur :
< <http://www.inist.fr/openaccess/> > (consulté le 14/09/2005).

JOINT INFORMATION SYSTEMS COMMITTEE, UNIVERSITY OF SOUTHAMPTON, BOULETREAU, V.(TRAD). *L'auto-archivage : foire aux questions* [EN LIGNE], 2005. Disponible sur :
<http://sophia.univ-lyon2.fr/boai/self-faq_fr.HTML> (consulté le 10/09/2005).

FÉDÉRATION DE DONNÉES STRUCTURÉES AVEC XML

COURTAUD,D., BROCHARD,P. *Xedix : une plate-forme XML pour la pérennisation des données.* CEA, réunion du groupe de travail PIN d'Aristote. 2001, Ile-de-France. [EN LIGNE], 2001. Disponible sur :
<http://vds.cnes.fr/pin/presentations/2001/xml_xedix_cea.pdf> (consulté le 10/08/2005).

DANG NGOC, T. T. *Fédération de données structurées avec XML.* [EN LIGNE]. Thèse PHD Informatique. Saint-Quentin: Université de Versailles Saint-Quentin en Yvelines, 2003, 274 p. Disponible sur :
<<http://tel.ccsd.cnrs.fr/documents/archives0/00/00/51/62/index.html>> (consulté le 20/08/2005).

ROUX, D. *Groupe Le monde : l'intégration de la technologie XML dans le processus documentaire.* [EN LIGNE], 2004. Disponible sur :
<<http://www.promidee.net/egide/presentations.html>> (consulté le 10/09/2005).

Table des annexes

ANNEXE 1. PRÉSENTATION DE L'INSTITUT CLAUDIUS RÉGAUD	79
ANNEXE 2. ORGANIGRAMME DU DÉPARTEMENT D'INFORMATION HOSPITALIÈRE	82
ANNEXE 3. GESTION DU PROJET, DIAGRAMME DE GANT	84
ANNEXE 4 . QUESTIONNAIRE D'ÉTUDE DE BESOINS.....	86
ANNEXE 5. MODÈLE DE TRAITEMENT DES DONNÉES DE REFERENCE MANAGER VERS ORACLE	89

Annexe 1. Présentation de l'Institut Claudius Régaud



INSTITUT CLAUDIUS REGAUD³⁹

STATUT

L'Institut Claudius Regaud est le Centre Régional de lutte contre le cancer en région Midi-Pyrénées. C'est un établissement privé à but non lucratif, participant au Service public. Il fait partie de la fédération Nationale des Centres de Lutte Contre le Cancer⁴⁰ (FNCLCC) qui regroupe les 20 centres français.

ACTIVITES- MISSIONS

L'établissement a pour but de dépister et de diagnostiquer les tumeurs, d'effectuer le soin auprès des patients, de participer à la mission de recherche pour développer de nouveaux médicaments et de réaliser des activités de formation continue auprès des professionnels de la santé.

ORGANISATION DE L'INSTITUT

Créé en 1923, l'Etablissement comprend 630 personnes, dont 70 médecins et chercheurs.

L'organisation de l'Etablissement s'articule autour du patient.

4 départements travaillent en collaboration avec les délégations et avec les services de l'Administration Générale.

Les départements se répartissent ainsi :

- Département d'Information Hospitalière (DIH)
- Chirurgie
- Médecine
- Radiations
- Biologie

Les délégations sont au nombre de 4 :

- Recherche
- Soins
- Qualité, risques
- Enseignement, formation

Les délégations ont une mission bien spécifique :

La mission du **Département d'Information Hospitalière (DIH)**, est de piloter l'informatisation des services de l'ICR selon un axe technique, un axe fonctionnel et un axe stratégique.

³⁹ <http://www.claudiusregaud.fr/>

⁴⁰ <http://www.fnclcc.fr/>

Il s'organise autour de trois pôles :

1. pôle infrastructure technique

Une équipe est chargée de l'architecture, des réseaux, de la sécurité, de l'administration des systèmes et de la convergence images-données

2. pôle infrastructure applicative

Une équipe chargée d'études et de développement, comme l'urbanisation du système d'information, les projets d'informatisation et d'accompagnement au changement

3. pôle production de données médicales

La production d'informations médicales exhaustives, vérifiées et validées : pour cela deux entités sont créées : l'une produit l'information, l'autre la contrôle (**DIM**, Département d'Information Médicale).

La médiathèque et le DIM s'intègrent dans le pôle de production de données médicales.

Le DIM a pour mission de contrôler l'information produite par l'assistanat médical.

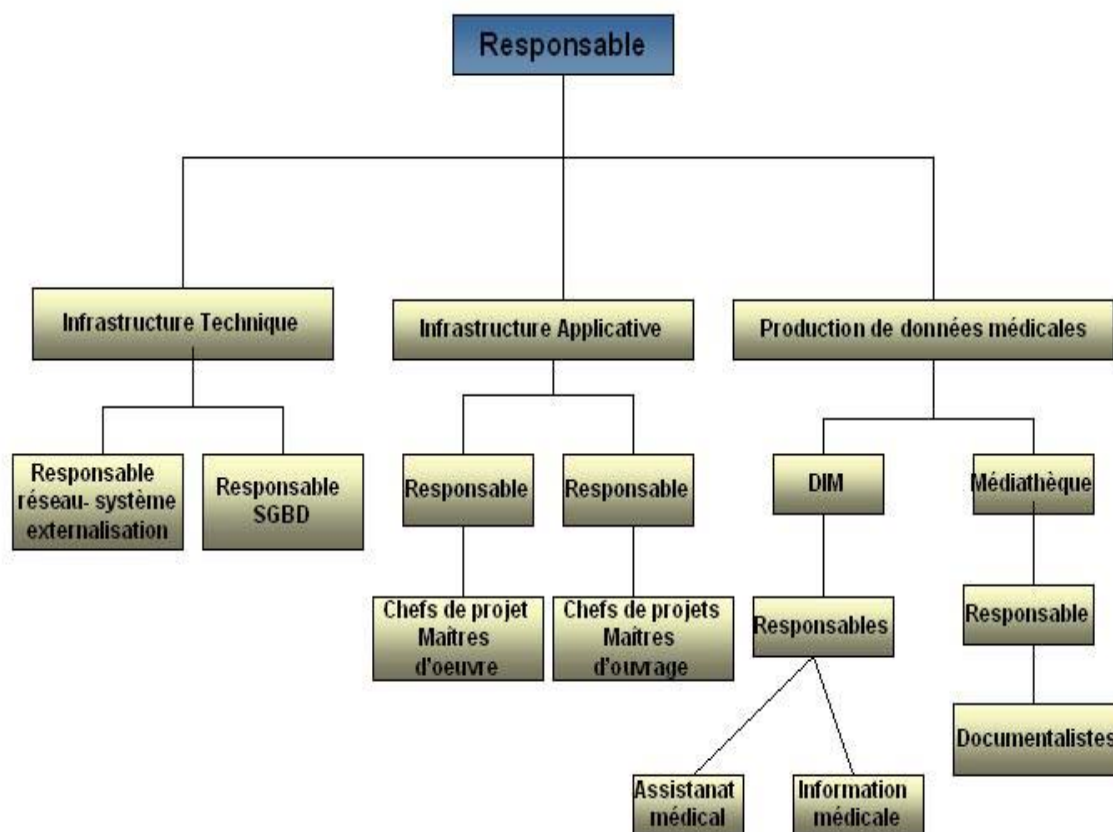
La médiathèque a pour mission :

- d'assurer la gestion du travail collaboratif et de la gestion des connaissances
- d'assurer la fourniture de documents scientifiques pour l'Institut

Annexe 2. Organigramme du Département d'Information Hospitalière



Département d'Information Hospitalière



Annexe 3. Gestion du projet, diagramme de Gant

Annexe 4 . Questionnaire d'étude de besoins

Enquête : entretien semi-dirigé d'étude de besoins

Nom de la personne interrogée :

Département :

Secteur d'activité, spécialité :

Présentation du cadre et d l'objet de l'enquête

Un besoin pour l'ICR :

- mutualiser les ressources électroniques scientifiques produites par l'ICR
- accéder à ces ressources à distance par le réseau local

L'objectif est de connaître les pratiques et de mieux cerner les besoins des chercheurs et praticiens.

1. Vos pratiques de gestion des publications scientifiques

1.1 Corpus de documents

- a) Quel genre de document avez-vous ? (références bibliographiques, texte intégral)

Type de document ? articles, autres

- a) Avez-vous stocké vos références bibliographiques en local sur un disque dur ou sur un disque partagé ? avez-vous un logiciel de gestion de références bibliographiques (lequel) ?

- b) Quel est le format de stockage des données ?

- c) Stockage exhaustif ? (nombre de références, nombre d'articles)

- d) Seriez-vous prêts à les exporter et à les donner à la médiathèque pour le projet ?

- e) Droits d'auteur : lorsque vous publiez un article, avez-vous cédé vos droits d'auteurs ? (cf contrat de l'éditeur)

- f) Connaissez-vous les pratiques des autres médecins de votre département et des autres départements ?

1.2 Recherche documentaire

- a) Avez-vous recours aux documents publiés par l'ICR? dans quel cadre ? fréquence

- b) Quelle est votre démarche pour y avoir recours ?

où ? (Intranet, rubrique médiathèque, bases de données comme Pubmed...)
comment ? recherche par auteur, par mots-clés,...

- c) Faites-vous des recherches documentaires sur vos propres publications où sur celles d' autres médecins (pour réaliser une épreuve de titres ?)

Vos besoins

1.1. Avantages , points à améliorer

- a. Quels avantages verriez-vous à cette nouvelle prestation proposée par la médiathèque ?
- b. Par rapport à l'existant, quels seraient les éléments du système actuel à améliorer ? quels seraient vos besoins en terme
 - de modalité d'accès,
 - de besoin en nature des informations,
 - d' organisation de l'information
- c. Voyez-vous un intérêt à publier les articles en texte intégral associé à la référence bibliographique ?
- d. Si l'information est centralisée dans une base de données unique, dans quel endroit serait-elle la mieux placée pour y accéder ? sur Intranet, dans quelle rubrique ?
- e. Verriez-vous un intérêt à étendre l'offre de service en offrant par exemple des articles de presse sur l'ICR ?

1.2. Alimentation de la future base

- a) Quel système serait le plus efficace d'après-vous pour signaler la parution de votre nouvel article ? (mél à la médiathèque, formulaire sur la base de données,..)
- b) Seriez-vous prêt à alimenter vous-même la base de données ?

si oui, sous quelle forme ?
si non, pourquoi ?

Conduite d'enquête

Verriez-vous d'autres personnes à interroger dans votre département ?

Annexe 5. Modèle de traitement des données de Reference Manager vers Oracle

Modèle présentant une solution possible de création de base de données présenté pour le choix d'une solution informatique.

Conception de la base de données : traitement des données avec Oracle/PHP

