

E.N.S.S.I.B
ÉCOLE NATIONALE SUPÉRIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHÈQUES

UNIVERSITÉ
CLAUDE BERNARD
LYON I

Consultation s/ place

DESS en INFORMATIQUE DOCUMENTAIRE

Mémoire de stage

Etude préliminaire pour la numérisation des
documents constituant les dossiers d' AMM.

Evelyne CHAPTAS

Sous la direction du

Docteur Jean-Claude ARNAUD
et Sophie LAFOURCADE
Institut de Recherches Internationales Servier
6 place des Pléiades
92415 Courbevoie Cedex

Année 1995-1996

1996
ED ST
18

E.N.S.S.I.B
ÉCOLE NATIONALE SUPÉRIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHÈQUES

**UNIVERSITÉ
CLAUDE BERNARD
LYON I**

DESS en INFORMATIQUE DOCUMENTAIRE

Mémoire de stage

**Etude préliminaire pour la numérisation des
documents constituant les dossiers d' AMM.**

Evelyne CHAPTAS

*Consultation
sur place*



Sous la direction du

**Docteur Jean-Claude ARNAUD
et Sophie LAFOURCADE
Institut de Recherches Internationales Servier
6 place des Pléiades
92415 Courbevoie Cedex**

1996

ED ST

18

Année 1995-1996

Etude préliminaire pour la numérisation des documents constituant les dossiers d' AMM.

Evelyne CHAPTAS

RESUME :

La reprise sous forme électronique des dossiers d'autorisation de mise sur le marché (AMM) existants répond à des besoins de sécurité et de diffusion facilité. Le choix du système à adopter est fait à la suite d'une étude préalable au projet et d'un test pilote effectué par une société de sous-traitance.

DESCRIPTEURS : (Thesaurus de Pascal)

Numérisation, Hypertexte.

ABSTRACT :

Digitizing of existing product licence application files replies to the needs of safety and easier diffusion. A preliminary study and an experimental test realized by a subcontractor permit us to choose the system we wish to install.

KEYWORDS : (Pascal Thesaurus)

Digitizing. Hypertext.

Remerciements pour leur accueil à tous les membres du D.D.S,
en particulier à Sophie Lafourcade et au Dr Arnaud.

Sommaire

INTRODUCTION	5
PARTIE I : NUMERISATION DES DOCUMENTS CONSTITUANT LES DOSSIERS D'AMM.....	6
I - LE CONTEXTE GENERAL.	6
1 - <i>Le groupe de Recherche Servier</i>	6
2 - <i>L'Institut de Recherches Internationales Servier et la D.R.D.</i>	8
3 - <i>Le Département de Documentation Scientifique et Traduction (D.D.S)</i>	8
II - LE PROJET : SAUVEGARDE ET DIFFUSION ELECTRONIQUES DES DOCUMENTS CONSTITUTIFS DES DOSSIERS D'AMM EXISTANTS.	9
III - DEFINITION DES BESOINS ET ANALYSE DE L'EXISTANT.....	9
1 - <i>Les dossiers d'autorisation de mise sur le marché (AMM).</i>	9
2 - <i>Le fond documentaire.</i>	11
3 - <i>Caractéristiques des dossiers d'AMM.</i>	11
4 - <i>Le système actuel de consultation des documents.</i>	13
5 - <i>Les utilisateurs.</i>	13
6 - <i>La faisabilité.</i>	14
IV - TEST DE FAISABILITE REALISE PAR JOUVE S.I.	14
1 - <i>Options de traitement proposées par Jouve S.I.</i>	15
2 - <i>Choix du traitement.</i>	17
3 - <i>Le traitement.</i>	18
V - ORIENTATION DU PROJET GLOBAL.	22
1 - <i>Trois scénarios possibles.</i>	22
2 - <i>Intégration dans le système.</i>	23
3 - <i>Sous-traitance.</i>	23
PARTIE II : INTERVENTIONS SUR LA BASE DE DONNEES IRISA.....	25
I - LA BASE DE DONNEES IRISA.....	25
II - CORRECTIONS EFFECTUEES.	26
BIBLIOGRAPHIE	28
ANNEXES.....	29
1 - PLAN GENERAL D'UN DOSSIER D'AMM	30
2 - DESCRIPTION DU DOSSIER D'AMM DU S6520 (FLUDEX)	31
3 - DESCRIPTION DU DOSSIER D'AMM DU S5614 (ISOMERIDE GELULES).....	34
4 - EXEMPLES DE PAGES TRAITEES PAR JOUVE S.I.	37
5 - PRESENTATION DE LA NORME NF X 66 - 022.....	55
6 - LISTE DE QUELQUES PRESTATAIRES DE SERVICE A CONTACTER.	56
7 - ABREVIATIONS UTILISEES.	57

Introduction

Les dossiers d'autorisation de mise sur le marché des médicaments (AMM) possèdent une valeur intrinsèque et juridique. Dans un souci de sécurité, de gain de place et d'une diffusion facilitée, l'Institut de Recherches Internationales Servier (I.R.I.S) a décidé de reprendre sous forme électronique les dossiers d'AMM des médicaments actuellement commercialisés par le groupe Servier. Le Docteur J.C Arnaud, responsable du Département de Documentation Scientifique et Traduction (D.D.S), m'a proposé de mener une réflexion à ce sujet afin d'aboutir à la rédaction d'un cahier des charges qui sera soumis à différentes sociétés de sous-traitance. La définition des besoins et l'analyse de l'existant effectuées, un test de faisabilité sur un nombre de pages significatif, actuellement en cours de réalisation, a été demandé à la société Jouve Systèmes d'Information.

D'autre part, le Département de Documentation Scientifique gère une base de données interne, IRISA, qui contient les références de documents portant sur les produits Servier. Dans le cadre d'une démarche de qualité, diverses corrections ont été apportées sur les enregistrements de cette base.

Partie I : Numérisation des documents constituant les dossiers d'AMM.

I - Le contexte général.

1 - Le groupe de Recherche Servier

Le groupe Servier est le premier groupe pharmaceutique français indépendant et le troisième dans le monde. Créé en 1954 par le Docteur Jacques Servier, actuel Président du groupe, le laboratoire orléanais ne reposait alors que sur une équipe de 9 employés et ne produisait que 2 ou 3 médicaments. Aujourd'hui, toujours à capitaux familiaux, le holding emploie 10 500 personnes à travers le monde et son chiffre d'affaires s'élève à 8 milliards de francs. Les objectifs affichés du groupe pharmaceutique Servier sont l'allongement de la durée de vie, la lutte contre la souffrance, et l'amélioration de la condition humaine. Satisfaire les malades et les médecins prescripteurs tout en apportant une contribution à la recherche et en s'assurant de l'épanouissement de tous les collaborateurs de l'entreprise sont les résultats recherchés par le Dr Servier pour son laboratoire.

En France, le groupe Servier est implanté dans trois régions : la région parisienne (Neuilly-sur-Seine, Courbevoie, Suresnes, et Croissy-sur-Seine), le Loiret (Orléans, Gidy et Fleury-les-Aubrais) et la Seine-Maritime (Bolbec). Il se compose de plusieurs laboratoires et sociétés dont les principales sont les Laboratoires Servier, Euthérapie, Therval, Biopharma, Ardix, Technologie Servier, I.R.I.S, Biologie Servier,...

Le problème de limitation des dépenses de santé existe en France, mais il n'est posé que dans une petite partie du monde. Présent dans 130 pays, en particulier à travers ses 60 filiales, le groupe Servier a développé une stratégie d'internationalisation maximale. Il est le quatorzième laboratoire européen et le vingt-neuvième mondial, pour les médicaments de prescription. Plus de 55 % de son chiffre d'affaires sont réalisés hors de France. Le groupe a observé une croissance de près de 15 % de ses ventes à l'international en 1995, en particulier en Europe de l'Est (Pologne, République Tchèque, Hongrie, Bulgarie) et en Asie du Sud-Est. Il est le deuxième groupe pharmaceutique français exportateur à destination des pays de l'Union Européenne et de l'Amérique du Nord et le premier à destination de l'Afrique. Servier est le seul groupe pharmaceutique français à avoir mondialisé - c'est-à-dire présent dans les sept grands pays industriels de la pharmacie - un médicament (Fludex) au cours des 15 dernières années. Parmi les 20 médicaments français actuellement admis aux Etats-Unis, quatre sont des produits Servier (Pondéral, Fludex, Coversyl et Isoméride). Deux autres produits font actuellement l'objet d'études en vue de leur enregistrement dans ce pays. Jusqu'à présent, le groupe Servier avait toujours bénéficié d'une croissance interne de l'entreprise. Mais en décembre 1995, il a réalisé la première opération de croissance externe de son histoire en acquérant 51 % du capital du premier laboratoire hongrois, Egis, permettant ainsi au laboratoire Servier de se lancer sur le marché des médicaments génériques.

Bien que le groupe Servier affiche une vocation internationale, 90 % de sa production se fait en France, à Gidy et à Bolbec. En 1995, 160 millions de boîtes ont été produites sous 1200 présentations différentes.

Une volonté de recherche intensive a conduit l'entreprise à lancer une trentaine de médicaments au cours des 30 dernières années et à synthétiser plus de 35 000 molécules. Le groupe Servier est reconnu comme l'une des maisons les plus dynamiques avec un des

meilleurs taux de progression dans le monde. Il faut noter que 25 % du chiffre d'affaires consolidé est consacré à la recherche, alors que la moyenne est de 12,5 % pour l'industrie pharmaceutique française (et de 3,5 % pour l'ensemble des industries françaises). 25 % du personnel se consacre à la recherche et développement. Ainsi, en dix ans, les sommes investies en recherche et développement ont été multipliées par cinq et ont doublé depuis 1990, assurant à long terme l'avenir du groupe. Le patrimoine de brevets s'élève à plus de 8 000 déposés, dont 7 300 à l'étranger. Aujourd'hui, le groupe intensifie ses efforts dans le développement de ses molécules (environ 37 molécules en développement) et mène quelques 300 études différentes par an. Quatre molécules font actuellement l'objet d'un développement en phase 3 (essais cliniques sur l'homme) et pourraient être lancées vers 2000.

Les efforts de recherche du groupe portent sur :

- la neuropsychiatrie (la schizophrénie et la dépression).
- les maladies dégénératives du cerveau (maladies d'Alzheimer et de Parkinson).
- la cancérologie.
- la pathologie cardiovasculaire.
- les maladies métaboliques (le diabète et l'obésité).
- la pathologie osseuse (l'arthrose et l'ostéoporose).

Le siège social du groupe est situé à Neuilly tandis que la recherche s'organise à partir de Suresnes. Les centres de Recherche et de Développement en France sont situés à Suresnes (Chimie, Pharmacologie), Croissy-sur-Seine (Pathologie cérébrale, Neuropsychopathologie, Neurobiologie), et Orléans (Pharmacocinétique, Toxicologie, Cancérologie, Tératologie, Galénique). Après l'ouverture du centre de recherche de Croissy-sur-Seine en 1994, une unité d'appui clinique a été inaugurée à Gidy en juin 1996. D'un investissement de 50 MF, elle est destinée à produire plus d'un million de boîtes par an pour l'expérimentation clinique en double aveugle.

Le groupe possède également sept centres de Recherche et Développement à l'étranger (les CIRT, Centre International de Recherche Thérapeutique), en particulier en Grande-Bretagne et en Australie. En outre, Servier entretient des relations étroites avec les universitaires, le CNRS et l'INSERM.

Actuellement 19 médicaments sont commercialisés par le groupe. Huit médicaments réalisent 80 % des ventes :

- Diamicon, un antidiabétique oral qui réalise 20 % des ventes mondiales du groupe.
- Coversyl, un inhibiteur d'enzyme de conversion, pour lutter contre l'hypertension artérielle. Présent dans 80 pays, il réalise 13 % du chiffre d'affaires.
- Vastarel, un traitement prophylactique de l'angine de poitrine. Enregistré dans 70 pays, il représente 12 % du chiffre d'affaires.
- Fludex, un antihypertenseur diurétique (12 % du chiffre d'affaires). Ce produit est en tête aux Etats-Unis avec 40 % du marché des diurétiques.
- Daflon 500mg, contre l'insuffisance veineuse (12 % du chiffre d'affaires).
- Isoméride, contre l'obésité (6 % du chiffre d'affaires).
- Hypérium, contre l'hypertension artérielle (3 % du chiffre d'affaires).
- Muphoran, pour le traitement du mélanome malin (1 % du chiffre d'affaires).

Le groupe compte sur le lancement en France l'année prochaine d'un autre antihypertenseur, baptisé le Biprel (demande d'AMM France déposée en été 1996).

Exclusivement intéressé depuis toujours par les spécialités dites de prescription (sur ordonnance), le groupe n'exclue plus la possibilité de se lancer dans les "OTC", les médicaments en vente libre dans les officines, qui nécessitent peu de dépenses de développement.

2 - L'Institut de Recherches Internationales Servier et la D.R.D.

L'Institut de Recherches Internationales Servier (I.R.I.S) est installée à Courbevoie depuis près de 6 ans, sur le site dit de " Fontanes ". Il comprend essentiellement les services de la D.R.D, la Direction Recherche et Développement. Environ 600 personnes travaillent sur ce site, dont près de la moitié pour I.R.I.S. Les principaux départements sont les différentes Divisions Thérapeutiques, les départements des Affaires Réglementaires, de la Planification, de la Pharmacovigilance, de la Communication Scientifique Internationale, de la Documentation Scientifique et de l'Information Scientifique.

3 - Le Département de Documentation Scientifique et Traduction (D.D.S).

Le Département de Documentation Scientifique et Traduction, dépendant de la D.R.D, est dirigé par le Docteur Jean-Claude Arnaud. Il comprend 15 personnes réparties en deux services : le service de la Documentation Scientifique et celui de la Traduction. Le D.D.S, est totalement indépendant de la Bibliothèque qui relève directement de la Direction Générale. Ainsi, il réalise des tâches différentes des activités de la Bibliothèque (il ne gère pas d'ouvrages, ni de revues).

Le D.D.S assure la gestion de l'ensemble des informations concernant les molécules Servier. Il assure l'entrée en Documentation des rapports d'études issus de la R & D et des publications. Les Responsables Produits ont pour mission de répondre aux demandes d'information et de recueillir des informations au sujet des produits Servier. Le D.D.S transmet des copies des documents aux demandeurs. Il interroge les banques de données internationales sur des questions biomédicales ponctuelles ou selon des profils personnalisés, le plus souvent à la demande du département de l'Information Scientifique. Des banques de données internationales sur CD-ROM (Medline, Europharm biblio, British Pharmacopeia, Pharmaprojects, Serline,...) sont mis à la disposition de certains départements sur le réseau local. Deux bases de données internes sont gérées par le D.D.S : DOCTHER, une base de données bibliographique alimentée par les utilisateurs, et IRISA, la base des produits Servier.

Le service de Traduction du Département assure la traduction de documents à la demande du département des Affaires Réglementaires ou des Divisions Thérapeutiques.

Le public client du D.D.S est large, mais il est essentiellement représenté par les personnes des Unités de Recherche (les Divisions Thérapeutiques), de l'Information Scientifique, des Affaires Réglementaires et de la Promotion France et Internationale des sites de Courbevoie et de Neuilly. Les CIRT ont également recours aux services du D.D.S.

Les produits du D.D.S sont IRIS-EXPRES, qui regroupe deux fois par mois les références des nouvelles publications concernant les produits Servier, et des bibliographies complètes sur chaque produit. A ces productions écrites, on doit ajouter une diffusion de l'information plus personnelle, par téléphone ou en personne, sur divers points (aide à la recherche dans IRISA,...).

II - Le projet : Sauvegarde et diffusion électroniques des documents constitutifs des dossiers d'AMM existants.

Les dossiers d'AMM en cours de rédaction, comme ceux qui suivront, sont construits directement sous forme électronique, à partir de divers logiciels de bureautique. Les dossiers plus anciens n'existent que sous une forme papier. Il serait cependant intéressant de reprendre sous une forme électronique ces derniers.

Les dossiers d'AMM originaux sont des documents de référence qui n'existent qu'en un seul exemplaire. Ainsi, un incendie ou une inondation causeraient une perte irréparable. La valeur des dossiers nous oblige à les préserver en diminuant leur manipulation qui peut entraîner des dégradations, des pertes ou des déclassements. D'autre part, les dossiers d'AMM étant confidentiels, il serait utile d'en contrôler l'accès par un module informatique.

Ces dossiers sont volumineux, et souvent consultés. Une optimisation de la diffusion des documents, par accès immédiat et partagé, est devenue nécessaire pour alléger les tâches des archivistes et diminuer le nombre grandissant des photocopies.

Un gain de productivité peut être espéré au niveau de l'exploitation des documents électroniques en les considérant comme de véritables outils de travail.

Un gain de place dans les locaux actuels pourrait être réalisé dans l'éventualité d'une délocalisation des archives.

L'obligation légale de conserver les dossiers d'AMM originaux sous forme papier (authentification des signatures) signifie que la mise en place d'un système de gestion électronique des dossiers d'AMM existants vise à améliorer le système actuel, mais pas à le remplacer.

Les utilisateurs visés par ce système sont les personnes qui consultent les dossiers : les personnes du département des enregistrements (les Affaires Réglementaires), les membres des différentes Divisions Thérapeutiques, et l'Agence du Médicament.

La reprise des dossiers d'AMM sous forme électronique sera confiée à un prestataire de services.

III - Définition des besoins et analyse de l'existant.

1 - Les dossiers d'autorisation de mise sur le marché (AMM).

Généralités.

La recherche permet de découvrir et de mettre au point de nouveaux moyens thérapeutiques nécessaires pour le traitement et la prophylaxie des maladies. Sept à dix ans s'écoulent entre l'isolement de la molécule et la sortie du médicament sur le marché. Pour 10 000 nouvelles molécules, 20 subiront les tests précliniques, 10 subiront les tests cliniques chez l'homme, et seule l'une d'entre elles obtiendra une autorisation de mise sur le marché. La R & D est de plus

en plus coûteuse, surtout dans les phases de développement clinique. En vingt ans, les délais nécessaires à la commercialisation d'une molécule ont été multipliés par deux et les coûts par trois.

Un dossier d'AMM du nouveau produit est élaboré et déposé auprès de l'Agence Française du Médicament, un établissement public de l'Etat créé en 1993. L'autorisation de mise sur le marché est délivrée après étude par une commission technique formée d'experts médicaux (rapporteurs désignés parmi les chercheurs des établissements de recherche et parmi des hospitalo-universitaires), qui donne son avis sur la qualité, l'efficacité et la sécurité d'emploi du produit. Accordée pour cinq ans, l'autorisation de mise sur le marché est renouvelable après diverses études. Les dossiers d'AMM sont mis à jour par ajouts de nouveaux documents, et non par l'évolution des documents existants.

L'Agence Européenne du Médicament est chargée depuis le 1er janvier 1995 de délivrer des AMM uniques pour tous les pays de la C.E.E.

Les dossiers sont construits par le département des Affaires Réglementaires des laboratoires pharmaceutiques selon un plan et des exigences précis variant d'un pays à un autre. En France, les dossiers d'AMM doivent comprendre cinq parties distinctes¹ dont chacune est constituée d'un ou plusieurs documents.

Soumission électronique des nouveaux dossiers d'AMM.

En 1995, l'Agence du Médicament a reçu 760 demandes d'AMM. Chaque dossier peut peser 1 à 2 tonnes, ce qui donne une idée de l'importance d'informations à manipuler. Le dossier d'AMM du médicament subit des mouvements au cours de sa vie : l'Agence du Médicament recense 3850 demandes de modifications par an. L'Agence n'a que 210 jours pour instruire le dossier, ce qui l'oblige à aller très vite sur les parties intéressantes du dossier. D'autre part, les industriels ont conscience que plus vite un dossier est réalisé et traité par l'Agence, plus vite le laboratoire conquiert des parts de marché et répond aux urgences de santé publique. La gestion électronique de documents va permettre de raccourcir la chaîne éditoriale de rédaction des dossiers d'AMM.

Pour ces diverses raisons, l'Agence du Médicament et les professionnels de l'industrie pharmaceutique (le SNIP, Syndicat National de l'Industrie Pharmaceutique) réfléchissent à la création du nouveau format d'échange électronique des dossiers : c'est le projet Sedamm (Soumission électronique des autorisations de mise sur le marché). Ils proposent aux laboratoires implantés en France de structurer leurs dossiers d'AMM au format de description HTML (Hyper Text Markup Language). HTML a été choisi comme un standard compatible avec des logiciels de navigation, comme Netscape, que l'on peut récupérer facilement sur Internet. La soumission électronique des dossiers complets d'AMM sur CD-WORM devrait s'opérer dès 1998.

D'autre part, l'Agence Européenne du Médicament, la France, la Grande-Bretagne et l'Allemagne réfléchissent au niveau européen à la soumission électronique des dossiers d'AMM. Le projet surnommé Mansev (Market authorization for network submission and evaluation) devrait inciter les agences des différents pays de l'Union à opter pour le stockage électronique des dossiers d'AMM.

Aux Etats-Unis, les premières soumissions électroniques à la FDA (Food and Drug Administration, l'agence américaine de contrôle des aliments et des médicaments) ont été introduites au milieu des années 80 selon le programme CAPLA (Computer Assisted Product License Application).

¹Cf Annexe 1

En France, un certain nombre de laboratoires pharmaceutiques, tels Sandoz et Pierre Fabre, réfléchissent sérieusement à la soumission électronique des nouveaux dossiers d'AMM. Servier, pour sa part, a récemment déposé la partie I du dossier d'AMM du Biprel sous forme de pages HTML. Cependant, il semble qu'il y ait peu de projets de reprise sous forme électronique des anciens dossiers, en dehors d'un simple objectif de stockage.

2 - Le fond documentaire.

La documentation gérée par le D.D.S est stockée au quinzième étage, le dernier du site de Courbevoie. Elle est répartie dans deux salles distinctes, une qui réunit les documents publiés (PU), l'autre qui regroupe les documents non publiés (NP). Les documents sont classés dans des armoires selon le code que leur a attribué automatiquement le système de la base IRISA. L'éventualité d'un déménagement des archives constitue un paramètre qui peut être pris en compte pour la réalisation du projet.

Les documents publiés sont des articles, abstracts, actes de congrès, posters, chapitres d'ouvrage,...

Les documents non publiés sont les rapports d'études, les résumés de rapports, les rapports d'experts, les rapports de chimie-pharmacie, les résumés des dossiers d'AMM, les brochures d'investigateurs, des notes, des monographies et des thèses.

Seuls certains rapports d'études, les rapports d'experts, les résumés de dossiers d'AMM, les rapports de chimie-pharmacie, et certaines publications composent les dossiers d'AMM.

Le contenu des documents est du domaine de la pharmacologie, de la chimie, de la toxicologie, de la pharmacocinétique, de la pharmacodynamie, ou des études cliniques.

Il y avait en août 1996, selon la base IRISA, 17500 publications et 4200 documents non publiés. Le 21 mars 1996, ce dernier chiffre s'élevait à 3879 documents, ce qui représentait près de 270000 pages (pour le corps des rapports uniquement).

Les documents concernés par la numérisation.

Des priorités de mise en oeuvre ont été définies. Il s'agit de reprendre les différents rapports composant les dossiers d'AMM déposés en France, dans leur version originale et définitive, en écartant les annexes et les publications insérées. Nous commencerons le traitement par les dossiers les plus consultés. Le choix de dossiers à traiter en priorité reste à faire.

Il est difficile de connaître le nombre exact de pages à traiter sans reprendre un par un les plans des dossiers et, document par document, comptabiliser les pages hors annexes. En effet, la base de donnée IRISA ne distingue pas le nombre de pages des annexes de celles du corps du document.

3 - Caractéristiques des dossiers d'AMM.

Les rapports actuels sont rédigés selon des standards qui formalisent la structure des documents. Mais les dossiers plus anciens ont été rédigés selon les inspirations de leurs auteurs. Il y a donc une très grande hétérogénéité de la composition des dossiers d'AMM.

Un dossier d'AMM est constitué d'un ensemble de documents multipages de quatre types : les résumés du dossier, les rapports d'études, les rapports d'experts, et des articles de journaux. Les documents sont confidentiels, définitifs et donc non modifiables. Cependant, un dossier d'AMM est régulièrement mis à jour ; certains documents sont enregistrés en complément d'un dossier d'AMM après le dépôt de celui-ci.

Chaque document est référencé dans la base IRISA. Il possède un code alphanumérique à 16 caractères (cote attribuée par IRISA) et un code-barre qui lui sont propres, apposés sur la page de garde. Un document est structuré, il peut comporter plusieurs volumes et peut être traduit en plusieurs langues. Les annexes sont insérées en fin de volume 1 ou constituent un ou plusieurs volumes séparés. Un document peut faire partie de plusieurs dossiers d'AMM.

Pour définir les caractéristiques des dossiers d'AMM déposés par le groupe Servier, deux dossiers ont été étudiés en particulier : celui du Fludex, le dernier déposé (1995), et celui plus ancien de l'Isoméride (1985). Ni les annexes, ni les publications n'ont été prises en compte.

La méthodologie adoptée a été la même pour les deux études : A partir du plan du dossier d'AMM comportant la liste des documents déposés, et la liste des documents ajoutés en complément du dossier, il a fallu retrouver le code IRISA correspondant à chaque document. Un jeu de photocopies de chaque document a été demandé pour ne pas travailler directement sur les originaux. En prenant les documents un par un, page à page, il a été possible de comptabiliser le nombre de pages réel à numériser et d'identifier les composants de chaque document. Les tableaux, les figures, les iconographies, les formules chimiques et mathématiques, les signatures et les logos, les cartouches, ainsi que les en-têtes et les pieds de page ont été comptés un à un. La qualité d'impression et la présence de couleurs ont été appréciées sur les originaux. Un échantillon de pages pris au hasard dans chaque document - un total de 125 pour le dossier du Fludex, 200 pour le dossier de l'Isoméride - a été analysé pour évaluer le nombre moyen de caractères par page, de caractères spéciaux, de gras, d'italiques, et de soulignés. Ceci en conformité avec la norme NF X 06-022 (Sélection de plans d'échantillonnage pour le contrôle par comptage de la proportion d'individus non conformes ou du nombre moyen de non-conformité par unité)².

Deux documents récapitulatifs de trois pages reprennent les résultats de ces analyses.³

Un aperçu rapide des autres dossiers d'AMM ajouté à ces analyses nous permet de dégager certaines observations quant à la structure des dossiers :

Le support est constitué de feuilles de papier A4, généralement non reliées et imprimées d'un seul côté. La qualité d'impression est très variable, en particulier à cause des machines à écrire utilisées pour la rédaction des documents plus anciens. Le contenu des annexes est varié : certains rapports n'en possèdent pas, tandis que certains auteurs mettaient toutes les données numériques, tableaux et figures en annexes. Certains documents ne possèdent pas de table des matières. Les langues utilisées sont le français en large majorité, suivi de l'anglais. On observe une hétérogénéité dans la pagination des documents : la majorité des pages possède une double pagination et les premières pages ne sont pas toujours numérotées, ni même comptées. Il y a de très nombreux tableaux et figures, surtout dans les documents récents, alors que les planches photographiques sont rares. Certains rapports, en particulier les plus anciens, sont signés en bas de chaque page.

² Cf Annexe 5.

³ Cf Annexes 2 et 3.

4 - Le système actuel de consultation des documents.

Le principe est le même qu'il s'agisse d'un document publié ou non. Les utilisateurs internes font une recherche dans la base IRISA (consultation sur PC ou console Vax) pour identifier les documents qui répondent à leurs besoins. A partir des références des documents sélectionnés (cote, volume ou titre), ils font une demande de sortie de document au D.D.S situé au deuxième étage du site de Courbevoie. La demande s'effectue en personne ou par téléphone. Une fiche de sortie du document (3 volets) est remplie par le demandeur lui-même ou par la personne qui accueille la demande. Le demandeur garde le troisième volet de la fiche. La demande doit être validée par la personne responsable des sorties de documents qui appose sa signature et un tampon. Une archiviste du D.D.S prend note de la demande et sort les documents stockés au quinzième et dernier étage. Un ou plusieurs jeux de photocopies, selon la demande, sont effectués au service de reprographie au dixième étage avant que l'original soit reclassé aux archives. Les photocopies sont alors transmises au demandeur, en mains propres ou par courrier interne, accompagnées du deuxième volet de la fiche de sortie. L'utilisateur garde les photocopies aussi longtemps qu'il estime en avoir besoin, puis il les détruit. Il peut aussi bien les stocker dans une de ses armoires.

Les utilisateurs externes, dépendant de l'Agence du Médicament, qui demandent à consulter certaines parties d'un dossier s'adressent au service des Affaires Réglementaires de Servier qui fera la commande de photocopies.

Le temps de satisfaire une demande nécessite moins de deux jours, selon la disponibilité des archivistes et la quantité de photocopies à effectuer. Certaines commandes urgentes sont traitées en priorité. Un même document peut être demandé plusieurs fois dans la même semaine par des personnes différentes. Un utilisateur peut demander plusieurs fois le même document à des périodes de temps rapprochées.

5 - Les utilisateurs.

Typologie des utilisateurs.

Les utilisateurs internes sont répartis sur les différents sites du groupe. Ils ont accès au réseau interne de l'entreprise, géré par un système d'exploitation Vax.

- Les personnes du département des Affaires Réglementaires (environ 40 personnes) qui se réfèrent aux dossiers d'AMM France originaux pour déposer soit des dossiers d'AMM pour de nouveaux produits en France, soit des dossiers d'AMM dans les pays étrangers. Elles sont réparties en quatre groupes selon le secteur géographique dont elles s'occupent.

- Les membres des différentes Divisions Thérapeutiques qui rédigent de nouveaux rapports d'études. Ils puisent des informations dans les anciens rapports pour leur éviter de refaire une étude, ou pour simplement accélérer la chaîne de rédaction des rapports.

- Bien qu'il ait déjà accordé l'autorisation de mise sur le marché des produits considérés, l'Agence du Médicament peut demander à consulter certaines parties d'un dossier, en particulier dans le cadre des revalidations.

Les besoins des utilisateurs.

Le fait d'avoir des utilisateurs externes à l'entreprise oblige la mise en place d'un système standard et évolutif.

Il doit être compatible avec les machines PC et MacIntosh, dans un environnement Windows. Selon les utilisateurs, tous les dossiers d'AMM méritent d'être repris sous forme électronique.

Le système devra offrir aux utilisateurs des services de recherche, de visualisation, de navigation dans le document et d'édition.

L'accès au document se fera par recherche à partir de la cote du document, ou à partir d'un menu reprenant le plan du dossier. L'accès aux documents sera immédiat (quelques secondes au plus) et géré par un code confidentiel.

La visualisation devra faire apparaître un texte lisible accompagné des différents éléments que contient le document original (tableaux, graphiques, formules,...) à leur place initiale. Le document renvoyé devra être aussi proche de l'original que possible.

A l'intérieur d'un document, la navigation doit être facilitée par un déplacement page à page, par bonds à partir de la table des matières du document, par recherche sur un mot ou une expression.

Il devra être possible de copier certaines parties de texte, de tableaux ou de graphiques pour une autre utilisation informatique. L'impression du document entier ou de certaines parties du document sera possible.

6 - La faisabilité.

Les documents qui constituent les dossiers d'AMM sont clairement identifiés et sont déjà enregistrés dans une base de données. Les dossiers sont organisés et réunis dans un même lieu. Les documents sont fréquemment utilisés, parfois en même temps par plusieurs personnes. Pour des problèmes de place, il est impossible de gérer des photocopies en plus des documents originaux. Leur diffusion est volumineuse et nécessite un trop grand nombre de photocopies. La diffusion accélérée et facilitée des documents est devenue nécessaire, en particulier si on projette un déménagement des archives.

Le projet de stockage des dossiers d'AMM sous forme électronique est envisagé depuis plusieurs années, ainsi la Direction et les utilisateurs sont sensibilisés et motivés par cette application.

Afin de savoir si le projet est viable, pour estimer l'étendue de la tâche à accomplir et pour faire des projections budgétaires pour l'ensemble du projet, nous avons demandé à la société Jouve Systèmes d'Information de réaliser un test pilote, de grande nature, sur un nombre de pages significatif.

IV - Test de faisabilité réalisé par Jouve S.I.

Nous avons demandé à la société Jouve S.I. de réaliser un test pilote en grande nature sur un dossier d'AMM comportant un nombre de pages significatif. Ce prestataire de services a été choisi pour ses références (il a effectué un travail semblable pour l'Office Européen des Brevets).

Pour le test, nous avons sélectionné le dossier d'AMM du Fludex déposé en 1995. Le dossier n'est pas représentatif de l'ensemble du fond documentaire, mais il est représentatif des dossiers d'AMM rédigés à la même époque. L'analyse approfondie du dossier a mis à jour un certain nombre de difficultés qu'il nous semblait intéressant d'affronter pour savoir ce qu'on pouvait

espérer comme résultat final. Nous avons donc fourni des photocopies du dossier, représentant 2370 pages sans les annexes.

1 - Options de traitement proposées par Jouve S.I.

La société Jouve S.I. nous a proposé 5 options de traitement. Elle nous a livré par ailleurs des fichiers de démonstration, réalisés à partir de quelques pages du dossier qui leur a été confié, pour nous aider à faire un choix parmi les options proposées.⁴

Option A : Simple numérisation du dossier pour une sauvegarde électronique.

Ce traitement permet d'obtenir un fac-similé de l'original. La scannérisation est réalisée avec une résolution de 300 dpi. Les images bitmap sont enregistrées sous un format TIFF, compressé selon le mode CCITT groupe IV. Les normes CCITT groupe III et IV sont les normes internationales de compression des images, édictées par le CCITT (Comité Consultatif International du Télégraphe et du Téléphone) et définies à l'origine pour la télécopie. Très utilisées pour les transmissions distantes, elles garantissent aucune perte d'information. Le format TIFF (Tagged Image File Format) est un format flexible, apte à s'adapter à différentes plates-formes, différents types d'images, de taille et de résolution différentes. Il est plus compact et d'une qualité équivalente au Postscript. Les photographies en noir et blanc sont numérisées en niveaux de gris. Les photographies couleurs sont stockées au format JPEG, le standard ISO dédié à la compression d'images fixes couleurs.

Un traitement visuel est exercé sur chaque page numérisée, ainsi que le traitement de calcul d'inclinaison et de redressement des pages. Il est possible de produire un fichier TIFF multipage ou un fichier TIFF par page du document. L'indexation du fichier se fait avec le code-barre qui identifie chaque document dans la base IRISA, présent sur la page de garde du document. Cependant, le code contient 16 caractères alors que le nom d'un fichier ne peut dépasser 8 caractères.

Jouve S.I. nous a fourni un shareware, Paint Shop Pro, pour visionner 32 fichiers TIFF correspondant à 32 pages de la partie II du dossier. La taille moyenne de ces fichiers est de 21,5 Ko par page. La restitution à l'écran, ainsi que leur impression, s'est avérée trop lente (peut-être un problème du poste de consultation). Les fichiers images ne sont pas agréables à lire à l'écran, l'utilisation du zoom ou de rotation permet une meilleure lecture mais la taille de la fenêtre ne permet alors plus l'affichage de la page complète. L'impression avec une imprimante laser retourne une qualité semblable à celle d'une photocopie. Le déplacement dans le document se fait de page en page, et à part la lecture, aucune exploitation du texte n'est possible.

Option B : Numérisation suivie d'une océrisation.

L'application des techniques d'OCR/ICR aux fichiers images obtenus par scannérisation nous renvoie un flux ASCII du texte.

Le format ASCII est le plus simple et le plus compact, mais la mise en page et la mise en valeur du texte sont très limitées. Toutefois, il est possible de tenir compte de la typographie du texte original et les éléments images (tableaux, figures, formules,...) peuvent ensuite être insérés à leur place. Certains tableaux contenant du texte peuvent être traités par océrisation afin d'en obtenir également un flux textuel.

Jouve S.I. nous a proposé une précision de saisie de 99,98 % garantie.

Aucun fichier de démonstration nous a été fourni.

⁴ Cf Annexe 4.

Option C : Numérisation et océrisation, suivie d'une structuration au format HTML ou SGML. Le flux texte ASCII provenant de l'étape de saisie OCR/ICR subit une restructuration logique. Pour cette option, une DTD (Document Type Definition) est développée afin de normaliser les documents. La DTD est la grammaire propre à une classe de documents. Elle décrit, avec des balises, la structure hiérarchique d'un document, de la plus grande entité (le document) à la plus petite (un mot).

Contrairement au format SGML (Standard Generalized Markup Language), le format HTML (Hyper Text Markup Language) permet la présence de liens hypertextes (possibilité de passer d'un sujet à un autre par activation d'un lien).

Un fichier HTML de démonstration, correspondant à deux pages du dossier, nous a été envoyé. Nous l'avons lu avec Netscape Navigator 2.01 (Navigator est disponible gratuitement sur <http://home.netscape.com>). Le fichier ne comprenait ni liens d'ancrage (figures, tableaux, molécules), ni liens internes, ni liens externes. La taille moyenne du fichier est de 1,8 Ko par page. La présentation du document, notamment la typographie, n'a plus rien à voir avec celle de l'original. Les tabulations, les sauts de pages ou de lignes ne sont pas respectés. L'exploitation du texte est possible en effectuant une recherche sur une chaîne de caractères ou de recopier certaines parties du texte. Il est impossible de faire une recherche sur le numéro de page. Le texte repris est bien lisible à l'écran et son impression est rapide.

Option D : Numérisation et océrisation suivie d'une génération de pages PDF.

Les étapes de numérisation et de saisie par les techniques d'OCR/ICR sont réalisées comme précédemment. Le format PDF restitue des pages aussi proches que possible des pages originales par leur présentation. Les pages PDF (Portable Document Format) peuvent comporter des liens entre le sommaire et le corps du document. D'autre part, il est possible de réaliser des fichiers structurés si une structuration SGML ou HTML a été effectuée auparavant. Le format PDF est un format d'échanges de fichiers entre ordinateurs, réalisables quelques soient le logiciel et la plateforme utilisés pour les créer.

Jouve S.I. nous ont fait parvenir 2 fichiers PDF, dont l'un comporte des liens internes.

Pour visualiser, parcourir et imprimer ces fichiers, le logiciel Acrobat Reader 2.0 - disponible gratuitement sur le serveur d'Adobe: <http://www.adobe.com/> - a été téléchargé.

Les pages renvoyées à l'écran et imprimées sont très proches des photocopies d'originaux. Le déplacement dans le fichier peut se faire de page en page ou grâce aux hyperliens. Il est possible de travailler sur le texte en recherchant un mot ou une expression, en copiant un passage ou une page entière, un tableau ou une figure. Le document est protégé, il est impossible de couper ou d'ajouter du texte. L'outil de visualisation utilisé nous permet, par un système de mosaïque, de voir plusieurs pages en même temps (mais de taille réduite). Quelques caractères spéciaux ont été mal saisis en raison d'un mauvais paramétrage du logiciel.

La taille moyenne des fichiers PDF transmis est de 4,8 Ko par page.

Option E : génération de fichiers Word structurés.

Le traitement consiste en une numérisation et une océrisation suivies d'une structuration SGML et d'une génération de fichiers Word structurés.

Nous avons lu avec Word 6 deux fichiers .doc de démonstration qui nous ont été donnés. La présentation du document diffère de celle des pages originales. Ni les logos, ni les code-barres ne sont repris. Les polices, les tabulations, les passages de lignes et les sauts de pages sont différents. Il n'y a pas de liens dans le document, ce qui n'autorise comme navigation dans le document que le page à page et la recherche sur un mot ou une expression. La recherche sur la pagination n'est pas possible. Les fichiers n'ont pas été protégés, il est donc possible de copier, de couper et de rajouter du texte.

2 - Choix du traitement.

D'après nos besoins, le devis proposé et la lecture des fichiers démonstratifs, nous avons opté pour une structuration du document au format HTML.

Les solutions de numérisation et de génération de flux texte ASCII ne nous paraissent pas suffisantes pour une pleine exploitation du texte. La création de fichier PDF aurait entraîné un dépassement du budget prévu pour ce test de faisabilité. Quant à la structuration au format Word, celui-ci étant un format propriétaire bien que très répandu, cette solution a été rapidement écartée.

Les formats de structuration.

Le format SGML (Standard Generalized Markup Language) est un modèle de structure logique pour coder l'information (texte, images, graphiques) défendu par le monde de l'édition et qui s'est imposé comme une norme internationale pour les échanges de documents (norme ISO 8879). Elle assure l'indépendance des informations vis à vis des matériels et des logiciels. SGML ne permet pas les liens hypertextes : c'est un standard qui s'applique aux textes lisibles séquentiellement. Les coûts du traitement et des logiciels, les analyseurs SGML, sont plus élevés que ceux des pages HTML. Cependant, le codage SGML est le plus économique en volume d'information stockées et transmises.

HTML (Hyper Text Markup Language) est un mode de représentation d'hyperdocuments défini selon le standard SGML. C'est donc un standard de fait. Le modèle HTML est étroitement associé à Internet qui est en pleine extension, il peut toutefois être utilisé indépendamment du web, dans une application locale. HTML est une application limitée du SGML, sa structure est donc simple avec un nombre réduit de balises. Le format des images (JPEG, GIF,...), des polices et des enrichissements sont prédéfinis. Les fichiers HTML ne sont guère plus gros que les fichiers ASCII correspondants. C'est un standard compatible avec les logiciels de navigation que l'on peut récupérer facilement sur Internet, d'où une facilité de consultation et à moindre coût. Les limites du HTML sont un obstacle au workflow où l'information est sans cesse remise à jour (impossible de modifier les documents natifs ni l'arborescence des fichiers sans perdre les liens). HTML est idéal pour les informations de référence stockées sur média non réinscriptibles. Il est possible d'évoluer vers d'autres standards de structuration, notamment vers le SGML, à partir d'un document HTML.

D'autres formats ont été écartés : ODA (Office Document Architecture), norme ISO 8613, qui n'est pas destiné aux grandes masses d'informations et qui est plus axé vers la création et la mise à jour des documents. CALS (Computer-aided Acquisition and Logistics Support) est une norme de structuration de la documentation technique selon la logique SGML, émanant du département de la Défense américain. HyTime (ISO/CEI 10744) est une notion apparue en 1992 autour de SGML pour définir son extension au multimédia et pour assurer un codage de la synchronisation temporelle (description de structures musicales). C'est généralement une application locale, qui n'est pas orientée vers le web.

Le support de stockage.

Les technologies de stockage optique, surtout de type WORM (Write Once, Read Many), ont atteint une fiabilité et une pérennité beaucoup plus grandes que les technologies magnétiques : une tête de lecture optique est mille fois plus éloignée de la surface qu'une tête magnétique, la

démagnétisation ou la contamination par un virus sont impossibles. Pour l'archivage à partir de gros systèmes, on utilise de préférence les DON ou la technologie WORM.

Le CD-ROM (Read Only Memory) et le CD-WORM sont des supports normalisés sur le plan physique et logique par la norme ISO 9660 (High Sierra). Les disques sont formatés de manière indépendante du système d'exploitation, PC, MacIntosh ou Unix. Le disque pourra éventuellement contenir une version de l'application spécifique à chaque système d'exploitation, mais les données seront communes. Les inconvénients de cette norme sont que l'arborescence des répertoires ne peut comporter que 8 niveaux, sans extension au nom des sous-répertoires, et que le nom des fichiers ne peut dépasser 8 caractères. Ces deux supports ont une capacité de stockage de 650 Mo, un format de 4,72 pouces (12 cm), et un faible coût (100 F environ pour un CD-WORM selon sa qualité). Leurs enregistrements sont irréversibles. Le CD-ROM est un support pré-enregistré en usine (gravure mécanique), généralement en de nombreux exemplaires.

Le CD-WORM peut être enregistré en une seule fois ou sous forme de plusieurs sessions (le prix d'un graveur de CD-WORM est maintenant inférieur à 10 000 F). Il peut être facilement dupliqué et diffusé pour un coût très abordable. Il est destiné aux éditions limitées, de 2 à 20 exemplaires, et peut être lu sur n'importe quel lecteur de CD-ROM (prix inférieur à 1 000 F). La durée de vie des disques compacts est supérieure à 30 ans. Le CD-WORM de haute capacité - le CD-HD de 4,7 Go sur une face et 9,4 Go sur deux faces - est annoncé pour 1998. Les DON-WORM (Disques Optiques Numériques - Write Once Read Many) ont une capacité de stockage supérieure aux CD mais sont plus coûteux (640 F pour un disque de 1,3 Go). Il existe en format 12" et 5"1/4. Un lecteur spécifique doit être utilisé (un lecteur-enregistreur vaut 15 000 F minimum). Il est impossible d'effacer les données en réécriture et leur durée de vie dépasse 30 ans. C'est un support inscriptible au fur et à mesure mais qui n'est pas normalisé. Cependant, il a une valeur légale dans certains pays (les Etats-Unis). Les DON ont un temps d'accès plus rapide que les CD-WORM car ils sont découpés en secteurs.

Le choix du CD-WORM a été fait car c'est un support normalisé, non effaçable, inscriptible en plusieurs fois (ce qui permettra l'ajout de nouveaux documents au dossier), peu coûteux, qui nécessite un simple lecteur de CD-ROM, et qui est préconisé par l'Agence du Médicament. Le DON-WORM a été détrôné par le CD-WORM qui s'affirme comme une solution de l'avenir.

La gestion de plusieurs CD-WORM peut se faire par un autochangeur (5 ou 6 média seulement), une tour d'enregistreurs/lecteurs, ou un juke-box (librairie optique). La capacité de ce dernier varie de 10 Go à plus d'un Téra octet (To). Les temps d'accès sont parfois un peu longs, selon le débit du lecteur. Assez coûteux, il n'est peut-être pas nécessaire pour moins de 100 000 pages. Sa mise en place peut être prévue dans une seconde phase du projet.

3 - Le traitement.

Les documents seront stockés selon deux modes : sous forme d'images TIFF et sous forme de pages structurées HTML. Le traitement se fait en deux étapes :

Les étapes du traitement.

- La préparation des documents, réalisée par le D.D.S.

Pour chaque dossier d'AMM à traiter, il nous faut d'abord retrouver le plan du dossier déposé et obtenir la liste des documents complémentaires. Identifier les documents cités, vérifier qu'ils

existent bien dans la base IRISA et noter les nouveaux codes si ils ne figurent pas sur le plan. Nous devons vérifier qu'il ne s'agit pas de traductions ni de versions obsolètes. Il faut faire la liste des documents et des volumes retenus, après avoir éliminé les publications et les annexes. On doit demander des photocopies de chaque document et vérifier la qualité de ces photocopies (nettes, droites, et sans agrafes).

- La phase de production, réalisée par le sous-traitant.

On doit trier les documents et supprimer les annexes internes aux documents.

Les documents sont scannés selon une résolution de 300 dpi pour obtenir une image bitmap des pages imprimées. Contrôler la qualité de la scannérisation, re-scannériser ou faire un redressement des pages éventuellement. Vérifier que le nombre de pages scannérisées correspond au nombre de pages attendu. Création d'un fichier image TIFF par document. Les fichiers sont compressés suivant le mode CCITT groupe IV. Le fichier est indexé à partir du code du document (code-barre). Les photographies en noir et blanc sont traitées en niveau de gris.

Le fichier image est segmenté pour séparer le texte des éléments visuels associés. Les images des figures, graphiques, tableaux,... sont stockées sur des fichiers séparés (un fichier GIF par élément). Les photographies couleurs sont codées en JPEG. Le texte est saisi par la technologie de l'OCR. Les code-barres sont également saisis par la technologie de l'ICR (reconnaissance intelligente de caractères). Le texte est corrigé à l'aide de dictionnaires.

L'écriture d'une DTD précède la structuration du texte au format HTML 3.0. Les liens hypertextes sont créés à l'intérieur du document, en particulier les liens reliant les titres à la table des matières. Les liens d'ancrage qui permettent l'insertion des éléments non textuels à leur place initiale sont ajoutés.

Les différents fichiers (.html, .tif, .jpeg, .gif) sont stockés provisoirement sur disque dur avant leur enregistrement sur CD-WORM.

Points particuliers de la demande.

Quelques particularités ont été demandées.

A un document physique (un volume) correspondra un fichier TIFF et un fichier HTML.

Le nom des fichiers sera constitué du code des documents tronqué à 8 caractères.

Les sauts de page doivent être indiqués dans le flux texte. Le foliotage original est repris si il est présent sur la page imprimée. Les pages blanches du document original sont conservées et signalées.

Certains tableaux textuels seront saisis - tableaux résumés et rapports d'experts sous forme de tableaux - afin de pouvoir exploiter leur contenu.

Les logos sont repris sur la première page et abandonnés dans les pages suivantes.

Les signatures et les tampons sont traités et conservés, au moins sur les premières pages et les certificats.

Les en-têtes et les pieds de page sont traités et conservés dans chaque page.

Les cartouches sont repris comme des en-têtes.

Appréciation du test.

Le test de faisabilité effectué sur les 2370 pages du dossier du Fludex devrait être achevé à la fin du mois de septembre. Seront évalués, d'une part la qualité du travail effectué par Jouve S.I., et les réponses apportées à nos attentes par la solution choisie d'autre part. En cours de production, les prestataires doivent nous rendre compte régulièrement de l'avancée du travail et nous envoyer des fichiers intermédiaires.

Différents points seront contrôlés dans l'exploitation des fichiers HTML : Le confort de la lecture du texte, ainsi que celui des tableaux et des figures. La possibilité de copier certaines parties du texte, une page entière, un tableau ou une figure. Le déplacement à l'intérieur des fichiers : recherche sur un mot ou une expression, la possibilité de "feuilleter" page à page le document (deux points qui dépendent des outils utilisés). L'insertion des liens internes (oubliés, inactifs, mauvaise correspondance) seront examinés avec attention.

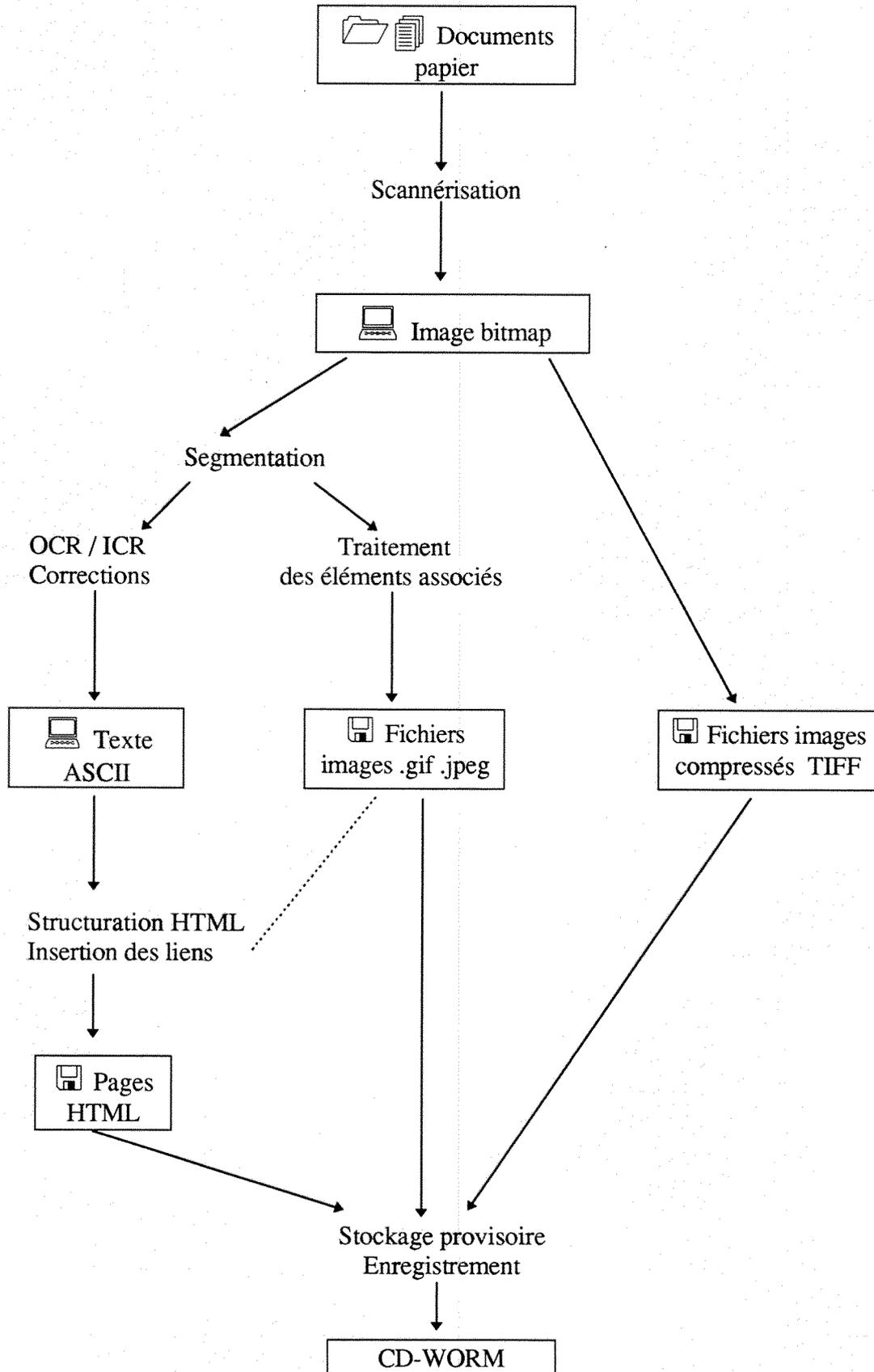
Le taux de reconnaissance de caractères attendu est de 99,98 % minimum. Le traitement des caractères spéciaux sera particulièrement surveillé, ainsi que les différents points particuliers demandés (reprise des en-têtes, pieds de page, du foliotage, des logos et des signatures).

L'identification des différents fichiers sera contrôlée. Les images TIFF seront elles aussi testées, notamment leur rendu à l'écran (sur des écrans de 14 pouces et plus) et leur impression (imprimante laser avec carte de décompression des fichiers TIFF). Il faudra évaluer la facilité d'accès des pages TIFF qui correspondent aux pages HTML affichées.

Plusieurs outils seront utilisés : Netscape Navigator 2.01, Hotmetal, Paint Shop Pro,...

Les tests seront effectués sur un PC monoposte et en réseau (en particulier pour estimer le temps de transfert et la rapidité de l'affichage).

Le traitement demandé.



V - Orientation du projet global.

En ce qui concerne le traitement de l'ensemble des dossiers d'AMM, trois scénarios sont envisageables. Le choix se fera en fonction du résultat du test de faisabilité réalisé par Jouve S.I. et du budget accordé à cette opération.

1 - Trois scénarios possibles.

Une simple numérisation.

Les documents sont scannés avec une résolution de 300 dpi et stockés sous forme de fichier bitmap (format TIFF, compressé selon la norme CCITT groupe IV). Cette solution correspond à la première option proposée par Jouve S.I. La diffusion peut se faire sur le réseau local ou par distribution de CD-WORM, notamment pour le Ministère.

- Avantages : La prestation est moins coûteuse et plus rapide. On peut visionner et imprimer des pages identiques aux pages imprimées originales. Un seul type de fichier est produit, donc un seul logiciel est nécessaire pour leur exploitation. Certains logiciels sont disponibles gratuitement ou à coût modique (Paint Shop Pro). Cette application est idéale pour un simple stockage électronique des documents. Elle peut être suffisante pour les dossiers les moins exploités.

- Inconvénients : Les fichiers images sont très lourds en octets. Ils sont compressés, mais le temps de compression/décompression s'ajoute aux temps de traitement du système et de transfert. Il ne s'agit pas de fichier texte, donc seule la lecture est possible (pas de copie, ni de recherche d'un mot). La lecture à l'écran est difficile (nécessité de posséder un écran de bonne qualité). Le déplacement dans le document se fait de page en page uniquement.

Certains paramètres sont déterminants pour l'impression, notamment la puissance des PC et des imprimantes utilisés. Les imprimantes doivent être équipées d'une carte de décompression des fichiers TIFF.

La génération de pages HTML.

Cette solution correspond à celle adoptée pour le test de faisabilité.

Les documents sont scannés, le texte est saisi par les techniques d'OCR/ICR puis structuré au format HTML avec l'intégration d'hyperliens internes et externes. Les fichiers sont stockés sur supports non-éffaçables et mis à disposition sur le réseau local.

- Avantages : On obtient des fichiers textes à partir desquels il est possible de copier des parties de texte ou des images (figures, tableaux, molécules,...). Au sein du document, il est possible de faire une recherche sur un mot ou une expression. Les liens hypertextes internes permettent des déplacements par bonds à partir de la table des matières ou des renvois entre différentes parties du document. Des liens externes (liens entre documents par exemple) peuvent être envisagés. HTML est un standard d'échange de documents qui est en pleine extension. Les documents HTML peuvent facilement être convertis vers d'autres formats. Les pages HTML peuvent être lues par des logiciels disponibles gratuitement sur Internet. La taille des fichiers est relativement faible. La lecture à l'écran est confortable et ne nécessite pas d'écran de qualité particulière.

- Inconvénients : La présentation des pages HTML n'a rien à voir avec celle des pages imprimées originales. Seul le contenu du texte et l'image numérisée des éléments associés (figures, tableaux, logos, signatures) sont respectés. Les polices ne sont pas reprises. Les entêtes et pieds de page, les sauts de page, les espacements, et les tabulations sont différents. Le

texte est continu, il n'est pas sectionné en pages. Il y a un grand nombre de fichiers à gérer (il y a autant de fichiers que d'éléments associés au texte).

Les différences entre les pages HTML et les pages originales nous obligent à gérer également les fichiers images TIFF des pages scannées. Il faut donc, soit trouver un logiciel permettant de visualiser ces deux types de fichiers, soit utiliser deux logiciels différents.

Les coûts de cette solution sont supérieurs à la solution précédente, en prestation et en place mémoire.

La production de pages PDF avec liens hypertextes.

Les documents sont scannés et OCRisés selon les techniques identiques aux solutions précédentes. Le texte est repris au format PDF avec des liens hypertextes.

- Avantages : Cette solution est assez peu gourmande en place mémoire. Le format PDF est un format standard, très proche du Postscript. Le document renvoyé est aussi proche du document initial que possible (proche d'une photocopie). On peut donc se passer de gérer en parallèle les images des pages scannées (les images des pages peuvent cependant être stockées indépendamment du système). Les fichiers sont consultables grâce à des logiciels disponibles gratuitement sur Internet (Acrobat Reader d'Adobe). Il s'agit de fichiers texte, donc exploitables (recopie, recherche sur un mot,...). Les liens hypertextes internes au document permettent une navigation facilitée à partir de la table des matières. Le déplacement de page à page est également possible.

- Inconvénients : C'est la solution dont la prestation est la plus coûteuse. Une imprimante Postscript est nécessaire.

2 - Intégration dans le système.

Il est essentiel d'anticiper les ressources budgétaires, matérielles et humaines nécessaires à l'intégration du nouveau système. Celle-ci devra se faire sans remettre en cause l'existant, notamment pour ne pas augmenter le temps d'accès aux applications déjà mises en place sur le réseau. Les problèmes matériels liés aux logiciels et aux machines à utiliser, ainsi que l'installation de lecteurs-distributeurs (Juke-Box ou autres systèmes) devront être envisagés. Le D.D.S jouera le rôle d'interface entre les services internes d'informatique et le sous-traitant. La garantie de confidentialité d'accès aux informations se fera selon le modèle de la base IRISA qui est la référence en terme de gestion de confidentialité.

On peut envisager des liens par une interface entre l'application mise en place et la base IRISA. Il est envisageable de diffuser les dossiers à l'Agence du Médicament via Internet en réseau semi-public, sur CD-WORM, ou sur tout autre support (bandes DAT, supports amovibles,...).

3 - Sous-traitance.

Un appel d'offre sera lancé à plusieurs prestataires de services.

La sécurité et le caractère vivant des documents obligent une production sur le site de Courbevoie. Les documents originaux seront prêtés une journée seulement et pourront être repris en cas de demande urgente. Un traitement hors site obligerait la production de photocopies à très grande échelle.

Le sous-traitant s'engage au respect de confidentialité et des échéances qui seront fixées.
Des contrôles de productivité et de qualité du travail effectué seront réalisés régulièrement par nos soins.

Les normes prises en compte seront, entre autres :

Norme ISO 9004-1. Gestion de la qualité et éléments de système qualité. Partie 1 : Lignes directrices.

Norme ISO 9004-2. Gestion de la qualité et éléments de système qualité. Partie 2 : Lignes directrices pour les services.

Partie II : Interventions sur la base de données IRISA

I - La base de données IRISA

La base IRISA est la base de données interne au groupe Servier qui gère les informations - références et valeur ajoutée (commentaire, descripteurs,...) - des documents publiés et non publiés sur les produits Servier. Son objectif est de permettre l'accès des différents collaborateurs à l'information concernant les produits du groupe. Elle fonctionne sur le système relationnel de gestion de données BASISplus (Information Dimensions), et comprend aujourd'hui plus de 21 000 enregistrements.

La base IRISA a été ouverte le 15 septembre 1994 pour gérer le nombre grandissant de documents entrant en documentation.

En réponse à un point du cahier des charges qui était l'exhaustivité des données, une reprise de l'existant a été effectuée. Ainsi, une base de structure identique à celle d'IRISA a été ouverte le 1er octobre 1994. Son contenu résulte de la récupération des informations enregistrées dans les anciennes bases de données internes et d'un enrichissement manuel. Ses enregistrements ont ensuite été basculés dans la base IRISA. La reprise de l'existant a donné lieu à un nouveau code de classement et à un ré-étiquetage des documents.

L'alimentation de la base se fait selon différentes sources : le téléchargement des références des bases de données internationales (puis commande des documents primaires à l'INIST), la transmission de documents par les Unités de Recherche, et l'obtention de publications à la Bibliothèque. L'analyse des documents enrichit les données.

Les documents publiés sont des articles, résumé, actes de congrès, posters, chapitres d'ouvrage,... Les documents non publiés sont les rapports d'études, les résumés de rapports, les rapports d'experts, les rapports chimie-pharmacies, les résumés des dossiers d'AMM, des monographies, des thèses,...

La situation d'un document dans la chaîne documentaire est suivie par l'indication d'une date spécifique et d'une action (code état).

La structure de la base comprend environ 40 champs, dont 28 sont visibles dans la vue utilisateur.

En plus des champs des références habituels (auteurs, affiliation, titre, date du document, titre de l'ouvrage ou du journal, pages, volume, numéro, résumé), l'analyse documentaire permet de renseigner les champs type de document, contenu général, produit concerné, marque, commentaires, mots-clés contrôlés,.... Une valeur de 1 à 4 est accordée à chaque document selon l'intérêt de l'information. La date d'entrée dans la base est indiquée (les notices reprises de l'existant ont pour date d'entrée le 31/08/94), ainsi que l'ancien code si il y a lieu.

Deux tables de correspondances créent un lien entre les références de documents appartenant à une même étude (numéro de protocole), et entre les références de documents d'un même congrès (numéro de congrès).

Les saisies dans la base IRISA sont contrôlées par 13 Thesaurus : les types de documents, le contenu général, les DCI (Dénomination Commune Internationale), les marques, les formes galéniques, les descripteurs, les journaux, les langues, les journaux, les bases téléchargées, les organismes, les iconographies et les supports.

Différentes vues sont possibles selon les droits accordés à l'utilisateur. Les recherches peuvent se faire en mode écran ou en mode commande FQM (Fundamental Query Manipulation, le langage BASIS).

II - Corrections effectuées.

La fiabilité des données contenues dans IRISA est un souci majeur pour le D.D.S. Homogénéiser l'indexation par des grilles d'indexation et des thesaurus permet d'améliorer les services offerts aux utilisateurs.

Les enregistrements de la base IRISA contiennent des imperfections qui sont apparues avec la reprise de l'existant, l'intégration de références téléchargées et la saisie de champs non contrôlés.

Les corrections ont porté essentiellement sur les champs descripteurs, titre d'ouvrage et éditeur.

Correction des descripteurs des rapports d'étude.

Le thesaurus des mots clés de l'ancienne base et celui d'IRISA sont différents. Au moment de la reprise de l'existant, les anciens descripteurs ont été repris dans la base IRISA selon la correspondance établie entre les mots clés des deux thesaurus. Deux mots clés anciens ont pu être remplacés par un même descripteur. Aucun contrôle n'a été fait pour éviter la répétition d'un mot clé dans le champ des descripteurs d'une notice.

Par exemple, le mot "métabolism" a été remplacé par "métabolism" et "pharmacokinetics" (le terme générique). Le terme "study-animal" a été remplacé par "animal", alors qu'à "monkey" correspond également le descripteur "animal".

Environ 2700 notices ont été reprises une à une pour corriger les mots clés qui se répétaient.

Homogénéisation des titres d'ouvrages.

Un même ouvrage pouvait être saisi sous des formes variées (orthographe, majuscules). Le champ titre d'ouvrage contenait des informations diverses (titre, chapitre, partie, tome, volume, numéro d'édition, date, nom d'éditeur). Certains documents étaient indexés comme un article de journal à la place d'un chapitre d'ouvrage, et inversement.

A l'avenir, le champ titre d'ouvrage ne contiendra plus que le titre proprement dit et le numéro d'édition. Les volumes et les tomes s'inscriront dans le champ volume, les chapitres et les parties dans le champ numéro. Un titre inconnu sera indiqué par "Not Referenced".

En se référant en cas de doute aux documents primaires, le contenu des champs titre d'ouvrage a été homogénéisé. Une liste des ouvrages actuellement cités dans la base IRISA a été éditée. Elle devra être consultée à chaque indexation d'un chapitre d'ouvrage ou d'un acte de congrès.

Par exemple, "Metabolic Activities of the lung (Ciba Foundation Symposium 78)" a été remplacé par "Metabolic Activities of the lung". "Ciba Foundation Symposium 78" a été reporté dans le champ commentaire.

Etablissement d'une liste des éditeurs.

Un travail identique au précédent a été fait sur le champ éditeur. Le contenu de ce champ était très irrégulier : nom des éditeurs, maison d'édition, adresse, numéro d'édition, date, orthographe varié, majuscules non respectées,... Un même ouvrage pouvait être référencé avec un éditeur différent ou avec un synonyme.

Une liste de référence a été établie. Dorénavant, le champ éditeur ne contiendra plus que le nom des éditeurs et/ou les maisons d'éditions. Deux maisons d'éditions seront séparées par un point. Les prénoms des éditeurs seront réduits à leurs initiales et suivront le nom. Les villes et les pays cités seront inscrits entre parenthèses. Un éditeur inconnu sera indiqué par "Not Referenced".

Une liste des éditeurs cités dans la base IRISA sera éditée.

Par exemple, "EXPANSION SCIENTIFIQUE FRANCAISE. G. Serratrice, JF Pellissier et J. Pouget. 1992" a été modifié en " Serratrice G., Pellissier J.F., Pouget J. Expansion Scientifique Francaise.".

Bibliographie

La gestion électronique documentaire. Jean-Yves Prax. Paris : Armand Colin Editeur, 1994.

Nouvelles technologies et communication de l'information. De l'analyse des besoins à l'ingénierie documentaire. Martine Sibertin-Blanc. Paris : ADBS Editions, 1994.

Le traitement électronique du document. Cours INRIA, 3-7 octobre 1994, Aix-en-Provence. INRIA. Paris : ADBS Editions, 1994.

SGML. Un outil pour la gestion électronique de document. Victor Sandoval. Paris : Hermes. Techniques de l'information, 1994.

Annexes

- 1 - Plan général d'un dossier d'AMM
- 2 - Caractéristiques du dossier d'AMM du Fludex
- 3 - Caractéristiques du dossier d'AMM de l'Isoméride
- 4 - Exemples de pages traitées par Jouve S.I.
- 5 - Présentation de la norme NF X 66 - 022.
- 6 - Liste de quelques prestataires de service à contacter.
- 7 - Abréviations utilisées.

Annexe 1

Plan général d'un dossier d'AMM

Partie I : Résumé du dossier.

- I A - Renseignements administratifs.
- I B - Résumé des caractéristiques du produit.
- I C - Rapports des experts sur la documentation chimique, pharmaceutique et clinique.

Partie II : Documentation chimique, pharmaceutique et biologique.

- II A - Composition.
- II B - Méthodes de préparation.
- II C - Contrôle des matières premières.
- II D - Contrôle des produits intermédiaires.
- II E - Contrôle des produits finis.
- II F - Stabilité.
- II G - Autres informations.

Partie III : Documentation toxicologique et pharmacologique.

- III A - Toxicité par administration unique.
- III B - Toxicité par administration répétée.
- III C - Etudes de reproduction.
- III D - Potentiel mutagène.
- III E - Potentiel oncogène / cancérogène.
- III F - Pharmacodynamie.
- III G - Pharmacocinétique.
- III Q - Autres informations.

Partie IV : Documentation clinique.

- IV A - Pharmacologie clinique.
- IV B - Expérience clinique.
- IV Q - Autres informations.

Partie V : Renseignements particuliers.

- V A - Présentation.
- V B - Echantillon.
- V C - Autorisation de fabrication.
- V D - Autorisation de mise sur le marché.
- V E - Déclarations d'intention.
- V F - Fiche destinée aux centres de pharmacovigilance et aux centres anti-poisons.

Annexe 2

Description du dossier d'AMM du S6520 (Fludex)

déposé en 1995

Le dossier d'AMM comprend 36 documents (mis à part les publications), soit 53 volumes. En éliminant les annexes, nous ne retiendrons que 41 volumes, soit 2370 pages.

Les documents sont structurés mais leur structure est variable.

Types de documents : résumés du dossier d'AMM, rapports d'études, rapports d'experts.

Il n'y a pas toujours de sommaire. Des annexes sont parfois incluses en fin de volume. Le contenu de ces annexes est variable d'un document à l'autre. Il arrive qu'une publication soit insérée au sein même d'un rapport.

Chaque document possède un code alphanumérique et un code barre, apposés sur la page de garde, qui lui sont propres

Le support.

Les pages ne sont pas reliées, mais sont simplement glissées dans une chemise cartonnée. Certains originaux sont perforés sur la gauche.

Papier de format A4, blanc, uniquement imprimé au recto. Certaines feuilles de couleur non imprimées servent d'intercalaire. Les pages sont parfois froissées, cornées ou déchirées (en particulier la première ou la dernière page). La page de garde peut être plus épaisse.

Rares sont les pages écrites sur deux colonnes (publications).

Ecriture noire, sauf les signatures et les en-têtes des pages de titre (bleue, grise,...).

Qualité d'impression variable : nous ne possédons parfois que de photocopies. (16,3 % des pages sont mal imprimées). Impression laser ou jet d'encre, mais pas d'impression matricielle.

La pagination.

Elle est de différente forme, en haut ou en bas de page : 10, 10., -10-, page 10, 10/79, page 10/79, 10 of 79, page 10 of 79, vol.1 p.10,...

Elle peut être double (cas où un document est constitué de plusieurs rapports) : 55 % des pages.

Les premières pages ne sont pas toujours numérotées, ou elles peuvent être paginées en chiffres romains.

La pagination peut être manuscrite (1,5 % des pages), ou mal imprimée.

Langue : Français (70,7 %) et Anglais (29,3 %). Il y a parfois un changement de langue au sein d'un même document.

Typographie.

Trois polices principales : Times New Roman (72,7 %), Arial (19,8 %), Courier (6,5 %).

La police peut varier au sein d'un document (pas seulement sur les titres). Certains rapports plus anciens ont été tapés à la machine à écrire (contour des lettres mal défini).

La taille est généralement homogène au sein d'un document (12 points le plus souvent), sauf à l'intérieur des tableaux.

Il y a en moyenne 716 caractères par page (sans le contenu et les légendes des tableaux et des figures).

Caractères : soulignés (15 caractères par page), gras (24 caractères par page), italiques (4 caractères par page).

Caractères spéciaux : \pm , \otimes , $^{\text{TM}}$, α , β , δ , μ , τ , λ , π , Δ , Σ , \equiv , \sim , \approx , \bullet , flèches, ...

Des indices et exposants, en particulier au niveau des formules chimiques et mathématiques.

Les encadrés.

Les titres des documents, chapitres ou tableaux sont rarement encadrés. Quelques cadres sont grisés.

Les tableaux.

Types de tableaux : tableaux de résumés, tableaux de rapports d'experts, liste des investigateurs, tableaux de données brutes ou de données statistiques, tableaux de formules unitaires, ... Ils peuvent contenir du texte, des nombres, des symboles ou des dessins de molécules.

24,1 % des pages contiennent un ou plusieurs tableaux. Les tableaux peuvent occuper une page entière ou une demi page. Certains tableaux se continuent sur les pages suivantes. Ils peuvent être imprimés en "mode paysage".

Leur impression n'est pas toujours nette (parfois tapés à la machine).

Les tableaux possèdent généralement un titre ou un numéro.

Les figures.

14,2 % des pages contiennent une ou plusieurs figures. Elles peuvent occuper une page entière ou une demi page et peuvent être imprimées en "mode paysage".

Types de figures : Courbes d'évolution, histogrammes, schémas de synthèse, schémas de l'étude, chromatogrammes, spectres infrarouges ou de RMN, ...

Leur qualité d'impression n'est pas toujours bonne (les chromatogrammes et certaines courbes sont issues de traitement de logiciels spécifiques, avant d'être insérés dans le document par photocopie).

Les figures possèdent généralement un titre ou un numéro.

Les planches photographiques. (20 planches dont 4 en couleur).

Elles sont peu nombreuses mais nous ne possédons pas toujours les originaux. Nous possédons généralement plusieurs exemplaires de ces planches.

Les signatures manuscrites (12 % des pages).

Il peut y en avoir plusieurs sur une même page. Certains rapports possèdent une signature en bas de chaque page. Une signature peut être accompagnée d'une date manuscrite ou d'un tampon.

Formules chimiques développées insérées dans le texte (dans 4,7 % des pages).

Formules chimiques brutes, avec indices ou exposants, insérées dans le texte (dans 7,8 % des pages).

Les formules mathématiques (1 pour 12 pages). Elles comprennent souvent un diviseur et des caractères spéciaux.

Les cartouches (15 % des pages).

Les cartouches possèdent des informations importantes (nom du produit, n° page, n° spécification,...). Situés en haut des pages, ils sont utilisés pour décrire les procédés de fabrication ou les méthodes de contrôle.

Les en-têtes et les pieds de pages (35,1 % des pages) : titre courant ou n° d'étude.

Les logos.

10 logos différents ont été identifiés.

13,4 % des pages possèdent un logo en pied de page.

On trouve également quelques logos "isolés", généralement sur la page de titre (1 logo pour 80 pages).

Annexe 3

Description du dossier d'AMM du S5614 (Isoméride gélules)

déposé en 1985

(excepté la partie I du dossier et les compléments)

Le dossier d'AMM comprend 79 documents (mis à part les publications), soit 118 volumes. En éliminant les annexes, nous ne retiendrons que 80 volumes, soit 3698 pages.

Les documents sont structurés mais leur structure est variable.

Types de documents : résumés du dossier d'AMM, rapports d'études, rapports d'experts.

Il n'y a pas toujours de sommaire, ou parfois très succinct.

Des annexes sont parfois incluses en fin de volume. Le contenu de ces annexes est variable d'un document à l'autre. Les annexes peuvent être annoncées dans le sommaire mais sans être distinguées dans le document.

Certains documents reprennent le terme Annexe (Annex) pour désigner les différentes parties du dossier d'AMM.

Le volume 1 indique rarement la table des matières des volumes suivants.

Il arrive qu'une publication soit insérée au sein même d'un rapport (12 pages).

Chaque document possède un code alphanumérique et un code barre, apposés sur la page de garde, qui lui sont propres. Ils ne sont pas souvent datés.

Le support.

Les pages ne sont pas reliées, mais sont simplement glissées dans une chemise cartonnée. Certains originaux sont perforés sur la gauche.

Papier de format A4, blanc, uniquement imprimé au recto. Certaines feuilles de couleur non imprimées servent d'intercalaire. Les pages sont parfois froissées, cornées ou déchirées (en particulier la première ou la dernière page). La page de garde peut être plus épaisse.

La qualité d'impression est très variable d'un document à l'autre, parfois même au sein d'un document (11,4 % des pages sont mal imprimées). Les feuillets sont souvent tachés. Nous ne possédons parfois que de photocopies. Quelques impressions laser ou jet d'encre, mais pas d'impression matricielle. La plupart des rapports ont été tapés à la machine à écrire (les lettres se touchent, ainsi que les accents).

Écriture noire, sauf les signatures et les en-têtes des pages de titre (bleue, grise,...).

On rencontre certaines corrections manuscrites ou au "correcteur liquide".

La pagination.

Elle est de différente forme, en haut ou en bas de page : 10, 7/15, -10, -10-, 10., 10-, 17 bis de 25,...

52,3 % des pages possèdent une double pagination (cas où un document est constitué de plusieurs rapports). Les premières pages ne sont pas toujours numérotées, ou elles peuvent être paginées en chiffres romains. La pagination peut être mal imprimée, ou sur le côté de la

feuille (cas des graphiques ou des tableaux imprimés en "mode paysage"). Elle est parfois cachée par une signature ou un tampon.

Langue : Français (79,8 %) et Anglais (20,2 %).

Typographie.

Deux polices principales : Times New Roman (13 %), Courier (87 %).

La police peut varier au sein d'un document (pas seulement sur les titres). On observe des variations typographiques selon les machines utilisées.

La taille est généralement homogène au sein d'un document (12 points le plus souvent).

Il y a en moyenne 622 caractères par page (sans le contenu et les légendes des tableaux et des figures).

Caractères : soulignés (19 caractères par page), gras (23 caractères par page), italiques (<1 caractère par page).

Les soulignés peuvent être simples, doubles, en pointillé ou en double pointillé.

Il y a peu de caractères spéciaux dans le texte : \pm , α , β , δ , μ , τ , λ , π , Δ , Σ , Ω , \sim , \approx , flèches,...Ceux-ci sont souvent manuscrits.

On trouve quelques indices et exposants, en particulier au niveau des formules chimiques et mathématiques.

Les encadrés. Les titres des documents, chapitres ou tableaux sont rarement encadrés. Il n'y a pas de cadre grisé.

Les tableaux.

Types de tableaux : tableaux de données brutes ou de données statistiques, tableaux de formules unitaires, calendriers,... (tableaux de résumés, tableaux de rapports d'experts, liste des investigateurs dans la partie I ?).

Ils peuvent contenir du texte, des nombres, des symboles ou des dessins de molécules.

22,3 % des pages contiennent un ou plusieurs tableaux. Les tableaux peuvent occuper une page entière ou une demi page. Certains tableaux se continuent sur les pages suivantes. Ils peuvent être imprimés en "mode paysage".

Leur impression n'est pas toujours nette. Le contour des tableaux est irrégulier, souvent en pointillé. Certains tableaux ne possèdent pas de contour (simples colonnes).

Les tableaux ne possèdent pas toujours un titre ou un numéro.

Les figures.

7,2 % des pages contiennent une ou plusieurs figures. Elles peuvent occuper une page entière (il n'y a pas toujours de marges) ou une demi page et peuvent être imprimées en "mode paysage".

Types de figures : Courbes d'évolution, histogrammes, schémas de synthèse, schémas de l'étude, chromatogrammes, spectres infrarouges ou de RMN,...

Leur qualité d'impression n'est pas toujours bonne (les chromatogrammes et certaines courbes sont issues de traitement de logiciels spécifiques, avant d'être insérés dans le document par photocopie). Certains graphiques sont achevés à la main.

Les figures ne possèdent pas toujours un titre ou un numéro.

Les planches photographiques. (8 planches).

Elles sont rares mais nous possédons les originaux (un seul jeu). Les photographies sont en noir et blanc.

Les signatures manuscrites (50,3 % des pages) et tampons.

Il peut y en avoir plusieurs sur une même page. Nombreux rapports possèdent une signature en bas (ou en haut) de chaque page. Une signature peut être accompagnée d'une date manuscrite ou d'un tampon : 4% des pages possèdent un tampon, qui parfois cache du texte.

Les formules.

Formules chimiques développées insérées dans le texte (dans 1 % des pages).

Les formules chimiques brutes, avec indices ou exposants insérées dans le texte, sont rares (1 pour 336 pages).

Les formules mathématiques (1 pour 20 pages). Elles comprennent souvent un diviseur et des caractères spéciaux. Il y en a généralement plusieurs sur la même page.

Les cartouches sont rares. (1 pour 217 pages).

Ils possèdent des informations importantes (nom d'un produit, n° page, n° spécification, date, signatures).

Les en-têtes et les pieds de pages (35,5 % des pages) : titre courant, n° d'étude, ou .../...

Les logos. (1 pour 740 pages).

5 logos différents ont été identifiés, situés sur les pages de titre.

Annexe 4

Exemples de pages traitées par Jouve S.I.

technologie servier

développement du médicament

FON15D57 DST: 11.04.95
NP0519700003FRE0
INDAPAMIDE

Page originale



S 6520

**comprimé enrobé à libération ralentie d'indapamide
dosé à 1,5 mg de principe actif**

PARTIE II : DOCUMENTATION CHIMIQUE ET PHARMACEUTIQUE

Volume 3

DC.R.CTR.006.R37.06520.01

Février 1995



27, rue Eugène Vignat - 45000 Orléans - SARL au capital de 50 000 F - RCS Orléans B 316 357 334
Tél. : 38 81 60 00 - Télex : TEC-SER 780029 F - Téléfax : 38 54 01 31
Adresse postale : BP 1749 - 45007 Orléans Cedex 01

technologie servier

développement du médicament

FONIS057 DST: 11.04.85
NP0519700003FRE0
INDAPAMIDE



Page TIFF

S 6520

comprimé enrobé à libération ralentie d'indapamide
dosé à 1,5 mg de principe actif

PARTIE II : DOCUMENTATION CHIMIQUE ET PHARMACEUTIQUE

Volume 3

DC.R.CTR.006.R37.06520.01

Février 1995



27, rue Eugène Vignat - 45000 Orléans - SARL au capital de 50 000 F - RCS Orléans B 316 357 334
Tél. : 38 81 60 00 - Télex : TEC-SER 780029 F - Téléfax : 38 54 01 31
Adresse postale : BP 1749 - 45007 Orléans Cedex 01

ETUDE DE STABILITE DE L'INDAPAMIDE DANS DES CONDITIONS SEVERES

1 INTRODUCTION

L'indapamide a été soumis en solution à des conditions sévères de pH et de température. Le but de ces études est :

- ▶ de tester la stabilité chimique de l'indapamide dans diverses conditions et de déterminer la nature des produits de dégradation formés.
- ▶ de justifier la sélectivité des conditions de chromatographie liquide utilisées pour les études de stabilité du principe actif sous forme de poudre et dans les formes pharmaceutiques.

Le principe actif a été maintenu pendant 20 heures en solution à + 4°C et à 80°C dans les conditions suivantes :

- ▶ dans un mélange méthanol-eau purifiée
- ▶ dans un mélange méthanol-eau purifiée-acide chlorhydrique
- ▶ dans un mélange méthanol-eau purifiée-hydroxyde de sodium

Les essais ont été effectués à partir du lot 36 008.

2 PREPARATION DES SOLUTIONS SOUMISES A DES CONDITIONS SEVERES

2.1 Préparation des solutions

- ▶ Solution d'indapamide dans le mélange méthanol-eau purifiée :
dissoudre 500 mg d'indapamide dans 60 ml de méthanol et compléter à 100,0 ml avec de l'eau purifiée, soit une solution à 5 mg/ml
- ▶ Solution d'indapamide dans le mélange méthanol-acide chlorhydrique :
dissoudre 500 mg d'indapamide dans 60 ml de méthanol, ajouter 10 ml d'acide chlorhydrique N et compléter à 100,0 ml avec de l'eau purifiée, soit une solution à 5 mg/ml.

▶ Solution d'indapamide dans le mélange méthanol-hydroxyde de sodium :
dissoudre 500 mg d'indapamide dans 60 ml de méthanol, ajouter 10 ml
d'hydroxyde de sodium N et compléter à 100,0 ml avec de l'eau purifiée,
soit une solution à 5 mg/ml.

2.2 Conditions des dégradations

Pour chacun des milieux, les solutions sont réparties en ampoules de 10 ml
par fractions de 5 ml. Les ampoules de chaque milieu sont conservées
pendant 20 heures :

- ▶ à + 4°C
- ▶ au bain-marie à 80°C

3 TECHNIQUES ANALYTIQUES

3.1 Dosage de produit restant par chromatographie liquide (CL)

La quantité d'indapamide non dégradé en solution dans les différentes
conditions est déterminée par chromatographie liquide dans les conditions
suivantes :

Colonne : acier inoxydable, longueur 15 cm, diamètre intérieur 4,6 mm

Phase stationnaire : Nucléosil 5C18 (Macherey-Nagel)

taille des particules 5 µm

Solution préliminaire à la préparation de la phase mobile :

solution aqueuse de laurylsulfate de sodium :

. lauryl sulfate de sodium	5 g
. acide acétique glacial	3 ml
. eau purifiée qsp	100 ml

Phase mobile :

. eau purifiée	69 volumes
. acétonitrile	31 volumes
. isobutanol	2 volumes
. triéthylamine	1 volume
. solution aqueuse de lauryl sulfate de sodium	0,6 volume
. acide orthophosphorique à 85 % (m/m) qsp	pH = 3

1/7 523 PA.R.STA.D02.R00.01520.02

ETUDE DE STABILITE DE L'INDAPAMIDE DANS DES CONDITIONS SEVERES

1 INTRODUCTION

L'indapamide a été soumis en solution à des conditions sévères de pH et de température. Le but de ces études est :

de tester la stabilité chimique de l'indapamide dans diverses conditions et de déterminer la nature des produits de dégradation formés.

de justifier la sélectivité des conditions de chromatographie liquide utilisées pour les études de stabilité du principe actif sous forme de poudre et dans les formes pharmaceutiques.

Le principe actif a été maintenu pendant 20 heures en solution à + 4°C et à 80°C dans les conditions suivantes :

- dans un mélange méthanol-eau purifiée
- dans un mélange méthanol-eau purifiée-acide chlorhydrique
- dans un mélange méthanol-eau purifiée-hydroxyde de sodium

Les essais ont été effectués à partir du lot 36 008.

2 PREPARATION DES SOLUTIONS SOUMISES A DES CONDITIONS SEVERES

2.1 Préparation des solutions

Solution d'indapamide dans le mélange méthanol-eau purifiée : dissoudre 500 mg d'indapamide dans 60 ml de méthanol et compléter à 100,0 ml avec de l'eau purifiée, soit une solution à 5 mg/ml

Solution d'indapamide dans le mélange méthanol-acide chlorhydrique : dissoudre 500 mg d'indapamide dans 60 ml de méthanol, ajouter 10 ml d'acide chlorhydrique N et compléter à 100,0 ml avec de l'eau purifiée, soit une solution à 5 mg/ml.

Solution d'indapamide dans le mélange méthanol-hydroxyde de sodium : dissoudre 500 mg d'indapamide dans 60 ml de méthanol, ajouter 10 ml d'hydroxyde de sodium N et compléter à 100,0 ml avec de l'eau purifiée, soit une solution à 5 mg/ml.

2.2 Conditions des dégradations

Pour chacun des milieux, les solutions sont réparties en ampoules de 10 ml par fractions de 5 ml. Les ampoules de chaque milieu sont conservées pendant 20 heures: > à + 4°C > au bain-marie à 80°C

3 TECHNIQUES ANALYTIQUES

3.1 Dosage de produit restant par chromatographie liquide (CL)

La quantité d'indapamide non dégradé en solution dans les différentes conditions est déterminée par chromatographie liquide dans les conditions suivantes : Colonne : acier inoxydable, longueur 15 cm, diamètre intérieur 4,6 mm

Phase stationnaire :

Nucléosil 5C18 (Macherey-Nagel)
taille des particules 5 µm

Solution préliminaire à la préparation de la phase mobile :
solution aqueuse de laurylsulfate de sodium :

- . lauryl sulfate de sodium 5 g
- . acide acétique glacial 3 ml
- . eau purifiée qsp 100 ml

Phase mobile:

- . eau purifiée 69 volumes
- . acétonitrile 31 volumes
- . isobutanol 2 volumes
- . triéthylamine 1 volume
- . solution aqueuse de lauryl sulfate de sodium 0,6 volume
- . acide orthophosphorique à 85 % (m/m) qsp pH = 3

anol-eau purifiée :

dissoudre 500 mg d'indapamide dans 60 ml de méthanol et compléter à

100,0 ml avec de l'eau purifiée, soit une solution à 5 mg/ml

<P>

 Solution d'indapamide dans le mélange méthanol-acide chlorhydrique :

dissoudre 500 mg d'indapamide dans 60 ml de méthanol, ajouter 10 ml

d'acide chlorhydrique N et compléter à 100,0 ml avec de l'eau purifiée,

soit une solution à 5 mg/ml.

<P>

 Solution d'indapamide dans le mélange méthanol-hydroxyde de sodium :

dissoudre 500 mg d'indapamide dans 60 ml de méthanol, ajouter 10 ml

d'hydroxyde de sodium N et compléter à 100,0 ml avec de l'eau purifiée,

soit une solution à 5 mg/ml.

<P>

<H3 ALIGN=LEFT>2.2 Conditions des dégradations</H3>

<P>Pour chacun des milieux, les solutions sont réparties en ampoules de 10 ml

par fractions de 5 ml. Les ampoules de chaque milieu sont conservées

pendant 20 heures:

> à + 4°C

> au bain-marie à 80°C</P>

<P>

<H2 ALIGN=LEFT>3 TECHNIQUES ANALYTIQUES</H2>

<H3 ALIGN=LEFT>3.1 Dosage de produit restant par chromatographie liquide (CL)</H3>

<P>La quantité d'indapamide non dégradé en solution dans les différentes

conditions est déterminée par chromatographie liquide dans les conditions

suivantes :

Colonne : acier inoxydable, longueur 15 cm, diamètre intérieur 4,6 mm</P>

Phase stationnaire :

Nucléosil 5C18 (Macherey-Nagel)

taille des particules 5 µm

<P>

<P>

Solution préliminaire à la préparation de la phase mobile :

solution aqueuse de laurylsulfate de sodium :

. lauryl sulfate de sodium 5 g

. acide acétique glacial 3 ml

. eau purifiée qsp 100 ml

<P>

Phase mobile:

. eau purifiée

69 volumes
. acétonitrile
31 volumes
. isobutanol
2 volumes
. triéthylamine
1 volume
. solution aqueuse de lauryl sulfate de sodium
0,6 volume
. acide orthophosphorique à 85 % (m/m)
qsp pH = 3

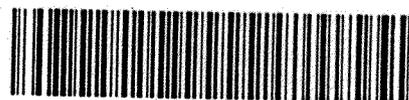
</BODY>
</HTML>

technologie servier

développement du médicament

FON15D57 DST: 11.04.95
NP0519700003FRE0
INDAPAMIDE

Page originale



NP0519700003FRE0

S 6520

**comprimé enrobé à libération ralentie d'indapamide
dosé à 1,5 mg de principe actif**

PARTIE II : DOCUMENTATION CHIMIQUE ET PHARMACEUTIQUE

Volume 3

DC.R.CTR.006.R37.06520.01

Février 1995



27, rue Eugène Vignat - 45000 Orléans - SARL au capital de 50 000 F - RCS Orléans B 316 357 334
Tél. : 38 81 60 00 - Télex : TEC-SER 780029 F - Téléfax : 38 54 01 31
Adresse postale : BP 1749 - 45007 Orléans Cedex 01

technologie servier

développement du médicament

FON15057 DST: 11.04.85
NP0519700003FRE0
INDAPAMIDE

Page PDF



NP0519700003FRE0

S6520

comprimé enrobé à libération ralentie d'indapamide
dosé à 1,5 mg de principe actif

PARTIE II: DOCUMENTATION CHIMIQUE ET PHARMACEUTIQUE

Volume 3

DC.R.CTR.006.R37.06520.01

Février 1995



27, rue Eugène Vignat - 45000 Orléans - SARL au capital de 50 000 F - RCS Orléans B 316 357 334
Tél. : 38 81 60 00 - Télex : TEC-SER 780029 F - Téléfax : 38 54 01 31
Adresse postale:BP 1749 - 45007 Orléans Cedex 01

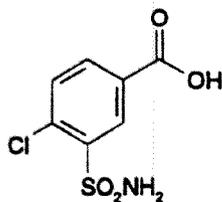
(1) Réactif de Dragendorff, formule de Munier et Macheboeuf :

. solution a : dissoudre 0,85 g de sous-nitrate de bismuth dans un mélange composé de 10 ml d'acide acétique glacial et de 40 ml d'eau purifiée.

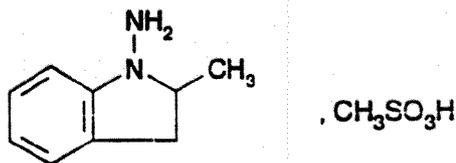
. solution b : dissoudre 8 g d'iodure de potassium dans 20 ml d'eau purifiée.
Mélanger 5 ml de solution a et 5 ml de solution b immédiatement avant emploi, puis ajouter 20 ml d'acide acétique glacial. Compléter à 100 ml avec de l'eau purifiée.

(2) Formules :

Y 36

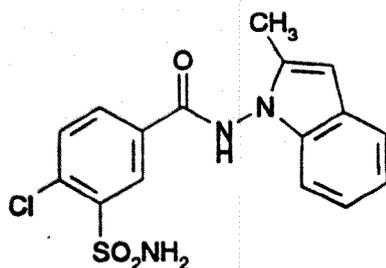


Y 37



. CH₃SO₃H

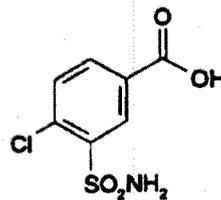
Y 38



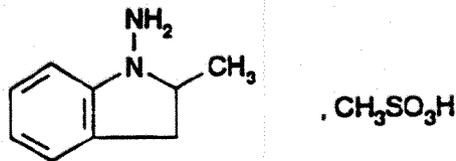
- (1) Réactif de Dragendorff, formule de Munier et Macheboeuf:
- . solution a : dissoudre 0,85 g de sous-nitrate de bismuth dans un mélange composé de 10 ml d'acide acétique glacial et de 40 ml d'eau purifiée.
 - . solution b: dissoudre 8 g d'iodure de potassium dans 20 ml d'eau purifiée.
- Mélanger 5 ml de solution a et 5 ml de solution b immédiatement avant emploi, puis ajouter 20 ml d'acide acétique glacial. Compléter à 100 ml avec de l'eau purifiée.

- (2) Formules :

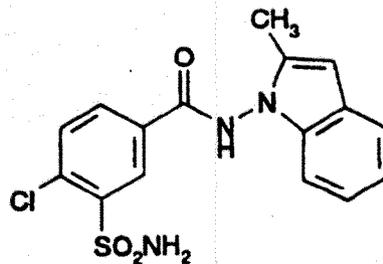
Y 36



Y 37



Y 38



technologie servier

développement du médicament

Page originale

FON1SD57 DST: 11.04.95
NP051970003FRE0
INDAPAMIDE



S 6520

comprimé enrobé à libération ralentie d'indapamide
dosé à 1,5 mg de principe actif

PARTIE II : DOCUMENTATION CHIMIQUE ET PHARMACEUTIQUE

Volume 3

DC.R.CTR.006.R37.06520.01

Février 1995



27, rue Eugène Vignat - 45000 Orléans - SARL au capital de 50 000 F - RCS Orléans B 316 357 334
Tél. : 38 81 60 00 - Télex : TEC-SER 780029 F - Téléfax : 38 54 01 31
Adresse postale : BP 1749 - 45007 Orléans Cedex 01

TABLE DES MATIERES

Volume 1

	<u>référence</u>	<u>page</u>
PREFACE		2
TABLE DES MATIERES		3
INDEX DES REFERENCES DES DOCUMENTS		9
PARTIE II A : COMPOSITION		13
<u>II A1 Composition de la spécialité pharmaceutique</u>	GL.R.FOR.042.R37.06520.02	15
<u>II A2 Récipient</u>	GL.R.CDT.036.R37.06520.01	17
<u>II A3 Formulations utilisées pour les essais cliniques</u>	GL.R.FOR.043.R37.06520.01	19
<u>II A4 Développement galénique</u>	GL.R.FAB.036.R37.06520.01	37
PARTIE II B : METHODE DE PREPARATION		99
<u>II B1 Formule de fabrication</u>		100
pour un lot de 2 000 000 de comprimés enrobés	GL.R.FAB.049.R37.06520.01	101
<u>II B2 Procédé de fabrication</u>		102
Chronologie des opérations de fabrication et des contrôles en cours de fabrication	GL.R.FAB.033.R37.06520.01	103
Procédé de fabrication industriel pour un lot de 2 000 000 de comprimés enrobés	GL.R.FAB.050.R37.06520.01	105
<u>II B3 Validation du procédé de fabrication</u>	GL.R.FAB.044.R37.06520.01	113
PARTIE II C : CONTROLE DES MATIERES PREMIERES		149
<u>II C1 Principe actif : Indapamide</u>		150
II C1.1 Spécifications et contrôle de routine :		151
II C1.1.2 Principe actif non décrit dans une pharmacopée		152
Spécifications de l'indapamide	PA.R.CTR.A05.R00.01520.01	153
Monographie de l'indapamide	PA.T.CTR.M02.R00.01520.02	154

technologie servier

développement du médicament

S6520

**comprimé enrobé à libération ralentie d'indapamide
dosé à 1,5 mg de principe actif**

PARTIE II: DOCUMENTATION CHIMIQUE ET PHARMACEUTIQUE

Volume 3

**DC.R.CTR.006.R37.06520.01
1995**

Février

Tél. : 38 81 60 00 - Télex : TEC-SER 780029 F - Téléfax : 38 54 01 31
Adresse postale:BP 1749 - 45007 Orléans Cedex 01

TABLE DES MATIERES

Volume 1

	<u>référence</u>	<u>page</u>
PREFACE		2
TABLE DES MATIERES		3
INDEX DES REFERENCES DES DOCUMENTS		9
PARTIE II A:COMPOSITION		
13		
<u>II A1 Composition de la spécialité pharmaceutique</u>		
	GL.R.FOR.042.R37.06520.02	15
<u>II A2 Récipient</u>	GL.R.CDT.036.R37.06520.01	17
<u>II A3 Formulations utilisées pour les essais cliniques</u>		
	GL.R.FOR.043.R37.06520.01	19
<u>II A4 Développement galénique</u>	GL.R.FAB.036.R37.06520.01	37
PARTIE II B:METHODE DE PREPARATION		
99		
<u>II B1 Formule de fabrication</u>		
100		
pour un lot de 2 000 000 de comprimés enrobés	GL.R.FAB.049.R37.06520.01	101
<u>II B2 Procédé de fabrication</u>		
102		
Chronologie des opérations de fabrication et des contrôles en cours de fabrication	GL.R.FAB.033.R37.06520.01	103
Procédé de fabrication industriel pour un lot de 2 000 000 de comprimés enrobés	GL.R.FAB.050.R37.06520.01	105
<u>II B3 Validation du procédé de fabrication</u>	GL.R.FAB.044.R37.06520.01	113
PARTIE II C: CONTROLE DES MATIERES PREMIERES		149

Annexe 5

Présentation de la norme NF X 66 - 022.

Norme Française NF X 66 - 022 (octobre 1991). Application de la Statistique. Sélection des plans d'échantillonnage pour le contrôle par comptage de la proportion d'individus non conformes ou du nombre moyen de non-conformité par unité.

Objet : Sélection de plans d'échantillonnage pour le contrôle du pourcentage d'individus non conformes ou du nombre moyen de non-conformités par 100 unités.

Domaine d'application : contrôle d'individus (unités) ou de produits susceptibles d'être individualisés.

Table 1 (extrait) : Effectif de l'échantillon en fonction de l'effectif des lots et du niveau de contrôle.

Cas d'un plan d'échantillonnage simple et d'un niveau de contrôle II.

Effectif des lots	Effectif d'échantillon
2 - 8	2
9 - 15	3
16 - 25	5
26 - 50	8
51 - 90	13
91 - 150	20
151 - 280	32
281 - 500	50
501 - 1 200	80
1 201 - 3 200	125
3 201 - 10 000	200
10 001 - 35 000	315
35 001 - 150 000	500
150 001 - 500 000	800
500 001 et plus	1250

Annexe 6

Liste de quelques prestataires de service à contacter.

Jouve S.I.

11, Bd Sébastopol B.P. 2734
75027 Paris cedex 01
Tel : (1) 44.76.86.00
(<http://www.jouve.fr>)

Berger-Levrault SA

Département Techniques Documentaires.
34, Avenue du Roule
92200 Neuilly-sur-Seine.
Tel : (1) 46.40.05.51
(<http://www.berger-levrault.fr>)

MC2

4, Chemin de Malacher
ZIRST 4401
38944 Meylan cedex
Tel : 76.90.22.00

E.I.S

117, rue Charles Michels
B.P. 184
93208 Saint Denis cedex.
Tel : (1) 48.13.25.00

Annexe 7

Abréviations utilisées.

AMM : Autorisation de Mise sur le Marché.
ASCII : American Standard Code for Information Interchange.
C.N.R.S : Centre National de Recherche Scientifique.
CALs : Computer-aided Acquisition and Logistics Support.
CAPLA : Computer Assisted Product License Application.
CCITT : Comité Consultatif International du Télégraphe et du Téléphone.
CD-ROM : Compact Disc - Read Only Memory.
CD-WORM : Compact Disc - Write Once Read Many.
CIRT : Centre International de Recherche Thérapeutique.
CNRS : Centre National de Recherche Scientifique.
D.D.S : Département de Documentation Scientifique et Traduction.
DCI : Dénomination Commune Internationale.
DON : Disque Optique Numérique.
DTD : Document Type Definition.
FDA : Food and Drug Administration.
FQM : Fondamental Query Manipulation.
FTP : File Transfer Protocol.
GED : Gestion Electronique de Documents.
GEIDE : Gestion Electronique de l'Information et des Documents Existants.
HTML : Hyper Text Markup Language.
INSERM : Institut National de la Santé et de la Recherche Médicale.
ICR : Intelligent Characters Recognition.
JPEG : Joint Photographic Expert Group.
OCR : Optical Characters Recognition.
ODA : Office Document Architecture.
OTC : Over The Counter.
PDF : Portable Document Format.
SGML : Standard Generalized Markup Language.
SNIP : Syndicat National de l'Industrie Pharmaceutique.
TIFF : Tagged Image File Format.
VAX : Virtual Address eXtension.



BIBLIOTHEQUE DE L'ENSSIB



810761E