

Chambre de Commerce et d'Industrie de Rouen

Institut Régional des Techniques Documentaires  
11 rue du Tronquet  
76130 Mont Saint Aignan

## Rapport de Stage

Création d'un produit documentaire électronique



## Sommaires des revues de l'IRD



Stage effectué du 07 septembre au 03 novembre 2000  
au sein de l'unité  
« INGENIERIE DE L'INFORMATION »  
du secteur documentation  
de l'Institut de Recherche pour le Développement de Bondy (93)

MICALETTI Marie -Angèle  
16<sup>ème</sup> promotion  
1999-2001

Février 2001

Chambre de Commerce et d'Industrie de Rouen

Institut Régional des Techniques Documentaires  
11 rue du Tronquet  
76130 Mont Saint Aignan

## Rapport de Stage

Création d'un produit documentaire électronique

Stage effectué du 07 septembre au 03 novembre 2000  
au sein de l'unité  
« INGENIERIE DE L'INFORMATION »  
du secteur documentation  
de l'Institut de Recherche pour le Développement de Bondy (93)

Tuteur de stage : Monsieur Pier Luigi ROSSI  
Responsable de l'unité  
« Ingénierie de l'information »  
32, avenue Henri Varagnat  
93 143 Bondy Cedex

MICALETTI Marie -Angèle  
16<sup>ème</sup> promotion  
1999-2001

Février 2001

## Remerciements

Je tiens à remercier toutes les personnes qui m'ont chaleureusement accueillie au sein de l'IRD et qui m'ont aidée à collecter des informations sur le secteur documentation, à savoir :

- Monsieur Pier Luigi ROSSI, pour son sens pédagogique et sa disponibilité à mon égard,
- ainsi qu' Emmanuelle ALDEBERT, Fortunée BRAMI, Valérie HOGNON, Evelyne MILLET, Marie-Hélène PERROT et Térésa PETESCH, pour les différents moments que nous avons partagés.

# Sommaire

<b>Introduction</b> .....	<b>1</b>
<b>1. Les missions de l'IRD et son réseau documentaire</b> ....	<b>2</b>
1.1. L'IRD : l'Institut de Recherche pour le Développement .....	2
1.1.1. Présentation de l'IRD .....	2
1.1.2. Missions et organisations des activités scientifiques de l'IRD ....	2
1.1.2.1. Missions de l'institut .....	2
1.1.2.2. Organisation des activités scientifiques de l'IRD.....	2
1.1.3. La place de la documentation à l'IRD.....	3
1.1.3.1. Le Secteur Documentation.....	3
1.1.3.2. Le projet « Infothèque » et les TIC (Technologies de l'Information et de la Communication).....	4
1.2. Le centre IRD à Bondy (93).....	6
1.2.1. Présentation du centre de Bondy .....	6
1.2.2. Organisation de l'activité scientifique .....	6
1.3. Le secteur documentation de l'IRD à Bondy.....	7
1.3.1. Présentation du service de documentation .....	7
1.3.1.1. Historique et situation actuelle.....	7
1.3.1.2. Composition du Secteur documentation à Bondy .....	7
1.3.1.3. Localisation et aménagement du service de documentation.....	9
1.3.2. Le fonctionnement du service de documentation.....	9
1.3.2.1. Présentation du personnel et des différentes fonctions pour chaque service .....	9
1.3.2.1.1. Le centre de documentation : la bibliothèque .....	9
1.3.2.1.2. Les unités techniques du Secteur Documentation présentes à Bondy.....	10
1.3.2.2. Fonctionnement en réseau avec les autres centres.....	12
1.3.3. Moyens du service de documentation .....	14
1.3.3.1. Moyens financiers.....	14
1.3.3.2. Moyens matériels .....	14
1.3.4. Utilisateurs internes et externes et leurs types de demandes....	15
1.3.4.1. Les utilisateurs internes .....	15
1.3.4.2. Les utilisateurs externes .....	16
1.3.5. Le fonds documentaire .....	16
1.3.5.1. Les domaines concernés.....	17
1.3.5.2. Evaluation du volume des divers types de documents et d'informations....	17
1.3.6. Activités de gestion du fonds documentaire .....	17
1.3.6.1. Les acquisitions.....	17
1.3.6.2. Le catalogage.....	17
1.3.6.3. Le traitement intellectuel des documents .....	18

1.3.7. Langages documentaires : Outils pour le traitement documentaire	18
1.3.8. Outils de recherche	18
1.3.8.1. Base de données bibliographique : Horizon	18
1.3.8.2. Bibliothèque électronique « Pleins Textes »	19
1.3.8.3. Autres supports électroniques : CD ROM	19
1.3.8.4. Autres Bases de données, désormais en ligne	19
1.3.9. Prestations : Produits et services	19
<b>2. Création d'un produit documentaire électronique au sein de l'unité "ingénierie de l'information"</b>	<b>21</b>
2.1. Historique du projet de stage	21
2.1.1. La numérisation : aspects techniques	22
2.1.2. Pleins Textes	24
2.1.3. Objet du stage	26
2.2. « L'information à mettre en ligne »	27
2.2.1. Contenu et présentation de l'information	27
2.2.2. Etat des lieux de l'existant	27
2.2.2.1. Rapide panorama de ce qui existe	28
2.2.2.2. Utilisation des informations trouvées sur les différents sites visités	29
2.2.2.3. Objectifs de cet état des lieux	29
2.2.3. Classement de l'information : « préparation de l'information »	30
2.2.3.1. Classement des notices	31
2.2.3.2. Découpage des fichiers informatiques	34
2.3. Création de la présentation Web : <a href="http://www.bondy.ird.fr/tdp">http://www.bondy.ird.fr/tdp</a>	36
2.3.1. Travail préparatoire à l'élaboration du site	36
2.3.1.1. Arborescence du site	37
2.3.1.2. Contraintes pratiques	40
2.3.2. Chronologie de la création du site	41
2.3.2.1. Les différentes étapes de la création de <a href="http://www.bondy.ird.fr/tdp">http://www.bondy.ird.fr/tdp</a>	41
2.3.2.2. Architecture du site Web « tdp » dans le contexte du réseau informatique de Bondy	45
2.3.3. Caractéristiques du matériel informatique utilisé	47
2.3.4. Utilisation d'éditeurs HTML et de logiciel de retouche d'image	48
2.3.5. Problèmes rencontrés	48
2.3.6. Caractéristiques de ce travail	49
<b>Conclusion</b>	<b>50</b>
<b>Bibliographie</b>	<b>54</b>
<b>Annexes</b>	Voir dossier : « Annexes »

## Introduction

J'ai effectué un stage de deux mois, en septembre et octobre 2000, au Secteur Documentation du centre IRD (Institut de Recherche pour le Développement) de Bondy (93), au sein de l'unité « Ingénierie de l'information ».

L'objectif du stage visait la mise en accès sur Internet d'articles de périodiques de l'IRD par l'intermédiaire de sommaires des revues et d'index d'auteurs. Les sommaires et les index donnent accès aux documents en texte intégral au format PDF. L'aspect informatique et créatif de cette activité m'a poussée à choisir ce stage.

Dans la première partie, on présentera l'IRD, ses missions et ses caractéristiques, ainsi que la place que tient le Secteur Documentation au sein de cette structure.

Le stage ayant été effectué dans le centre de Bondy, on s'intéressera ensuite aux différentes spécialités développées par les scientifiques et les différents métiers présents sur ce site.

Enfin, on présentera le Secteur Documentation à Bondy : sa composition, ses fonctions, ses moyens humain, financier et matériel, son fonctionnement en réseau et les caractéristiques des utilisateurs.

Dans la deuxième partie, on développera et analysera la mission du stage : les différentes tâches effectuées, les problèmes rencontrés et les questions posées.

Plus largement, on réfléchira sur les apports de ce stage en complément à la formation.

# 1. Les missions de l'IRD et son réseau documentaire

## 1.1. L'IRD : l'Institut de Recherche pour le Développement

### 1.1.1. Présentation de l'IRD

L'Institut de Recherche pour le Développement (IRD, ex-ORSTOM) est un établissement public à caractère scientifique et technologique, placé sous la double tutelle des ministères chargés de la Recherche et de la Coopération. (voir annexe 1 : organigramme général de l'IRD).

Cet institut mène des recherches en Afrique, dans l'Océan Indien, en Amérique latine, en Asie et dans le Pacifique, dans les domaines suivants : milieux et environnement, ressources vivantes, sociétés et santé. Il dispose pour ce faire de 36 implantations dont cinq en France métropolitaine (Paris, Bondy, Montpellier, Brest et Orléans), cinq dans les collectivités territoriales d'outre-mer (Guyane, Nouvelle-Calédonie, Polynésie française, Martinique, La Réunion) et de 26 centres situés essentiellement dans la zone intertropicale. (voir annexe 2 : Carte du monde présentant les différentes implantations de l'IRD)

Son budget est de 1,1 milliard de francs. Son effectif comprend 2300 agents dont 800 chercheurs, 750 ingénieurs, techniciens et agents administratifs et des personnels locaux. 40% de ses agents travaillent hors métropole, de nombreux techniciens sont originaires du Sud et ses programmes de recherche associent directement des chercheurs étrangers. L'IRD participe à de nombreux programmes scientifiques internationaux.

### 1.1.2. Missions et organisations des activités scientifiques de l'IRD

#### 1.1.2.1. Missions de l'institut

L'IRD remplit trois missions fondamentales : la recherche, l'expertise et la formation. Il conduit des programmes scientifiques centrés sur les relations entre l'homme et son environnement dans les pays du Sud, dont l'objectif est de contribuer à leur développement durable.

#### 1.1.2.2. Organisation des activités scientifiques de l'IRD

Les activités scientifiques de l'IRD s'organisent en cinq départements : trois pour la recherche, un pour l'expertise et un pour la formation.

- La recherche s'organise en trois départements : Milieux et environnement, Ressources vivantes et Sociétés et Santé.

Le département *Milieus et environnement* développe des recherches sur les variations climatiques tropicales, les milieux littoraux et terrestres (sols, couverture végétale) et les ressources en eau, les risques naturels (séismes, inondations, etc.), les ressources minérales et les impacts des activités humaines sur l'environnement.

Le département *Ressources vivantes* se consacre à l'étude de la biodiversité, des écosystèmes aquatiques (marins, littoraux et continentaux) et des systèmes agricoles tropicaux (amélioration génétique, défense des plantes cultivées, biotechnologies, etc.) afin d'assurer la viabilité de leur exploitation par une gestion appropriée.

Le département *Sociétés et Santé* conduit des programmes de recherche sur la santé, (paludisme, sida et autres maladies parasitaires et virales, malnutrition, santé publique), sur les questions urbaines, les dimensions économiques, sociales et culturelles du développement.

- Le département *Expertise et valorisation* réalise des expertises sur des thèmes scientifiques en discussion avec d'autres organismes de recherche publics ou privés à l'échelle internationale. Ces synthèses de connaissances (expertises collégiales) sont valorisées dans la mise en place d'actions concrètes et dans les dépôts de brevet (créations d'entreprises locales et protection de la propriété intellectuelle).
- Le département *Soutien et formation des communautés scientifiques du Sud* a pour mission de former des chercheurs et des scientifiques de hauts niveaux et de développer de véritables partenariats entre les équipes du Nord et du Sud.

Ces différentes missions impliquent nécessairement l'existence d'un secteur documentation à part entière spécialisé dans les divers domaines de recherche de l'IRD. Par ailleurs, le secteur documentation doit s'adapter à la dispersion géographique qui caractérise cet institut de recherche.

### **1.1.3. La place de la documentation à l'IRD**

#### **1.1.3.1. Le Secteur Documentation**

Les activités documentaires sont coordonnées par le secteur Documentation de la Délégation à l'Information et à la Communication (DIC)<sup>1</sup>. Il intervient à la première et à la dernière étape de l'activité de recherche, dans une double démarche de recherche de l'information (pour les scientifiques) et de restitution des résultats de recherche (des chercheurs). Sa mission est donc de contribuer à une circulation efficace de l'information scientifique en assurant :

- l'accès à l'information au bénéfice des scientifiques de l'IRD

---

<sup>1</sup> Voir page 5 : Schéma intitulé : « Place du Secteur documentation à l'IRD »

- la diffusion de l'information produite par l'Institut (diffusion des résultats des recherches de l'IRD par l'intermédiaire d'un référentiel bibliographique et par l'accès aux documents originaux)
- la diffusion de l'information scientifique vers et à partir des pays du Sud, et donc participer au désenclavement des scientifiques de cette partie du monde
- des services d'appui documentaire aux programmes de recherche de la zone tropicale
- des opérations en coopération et l'appui aux structures documentaires des pays partenaires (formation des personnels, équipement des structures, etc.)

La base de données bibliographique multidisciplinaire Horizon recense les publications des équipes de recherche de l'IRD.

### 1.1.3.2. Le projet « Infothèque » et les TIC (Technologies de l'Information et de la Communication)

#### ➤ Le projet « Infothèque »

La DIC a lancé en 1999 un projet intitulé « Infothèque des Milieux et Sociétés tropicales », qui vise à mettre en ligne et à rendre accessibles par Internet les connaissances acquises et les travaux en cours dans les différents domaines scientifiques de l'IRD. Le principal objectif est de mettre ces savoirs à la disposition des équipes scientifiques du Sud, en les aidant parallèlement par des actions de formation et d'équipement à maîtriser l'usage des nouvelles technologies.

« Infothèque » concerne tous les secteurs de l'information et de la communication traités par la DIC et réunira des articles et ouvrages, des cartes, des photographies, des bases de données complexes, comme des herbiers, des statistiques démographiques ou hydrologiques. Les partenaires du Sud seront associés à la constitution de cette bibliothèque numérique. « Infothèque » deviendra un portail<sup>1</sup> d'accès à l'information concernant les pays en développement.

Le Secteur Documentation s'est fortement investi dans ce projet (architecture générale du projet, numérisation du fonds documentaire, mise en place d'outils documentaires).

#### ➤ Les TIC : Technologies de l'Information et de la Communication à l'IRD

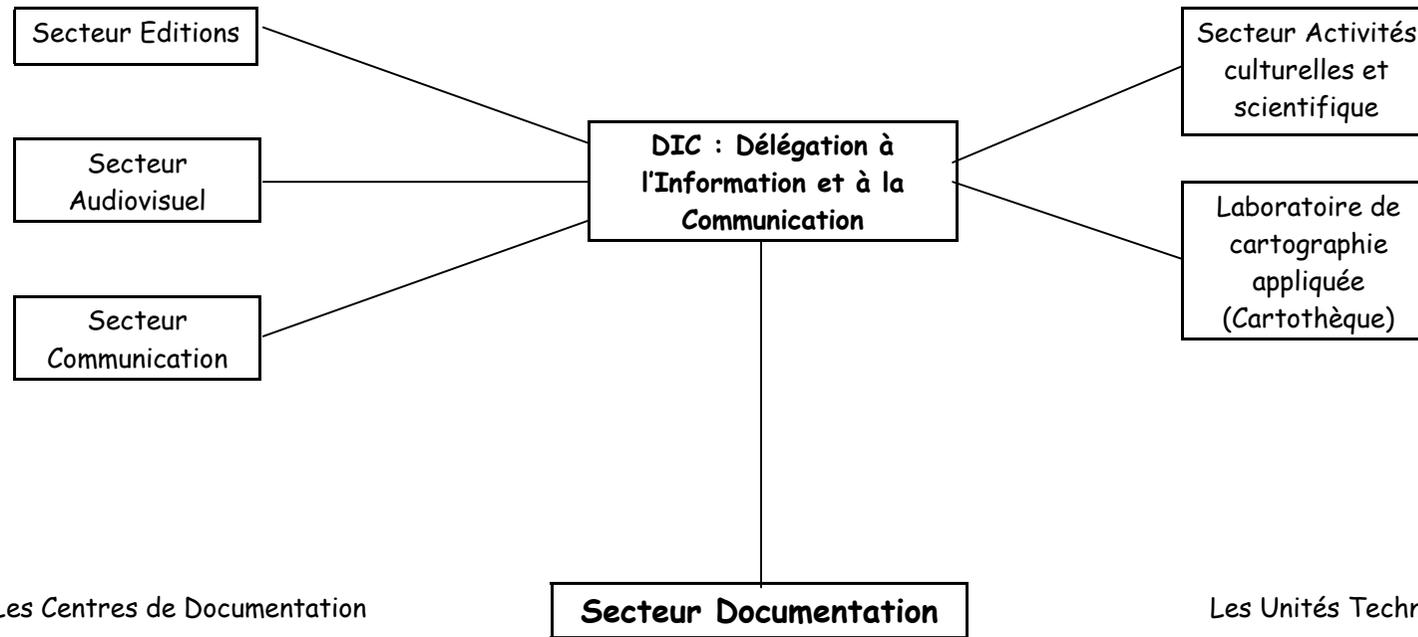
En 1999, le Secteur Documentation s'est efforcé d'améliorer ses services grâce à une utilisation accrue des TIC. Son site Internet est en cours d'achèvement ; il permettra un accès facile à l'ensemble des informations détenues et diffusables par réseau. Compte tenu de la dispersion géographique de l'IRD, les TIC sont indispensables dans la diffusion de l'information. Ainsi, le serveur Web de l'IRD : <http://www.ird.fr> permet un accès à l'information scientifique et technique. (Rubrique : « centres de ressources »)

---

<sup>1</sup> Portail : voir annexe 3 : lexique informatique

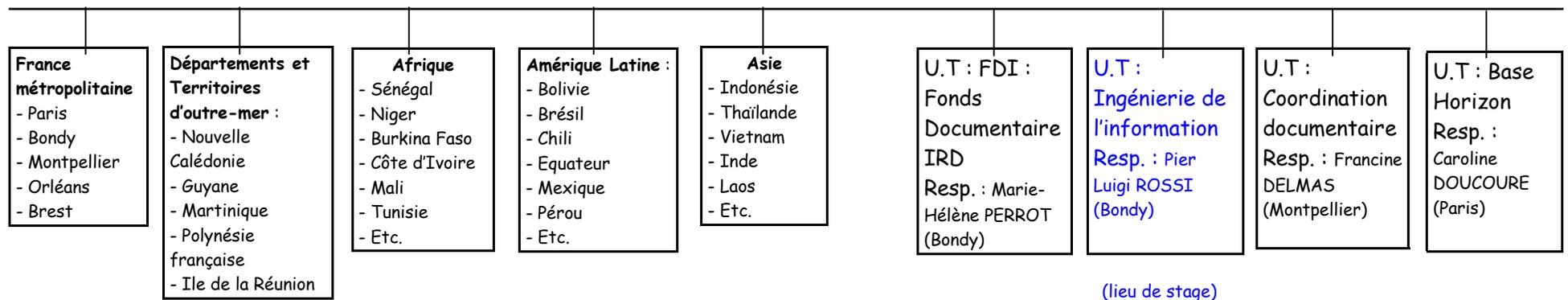
NB : Les différents renvois à cette annexe se feront sous forme de note de bas de page, en raison du nombre de termes figurant dans ce lexique.

**Le secteur documentation dans le département DIC et ses implantations dans le monde.**



Les Centres de Documentation

Les Unités Techniques (U.T.)



## 1.2. Le centre de recherche en Ile de France de l'IRD : Bondy (93)

### 1.2.1. Présentation du centre de recherche

Le centre de recherche de Bondy comporte environ 200 personnes (chercheurs et ITA : ingénieurs, techniciens et agents). Les domaines scientifiques d'intervention du centre sont : les Sciences de la Terre, les Sciences Sociales et l'Informatique appliquée. Environ 80 ITA relèvent de la DIC.

Ce centre de recherche (voir annexe 4 : Organigramme du centre de Bondy) bénéficie d'une position internationale favorable au plan scientifique grâce aux programmes de recherche et aux nombreux réseaux de coopération scientifique auxquels appartiennent les chercheurs et ingénieurs présents dans le centre. Des programmes comme ECOFIT (Ecosystèmes forestiers intertropicaux), Programme international GEOSPHERE-BIOSPHERE, GEOSCOPE ou ADOC (Amélioration des outils des connaissances) y sont développés.

Il entretient également de nombreuses relations avec les pays industrialisés du Nord (Canada, USA, Allemagne, Grande Bretagne, Italie, etc.) pour intervenir en partenariat dans les pays du Sud.

Ce centre de recherche développe des actions de formation liées à son domaine pluridisciplinaire et à ses relations universitaires dans la région Ile-de-France.

Les chercheurs et techniciens du centre sont engagés dans 14 D.E.A., parmi lesquels 4 sont habilités au titre de laboratoire d'accueil universitaire. Quelques 120 stagiaires passent annuellement, par le centre de Bondy, et bénéficient de ces diverses relations.

### 1.2.2. Organisation de l'activité scientifique

L'activité scientifique développée à Bondy s'organise autour de six « Unités - Laboratoires » :

- Laboratoire des Formations Superficielles (Etude des sols)
- Laboratoire de Géodynamique Interne (Géophysique)
- Laboratoire d'Informatique Appliquée
- Laboratoire de Cartographie Appliquée
- Laboratoire Sociétés du Sud et Développement (Sciences sociales : sociologie, anthropologie, géographie humaine etc.)
- Laboratoire d'Écologie des Sols tropicaux

Ces six laboratoires travaillent en relation avec d'autres services présents sur le centre de Bondy, tels que :

- Le Secteur Documentation
- Le Secteur Editions
- Le Secteur Audiovisuel

## 1.3. Le secteur documentation de l'IRD à Bondy

### 1.3.1. Présentation du service de documentation

#### 1.3.1.1. Historique et situation actuelle

A partir du début des années 60, l'IRD a constitué sur le site de Bondy un fonds documentaire de référence comportant l'ensemble de la production scientifique de ses agents. Ce fonds documentaire constitue la mémoire scientifique de l'institut.

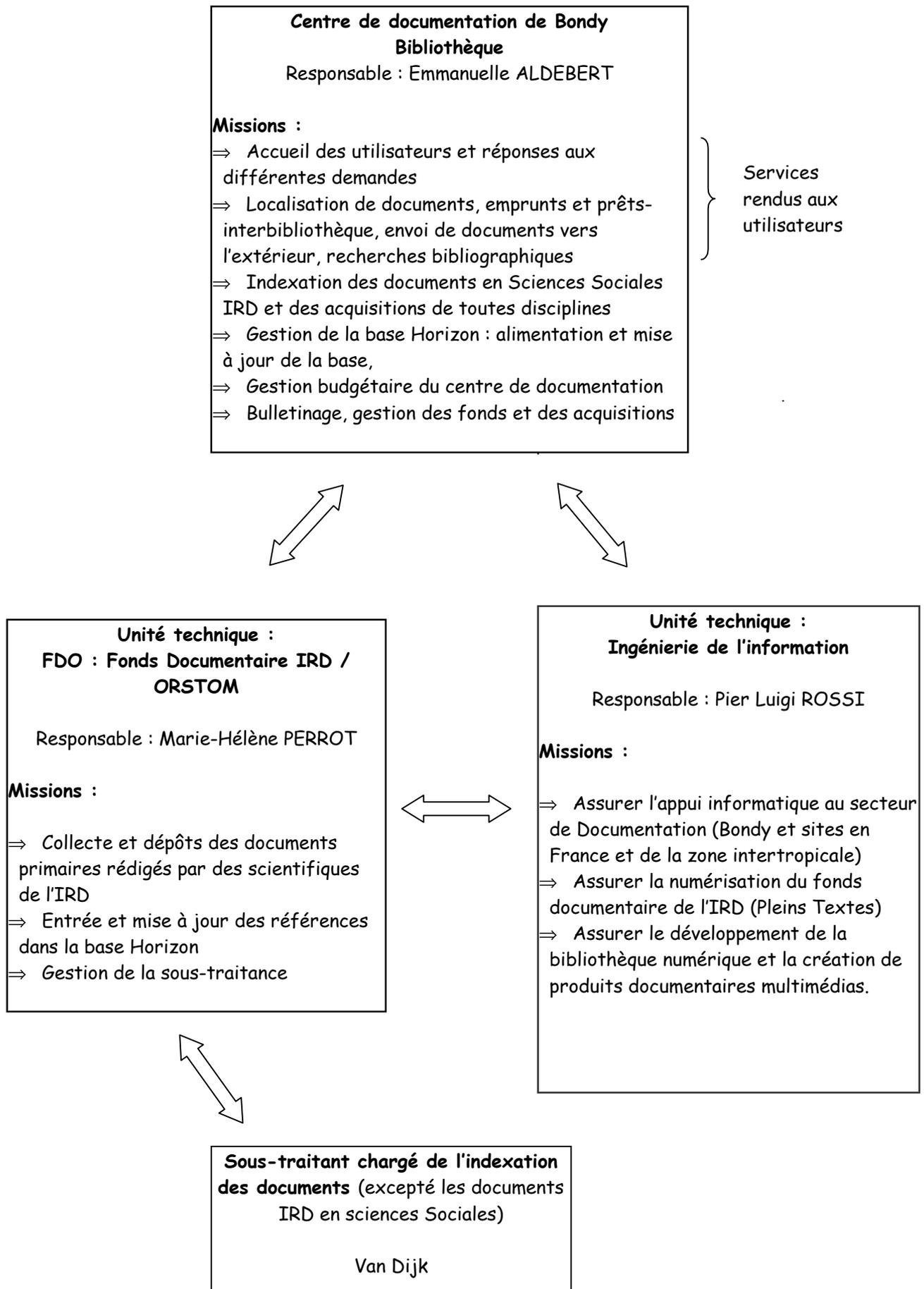
Après une première étape avec un enregistrement manuel des documents et une production de fiches cartonnées, le secteur documentation a été informatisé en 1986 sur des crédits du Ministère de la Recherche. C'est à cette date que la base bibliographique Horizon a été produite. Elle répertorie et décrit chaque unité documentaire (un article, un ouvrage, un chapitre, etc.) présente dans le fonds documentaire.

Actuellement, le secteur documentation continue à alimenter la base Horizon tout en mettant en place une bibliothèque électronique (Pleins Textes) qui permet l'accès aux documents en texte intégral.

#### 1.3.1.2. Composition du Secteur documentation à Bondy

A Bondy, le service de documentation se compose du centre de documentation (la bibliothèque), et de deux unités techniques : le FDI (Fonds Documentaire IRD), et le service d'ingénierie de l'information. Ces trois services travaillent en relation les uns avec les autres, comme on peut le voir sur le schéma page 8, intitulé « Organisation du service de documentation de Bondy ».

## Schéma sur l'organisation du service de documentation de Bondy



### 1.3.1.3. Localisation et aménagement du service de documentation

Bien que les différents bâtiments constituant le centre de Bondy soient implantés sur une assez grande étendue, on peut considérer que le secteur documentation bénéficie d'une position favorable puisqu'il est localisé au rez-de-chaussée du bâtiment principal. C'est donc un lieu facile d'accès.

Le centre de documentation (bibliothèque) est ouvert toute la journée aux chercheurs et scientifiques travaillant à Bondy et à l'IRD. La salle de lecture reste ouverte pendant l'heure du déjeuner. (Voir annexe 5 : plan des locaux du service de documentation de l'IRD Bondy)

### 1.3.2. Le fonctionnement du service de documentation

Bien qu'ayant effectué mon stage au sein de l'unité technique « Ingénierie de l'information », je présente dans ce chapitre l'ensemble des services composant le secteur documentation à Bondy.

#### 1.3.2.1. Présentation du personnel et des différentes fonctions pour chaque service

##### 1.3.2.1.1. Le centre de documentation : la bibliothèque

Le personnel y travaillant est composé de personnel permanent (titulaires de la Fonction Publique) et de personnel temporaire.

#### ➤ Le personnel permanent

**Emmanuelle ALDEBERT** : Responsable du centre de documentation (formation en documentation)

Ses différentes fonctions :

- ✓ Responsable du centre
- ✓ Accueil des utilisateurs, localisation de documents, emprunts
- ✓ Indexation des documents de Sciences Sociales et des acquisitions de toutes disciplines
- ✓ Recherches bibliographiques sur HORIZON
- ✓ Recherches / Exploitation des CD Rom (Pascal, Francis etc.) et des bases de données, telles que Georef, Current Contents, etc.

**Evelyne MILLET** :

Ses différentes fonctions :

- ✓ Accueil des utilisateurs, localisation de documents, emprunts
- ✓ Gestion budgétaire du centre de documentation
- ✓ Gestion des périodiques et des acquisitions
- ✓ Recherches / Exploitation des CD Rom (Pascal, Francis, etc.) et des Current Contents
- ✓ Entrée des acquisitions dans la Base Horizon

### **Fortunée BRAMI :**

Ses différentes fonctions :

- ✓ Localisation de documents, emprunts
- ✓ Entrée des références et gestion des documents en Sciences Sociales pour la base Horizon
- ✓ Fournitures de documents à l'extérieur (prêt - interbibliothèque)
- ✓ Secrétariat

### ➤ **Le personnel temporaire**

**Térésa PETESCH**, en Contrat Emploi Consolidé, à 75% du temps de travail

Ses différentes fonctions :

- ✓ Périodiques : bulletinage, sommaires, aide à la gestion des fonds
- ✓ Participation à la localisation de documents, demandes de photocopies (emprunts et prêts - interbibliothèques)
- ✓ Secrétariat divers

**Valérie HOGNON**, en Contrat Emploi Solidarité, à 50% du temps de travail

Ses différentes fonctions :

- ✓ Participation à la localisation de documents
- ✓ Emprunt et envoi de documents

### **1.3.2.1.2. Les unités techniques du Secteur Documentation présentes à Bondy**

#### ➤ **Le FDI : Fonds Documentaire de l'IRD (ex-ORSTOM)**

Il est dirigé par Marie-Hélène PERROT qui est chargée de collecter tous les documents rédigés par les scientifiques travaillant à l'IRD (Dépôt des documents primaires de l'IRD). Les archives les plus complètes se trouvent à Bondy. Le fonds documentaire débute en 1962.

Je traiterai des différents modes d'acquisitions des documents dans la partie 1.3.6., page 17, concernant les activités de gestion du fonds documentaire.

Mme PERROT est également chargée de la gestion de la sous-traitance.

#### ➤ **Le Service d'ingénierie de l'information**

Ce service est dirigé par Pier Luigi ROSSI. C'est au sein de cette unité technique que j'ai effectué mon stage.

Géologue de formation et ancien documentaliste spécialisé dans les Sciences de la Terre, il s'est davantage orienté vers l'informatique documentaire et est devenu le responsable informatique du Secteur Documentation.

Ses différentes fonctions :

✓ **Assurer l'appui informatique au Secteur Documentation**

Il assure l'appui informatique au Secteur Documentation aussi bien à Bondy, qu'à tous les sites IRD en France et dans la zone intertropicale, en collaboration avec l'unité technique « Coordination documentaire », dirigée par Francine DELMAS à Montpellier.

Les autres centres de documentation peuvent donc faire appel à ces deux unités techniques. Concernant plus particulièrement les centres de documentation situés dans la zone intertropicale, toutes les recommandations sur les outils (logiciels, etc.), leur installation et l'organisation technique en général sont apportées par ces deux unités techniques du secteur documentation.

Le responsable de cette unité technique est donc régulièrement amené à se déplacer dans les structures documentaires situées à l'étranger pour :

- y assurer l'informatisation des bibliothèques et des centres de documentation,
- y assurer le suivi informatique et l'évolution des pratiques et des compétences,
- y organiser des formations sur les nouvelles technologies destinées aux documentalistes et aux différents utilisateurs (sur Internet : méthodologie de recherche, ainsi que sur la numérisation de documents). Ces formations sont menées afin que les personnels des différentes structures maîtrisent les différentes possibilités qu'offrent les nouvelles technologies.

✓ **Assurer la numérisation du fonds documentaire IRD - ORSTOM**

Parallèlement à cette mission d'appui et de conseil informatique, cette unité technique est également chargée d'assurer la numérisation du fonds documentaire IRD - ORSTOM.

Par ailleurs, Pier Luigi ROSSI travaille sur le projet « Infothèque » étroitement lié à la numérisation du fonds. (création d'une bibliothèque numérique)

La numérisation des publications de l'IRD a été décidée essentiellement pour des raisons de diffusion de l'information. En effet, à partir du milieu des années 1990, l'évolution conjointe des technologies de stockage et de diffusion de l'information a rendu possible une mise en accès des documents sur des supports électroniques (Internet, CD Rom). Ce nouveau contexte de diffusion de l'information permet donc à l'utilisateur final de pouvoir consulter les documents sans nécessairement se déplacer (physiquement) dans une bibliothèque : la consultation peut se faire soit en ligne (via Internet ou en intranet), soit sur un CD Rom.

Compte tenu de la dispersion géographique de l'IRD, ceci est essentiel pour les scientifiques et leurs différents partenaires des pays du Sud. En effet, la numérisation du fonds documentaire de l'IRD vise l'élargissement et la simplification de la diffusion des documents en facilitant l'accès direct aux documents en texte intégral.

✓ **Assurer la création et le développement de « Pleins Textes »**

La numérisation massive des publications IRD a permis la création, le développement et la mise en œuvre d'un système de recherche au niveau du contenu des documents, appelé « Pleins Textes ». Cette bibliothèque électronique de référence en matière de recherche dans les pays de la zone intertropicale est accessible sur Internet à l'adresse suivante : [http://www.bondy.ird.fr/pleins\\_textes/](http://www.bondy.ird.fr/pleins_textes/).

Le stage que j'ai effectué émanait donc d'une réflexion sur un autre accès possible aux documents en texte intégral, via une présentation différente de l'information, sous la forme de « tables annuelles des sommaires de revues de l'IRD ».

Dans la deuxième partie concernant le stage effectué, je développerai l'aspect technique de la numérisation et du fonctionnement de « Pleins Textes ».

### **1.3.2.2. Fonctionnement en réseau avec les autres centres**

Le schéma intitulé « Place du Centre de Bondy dans le réseau documentaire IRD », page 13, représente l'architecture informatique du réseau documentaire des différentes structures de l'IRD (centres de documentation et unités techniques) en France et à l'étranger. Je n'ai présenté que quelques centres en France et un en Afrique (Niamey - Niger) pour illustrer ce schéma, mais tous les autres centres fonctionnent de façon identique.

L'architecture informatique est basée sur une interconnexion de réseaux informatiques. En France, les différents réseaux sont interconnectés par RENATER<sup>1</sup> : Réseau National de la Recherche mettant un ensemble de lignes et de câbles loués à France Télécom à la disposition de différentes structures telles que le CNRS<sup>2</sup>, l'IRD, le CIRAD<sup>3</sup>, des instituts de formation, etc. Ce réseau donne l'accès à Internet. La liaison entre les différents centres IRD en France est de type Internet (utilisation du même protocole qu'Internet : TCP/IP<sup>4</sup>).

En ce qui concerne les centres IRD basés dans la zone intertropicale, ils doivent se connecter à Internet pour avoir accès au réseau documentaire IRD en France. Parallèlement à la base HORIZON, ils ont créé des bases de données locales.

---

<sup>1</sup> RENATER : voir annexe 3 : lexique informatique

<sup>2</sup> CNRS : Centre National de Recherche Scientifique

<sup>3</sup> CIRAD : Centre International de Recherche pour l'Agronomie du Développement

<sup>4</sup> TCP/IP : voir annexe 3 : lexique informatique

# Architecture informatique du secteur documentation (Place du centre de Bondy dans le réseau documentaire)

**Brest**

Postes de travail

Réseau informatique  
Brest

Renater

**Bondy**

Station SUN (Unix)

Environ 15 Postes de travail

2 Imprimantes réseau

Réseau informatique  
Bondy

Renater

Texto  
(Unix)

- base Horizon (et son système de production)
- système de prêt de Bondy
- système de prêt d'Orléans
- système de prêt Hydrologie Mpt (34)
- fichier documentaire de Brest
- bibliothèque électronique

Verity (Information Server)

Terminaux graphiques

PC

**Montpellier**

Réseau informatique  
Montpellier

Station SUN

Postes de travail

Imprimantes réseau

Texto

- duplication de la base Horizon (interrogation)
- système de prêt

- Network Computer
- PC

**Niamey (Niger)**

Réseau informatique  
Niamey

Internet  
consultation

Station SUN

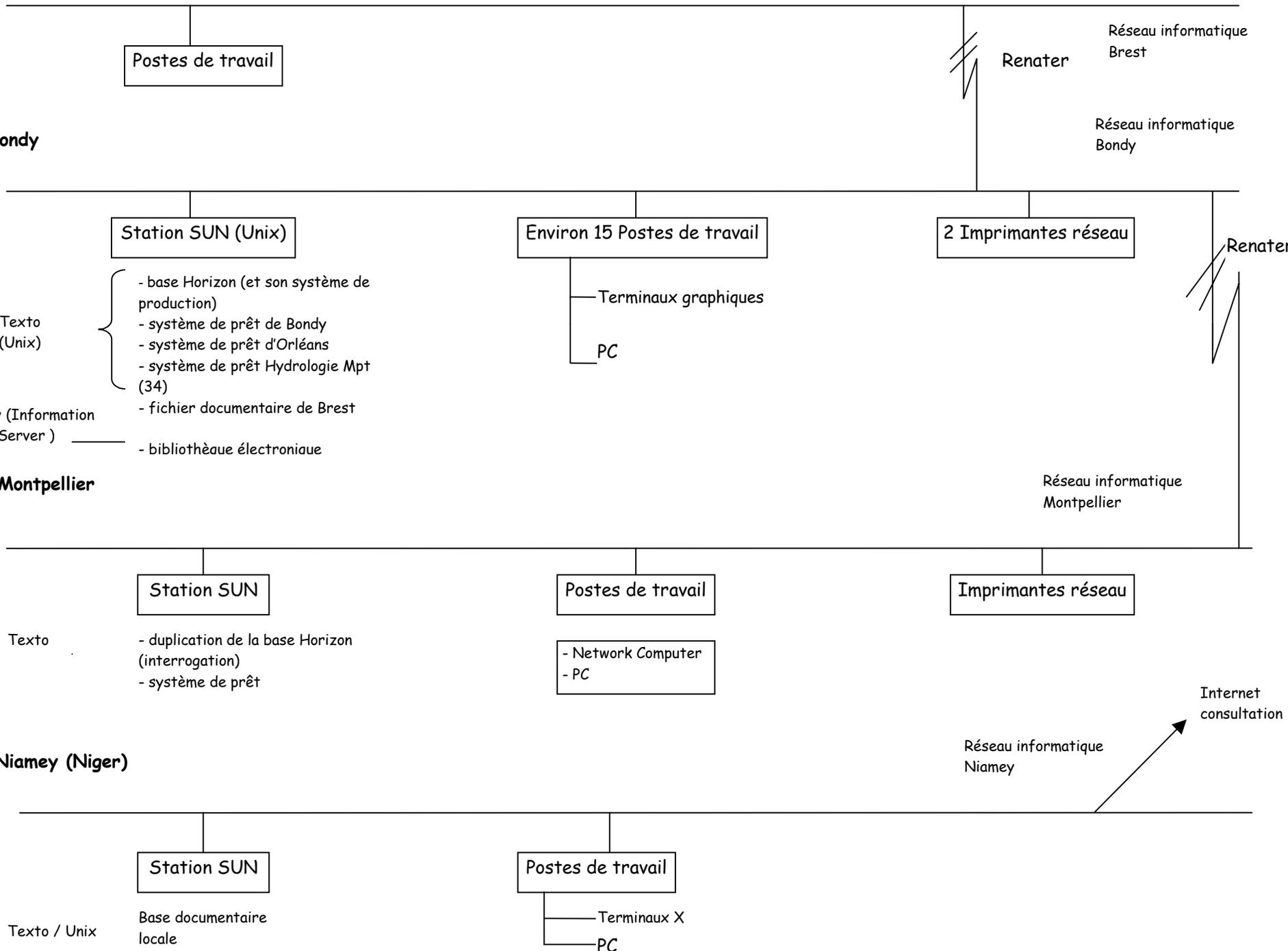
Postes de travail

Texto / Unix

Base documentaire  
locale

Terminaux X

PC



### 1.3.3. Moyens du service de documentation

#### 1.3.3.1. Moyens financiers

Le budget du Secteur Documentation s'élève à 2.000.000 de francs répartis sur tous les centres de documentation en France et les quatre unités techniques.

Pour 1999, le budget alloué à l'unité « ingénierie de l'information » s'élevait à environ 400.000 francs (voir annexe 6 : Bilan budgétaire 1999 de l'unité : ingénierie de l'information). Ce budget vient en partie du secteur documentation et du projet « Infothèque » de la DIC.

Le centre de documentation dispose également d'une somme d'environ 400.000 francs (voir annexe 7 : Rapport d'activité annuel 1999 du centre de documentation).

#### 1.3.3.2. Moyens matériels

##### ➤ Les locaux

Comme je l'ai déjà indiqué, le service de documentation bénéficie d'une position assez favorable au sein des bâtiments constituant le centre de recherche de Bondy. Les deux unités techniques sont situés dans deux bureaux voisins et les cinq personnes travaillant dans le centre de documentation sont réparties dans deux espaces différents : le centre de documentation, lui-même et un autre bureau. L'espace consacré à la salle de lecture n'est pas très important par rapport à l'espace réservé aux archives : « le Nautilus » qui se trouve au rez-de-chaussée du bâtiment et les archives situées dans les sous-sols. (voir annexe 5 : plan des locaux du service de documentation de l'IRD Bondy)

##### ➤ Le matériel informatique

Il est difficile de parler du matériel informatique sans dire quelques mots sur le système informatique utilisé dans le centre de Bondy comme dans l'ensemble des centres IRD.

##### - Système informatique :

Les différents services de Bondy travaillent en réseau sous la plateforme Unix<sup>1</sup>. Ce système a été choisi car il est multitâches et multi-utilisateurs. Par ailleurs, il est apprécié pour sa stabilité.

En fonction des services, on trouve des stations de travail SUN<sup>2</sup>, des PC ou des MAC. Ainsi, les MAC sont très appréciés au Laboratoire de Cartographie Appliquée pour le graphisme.

---

<sup>1</sup> Unix : voir annexe 3 : lexique informatique

<sup>2</sup> station de travail SUN : voir annexe 3 : lexique informatique

#### - Equipement informatique :

Le service de documentation de Bondy compte quinze postes informatiques dont deux stations de travail SUN, sept terminaux X et cinq PC. Je décrirai les caractéristiques du PC sur lequel j'ai travaillé au cours de mon stage dans la deuxième partie.

Le logiciel documentaire utilisé par l'ensemble des structures documentaires du secteur documentation est TEXTO en version 4.6 sous Unix.

Le service de documentation dispose de deux imprimantes réseaux (de marque EPSON) et l'unité Ingénierie de l'information dispose de trois scanners.

### **1.3.4. Utilisateurs internes et externes et leurs types de demandes**

Leur place dans le service de documentation

#### **1.3.4.1. Les utilisateurs internes**

Les utilisateurs internes sont pour la plupart des chercheurs et des stagiaires en DEA ou en thèse.

##### ➤ **Les utilisateurs internes et le centre de documentation (bibliothèque)**

Quand ils sont à la recherche de documents qu'ils ne parviennent pas à localiser eux-mêmes, ils s'adressent au centre de documentation en y déposant une liste de références à rechercher.

Quand ils n'ont pas les références précises des documents, leurs demandes sont souvent thématiques.

En ce qui concerne les recherches dans les bases de données bibliographiques, elles sont centralisées auprès d'Elisabeth AMBERT, documentaliste du centre IRD de Montpellier. Elle se charge également de trouver des sites Internet susceptibles d'intéresser les documentalistes des différents sites IRD (veille).

Dans l'ensemble, les chercheurs ne se déplacent pas souvent au centre de documentation pour différentes raisons :

- Ils ont la documentation nécessaire à leur travail dans leur laboratoire ; ils ne se rendent donc au centre de documentation que lorsqu'ils ne trouvent pas ce qu'ils cherchent, tels que les anciens documents ou les documents difficiles d'accès (se trouvant à l'étranger, par exemple), ou lorsqu'ils souhaitent commander un ouvrage.
- Les chercheurs sont en général autonomes sur le plan de la maîtrise de l'accès à l'information. Ils vont facilement chercher eux-mêmes de l'information sur Internet.

##### ➤ **Les utilisateurs internes et l'unité technique : FDI**

Dans le cadre de la collecte des documents rédigés par les scientifiques de l'IRD, les chercheurs et la responsable de l'exhaustivité du FDI (Mme PERROT) entretiennent

des relations assez privilégiées. En effet, ils réfléchissent ensemble sur la finesse de l'indexation et de l'analyse documentaire en général, sur l'exactitude des informations contenues dans les notices bibliographiques, sur la pertinence d'intégrer ou pas dans le FDI certains documents.

#### ➤ **Les utilisateurs internes et l'unité technique : Service ingénierie de l'information**

Le format PDF<sup>1</sup> étant très utilisé par les scientifiques, certains chercheurs diffusent déjà leurs travaux sous ce format. Parfois confrontés à des problèmes techniques, ils viennent demander des conseils sur la création de fichiers PDF auprès de M. ROSSI.

On constate que les chercheurs se retrouvent au début de la chaîne documentaire en tant que producteurs d'articles scientifiques et en fin de chaîne en tant qu'utilisateurs des bases de données.

#### **1.3.4.2. Les utilisateurs externes**

Le public des utilisateurs externes est aussi bien constitué de chercheurs, d'étudiants que de salariés d'entreprises effectuant des missions dans la zone intertropicale.

Leurs demandes sont thématiques. Les recherches se font sur la base Horizon, si les demandes sont trop pointues dans un domaine, les documentalistes les orientent vers des bibliothèques spécialisées (elles leur donnent les coordonnées des bibliothèques pouvant les aider).

Ces utilisateurs viennent sur place : le centre de documentation leur est ouvert quatre après-midi par semaine sauf le mercredi. (voir annexe 7, page 4 : Rapport annuel 1999 du centre de documentation : statistiques de fréquentation). Ils ne peuvent pas emprunter les documents mais les photocopier (photopies payantes).

Ils peuvent également téléphoner et indiquer leur requête. Les réponses, c'est-à-dire la liste des notices bibliographiques des documents trouvées à l'interrogation de recherche, sont alors envoyées par fax, par courrier ou par courrier électronique. La personne peut ensuite venir sur place munie de cette liste pour consulter les documents. Les documents peuvent également être empruntés par le service des prêts inter-bibliothèques.

#### **1.3.5. Le fonds documentaire**

En France (et à l'étranger), le fonds documentaire de chaque centre de documentation est spécialisé en fonction des domaines de recherche développés par les équipes scientifiques.

Le fonds documentaire IRD-ORSTOM (FDI) et le fonds de la bibliothèque sont deux fonds distincts. Je traiterai essentiellement du FDI. Le fonds de la bibliothèque est

---

<sup>1</sup> format PDF : voir annexe 3 : lexique informatique

constitué de documents généralistes sur les différents domaines intéressant les utilisateurs internes de l'IRD, ainsi que d'usuels (dictionnaires, encyclopédies), d'ouvrages de vulgarisation et de périodiques.

#### **1.3.5.1. Les domaines concernés**

Le Centre de Bondy est spécialisé avant tout dans les Sciences de la Terre et les Sciences Sociales. Les archives les plus complètes de l'IRD se trouvant à Bondy, le fonds documentaires regroupe des travaux de recherche sur d'autres domaines d'intervention.

#### **1.3.5.2. Evaluation du volume des divers types de documents et d'informations**

Le fonds documentaire de l'IRD est constitué d'environ 51 000 documents, composés de

- plus de 35 000 tirets à part de périodiques,
- 10 000 documents de « littérature grise » (thèses, rapports, actes de congrès, etc.),
- plus de 5 000 ouvrages.

#### **1.3.6. Activités de gestion du fonds documentaire**

##### **1.3.6.1. Les acquisitions**

###### **➤ Les documents non produits par l'IRD**

Les acquisitions sont centralisées par le centre de documentation. Les chercheurs peuvent y passer des commandes. Après le traitement documentaire, le document est intégré dans la bibliothèque ou dans la « bibliothèque » du laboratoire scientifique. Quand les chercheurs passent des commandes, l'achat du document se fait sur le budget du laboratoire. Parfois, les laboratoires font des dons de documents au centre de documentation.

Les acquisitions de documents plus généralistes sont faites par les documentalistes. Ces documents restent à la bibliothèque.

###### **➤ Les documents IRD**

Les différents modes d'acquisition de ces publications sont les suivants : le dépôt du document par l'auteur (environ un quart des documents collectés), la veille documentaire et la récupération (dans les centres de documentation IRD en France et dans la zone intertropicale, ainsi que dans d'autres centres de recherche quand les scientifiques travaillent en collaboration ailleurs).

### 1.3.6.2. Le catalogage

A la réception du document, on procède à son enregistrement dans la base Horizon, on lui attribue un numéro d'inventaire et une cote. On effectue une saisie minimum de la notice sur un bordereau de saisie.

Le catalogage des documents IRD est effectué par le sous-traitant Van Dijk, excepté les documents IRD en Sciences Sociales (documents multidisciplinaires).

### 1.3.6.3. Le traitement intellectuel des documents

#### ➤ Les documents IRD

Tous les documents IRD, sauf les documents multidisciplinaires en Sciences Sociales, sont indexés par le sous-traitant Van Dijk. Francine DELMAS, responsable de la « Coordination documentaire » procède ensuite à une relecture des analyses documentaires. Les notices intègrent ensuite la base Horizon, à partir des disquettes fournies par le sous-traitant.

L'indexation des documents multidisciplinaires est traitée par Emmanuelle ALDEBERT. Les centres de Nouméa, Quito, Cayenne et Brest indexent eux-mêmes les documents.

#### ➤ Les documents non produits par l'IRD

Toutes les acquisitions de documents non IRD, ainsi que les ouvrages commandés par les laboratoires scientifiques, sont indexés par Emmanuelle ALDEBERT. Les notices sont ensuite saisies dans la base Horizon par Fortunée BRAMI et Evelyne MILLET.

### 1.3.7. Langages documentaires : Outils pour le traitement documentaire

Les outils utilisés pour le traitement intellectuel des documents sont les suivants :

#### ➤ Plan de classement : deux plans de classement

*Le plan de classement Horizon* : il reprend le plan de classement de la base Horizon pour les documents FDI et les autres documents.

Il existe un *plan de classement secondaire* : utilisé pour attribuer une cote aux documents non FDI ( pour leur emplacement physique) dans la bibliothèque.

#### ➤ Lexique

C'est un lexique intitulé : « Vocabulaire multidisciplinaire » qui a été élaboré par des groupes de travail réunissant des documentalistes et des chercheurs. Les termes ne faisant pas partie de ce lexique sont considérés comme des candidats descripteurs et gérés dans un champ spécifique.

#### ➤ Thésaurus géographique IBISCUS

### **1.3.8. Outils de recherche**

#### **1.3.8.1. Base de données bibliographique : Horizon**

Base de données bibliographique multidisciplinaire, Horizon recense l'ensemble des publications depuis 1944 des équipes de recherche de l'IRD consacrées aux différents aspects de l'environnement physique, biologique et humain des pays en développement de la zone intertropicale. Les sources comprennent des articles de périodiques, des ouvrages, des communications de congrès et la littérature grise.

La base Horizon couvre les thèmes suivants :

Sciences de la terre, Santé, Océanographie et Hydrobiologie ; Sciences du monde végétal et animal ; Sciences fondamentales et techniques ; Sciences économiques et sociales ; Sciences de l'ingénieur et de la communication.

Les mises à jour sont bimestrielles. La base Horizon répertorie également les acquisitions des bibliothèques de l'IRD en France à partir de 1986.

Le nombre de références s'élève à environ 51 000 pour les travaux IRD et 63 405 pour l'ensemble des fichiers.

#### **1.3.8.2. Bibliothèque électronique « Pleins Textes »**

« Pleins Textes », que j'aborderai plus en détail dans la partie 2.1.2., page 24, est une bibliothèque électronique créée par l'unité « ingénierie de l'information » qui permet d'accéder aux documents du FDI en texte intégral. Le format des documents est le PDF en mode « image + texte caché » (voir annexe 8 : Les modes de numérisation).

#### **1.3.8.3. Autres supports électroniques : CD ROM**

Francis et Pascal de l'INIST pour les sciences humaines

#### **1.3.8.4. Autres Bases de données, désormais en ligne**

- Géoref, BDD américaine spécialisée dans les Sciences de la Terre.
- CAB : Commonwealth Agricultural Bureau, BDD britannique
- Current Contents, portant sur les Sciences de la Terre, l'Agriculture, les Sciences de la Vie et les Sciences sociales.

L'ensemble de ces bases est accessible en ligne via Internet pour les personnels de l'IRD des centres de Bondy, Paris, Montpellier, etc.

### 1.3.9. Prestations : Produits et services

#### ➤ **Produits documentaires**

Un bulletin bibliographique a été publié semestriellement par le secteur documentation de l'IRD jusqu'en 1998. Son édition a été arrêtée dans la mesure où le secteur documentation estime trop élevés les coûts de production de ce support et l'accès via Internet largement suffisant pour l'accès aux références bibliographiques.

#### ➤ **Services rendus par le centre de documentation**

- ✓ Accueil des utilisateurs
- ✓ Réponses aux demandes des utilisateurs.
- ✓ Mise à disposition du fonds documentaire
- ✓ Acquisitions thématiques de périodiques ou d'ouvrages
- ✓ Recherches bibliographiques
- ✓ Commandes d'articles et de documents par l'INIST et par le système de prêt interbibliothèque

Dans cette première partie, j'ai décrit et analysé le réseau documentaire et informatique de Bondy, ainsi que l'organisation du service de documentation.

On peut constater que l'informatique tient une place importante au sein de ce service, en raison :

- de la dispersion géographique des différents services de documentation en France et à l'étranger,
- des besoins spécifiques des utilisateurs,
- et de la numérisation de masse du fonds documentaire de l'IRD dans le cadre du projet « Infothèque ».

Au cours du stage, mon travail a été axé sur la conception d'une présentation Web pour réutiliser et mettre en valeur des documents numérisés.

## 2. Création d'un produit documentaire électronique au sein du service informatique du secteur documentation

J'ai effectué mon stage au sein de l'unité « Ingénierie de l'information », ce qui a orienté mon travail vers l'informatique documentaire. La numérisation du fonds documentaire et la création de « Pleins - Textes » sont en partie à l'origine du stage.

### 2.1. Historique du projet de stage

#### Raisons du stage :

##### ➤ Pour l'IRD

Plusieurs raisons expliquent la décision de l'IRD de mettre en ligne des revues qu'il a éditées et numérisées en créant des sommaires des différentes années, ainsi que des volumes et numéros.

- D'une part, la particularité de son principal public, à savoir les chercheurs et les scientifiques, explique ce choix.

Comme on l'a expliqué dans la partie 1.3.4.1., page 15, les chercheurs se caractérisent par une importante autonomie sur le plan de la maîtrise des technologies de l'information. Ils vont facilement chercher eux-mêmes l'information sur Internet.

Par ailleurs, ils possèdent leurs propres documentations dans leur laboratoire, ils se déplacent donc rarement au centre de documentation pour demander des informations.

- D'autre part, l'évolution des technologies de l'information permet la création de nouveaux produits documentaires.

Le développement d'Internet a provoqué un bouleversement dans le rapport à l'information. Ceci a bien évidemment eu des répercussions sur le rôle des documentalistes. De par les caractéristiques propres du public constitué de scientifiques, il est nécessaire pour les documentalistes de leur proposer des produits documentaires sur un nouveau type de support. Ces nouveaux produits documentaires sont élaborés grâce aux technologies de l'information, à savoir la possibilité de numériser des documents, de les mettre en ligne sur Internet et ainsi de les rendre accessibles à une grande échelle. Les chercheurs travaillant dans les pays en développement peuvent, par ce biais-là, également y avoir accès.

D'autres organismes et instituts de recherche proposent déjà des produits documentaires électroniques que nous avons eu l'occasion de découvrir au cours de l'état des lieux, mais on étudiera ce point dans la partie 2.2.2.1., page 28.

## ➤ Pour moi

En ce qui me concerne, j'ai choisi d'effectuer un stage plutôt orienté en informatique documentaire car cela me permettait d'utiliser davantage l'outil informatique. Par ailleurs, l'idée de créer ce qu'on pourrait appeler « un produit documentaire électronique » m'intéressait particulièrement.

Je ne peux pas parler de mon stage sans aborder la question de la numérisation du fonds documentaire IRD et de la création de Pleins Textes (bibliothèque numérique). En effet, la numérisation permet de disposer des productions scientifiques sur un autre support par rapport au papier, ce qui a pour avantage d'enrichir les bases de données en texte intégral.

J'aborderai l'aspect technique de la numérisation et du fonctionnement de Pleins Textes car cela permet d'éclaircir quelques points sur l'aspect technique de mon stage, notamment sur le traitement de l'information à mettre en ligne.

### 2.1.1. La numérisation : aspects techniques

Comme je l'ai déjà indiqué, le projet de la numérisation a été lancé pour des raisons d'élargissement et de simplification de la diffusion de l'information vers un plus large public, notamment vers les partenaires du Sud.

C'est l'unité technique « Ingénierie de l'information », composée de Pier Luigi ROSSI et de Marcel NGOMA-MOUAYA, alors vacataire à l'IRD, qui assure la numérisation des documents IRD, ainsi que son suivi. En effet, une partie des opérations est soustraitée : les traitements « scanner » et « reconnaissance optique des caractères » le sont à 50% des volumes.

Composition du FDI : documents à numériser

Le fonds documentaire IRD (FDI) se compose de 51350 unités documentaires (articles, ouvrages, chapitres, etc.), ce qui représente 2.044.346 pages à numériser.

Les documents édités par l'IRD ont été numérisés en priorité, notamment ceux qui sont disponibles en surnombre sur le site de Bondy. Etant donné que ce sont des publications IRD, il n'est pas nécessaire d'obtenir une autorisation sur les droits d'auteurs et de copyright.

Les cartes thématiques et les couvertures en couleur sont numérisées dans un deuxième temps.

Critères techniques retenus pour effectuer la numérisation : (voir annexe 8 : Les modes de numérisation)

Après avoir réalisé une veille et des tests technologiques sur différents types de documents constituant le FDI, entre 1996 et 1997, l'unité « Ingénierie de l'information » du secteur documentation a pu définir une approche générale de la numérisation selon ces critères techniques :

➤ Création des fichiers des documents numériques au format PDF,

- traitement de reconnaissance optique de caractères en mode image plus texte caché pour les pages numérisées en noir et blanc,
- numérisation des documents en mode bitonal (noir et blanc) à 300dpi,
- numérisation en mode « bitonal - image » pour les pages en couleur,
- numérisation en couleur à 150 dpi pour les planches hors texte et les couvertures en couleur.

Au cours de la phase de veille, le service informatique a mis en place un petit atelier de numérisation tout en se familiarisant avec les différentes technologies mises en œuvre. A partir de cette expérience, il a soumis le cahier des charges à des prestataires pour obtenir des propositions financières permettant de lancer la numérisation de masse. Actuellement la numérisation est sous-traitée auprès d'un prestataire de service choisi en fonction de son savoir-faire et de ses propositions financières.

Environ 10% du fonds numérique actuel a été entièrement traité en interne.

#### ✓ **Création de fichiers des documents numériques au format PDF**

Le format PDF a été choisi en raison de :

- sa « portabilité multiplateforme »<sup>1</sup>  
En effet, un document au format PDF se présente de la même façon sur l'écran et en imprimé, indépendamment du type d'ordinateur ou d'imprimante que l'on utilise et indépendamment du logiciel avec lequel le document a été créé à l'origine.
- son optimisation pour la diffusion sur Internet  
Bien que les fichiers PDF contiennent la mise en forme complète du document original, ce sont des fichiers très condensés. On peut donc télécharger efficacement et rapidement des informations complètes. Ce format permet donc de visualiser et d'imprimer très facilement des documents assez conséquents pouvant comporter des graphiques ou des cartes et pour lesquels le respect de la mise en page est important.
- ses technologies de reconnaissance optique de caractères associant l'image originale au texte reconnu (conservation de la mise en page originale avec les polices de caractères et les images).  
Le mode « image + texte caché » a paru essentiel pour le projet de numérisation en masse : la représentation du document est conforme à l'original et le texte reconnu est utilisé pour l'indexation et la recherche dans le cadre de la bibliothèque électronique, « Pleins Textes », ainsi que par l'utilisateur final. En effet, ce dernier doit pouvoir imprimer le document et éventuellement récupérer le texte pour l'exploiter.
- sa très large utilisation par les communautés scientifiques

---

<sup>1</sup> Portabilité multiplateforme : voir annexe 3 : lexique informatique

## ✓ Le problème de la couleur dans les documents à numériser

S'agissant d'un fonds documentaire scientifique, la problématique de la numérisation des documents comportant des illustrations en couleur s'est posée. Il a donc fallu passer en revue les différents matériels et technologies existant et tenir compte des modes opératoires (manipulations) pour la numérisation ainsi que les contraintes imposées par les traitements de reconnaissance optique de caractères.

Le choix s'est porté sur le mode « bitonal » (pour le noir et blanc) et « bitonal-image » (pour la couleur). En effet, lors de la compression des fichiers numérisés, ce mode n'entraîne pas une perte de qualité pour les fichiers.

La mise au point d'une procédure de numérisation pour les cartes thématiques et les couvertures en couleur garantit la conservation des hauts niveaux d'information et de signification caractérisant ces composants des documents scientifiques.

### **Bilan de la numérisation dressé par le service informatique**

Actuellement, environ 6000 documents ont été numérisés, ce qui représente environ 300.000 pages.

L'étude des procédures et des technologies de la numérisation a permis au service informatique de mieux négocier les coûts des traitements auprès des prestataires. Le coût moyen de la numérisation d'une page est d'environ 1,10 F, ce qui porte le coût global de la numérisation du fonds documentaire à environ 2 250 000 F.

Elle lui fournit également l'expérience et le savoir faire nécessaires à la mise en œuvre de projets de numérisation auprès des organismes partenaires de l'IRD, particulièrement dans les pays en développement.

La numérisation des documents a permis la mise en ligne de « Pleins Textes », la bibliothèque électronique de l'IRD.

### **2.1.2. Pleins Textes**

La bibliothèque électronique « Pleins textes » est accessible à l'adresse suivante : [http://www.bondy.ird.fr/pleins\\_textes/](http://www.bondy.ird.fr/pleins_textes/). Elle a été créée à l'aide du progiciel Verity<sup>1</sup> (Information Server). Celui-ci fonctionne comme un moteur de recherche et d'indexation en texte intégral et a été appliqué à la base bibliographique Horizon et aux documents numérisés. (voir annexe 9 : Article de Pier Luigi ROSSI sur Pleins Textes)

Actuellement, 300 000 pages numérisées peuvent être consultées en ligne.

Il me semble important de décrire le fonctionnement (technique) de cette base (voir annexe 10 : Schéma intitulé « Fonctionnement de Pleins Textes ») car le traitement des fichiers informatiques de <http://www.bondy.ird.fr/tdp> s'en inspire.

---

<sup>1</sup> Progiciel Verity : voir annexe 3 : lexique informatique

La base de cette bibliothèque électronique offre la possibilité d'effectuer des recherches sur les notices bibliographiques (comme dans la base Horizon), mais aussi sur le contenu des documents.

Pour cela, des relations basées sur un identifiant unique ont été établies entre les fiches bibliographiques de la base Horizon et les documents électroniques (numérisés). L'identifiant unique est le numéro d'inventaire du document présent dans la notice bibliographique et il devient le nom du fichier électronique (fichier PDF).

Un programme informatique, diffusé gratuitement, a été développé par l'unité « Ingénierie de l'information » pour permettre de transférer les informations documentaires dans les fichiers électroniques.

En effet, les champs (structurés) : « auteurs, titre, sujet et mots clés » sont récupérés automatiquement à partir de la base bibliographique et sont injectés dans les fichiers PDF et dans le système d'indexation d'Information Server (Verity).

En fonction de ses besoins, l'utilisateur peut opter pour la recherche sur l'information structurée en champs (auteurs, titre, etc.) et la recherche sur le contenu des documents. Ce dernier mode de recherche lui permet de repérer les parties les plus pertinentes des documents.

Pour chaque document électronique au format PDF, le positionnement (de la réponse) se fait dans la page contenant l'élément pertinent (mot, phrase, paragraphe, page) et non plus sur l'intégralité du ou des documents.

« L'élément pertinent » est mis en surbrillance et seule la page le contenant peut être téléchargée par l'utilisateur.

Le développement de Pleins Textes progressant au fur et à mesure de la numérisation, l'accès en ligne aux notices bibliographiques sur la base Horizon est assuré pour que les utilisateurs disposent de l'intégralité des références des unités documentaires.

Bien que Pleins Textes et la base bibliographique Horizon aient un format de fichiers différents (HTML<sup>1</sup> pour Horizon et PDF pour Pleins Textes), ils sont accessibles en même temps pour une requête identique. Il suffit de préciser sur le formulaire de recherche qu'on veut également interroger la base Horizon en cochant la case correspondante.

Des notions de catégories thématiques, d'associations de documents par groupes de mots communs, par thésaurus géographiques ont été intégrées dans le mode de recherche de Pleins Textes pour affiner la recherche et augmenter la pertinence des réponses obtenues.

---

<sup>1</sup> HTML : voir annexe 3 : lexique informatique

La figure ci-dessous présente le formulaire d'interrogation élaboré offrant en ligne ces différentes possibilités d'explorations.

Pleins Textes

Vos commentaires

Entrez des mots ou des phrases séparés par des virgules

Recherche dans :

Le texte  Operateurs entre les mots Proche à 5 mots

Les auteurs  OU

Les mots clef  OU

Le titre  OU

Opérateur de recherche entre les champs : ET

Disciplines : Aucune

Liste des collections disponibles:

Base bibliographique HORIZON  Pleins\_Textes

Classement global des documents :  
triés par : La pertinence  Ordre décroissant

Classement des documents par ensembles (clusters)  
(Traitement un peu lent, résultat intéressant ...)  
Finesse des ensembles : Par défaut

Rechercher Annuler

Vos commentaires

L'identifiant unique permet de réaliser des extractions fines de documents numériques à partir des résultats de recherche dans la base Horizon. Ce système de relations entre les notices et les fichiers permet également de générer différents types de pages de présentation ou de classification pour lesquelles les liens vers les fichiers PDF sont générés automatiquement, comme cela a été le cas pour <http://www.bondy.ird.fr/tdp>.

### 2.1.3. Objet du stage

L'objectif du stage, tel qu'il a été formulé, « est de réaliser des pages de présentation des revues et des collections éditées par l'IRD. Chaque page donnera accès aux tables annuelles et/ou aux titres des ouvrages. Chaque entrée des tables donnera accès aux documents en texte intégral. Le modèle est celui des présentations des revues type Elsevier (« Science Direct ») ou Springer (« Link »). Les pages réalisées seront accessibles sur Internet. »

On peut ajouter que par rapport à la bibliothèque électronique « Pleins Textes », la création des sommaires électroniques des différents numéros de revues est une autre façon de traiter et de présenter l'information.

## 2.2. « L'information à mettre en ligne »

Avant la création de pages destinées à Internet, il est indispensable de réfléchir sur l'information à mettre en ligne (contenu, présentation, etc.) et d'établir un plan bien défini avant de travailler avec le texte et les images et de créer des liens car le document est ensuite difficile à mettre à jour et à développer.

Voici la présentation du plan de travail tel qu'il a été fixé et suivi pour l'élaboration de la présentation « tdp » :

- Réfléchir sur le contenu de l'information à mettre en ligne : Que veut-on mettre en ligne ? Quel type d'information ?
- Etablir un « état des lieux » de ce qui existe dans le domaine pour avoir une meilleure idée sur la présentation de l'information
- Faire le bilan de cet état des lieux : « réutilisation » des informations trouvées chez les autres : comment et dans quel but ?
- « Préparer l'information » à mettre en ligne : traitement de l'information pour la présenter d'une certaine façon
- Gérer les fichiers informatiques comprenant les informations à mettre en ligne

Une fois ce travail préliminaire effectué concernant l'information et son environnement, on peut se lancer dans la création de la présentation Web, mais un travail préparatoire à l'élaboration du site est également nécessaire :

- Réfléchir sur l'arborescence du site : les différentes étapes
- Tenir compte des contraintes pratiques rencontrées et à respecter : charte graphique de l'IRD (voir annexe 11 : Charte graphique de l'IRD)

### 2.2.1. Contenu et présentation de l'information

Ce sont des revues IRD numérisées qu'on veut mettre en ligne, avec les différentes notices fournissant un accès aux documents en texte intégral.

Quant à la présentation de cette information, il était prévu de s'inspirer de la présentation des sommaires des revues de deux éditeurs scientifiques, à savoir : Elsevier (« Science direct ») et Springer (« Link »).

### 2.2.2. Etat des lieux de l'existant

Au début du stage, nous avons établi un état des lieux sur d'autres centres et organismes de recherche. Ceci est nécessaire pour réfléchir sur la présentation de l'information à mettre en ligne.

Afin de trouver rapidement l'adresse de différents sites d'organismes scientifiques, j'ai utilisé le site du Ministère de la Recherche, qui les répertorie sur une liste. Ainsi j'ai pu analyser et comparer les présentations de sommaires des publications des différents organismes.

Ce travail de recherche sur Internet m'a permis de voir ce qui est fait dans le domaine et si les articles sont accessibles en texte intégral.

### 2.2.2.1. Rapide panorama de ce qui existe

Cette liste n'est pas exhaustive car je n'ai retenu que quelques adresses parmi tous les sites visités dans différents établissements et éditeurs scientifiques.

#### ➤ Quelques centres de recherche

- le CNRS : Centre National de Recherche Scientifique : par l'INIST

Les sommaires des périodiques ainsi que les notices des articles (numéro de la revue, année, titre de l'article) sont accessibles à l'adresse suivante : <http://services.inist.fr/public/fre/conslt.htm>. L'arborescence est la suivante : une notice pour présenter la revue et une page de sommaires avec les différentes années. L'accès aux articles en texte intégral est payant.

- l'INRA : Institut National de Recherche Agronomique : <http://www.inra.fr>

propose un « inventaire des publications » des chercheurs de l'INRA.

La recherche peut se faire en interrogeant plusieurs champs : mots-clés, auteurs, titre de l'article, date, etc. Les réponses trouvées se présentent sous forme de notices bibliographiques. Le titre du document (présent sur la notice) permet d'accéder au résumé, mais pas au texte intégral. Ce site ne propose pas de sommaire des périodiques ou des publications des chercheurs de l'INRA.

- le CIRAD : Centre International de Recherche pour l'Agronomie de Développement <http://www.cirad.fr>

On peut accéder aux publications du CIRAD en texte intégral par la rubrique : « Prestations et produits » - IST. La plupart des documents en texte intégral sont des pages HTML (documents HTML). Seuls quelques articles sont accessibles au format PDF.

- L'Institut Pasteur : <http://www.pasteur.fr>

Site très intéressant, sur lequel nous avons trouvé des « produits documentaires » tels qu'une liste d'éditeurs scientifiques, des listes de périodiques scientifiques, des annuaires des thèses en ligne, etc. Ces informations sont importantes et utiles pour les chercheurs et scientifiques qui ont ainsi directement accès à l'information.

L'accès aux articles en texte intégral (au format PDF) est payant.

➤ L'Agence Universitaire de la Francophonie : <http://www.aupelf-uref.org/revues/>.

On y trouve les articles en texte intégral composant les Cahiers Sécheresse, Cahiers Santé et Cahiers Agriculture. Les périodiques en ligne sont accessibles au format HTML mais pas au format PDF.

➤ Deux éditeurs scientifiques : Elsevier et Springer.

Ces éditeurs proposent déjà des sommaires, des index d'auteurs, des index de mots-clés en ligne.

La visite des sites Internet de ces éditeurs a soulevé la question concernant le coût de l'information, dont M. ROSSI et moi avons discuté.

Elsevier a été le premier éditeur à rendre payant l'accès à certains documents. Or on trouve sur le site de cet éditeur plus de textes intégraux gratuits que sur le site de Springer. Est-ce une nouvelle politique menée par Elsevier ? L'information « redevient-elle » accessible gratuitement ? Sur le site de Springer, tous les textes intégraux des articles sont payants.

La position de l'IRD sur ce sujet est la suivante : L'IRD est un service public et il souhaite mettre en ligne l'information gratuitement au format PDF. L'IRD étant un établissement scientifique il est par conséquent nécessaire de mettre en ligne l'information dont les scientifiques ont besoin et la rendre accessible aux pays en développement.

#### **2.2.2.2. Utilisation des informations trouvées sur les différents sites visités**

Ces informations ont été utilisées en suivant ces différentes étapes :

- Dépouillement des différents périodiques et de leurs sommaires, au moyen d'Acrobat Reader<sup>1</sup> :  
Enregistrement, sous le format PDF, des pages HTML intéressantes des sites visités (à l'aide d'Acrobat), visionnage et téléchargement des pages au moyen d'Acrobat. Cette méthode a pour avantage de pouvoir garder la mise en forme initiale du document et de n'imprimer qu'une page à la fois (celle qui nous intéresse) contrairement aux pages HTML qu'on peut plus difficilement imprimer individuellement.
- Impressions sur papier des modèles intéressants.
- Analyse des pages sources (en langage HTML) des exemples pertinents en réfléchissant sur la difficulté et la façon de refaire les pages.
- Choix d'une présentation possible (**voir annexe 12 : modèle de page choisi**)

#### **2.2.2.3. Objectifs de cet état des lieux**

Ce travail de recherche sur Internet, qui a duré environ deux jours, avait pour objectif de :

- ✓ Recenser et analyser ce qui existe dans le domaine (« présentation des tables annuelles de sommaires de périodiques »)
- ✓ Voir si les documents proposés sont accessibles en texte intégral et sous quel format (HTML ou PDF).

Cet état des lieux m'a permis d'analyser la façon de présenter et d'organiser l'information scientifique et technique par différents instituts de recherche et des

---

<sup>1</sup> Acrobat Reader : voir annexe 3 : lexique informatique

éditeurs scientifiques. Par la même occasion, j'ai été amenée à constituer des signets (favoris) des pages et sites pertinents.

Ceci m'a éclairée pour la suite de mon stage car je n'étais pas du tout familiarisée avec le domaine scientifique.

J'ai constaté que les sommaires de périodiques scientifiques mis en ligne sur Internet sont présentés de façon assez identique. C'est une présentation claire permettant un accès rapide à l'information. Par ailleurs, de nombreuses entrées sont généralement proposées pour accéder à l'information, par le biais d'index des auteurs, d'index des mots - clés, etc.

Les sites Internet d'organismes scientifiques proposant l'accès aux articles de périodiques en texte intégral et au format PDF ne sont pas très nombreux. D'une part, quand les documents sont accessibles, ils sont encore fréquemment au format HTML. Les documents au format PDF supposent la numérisation du document primaire. D'autre part, quand les articles sont accessibles en texte intégral et au format PDF, ils sont souvent payants. Il faut s'abonner ou payer en ligne pour pouvoir les imprimer.

Cet état des lieux nous a donc été très utile puisqu'il nous a servi de base de travail pour créer une maquette de présentation de tables annuelles des revues.

Nous avons décidé de faire apparaître ces différentes informations dans un tableau constitué de deux colonnes :

Titre de la revue - Année, volume, numéro	
Pagination	Titre de l'article Auteur Lien vers le document au format PDF

Nous avons du traiter l'information que nous souhaitons mettre en ligne pour qu'elle soit présentée de cette façon et qu'elle soit correctement classée en fonction de l'année, du volume, du numéro, ainsi que de la pagination.

### **2.2.3. Classement de l'information : « préparation de l'information »**

Avec M. ROSSI, nous avons interrogé la base Horizon sous Texto et procédé à diverses manipulations pour créer de nombreux fichiers informatiques comportant les informations que nous voulions intégrer dans les tables annuelles (sommaires et index des auteurs).

#### **2.2.3.1. Classement des notices**

Nous avons travaillé sur quatre revues, faisant partie des documents numérisés, les unes après les autres. (Cahiers ORSTOM, série Biologie, Cahiers ORSTOM, série Pédologie, Revue de Nématologie et Cahiers des Sciences humaines)

➤ **Pour générer le fichier « tables\_annuelles » (sommaires des revues)**

Voici les différentes étapes :

- ◆ Interrogation de la base Horizon sous Texto pour chaque revue.
- ◆ Création d'un index de classement des différentes notices de la revue, selon les critères suivants : année, volume, numéro et pagination.

Nous n'extrayons que certains champs, qui nous intéressent, contenus dans les notices, à savoir le titre de l'article et l'auteur que nous classons selon les critères cités ci-dessus. Nous avons ainsi créé un index de classement qui constitue un fichier de travail. Tous les champs contenus dans la notice complète ne sont pas repris, comme les mots-clés, car on ne peut pas retrouver les articles des tables annuelles par une interrogation de recherche.

Chaque notice est associée au document numérisé (au format PDF) correspondant grâce à son numéro d'inventaire (l'identifiant unique d'un document dans le fonds documentaire).

L'index de classement des notices par année, volume, numéro et pagination a donc permis la création des fichiers composant les sommaires.

Nous avons effectué ces interrogations de la base horizon et ces index pour chaque revue mise en ligne.

- ◆ Lancement d'un programme Logotel<sup>1</sup> pour générer un fichier HTML sous une présentation particulière et définie en fonction de ce que nous voulions faire sur les pages web, pour ensuite éviter de procéder à trop de rectifications « manuelles » sur les pages HTML.

Nous avons modifié le programme Logotel : « tables annuelles » au fur et à mesure des différentes revues traitées, dans le but d'automatiser le plus possible la création de ces fichiers HTML.

Nous avons amélioré la présentation des pages en rajoutant des balises HTML<sup>1</sup>, telles que <br> qui représente un retour chariot (saut de ligne). Nous avons également modifié la numérotation des fichiers PDF qui n'était pas homogène. En effet, le nom (numéro d'inventaire) de certains de ces fichiers commençait par le chiffre « 0 » (ex : 012456.pdf) et d'autres pas. Ceci entraînait l'existence de liens non valides vers certains documents au format PDF.

Nous obtenons donc un fichier HTML composé d'une succession de notices.

---

<sup>1</sup> Programme Logotel : voir annexe 3 : lexique informatique

<sup>1</sup> Balise HTML : voir annexe 3 : lexique informatique



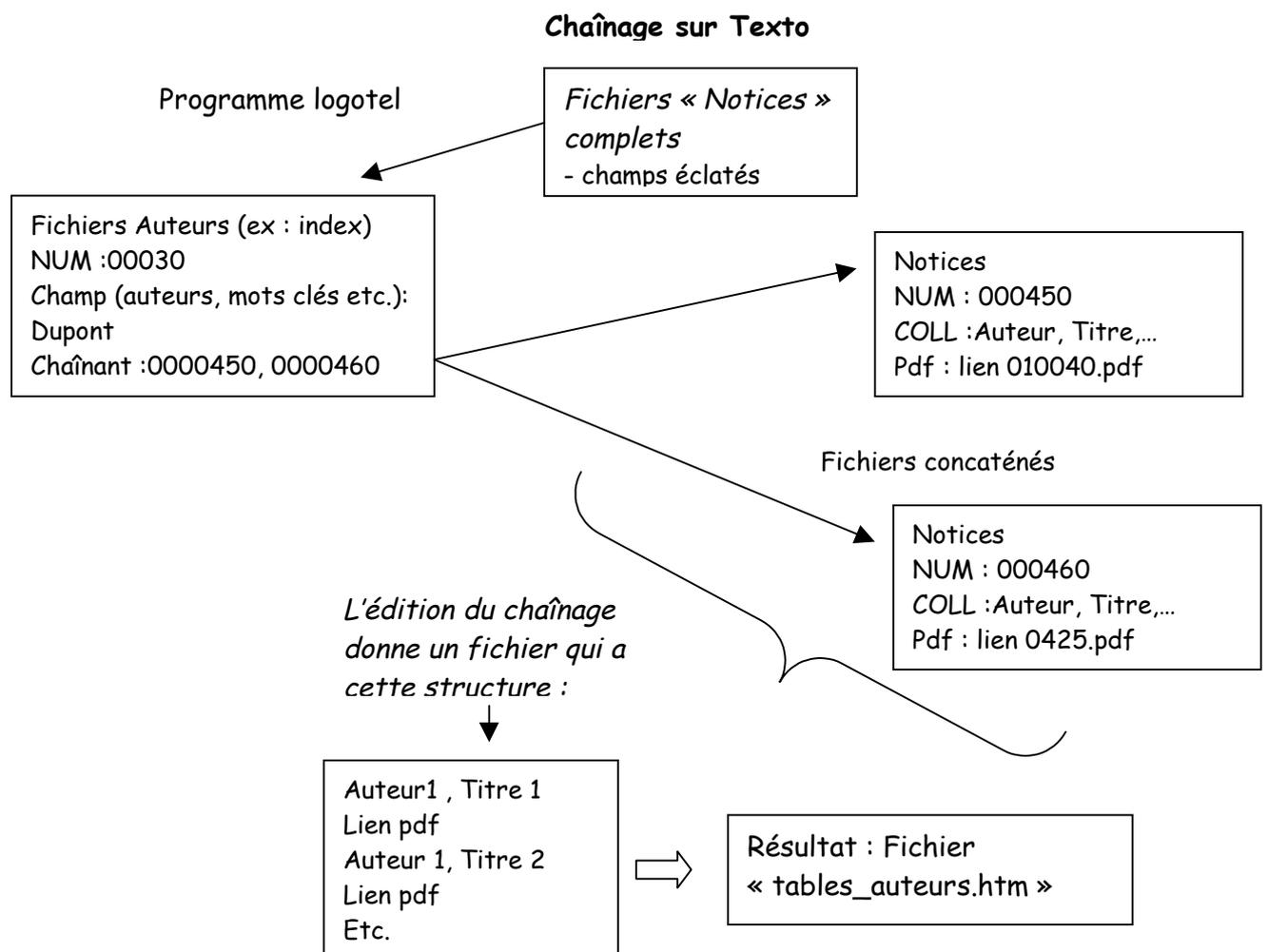
➤ Pour générer le fichier « tables\_auteurs » (index des auteurs)

En ce qui concerne la création de l'index des auteurs, nous avons suivi les mêmes étapes que pour la création des tables annuelles des sommaires. La seule différence se trouve dans la nature du fichier « tables\_auteurs ».

En effet, nous avons utilisé le fichier « tables\_annuelles » comme point de départ. Nous avons ensuite créé un fichier factice (fichier de travail) qui procède à la concaténation<sup>1</sup> des différents champs : auteurs, titre de l'article, titre de la revue, année, volume, numéro et pages dans un seul champ pour une meilleure présentation dans le format HTML. Ce nouveau champ regroupe ces informations, sans changer le champ « Auteur », ni la clé d'entrée de chaque notice. Ensuite un index sur les auteurs est créé.

La création du fichier « tables\_auteurs » a été possible grâce au chaînage des champs « Auteurs » du fichier « tables\_annuelles », comme le montre le schéma ci-dessous.

### Chaînage sur Texto pour la création d'index des auteurs



<sup>1</sup> Concaténation : voir annexe 3 : lexique informatique

Les tables annuelles et les index des auteurs constituent deux moyens différents d'accéder aux notices bibliographiques et indirectement aux documents en texte intégral. Ce sont deux entrées différentes vers l'information.

#### ➤ Difficultés rencontrées :

Lors de la création du fichier « tables\_annuelles », nous avons été confrontés à des problèmes de classement des notices composant cet index. Ceci était dû aux différences d'écriture : ex : la pagination : « p. 1-12 » était parfois notée de la façon suivante : « P. 1-12 ». Par ailleurs, nous avons constaté à plusieurs reprises que la pagination du document avait été oubliée dans la notice lors de l'analyse documentaire.

**Question soulevée à cette occasion :** La nécessité de mettre en œuvre un contrôle de contenu des notices (notion de qualité de l'information).

### 2.2.3.2. Découpage des fichiers informatiques

Le fichier global des notices que obtenu après la création de l'index de classement pour chaque revue (ex : « tables\_biologie.htm ») comporte toutes les notices des Cahiers ORSTOM, série Biologie les unes à la suite des autres, ainsi que des balises HTML qui structurent la présentation de cette liste.

Ce fichier est généré directement par un programme Logotel et un document d'édition Textó.

```

<table>
<tr>
<td align=left valign=top>
<NOBR>39 p.</NOBR>
</td>
<td align=left>
<b><font face=arial>Donn&eacute;es r&eacute;centes sur la physiologie du palmier &agrave; huile<br>Pr&eacute;vot, Pierre<br>
<a href="/pleins_textes/pleins_textes_4/biologie/17502.pdf">Document au format PDF</a></td>
</tr>

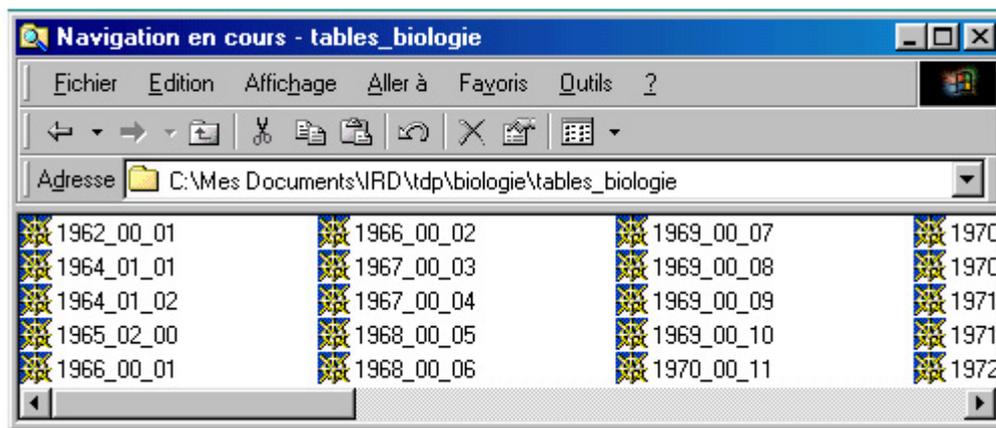
<tr>
<td align=left valign=top>
<NOBR>p. 1-63</NOBR>
</td>
<td align=left>
<b><font face=arial>La physiologie du caf&eacute;ier. 1. De la germination &agrave; la morphog&eacute;
Rab&eacute;chault, Henri; Cambrony, H.; Pr&eacute;vot, Pierre (pref.)<br>
<a href="/pleins_textes/pleins_textes_4/biologie/17503.pdf">Document au format PDF</a></td>
</tr>

<tr>
<td align=left valign=top>
<NOBR>p. 5-34</NOBR>
</td>
<td align=left>
<b><font face=arial>La physiologie du caf&eacute;ier. 2. La croissance (syst&egrave;me radica
Rab&eacute;chault, Henri; Cambrony, H.<br>
<a href="/pleins_textes/pleins_textes_4/biologie/17504.pdf">Document au format PDF</a></td>
</tr>

```

Il est nécessaire de procéder au **découpage informatique du fichier global** en différents fichiers correspondant aux différentes pages des tables annuelles. Ces fichiers sont classés par année, volume, numéro et pages.

A partir des informations contenues dans ce fichier, on crée différents fichiers nommés par l'année, le volume et le numéro. Ex : 1969\_00\_07 : Année 1969, Volume 0 : pas de volume et Numéro 7. Tous les fichiers informatiques sont nommés de la même façon même si les numéros de revues n'ont pas de volume. Il est indispensable de définir une logique dans l'attribution du nom des fichiers car ils sont très nombreux.



L'opération de découpage du fichier global va de paire avec la phase de création des différentes pages HTML. Cependant, je le place volontairement avant la partie sur la création de la présentation web car j'ai préparé ces fichiers avant de créer réellement les pages. J'ai regroupé les notices faisant partie du même numéro, volume et année en les « pré-nommant » par leur futur nom de fichier. Ceci me permettait par la suite, lors de la création finale des pages HTML, de repérer rapidement les notices faisant partie du même fichier et de procéder à un « copier-coller » sur le modèle de page HTML précédemment créé. Nous reparlerons de la page HTML - type comportant tous les éléments obligatoires figurant dans la charte graphique dans la partie 2.3.1.2, page 40.

### Vérification des notices dans les revues elles-mêmes

Tout en découpant le fichier global, je vérifiais également les notices dans les revues (dans les documents primaires) afin de noter les éventuels problèmes.

En voici quelques-uns dont j'ai fait part à Marie-Hélène PERROT, responsable du FDI pour qu'elle puisse corriger les notices de la base Horizon :

- problèmes de saisie : utilisation de majuscule au lieu de minuscule : Ex : P. 2-20 au lieu de p. 2-20 ; la notice d'un article faisant partie des Cahiers ORSTOM, série Géologie était saisie dans les Cahiers ORSTOM, série Pédologie
- problèmes de classement : parfois des problèmes de date de publication d'un document figuraient sur les notices
- l'absence de pagination : la pagination de l'article n'est pas indiquée
- la présence de doublons de notices dans la base : l'erreur est due au fait que le prénom de l'auteur apparaissait une fois par son initiale et une autre fois par le prénom en entier.
- l'absence de notice enregistrée dans la base pour un article des Cahiers ORSTOM, série Biologie
- problème d'orthographe sur le nom de certains auteurs : ex : HOMONICK pour HOMINICK

La remontée des différentes erreurs pour chaque revue auprès de la responsable du FDI (collecte des documents primaires rédigés par les scientifiques de l'IRD) a permis de les corriger dans la base Horizon.

Au final, même si la création d'une présentation web peut parfois paraître un peu éloignée des activités documentaires « classiques », on constate :

- qu'on travaille sur la même base bibliographique et que cela peut par conséquent servir à noter certaines erreurs s'il y en a,
- qu'on extrait des informations contenues dans cette base, qu'on les classe, organise et met en forme en mettant en évidence leur pertinence par rapport à ce que l'on souhaite présenter.

## 2.3. Création de la présentation Web : <http://www.bondy.ird.fr/tdp>

Bien que j'utilise le terme « site Internet » en me référant à <http://www.bondy.ird.fr/tdp>, c'est en fait une « présentation Web ». En effet, celle-ci est située dans la hiérarchie du site Internet du centre IRD de Bondy à l'adresse suivante : <http://www.bondy.ird.fr>. Un même site Internet peut contenir plusieurs présentations Web traitant de sujets bien différents et pouvant être développés par des personnes de services différents.

Voici deux exemples sur le site de l'IRD Bondy :

- [www.bondy.ird.fr/carto](http://www.bondy.ird.fr/carto)
- [www.bondy.ird.fr/editions](http://www.bondy.ird.fr/editions)

### 2.3.1. Travail préparatoire à l'élaboration du site

Avant de se lancer dans la création d'un site, il est indispensable de réfléchir sur un certain nombre de points, tels que l'arborescence du site et les contraintes pratiques dont on devra tenir compte.

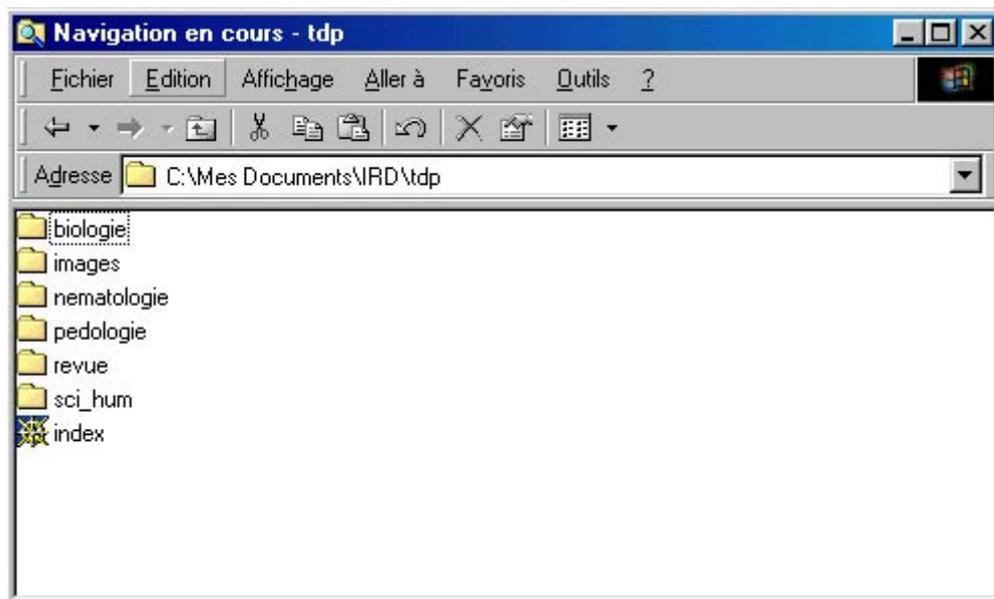
Je me suis rendue compte, au cours de ce stage, qu'il est primordial de bien prendre le temps d'analyser l'environnement de la présentation Web qui doit s'insérer dans un site déjà existant, les contraintes que cela entraîne et la structure - même de cette présentation.

#### 2.3.1.1. Arborescence du site

Nous avons passé environ une demi-journée à réfléchir et discuter sur l'arborescence du site et tout ce que cela englobe, avec un informaticien (Françoise Pelletier) assurant la gestion informatique du site de Bondy. Voici tous les problèmes et questions que nous nous sommes posés :

- L'intégration de notre présentation Web au sein d'un site existant : <http://www.bondy.ird.fr/>  
Faut-il la rattacher à la rubrique du Secteur Documentation ? Mais notre présentation va être créée à titre expérimental, on ne pouvait donc pas vraiment la rattacher à une rubrique officielle. Il a finalement été décidé de la faire apparaître au même niveau que les présentations d'autres services présents à Bondy.
- Le choix du nom de notre présentation pour pouvoir créer une adresse sur le serveur de l'IRD Bondy.  
Nous avons choisi « tdp » pour « Tables des Publications de l'IRD » laissant la possibilité d'y intégrer d'autres titres de périodiques et de collections d'ouvrages. Pour notre site, nous avons préféré intituler la page d'accueil de la présentation : « Sommaires des Revues de l'IRD » s'agissant seulement d'une présentation de périodiques. Nous avons du trouver ce nom rapidement sinon nous ne pouvions pas travailler directement sur le serveur Internet.

- L'organisation des fichiers à l'intérieur du site



- Le nom obligatoire attribué aux pages d'accueil du site et des différents dossiers, à savoir « index.htm ».
- La nécessité de respecter la charte graphique de l'IRD, que nous décrivons dans la partie 2.3.1.2., page 40.
- **Explication de l'arborescence du site :** <http://www.bondy.ird.fr/tdp>, page suivante :

La page d'accueil de la présentation Web « index » s'intitule « Sommaires des revues de l'IRD ». Elle comporte les quatre titres de revues mises en ligne, à savoir :

- Cahiers des Sciences Humaines
- Cahiers ORSTOM, série Biologie
- Cahiers ORSTOM, série Pédologie
- Revue de Nématologie

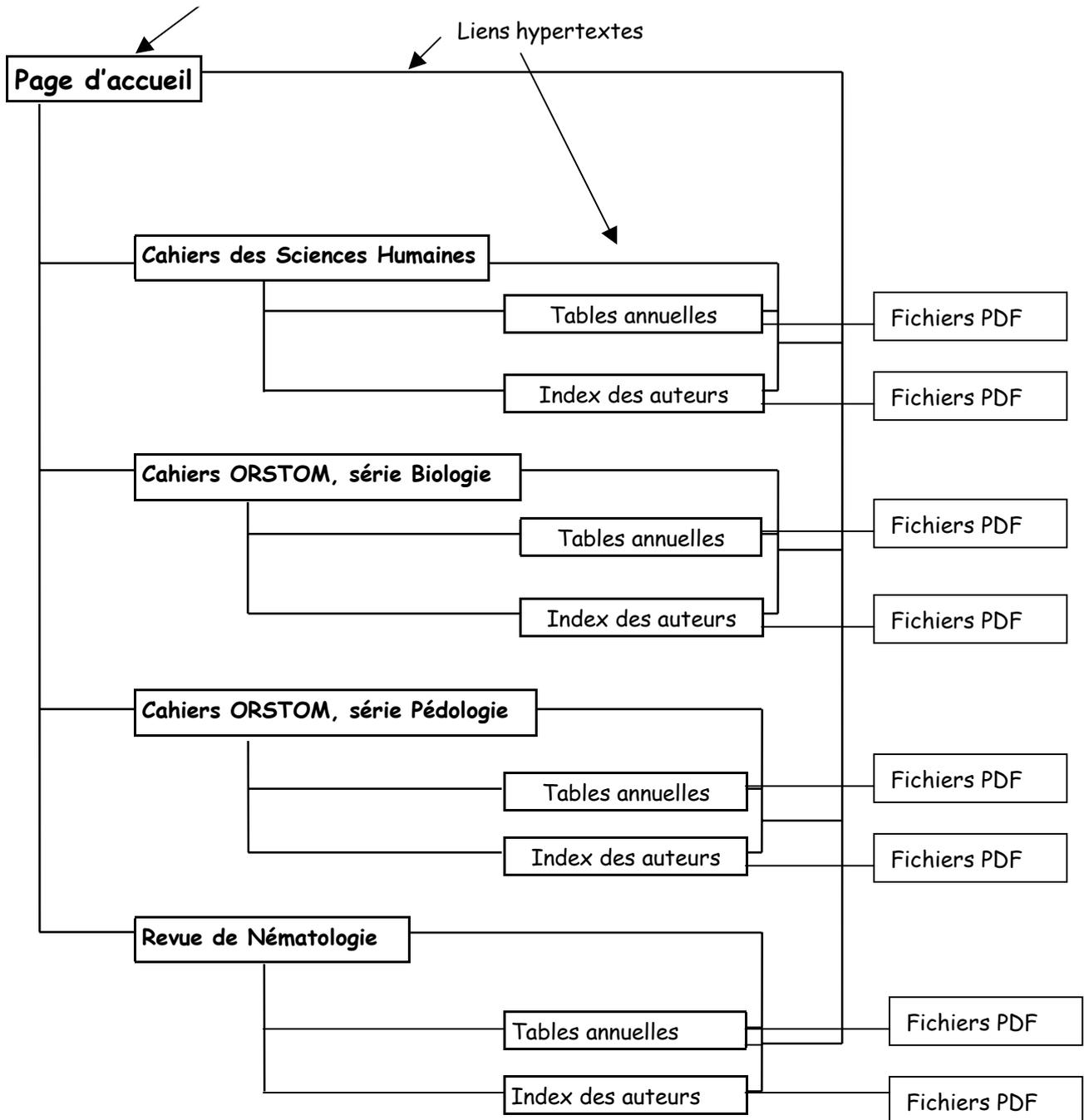
Chaque titre de revue donne accès aux tables annuelles et à l'index des auteurs. Ces derniers proposent des retours à la page de présentation de la revue et à la page d'accueil du site (« Sommaires des revues de l'IRD »). Le lien vers le document en texte intégral au format PDF est également indiqué.

# Arborescence de la présentation Web

URL : <http://www.bondy.ird.fr/tdp>

« Sommaires des revues de l'IRD »

Sommaires des revues de l'IRD



### 2.3.1.2. Contraintes pratiques

Au cours du travail préparatoire, nous avons été confrontés à des contraintes pratiques dont nous avons du tenir compte.

➤ **Le respect de la charte graphique de l'IRD (voir annexe 11 : charte graphique)**

Toutes les pages HTML des différents sites de l'IRD étant soumises à la charte, Françoise Pelletier nous a conseillé de créer une page web - type pour nous faciliter le travail par la suite. Nous en avons donc élaboré une, pour ensuite créer les différentes pages des sommaires des revues de l'IRD. Cette page doit respecter la charte graphique de l'IRD, en comprenant tous les éléments obligatoires imposés par la charte et qui doivent apparaître sur toutes les pages du site, à savoir :

- le bandeau supérieur de présentation comportant les différents logos de l'IRD (institutionnel et « local »), des liens vers différentes rubriques du site institutionnel, et titre de la page ;
- la partie centrale réservée au contenu et
- le bandeau inférieur comportant des informations pratiques.

Françoise Pelletier m'a conseillée de consulter la partie Intranet du site <http://www.ird.fr> concernant cette charte, où figurent des conseils : police de style (arial), taille des caractères, textes entrant dans un tableau ne pouvant dépasser 570 pixels<sup>1</sup> de largeur, le format des images (GIF<sup>2</sup>), des modèles, des images à reprendre telles que les flèches de retour ou hauts de page à utiliser.

Cette charte a pour objectif de tenir compte de l'évolution des techniques actuelles (typologie des serveurs, existence de différents navigateurs) pour un rendu similaire sur toutes les machines (systèmes d'utilisateurs MAC, Windows, Unix), sur toutes les tailles d'écran (écran de 15 pouces et plus), et dans tous les pays.

Une autre contrainte pratique :

➤ **La présentation finale des pages**

Il nous a fallu environ deux jours de travail pour parvenir à un modèle de page dont nous soyons satisfaits. Cette partie du travail a été assez longue par rapport à la création « réelle » des pages car nous étions souvent tentés de modifier des éléments afin de les améliorer, pour ne plus avoir à les retoucher par la suite et parvenir à un modèle de page facilement réutilisable.

---

<sup>1</sup> Pixel : voir annexe 3 : lexique informatique

<sup>2</sup> Format GIF : voir annexe 3 : lexique informatique

Nous avons donc créé une page - type pour les sommaires de revue, ci-dessous la copie d'écran. Dans un deuxième temps, nous avons créé une page - type pour l'index des auteurs quand nous avons décidé d'en faire.



## 2.3.2. Chronologie de la création du site

### 2.3.2.1. Les différentes étapes de la création de <http://www.bondy.ird.fr/tdp>

Une fois l'élaboration des pages web-type terminées, le travail de création des différentes pages HTML a été relativement rapide (voir annexe 13 : pages HTML « tdp »). Au début du stage, M. ROSSI avait prévu de mettre en ligne les tables annuelles de quatre revues mais comme cela a été assez rapide, nous avons pu également créer un index des auteurs pour chaque revue. Pour l'index des auteurs, nous nous sommes également inspirés d'exemples de présentation trouvés lors de l'état des lieux.

Voici les différentes étapes du travail, tel qu'il a été réalisé :

- Elaboration et mise en ligne des différentes pages des Cahiers ORSTOM, séries Biologie et Pédologie, ainsi que la Revue de Nématologie.
- Elaboration d'index des auteurs pour les différentes revues.
- Mise en ligne du Cahier des Sciences Humaines, sommaires et index des auteurs.

Les liens entre les pages HTML et les fichiers PDF ont été systématiquement vérifiés au cours de la construction du site.

➤ Voici quelques exemples des différents types de pages HTML composant la présentation Web « tdp »

## La page d'accueil du site

**IRD**  
Institut de recherche  
pour le développement

Qu'est-ce que l'IRD ?   Actualités   Centres de ressources   Infothèque

# Sommaires des revues de l'IRD

[hydrologie continentale](#)   [GEODYNAMIQUE](#)   [REVUE D'HYDROLOGIE TROPICALE](#)   [PEDOLOGIE](#)   [Cahiers des Sciences Humaines](#)

- ⇒ [Cahiers des Sciences Humaines](#)  
Les Cahiers des Sciences Humaines présentent des articles consacrés à différentes disciplines des sciences sociales, telles que l'anthropologie, la sociologie, la démographie, l'économie, l'archéologie, la géographie, la linguistique, etc.
- ⇒ [Cahiers ORSTOM, série Biologie](#)  
Les Cahiers ORSTOM, série Biologie présentent des articles consacrés à diverses branches de la biologie végétale et animale (agronomie, biologie et amélioration des plantes utiles, entomologie agricole, microbiologie du sol).
- ⇒ [Cahiers ORSTOM, série Pédologie](#)  
Les Cahiers ORSTOM, série Pédologie présentent des articles consacrés à l'étude des sols : morphologie, caractérisations physico-chimique et minéralogique, classification, relation entre sols et géomorphologie, problèmes liés aux sels, à l'eau, à l'érosion, à la fertilité.
- ⇒ [Revue de Nématologie](#)  
La Revue de Nématologie présente des articles consacrés à l'étude des nématodes parasites de plantes ou d'insectes, des nématodes libres du sol et des nématodes d'eaux douces ou marins.

## La table annuelle des sommaires de la « Revue de Nématologie »

**Revue de Nématologie - Microsoft Internet Explorer**

Adresse: <http://www.bondy.ird.fr/tdp/nematologie/index.htm>

IRD Institut de recherche pour le développement

Qu'est-ce que l'IRD ? Actualités Centres de ressources Infothèque

Sommaires des revues de l'IRD **Revue de Nématologie**

hydrologie continentale GEODYNAMIQUE REVUE D'ETHNOLOGIE TROPICALE PEDOLOGIE

La Revue de Nématologie présente des articles consacrés à l'étude des nématodes parasites de plantes ou d'insectes, des nématodes libres du sol et des nématodes d'eaux douces ou marins. Cette revue a été publiée de 1978 à 1991. [L'index des auteurs](#) est également disponible.

**Revue de Nématologie**

1978 ⇒ Volume 1, Numéro [1](#), [2](#)

1979 ⇒ Volume 2, Numéro [1](#), [2](#)

1980 ⇒ Volume 3, Numéro [1](#), [2](#)

## Page de présentation du sommaire de : 1964 - Vol. 1, N°1 des Cahiers ORSTOM, série Biologie

**Cahiers ORSTOM, Série Biologie - Microsoft Internet Explorer**

Adresse: [http://www.bondy.ird.fr/tdp/biologie/tables\\_biologie/1964\\_01\\_01.html](http://www.bondy.ird.fr/tdp/biologie/tables_biologie/1964_01_01.html)

IRD Institut de recherche pour le développement

Qu'est-ce que l'IRD ? Actualités Centres de ressources Infothèque

Sommaires des revues de l'IRD **Cahiers ORSTOM, série Biologie**

hydrologie continentale GEODYNAMIQUE REVUE D'ETHNOLOGIE TROPICALE PEDOLOGIE

**Physiologie des plantes tropicales cultivées**

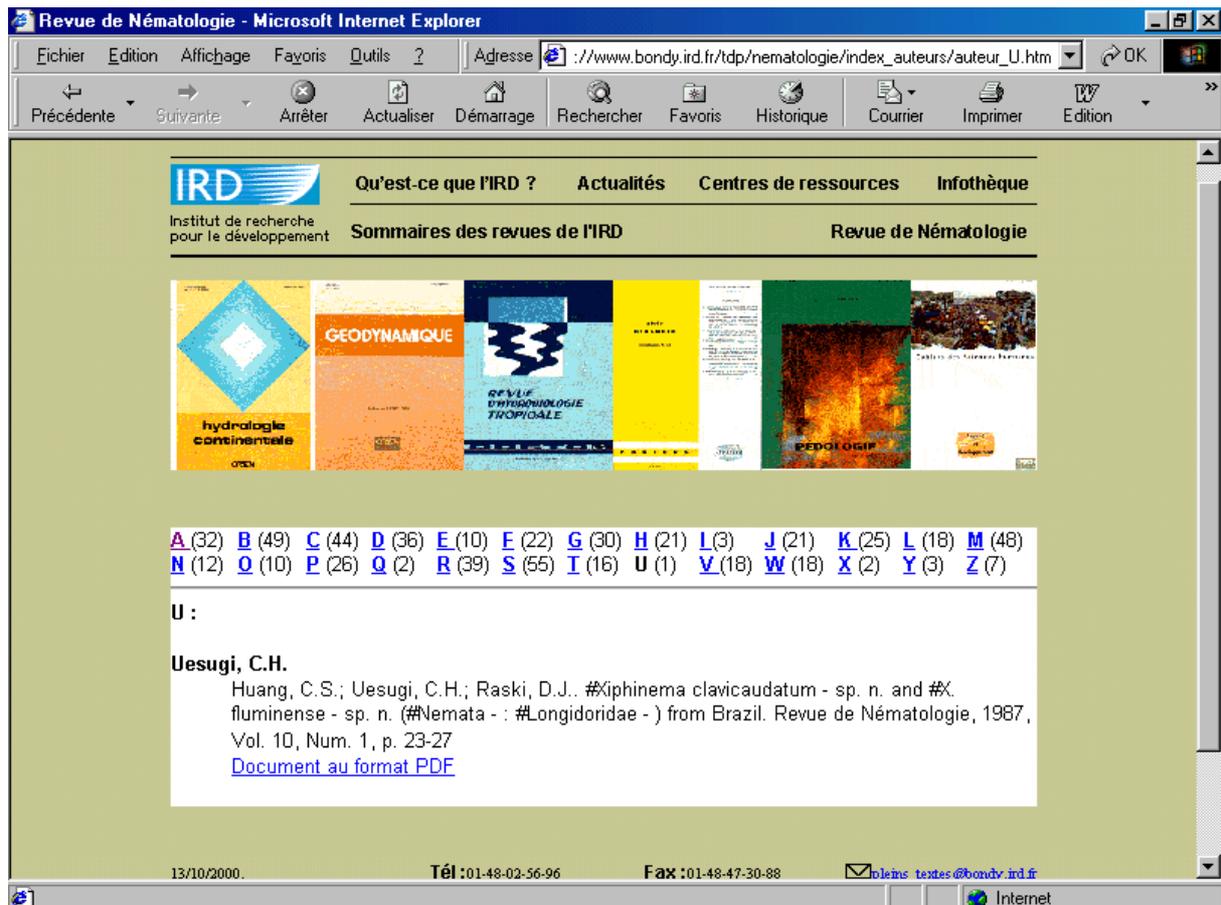
1964 - Volume 1, Numéro 1

p. 1-63 **La physiologie du caféier. 1. De la germination à la morphogenèse**  
Rabéchault, Henri; Cambrony, H.; Prévot, Pierre (pref.)  
[Document au format PDF](#)

20/09/2000. Tél : 01-48-02-56-96 Fax : 01-48-47-30-88 [pages\\_textes@bondy.ird.fr](mailto:pages_textes@bondy.ird.fr)

IRD Recherche L'IRD dans le monde Pratique

## Exemple de page de présentation d'index des auteurs de la « Revue de Nématologie »



### ➤ Bilan des pages HTML créées pour cette présentation Web

Voici deux tableaux présentant, à titre indicatif, le nombre total de fichiers informatiques (pages HTML) créés, ainsi qu'une estimation du temps de travail passé pour chaque revue IRD.

#### Nombre total de fichiers informatiques créés (pages HTML)

Titres	Pages d'accueil	Tables annuelles	Index des auteurs	Total des fichiers
Page d'accueil générale	1	-	-	1
Cahiers ORSTOM, série Biologie	1	49	20	70
Cahiers ORSTOM, série Pédologie	1	101	24	126
Revue de Nématologie	1	44	26	71
Cahiers des Sciences humaines	1	112	24	137
<b>Total des fichiers</b>	<b>5</b>	<b>306</b>	<b>94</b>	<b>405</b>

## Estimation du temps de travail passé pour chaque revue (en nombre de jours)

Titres	Pages d'accueil + modèle de page-type	Tables annuelles	Index des auteurs	Total en jours
Page d'accueil générale + modèle de page-type	1,5	-	-	1,5
Cahiers ORSTOM, série Biologie	1	2,5	2,5	6
Cahiers ORSTOM, série Pédologie	0,5	3,5	2,5	6,5
Revue de Nématologie	0,5	2,5	2,5	5,5
Cahiers des Sciences humaines	0,5	3,5	2,5	6,5
<b>Total en nombre de jours</b>	<b>4</b>	<b>12</b>	<b>10</b>	<b>26</b>

Le temps de travail passé pour chaque revue englobe :

- la création des fichiers « tables\_annuelles » et « tables\_auteurs » sur Texto
- le découpage du fichier informatique global (« tables\_annuelles » et « tables\_auteurs »)
- la vérification de l'exactitude des informations contenues dans les notices en comparaison avec les documents primaires
- la création des différentes pages HTML

### 2.3.2.2. Architecture du site Web « tdp » dans le contexte du réseau informatique de Bondy

Explication du schéma intitulé : « Architecture des répertoires de travail et du serveur Web », page suivante.

Le serveur Internet de Bondy héberge deux serveurs distincts :

- [www.ird.fr](http://www.ird.fr) : site institutionnel,
- [www.bondy.ird.fr](http://www.bondy.ird.fr) : serveur de Bondy

Sur le réseau informatique de Bondy, mon travail s'est concrétisé en un montage de répertoires de travail vers le disque dur accessible via Internet.

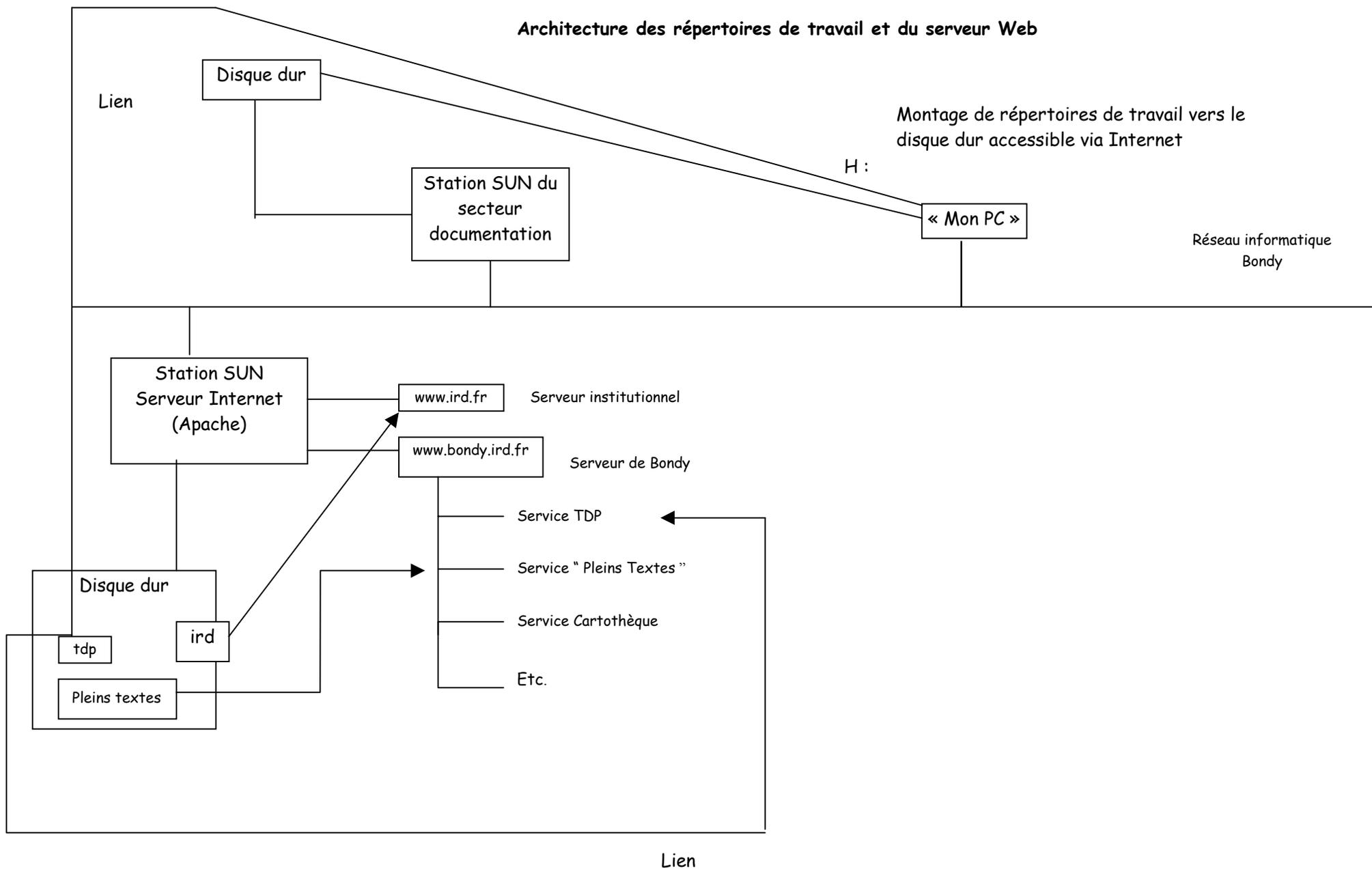
D'une part, le disque dur de « mon PC » est relié à la station SUN du secteur documentation. D'autre part, « mon PC » est relié au disque dur de la station SUN du serveur Internet de Bondy par le répertoire H :

Sur « mon PC » dans le répertoire H :, je crée un dossier « tdp » dans lequel j'intègre tous les fichiers au fur et à mesure, selon l'arborescence du site qu'on a établie, en laissant le fichier « index » comme point de démarrage du site.

Le fichier « index » correspond à la page d'accueil, c'est-à-dire à la page vers laquelle va pointer le navigateur sur l'URL suivant : <http://www.bondy.ird.fr/tdp/>.

La présentation « tdp » est donc accessible sur le répertoire H : en interne et sur le site Internet de Bondy en interne et en externe.

# Architecture des répertoires de travail et du serveur Web



### 2.3.3. Caractéristiques du matériel informatique utilisé

Au cours de mon stage, j'ai travaillé dans l'atelier de numérisation sur un PC de marque NEC.

#### ➤ **Le PC**

Le PC sur lequel j'ai travaillé a accès au réseau comme tous les postes informatiques (Mac, PC, etc.) du site de Bondy. Le système de connexion au réseau est le suivant : « login + mot de passe ».

Voici les grandes caractéristiques du PC :

Nec MultiSync

Pentium II

300 Mhz

64 Mo de mémoire vive

Disques durs de 6 et de 15 Mo

Windows 95

Ecran 17"

Antivirus F-Secure

#### ➤ **Logiciel de navigation ou navigateur<sup>1</sup>**

Netscape Communicator<sup>2</sup>

#### ➤ **Logiciel de messagerie utilisé**

Eudora Pro<sup>3</sup>

#### ➤ **Imprimante (réseau)**

EPSON 800

Jet d'encre couleur

#### ➤ **Scanner**

J'ai utilisé un scanner pour numériser des couvertures de revues dans le but de créer une image représentant un bandeau de présentation. Ceci m'a permis d'apprendre à utiliser un scanner.

Caractéristiques du scanner :

Mustek SCANEXPRESS A3 SP

Type de scanner : couleur à plat

Numérisation sur 36 bits (permet d'améliorer la précision des couleurs)

Zone de numérisation : 297 x 431 mm (taille maximum des documents)

Résolution optique : 300 pp (H) x 600 ppp<sup>4</sup> (V)

---

<sup>1</sup> Logiciel de navigation ou navigateur : voir annexe 3 : lexique informatique

<sup>2</sup> Netscape Communicator : voir annexe 3 : lexique informatique

<sup>3</sup> Eudora Pro : voir annexe 3 : lexique informatique

<sup>4</sup> ppp : voir annexe 3 : lexique informatique

### 2.3.4. Utilisation d'éditeurs HTML et de logiciel de retouche d'image

Au cours du stage, j'ai travaillé sur un certain nombre de logiciels pour créer la présentation Web, comme des éditeurs HTML et des logiciels de retouche d'images.

#### ➤ Utilisation d'éditeurs HTML

Au départ, nous avons utilisé l'éditeur Pagemil<sup>1</sup> 3.0 d'Adobe pour créer les pages Web - type mais nous avons été rapidement confrontés à quelques problèmes d'écriture dans le langage HTML générés par ce logiciel que je décrirai dans la partie 2.3.5.

Nous avons donc essayé de créer les mêmes pages sur l'éditeur Netscape Composer 4.5<sup>2</sup>. Le résultat nous paraissant plus satisfaisant, nous avons donc décidé d'utiliser Netscape Composer.

Plus largement, nous avons eu recours à un éditeur de texte (type Wordpad<sup>3</sup>) car une partie du travail sur le langage HTML a été réalisé manuellement par des « copier-coller » et des rajouts de balises HTML.

#### ➤ Utilisation de logiciel de retouche d'image

Nous avons utilisé le logiciel Adobe Photoshop 5.5 pour créer les images du bandeau de présentation de différentes couvertures de revues IRD, ainsi que les petites images comportant le titre des différentes revues que nous avons placées dans le bandeau supérieur. Ceci nous a permis de les retoucher notamment en les recadrant à la taille souhaitée.

### 2.3.5. Problèmes rencontrés

Les problèmes rencontrés lors de la création de la présentation web sont surtout liés à l'utilisation d'éditeurs HTML. En effet, ils génèrent parfois automatiquement une mise en forme particulière, d'où la nécessité de travailler directement sur la page source (sur le langage HTML) pour apporter certaines modifications.

En ce qui concerne Pagemil, nous l'avons très peu utilisé car il génère des espaces dans le bandeau de présentation des différentes rubriques du site en haut de la page. Ceci entraînait des différences de taille et d'espace à l'intérieur du bandeau par rapport au bandeau imposé par la charte graphique.

En ce qui concerne Netscape Composer version 4.5, il paraît plus fidèle dans l'écriture HTML, malgré la création d'un espace insécable au-dessus du bandeau supérieur de présentation des différentes rubriques au tout début de la page HTML. Ceci m'a obligée à travailler directement sur la page source (en langage HTML) pour y apporter certaines modifications.

---

<sup>1</sup> Pagemil 3.0 d'Adobe : voir annexe 3 : lexique informatique

<sup>2</sup> Netscape Composer 4.5 : voir annexe 3 : lexique informatique

<sup>3</sup> Wordpad : voir annexe 3 : lexique informatique

Au début, j'étais parfois déroutée car j'avais quelques notions de (langage) HTML mais je ne le maîtrisais pas vraiment, ce qui m'a fait perdre du temps dans les premiers jours. A l'aide de M. ROSSI et d'un ouvrage très complet sur le HTML, je me suis sentie beaucoup plus à l'aise.

### **2.3.6. Caractéristiques de ce travail**

#### **➤ Travail assez répétitif et automatique**

Quand on prend l'habitude, ce travail devient rapidement assez répétitif et automatique. Ceci a bien évidemment l'inconvénient d'entraîner parfois un certain ennui mais cela présente également des avantages. En effet, je créais les différentes pages assez rapidement ce qui m'a laissé un peu de temps pour

- comprendre le fonctionnement de l'IRD et le statut du personnel
- comprendre le fonctionnement du réseau informatique et documentaire, ainsi que de la base Pleins Textes.

#### **➤ Possibilité de vérifier immédiatement le résultat obtenu sur le serveur Internet**

Si le résultat de la page ne correspondait pas à ce que j'attendais, je pouvais la rectifier immédiatement.

Par ailleurs, M. ROSSI a annoncé aux personnes travaillant à Bondy, via la messagerie électronique, qu'il était possible de consulter une nouvelle présentation Web. Ceci m'a motivée à m'appliquer d'autant plus dans mon travail.

## Conclusion

### ➤ Bilan du travail effectué au cours du stage

Ce stage m'a énormément intéressée au niveau de la réflexion sur la présentation de l'information dans les produits documentaires électroniques. Cela suppose également la réflexion sur les modes de navigation dans un document électronique, sur les différentes clés d'entrée vers l'information. Proposer des clés d'entrée à l'utilisateur pour naviguer facilement dans un document correspond à une façon de traiter l'information. En effet, on lui donne des outils, des indices pour qu'il accède facilement à l'information et pour qu'il puisse l'exploiter.

Dans la présentation « tdp », les deux clés d'entrée sont les sommaires des revues classés par année, volume, numéro, page et les index d'auteurs.

Nous avons « paramétré » le programme Logotel de façon à automatiser le plus possible les fichiers (« tables\_annuelles » et « tables\_auteurs ») pour qu'ils soient très facilement exploitables par une tierce personne. La manipulation est alors simple et assez rapide.

Ce produit documentaire pourrait évoluer en lui rajoutant, par exemple, un index de mots-clés. Cet index serait « statique » comme toutes les pages HTML qui ont été créées. En effet, on ne pourrait pas accéder à des articles par une interrogation de recherche.

### **Le problème des pages statiques (« tdp »)**

Les pages HTML créées pour présenter les tables annuelles et les index d'auteurs ne sont pas des pages dynamiques<sup>1</sup>, générées par une simple interrogation de la base bibliographique (et par le biais d'un programme). Les outils dont on dispose ne le permettent pas.

La nécessité de mettre à jour les notices des quatre revues mises en ligne poserait un problème car il faudrait créer de nouvelles pages HTML. La mise à jour ne se ferait pas automatiquement. A priori, ce sont des revues « mortes » dont les notices ne seront pas modifiées.

Par contre, si la présentation des pages imposée par la charte graphique de l'IRD évolue, toutes les pages HTML de « tdp » seront à refaire puisque la mise à jour n'est pas automatique.

N'ayant pas d'expérience auparavant dans la documentation, ce stage m'a permis de mettre en comparaison les notions apprises en cours avec la réalité du travail au sein d'un service de documentation. J'ai pris conscience à quel point il est nécessaire d'être rigoureux et méthodique dans son travail pour proposer une information claire et accessible aux utilisateurs.

---

<sup>1</sup> Pages dynamiques : page qui n'existe qu'au moment où on la demande (page créée en fonction du besoin)

## ➤ Différents apports et découvertes de ce stage

Ayant choisi ce stage en raison de son aspect informatique, je ne suis pas du tout déçue car il m'a énormément apporté sur ce plan-là. Outre une assurance plus importante, j'ai acquis différentes compétences.

Le secteur documentation de l'IRD travaillant sur le logiciel Texto sous Unix, j'ai eu une nouvelle approche de ce logiciel documentaire, par rapport à la formation. Plus largement, j'ai découvert le réseau informatique de l'IRD fonctionnant sous une plateforme Unix et grâce aux explications de M. ROSSI, j'en ai eu une vision d'ensemble.

J'ai également pu développer mes connaissances du langage HTML, que nous avons eu l'occasion d'aborder en cours. A présent, je suis capable de comprendre la structure d'une page HTML et la logique de ce langage.

Par ailleurs, j'ai travaillé sur Photoshop 5.5 et Netscape Composer 4.5 qui me sont devenus assez familiers et j'ai pu comparer ce dernier à Pagemil que nous n'avons pas retenu pour l'élaboration du projet « tdp ».

Pour finir, j'ai découvert l'éventail de possibilités qu'offre le logiciel Acrobat.

Dans l'ensemble, j'ai pris conscience :

- qu'il faut essayer d'utiliser au maximum les logiciels pour automatiser et faciliter le travail
- et que la maîtrise d'un logiciel quel qu'il soit s'acquiert essentiellement par la pratique.

Ce stage a également été pour moi l'occasion de découvrir le milieu de la recherche (multidisciplinaire) qui m'était inconnu auparavant. Cela m'a beaucoup apporté sur le plan personnel, en raison de la richesse des rencontres.

Par ailleurs, M. ROSSI m'a laissée autonome dans mon travail, ce que j'ai apprécié, tout en étant très présent lorsque j'étais confrontée à certains problèmes.

Ce stage et l'environnement de l'IRD m'ont beaucoup apporté sur la variété des activités que propose le métier de documentaliste. En effet, je n'imaginais pas l'éventail des possibilités qu'offre la documentation et je n'avais pas vraiment pris conscience de cet important « champ d'application » auparavant.

## ➤ Décalages par rapport à la formation

Bien que la formation de l'IRTD aborde tous les aspects de la documentation et qu'elle permette de nous adapter assez facilement sur le terrain, je pense que certains points techniques (informatiques) pourraient être davantage développés, comme :

- l'utilisation d'un logiciel tel qu'Acrobat qui permet de créer des documents au format PDF et qui offre de nombreux avantages,
- et l'approche de la numérisation de documents.

Il est vrai que le programme de la formation est bien chargé et que le temps est limité, mais ces aspects techniques me semblent importants quand on considère l'évolution du métier de documentaliste.

En effet, le bouleversement provoqué par le développement d'Internet dans le rapport à l'information a bien évidemment des répercussions sur le rôle des documentalistes. A l'IRD, malgré l'existence de deux types de documentation, une qu'on pourrait qualifier de « classique » et l'autre essentiellement liée aux avancées des technologies de l'information, les documentalistes réfléchissent sur leur rôle dans les années à venir.

Ce stage m'a permis de réfléchir plus largement sur :

- le rôle des documentalistes face à l'évolution des technologies de l'information,
- les changements auxquels ce métier est confronté,
- les problèmes que cela pose,
- la nécessité d'évoluer également en proposant de nouveaux produits documentaires accessibles sur Internet ou sur CD Rom,
- la notion de « bibliothèque électronique » offrant l'accès aux documents en texte intégral,
- et l'évolution des relations avec les utilisateurs. Lors de la création de produits documentaires électroniques, on est moins en contact direct avec les utilisateurs mais on en touche davantage lors de la diffusion de l'information.

Etant débutante dans la profession, je suis soucieuse d'acquérir une expérience significative. C'est pourquoi, j'ai accepté une vacation à l'IRD pour participer à la création d'un CD Rom sur la présentation de notices explicatives et de cartes thématiques.

Je suis très intéressée par l'évolution du métier liée aux NTIC. Selon moi, les documentalistes ont un rôle important à jouer grâce à ce que permettent les nouvelles technologies et en mettant en évidence leur savoir-faire. Cela suppose que les documentalistes soient de plus en plus qualifiées dans le domaine informatique. Par ailleurs, cela permet également d'être moins dépendant des informaticiens qui ne comprennent pas toujours les besoins liés à la documentation.

### ➤ **Projets développés par le service informatique du secteur documentation**

Le stage au sein de l'unité « Ingénierie de l'information » du secteur documentation m'a permis de comprendre l'évolution d'un service de documentation face aux avancées des technologies de l'information.

Malgré la difficulté de recruter du personnel permanent (par le biais d'une création de poste), l'unité « Ingénierie de l'information » est très active et entreprend de nombreux projets. Ces derniers tournent essentiellement autour de la numérisation et de la mise en œuvre de la bibliothèque électronique de l'IRD.

Parallèlement à cela, M. ROSSI participe au développement et à la mise en œuvre du projet « Infothèque ». Celui-ci donne lieu à de nombreuses réunions entre les différents services de la DIC qui mènent une importante réflexion sur la création d'un portail d'accès à l'information concernant les pays en développement. Cela constitue un enjeu important pour les pays du Sud, car l'accès à l'information (via Internet) est primordial.

La base bibliographique Horizon et Pleins Textes étant accessibles simultanément, on se retrouve avec des doublons de notices quand on interroge ces bases. A terme, les notices bibliographiques Horizon devraient être peu à peu remplacées par les notices de Pleins Textes qui offrent également l'accès aux documents en texte intégral.

Intervenant régulièrement dans les pays du Sud pour assurer des formations et des appuis informatiques, M. ROSSI a une vision d'ensemble des enjeux de la numérisation des différents fonds documentaires et surtout de la réunion de tous les documents IRD dans une même base accessible sur Internet.

En effet, les centres IRD à l'étranger développent des bases locales, également sous Texto. Certains d'entre eux, comme celui situé au Niger, ont rendu leur base locale accessible sur Internet. Ainsi, un étudiant (chercheur) travaillant sur ce pays peut déjà se constituer une bibliographie importante avant de se déplacer dans le pays. Cela reviendrait à créer un portail sur le développement regroupant toutes les bibliothèques électroniques de l'IRD.

Pour finir, M. ROSSI réfléchit également sur la mise en place d'un atelier de formation destiné aux chercheurs pour les aider à produire leurs travaux au format PDF. Cela éviterait, une fois la numérisation de l'ensemble du FDI terminée, de numériser les nouveaux travaux de recherche. Ces derniers intégreraient directement Pleins Textes, après leur traitement documentaire.

La création de sommaires des revues de l'IRD et des index d'auteurs pendant mon stage a été l'occasion de réutiliser les documents numérisés.

Puisque ces documents sont présentés sur un support différent que le papier, il est nécessaire de réfléchir sur l'organisation de l'information en l'adaptant au support électronique (Internet, CD Rom). Cela a constitué une partie importante de mon travail pendant le stage.

## Bibliographie

- LEMAY Laura, *HTML 4 « Le programmeur »*, Paris, Campus Press, 2000, 803 p.
- Rapport d'activité annuel de l'Institut de Recherche pour le Développement 1999
- ROSSI Pier Luigi , « *Pleins textes* », : *bibliothèque électronique de l'IRD*, octobre 2000, 7 p.

### Adresses des sites Internet visités

#### Instituts de recherche

- **IRD** : Institut de Recherche pour le Développement : <http://www.ird.fr>  
et tous les liens vers les différents centres de recherche de l'IRD en France et à l'étranger
- **CIRAD** : Centre International de Recherche pour l'Agronomie de Développement  
<http://www.cirad.fr>
- **CNRS** : Centre National de Recherche Scientifique : <http://www.cnrs.fr>
- **INRA** : Institut National de Recherche Agronomique : <http://www.inra.fr>
- **Institut Pasteur** : <http://www.pasteur.fr>
- **Institut de recherche Pierre et Marie Curie** : <http://www.curie.fr>
- **INSERM** : Institut National de la Santé et de la Recherche Médicale :  
<http://www.inserm.fr>
- **INED** : Institut National des Etudes Démographiques : <http://www.ined.fr>
- **INIST** : Institut National de l'Information Scientifique et Technique :  
<http://www.inist.fr>

#### Administrations

- **Ministère de la Recherche** : <http://www.recherche.gouv.fr>
- **Agence Universitaire de la Francophonie** : <http://www.aupelf-uref.org>

#### Editeurs scientifiques

- **Elsevier** : <http://www.elsevier.com> (Science Direct)
- **Springer** : <http://www.springer.de> (Link)

#### Informations sur le vocabulaire informatique

- **Format PDF** : <http://emala.free.fr/acrobat.htm>
- **Terminologie informatique** : <http://www.inria.fr>

Chambre de Commerce et d'Industrie de Rouen

Institut Régional des Techniques Documentaires  
11 rue du Tronquet  
76130 Mont Saint Aignan

## Annexes du Rapport de Stage

Création d'un produit documentaire électronique



## Sommaires des revues de l'IRD



MICALETTI Marie -Angèle  
16<sup>ème</sup> promotion  
1999-2001

Février 2001

## Table des annexes

→ <b>Annexe 1</b> : Organigramme général de l'IRD	<b>I</b>
→ <b>Annexe 2</b> : Carte du monde présentant les différentes implantations de l'IRD	<b>II</b>
→ <b>Annexe 3</b> : Lexique des termes informatiques	<b>III</b>
→ <b>Annexe 4</b> : Organigramme du centre IRD de Bondy	<b>IV</b>
→ <b>Annexe 5</b> : Plan des locaux du service de documentation de l'IRD Bondy	<b>V</b>
→ <b>Annexe 6</b> : Bilan budgétaire 1999 de l'unité technique « Ingénierie de l'information »	<b>VI</b>
→ <b>Annexe 7</b> : Rapport d'activité annuel 1999 du centre de documentation	<b>VII</b>
→ <b>Annexe 8</b> : Les modes de numérisation	<b>VIII</b>
→ <b>Annexe 9</b> : Article de P.L. ROSSI sur la bibliothèque électronique « Pleins Textes »	<b>IX</b>
→ <b>Annexe 10</b> : Schéma sur le fonctionnement de Pleins Textes	<b>X</b>
→ <b>Annexe 11</b> : Charte graphique de l'IRD	<b>XI</b>
→ <b>Annexe 12</b> : Modèle de page choisi lors de l'état des lieux de l'existant	<b>XII</b>
→ <b>Annexe 13</b> : Quelques pages réalisées pour la présentation Web : <a href="http://www.bondy.ird.fr/tdp">http://www.bondy.ird.fr/tdp</a>	<b>XIII</b>

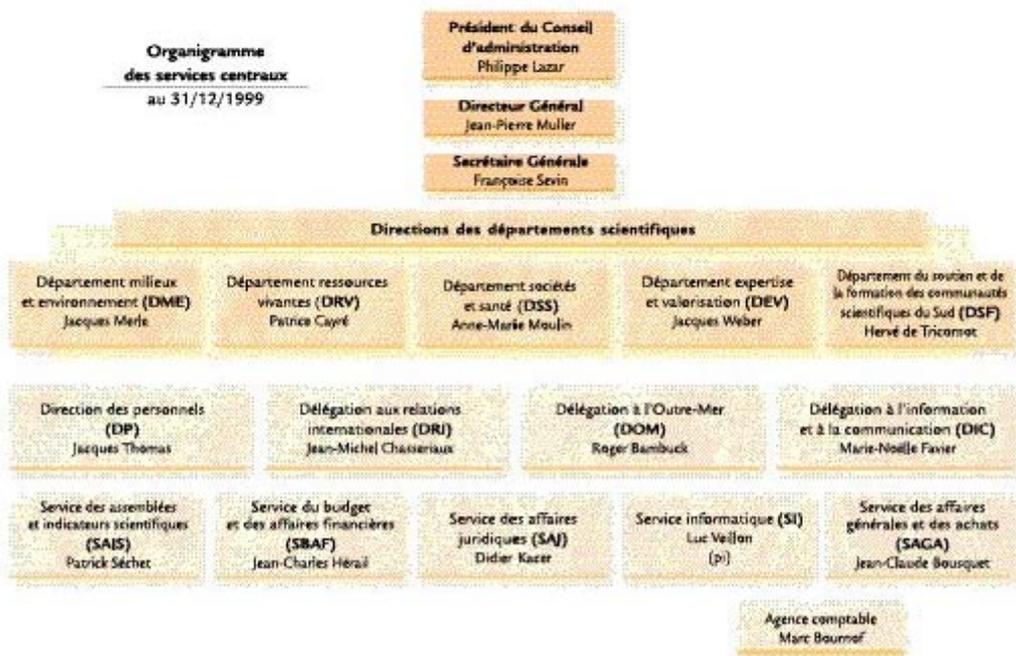
# ANNEXE 1

## Organigramme général de l'IRD

**Organigramme général**  
au 31/12/1999

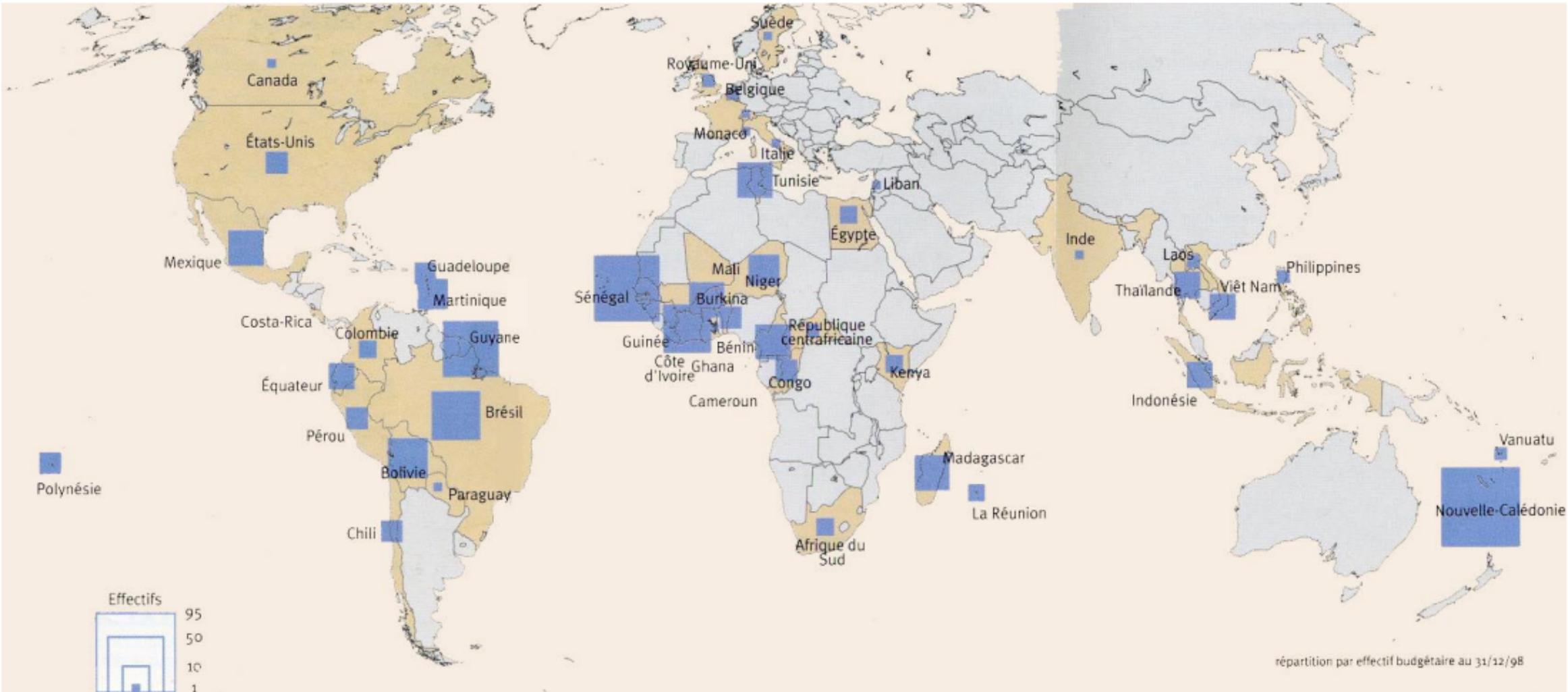


**Organigramme des services centraux**  
au 31/12/1999



## ANNEXE 2

Carte du monde présentant les  
différentes implantations de l'IRD



## ANNEXE 3

### Lexique des termes informatiques

## Lexique informatique

### A

#### **Acrobat Reader**

Logiciel, téléchargeable gratuitement sur Internet, qui affiche à l'écran les documents tels qu'ils ont été conçus dans leur mise en page initiale, sans avoir le logiciel qui a servi à cette mise en page. Acrobat Reader offre une taille de stockage réduite, ainsi qu'une bonne qualité d'impression et des fonctions de base très utiles pour parcourir le document (zoom, recherche, impression partielle ou globale du document).

#### **Acrobat 4.0**

Logiciel, qui permet de créer des fichiers PDF (cf. PDF) à partir de documents HTML, de fichiers numérisés. Il a la particularité de reproduire convenablement des documents numérisés (HTML, texte, image, carte etc.).

#### **Anti-virus :**

Programme chargé de lutter contre les virus informatiques.

#### **Apache : Station Server**

Le mot Apache viendrait de l'expression « a patchy server » (un serveur inégal ). Apache est un serveur Web gratuit fonctionnant sous UNIX et Windows, il est aujourd'hui le serveur le plus utilisé, il héberge , en effet, plus d'un million de sites. Il est bien plus stable que le serveur Web de Microsoft.

### B

#### **Bit :**

C'est la plus petite unité informatique, binaire, de représentation de données, dont la valeur est 0 ou 1. Huit bits forment un octet.

### C

#### **Concaténation de fichiers :**

Enchaînement de deux ou plusieurs fichiers provenant de logiciels différents.

### D

#### **DPI : ( Dots Per Inch)**

Unité équivalente aux ppp (points par pouce) utilisée pour mesurer la résolution des imprimantes, des scanners et des images.

## E

### **Eudora Pro:**

Logiciel de messagerie client très populaire dans les environnements PC et Mac.

## G

### **GIF : Graphique Interchange Format**

Format comprimé d'images de 256 couleurs maximum, inventé par CompuServe et exploitable avec un Mac ou un PC. Très souvent utilisé sur le Web en raison de son faible encombrement et ses possibilités de transparence.

## H

### **HTML : Hypertext Markup language / langage de balise hypertexte.**

Le HTML est le langage de programmation adopté dans le monde entier, utilisé pour créer des documents hypertextes et construire des pages Web. Il utilise une liste déterminée de balises qui décrivent la structure d'un texte (caractères, couleurs, paragraphe, titre) ou l'emplacement d'éléments incorporés à la page (photo, graphique, applet, tableau). Par exemple la phrase : « le centre de documentation est ouvert sur le monde. » s'écrit en HTML :

```
<html>
<body>
<b>le centre de documentation est ouvert sur le monde.< /b>
</body>
</html>
```

Aujourd'hui de nombreux logiciels permettent de créer ces pages sans connaître le langage. Cependant la maîtrise du langage (HTML) permet de corriger les erreurs que les logiciels produisent automatiquement. Nous avons utilisé trois logiciels pour « TDP », Pagemil 3.0 d'Adobe, Netscape Composer, et l'éditeur de texte WordPad de Windows.

### **Hypertexte :**

Présentation d'un ensemble de textes et d'autres documents (graphique, base de données, vidéos, etc.) qui permet une consultation non linéaire. Un renvoi (hyperlien) permet d'afficher un autre document.

## I

### **Intranet :**

Réseau interne d'une entreprise qui utilise la technologie d'Internet. Le système est ouvert ou fermé vers Internet mais les protocoles utilisés sont les mêmes. L'IRD, par exemple, possède des liaisons avec l'Internet et un réseau Renater pour échanger des informations entre ses différents centres.

## L

### **Logotel :**

Langage de Texto qui permet de programmer la réalisation d'applications périphériques, telles les interfaces utilisateurs, la gestion des langages documentaire dont les thesaurus, le contrôle de valeurs des champs ou l'automatisation des processus répétitifs.

## M

### **Mode bitonal (numérisation en)**

Le mode bitonal code chaque pixel sur un bit, lui donnant ainsi la possibilité de ne prendre que deux valeurs, le noir et le blanc. Ce format est très économique en mémoire.

## N

### **Navigateur (ou logiciel de navigation)**

Logiciel utilisé pour visualiser des pages (HTML) et naviguer sur le Web.

### **Netscape Navigator :**

Ce navigateur est un logiciel gratuit (freeware) qui permet de visualiser et de naviguer sur le Web. Il permet de lire des documents HTML.

### **Netscape Composer :**

Logiciel gratuit qui permet de créer des pages Web sans connaître le langage HTML.

### **Numérisation :<sup>1</sup>**

Conversion d'un signal analogique en un signal numérique.

---

<sup>1</sup> voir annexe 8 : « modes de numérisation »

## O

**OCR** : voir **ROC**

## P

### **Pagemil 3.0:**

Editeur de HTML, Logiciel de la société Adobe qui permet de créer des pages Web sans connaître le langage HTML.

### **PDF : Portable Document Format.**

Format de fichier conçu par la société Adobe qui permet de mettre à disposition des internautes des documents sur leur forme originale via le logiciel Acrobat Reader qui est disponible gratuitement.

### **Pixel :**

Plus petit élément d'une présentation visualisée auquel peuvent être affectés séparément des attributs tels que luminosité, couleur, clignotement.

### **Portabilité :**

Aptitude d'un programme à être utilisé sur des systèmes informatiques de types différents.

### **Portail :**

Une page d'entrée de site Web offrant le maximum de liens vers les différentes parties de ce site ou d'autres sites.

### **Progiciel :**

Ensemble complet et documenté de programmes conçu pour être fourni à plusieurs utilisateurs, en vue d'une même application ou d'une même fonction.

## R

### **ROC : Reconnaissance optique de caractère / OCR : Optical Character Recognition**

Logiciel qui reconnaît automatiquement le contenu d'une page de texte numérique (issue d'un scanner ou d'un fax), pour le transformer en fichier lisible par un logiciel de traitement de texte.

**Renater** : Réseau National de télécommunication pour la technologie l'Enseignement et la Recherche.

Groupement français d'intérêt public gérant le réseau d'interconnexion entre les établissements d'enseignements et les centres de recherche.

## S

**Station Sun (Sun Microsystems) :**

Ordinateur opérant sous UNIX construit par la société américaine Sun Microsystems.

**Serveur : (anglais : server)**

Ordinateur dédié à l'administration d'un réseau informatique. Il gère l'accès aux ressources et aux périphériques et les connexions des différents utilisateurs. Il est équipé d'un logiciel de gestion de réseau sous le protocole TCP/IP. Un serveur de fichier gère l'emplacement des fichiers, un serveur d'impression gère et exécute les sorties sur imprimantes du réseau, enfin un serveur d'applications rend disponible à partir d'un disque dur, les programmes pouvant être appelés à travers le réseau.

Système informatique destiné à fournir des services à des utilisateurs connectés et par extension, organisme qui exploite un tel système. Un serveur peut permettre la consultation et l'exploitation de banques de données.

## T

**TCP / IP : Transmission Control Protocol / Internet Protocol**

Il s'agit du couple de protocoles de base le plus important pour Internet. Le protocole de contrôle de transmissions (TCP) est un langage qui permet à deux ordinateurs dissemblables d'établir une liaison entre eux et de contrôler l'émission et la réception de messages, quel qu'en soit le contenu.

Le protocole Internet (IP) régit l'attribution des adresses Internet et la manière dont les informations circulent depuis l'émetteur jusqu'au destinataire.

Ces normes régissent le fonctionnement du réseau mondial des réseaux, indépendamment des différences qui peuvent exister entre les ordinateurs.

## U

### **UNIX : Uniplexed Information and Computer Service**

Système d'exploitation multitâche et multi-utilisateurs très performant que les « Bell laboratoires » ont développé à la fin des années 60 pour la mini-informatique. Ce système d'exploitation ouvert a été porté sur la plupart des ordinateurs, de l'ordinateur central au PC à processeur Intel en passant par les stations de travail, ce qui fait qu'il en existe maintenant beaucoup de variantes. Le dérivé le plus connu est actuellement Linux.

### **URL (adresse) Universal Resource Locator**

C'est l'adresse d'un site Internet particulier (ou d'une page d'un site). L'URL est composé d'un nom de domaine et éventuellement du chemin d'accès à une page particulière. On tape l'adresse de l'URL dans la ligne de saisie du navigateur.

Ainsi, pour se connecter au site TDP, il faut taper : <http://www.bondy.ird.fr/tdp>

## V

### **Verity Information Server :**

Progiciel et Moteur d'indexation et de recherche en texte intégral, il permet d'accéder à l'information personnalisée et pertinente (200 formats supportés : HTML, PIF, Word, PDF etc.). Il est destiné aux Intranet volumineux et fédère de nombreuses sources d'informations autour d'un portail. Il est construit et commercialisé par la société Verity.

## W

### **WordPad :**

Logiciel de Windows qui permet notamment d'éditer des fichiers. Il peut être utilisé pour éditer et corriger des fichiers au format HTML.

## ANNEXE 4

### Organigramme du Centre IRD de Bondy

# Organigramme du Centre IRD-Ile de France

## Centre de Recherche d'Ile de France

### DIRECTION ADMINISTRATION DU CENTRE D'ILE DE FRANCE



**DIRECTION  
GENERALE**

Régie - Gestion financière et Comptable - Unité Technique - Division du Personnel

Comité Technique Paritaire Local - Comité Hygiène et Sécurité

	<b>DME</b>	<b>DRV</b>	<b>DSS</b>
<b>SECRETARIAT GENERAL</b>	<b>R027</b> -Interaction entre couvertures d'altération et aquifères	<b>Groupe de recherche IRD</b> - Université de Paris IV, Ecologie des sols tropicaux	<b>R003</b> -Travail et mondialisation <b>R029</b> -Environnement urbain
<b>SERVICES CENTRAUX</b>	<b>R055</b> - Paléoenvironnements tropicaux et variabilité climatique	<b>R064</b> -Biodiversité et écologie des sols	<b>R105</b> -Savoir et développement
<b>DP, DRI, DOM DIC, SAIS, SBAF, SAJ, DSI, SAGA, AC</b>	<b>R079</b> -Géométrie des Espaces Organisés, Dynamiques Environnementales et Simulations	<b>R049</b> -Erosion et changement d'usages des terres	<b>R103</b> -Mobilités et recompositions urbaines
	<b>S018</b> -Actualisation et valorisation des données pédologiques en milieu tropical et méditerranéen	<b>R075</b> -Physiologie et génétique de la résistance des plantes aux parasites	<b>R107</b> -Constructions identitaires et mondialisation
<b>DEPARTEMENTS</b>	<b>S094</b> -Géosciences des environnements intertropicaux	<b>S017</b> - Jachère en Afrique tropicale	
<b>DME, DRV, DSS, DEV, DSF</b>	<b>R058</b> -Processus d'altération et de pédogénèse et bilan des transferts dans la géosphère		

**DSI****Centre de  
Ressources  
Informatiques**

- Réseau et systèmes
- Messagerie
- Accès Internet
- Gestion Parc Informatique

**Service  
Informatique de  
Gestion**

- Etudes
- Réseau et systèmes (Siège)
- Exploitation
- Bureautique

**Information  
géographique****Diffusion****Documentation****Audiovisuel****DIC****Atelier de Cartographie  
Appliquée**

- Edition et multimédia
- Base de données géoréférencées
- Formation, appui aux unités

**Diffusion d'ouvrages et  
de produit multimédia**

- Centre de documentation
- Base Horizon
- Numérisation, Accès WEB
- Fonds documentaire IRD

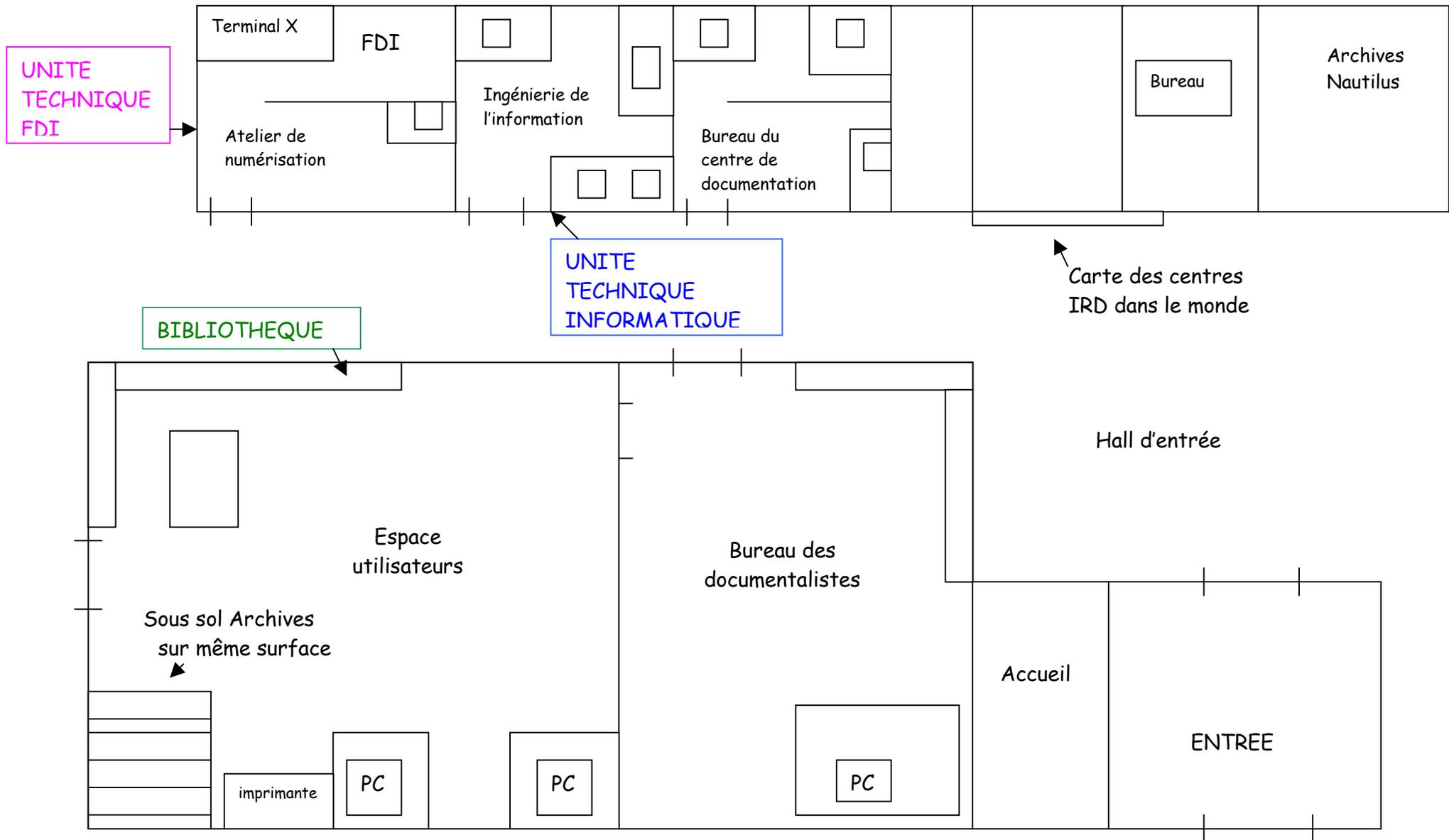
- Atelier Audiovisuel**
- Production cinéma, vidéo et multimédia
  - Production audiophonique
  - Fonds audiovisuel

**CENTRES,  
REPRESENTATIONS****RESEAUX**

<b>DP</b>	Direction des personnels	<b>SAGA</b>	Service des Affaires Générales et c Achats
<b>DRI</b>	Délégation aux Relations Internationales	<b>AC</b>	Agence Comptable
<b>DOM</b>	Délégation à l'Outre -Mer	<b>DME</b>	Département Milieux et Environne
<b>DIC</b>	Délégation à l'Information et à la Communication	<b>DRV</b>	Département des Ressources Vivar
<b>SAIS</b>	Service des Assemblées et Indicateurs Scientifiques	<b>DSS</b>	Département Société et Santé
<b>SBAF</b>	Service du Budget et des Affaires Financières	<b>DEV</b>	Département de l'Expertise et de la Valorisation
<b>SAJ</b>	Service des Affaires Juridiques	<b>DSF</b>	Département du Soutien et de la Formation des communautés scientifiques du Sud
<b>DSI</b>	Délégation au Service de l'Information		

## ANNEXE 5

### Plan des locaux du service de documentation de l'IRD de Bondy



**PLAN DES LOCAUX DU SERVICE DE DOCUMENTATION DE L'IRD BONDY**

Bilan budgétaire 1999 de  
l'unité technique  
« Ingénierie de l'information »  
du secteur documentation de l'IRD

## ANNEXE 7

# Rapport d'activité annuel 1999 du Centre de documentation de l'IRD de Bondy

## ANNEXE 8

### Les modes de numérisation

## **Les modes de numérisation**

### Définition de la numérisation

### Le mode image

### Les avantages et inconvénients du mode image

### Le mode texte

### Le mode vectoriel

### Bibliographie

### Sites à consulter

#### **▲ Définition de la numérisation**

La numérisation est la codification numérique des intensités lumineuses et de la colorimétrie d'un document. La représentation numérique d'un texte peut se faire selon deux modalités :

- le texte est considéré comme une image qui permet de rendre seulement la forme des caractères, il est alors représenté sur un mode photographique. Ce type de document est obtenu par numérisation directe du document (mode image).
- chaque caractère a sa représentation unique sous forme numérique, ce type de document en mode caractère est obtenu soit par saisie directe par des outils de type traitement de texte, soit par reconnaissance optique de caractères à partir d'un document en mode image.

## ▲ Le mode image

Les images bitmap (aussi appelées images raster ou image en mode point) sont constituées d'une grille de points appelés pixels. Chaque pixel peut être codé :

- sur un bit pour une image en noir et blanc ;
- sur 8 bits pour une image en niveaux de gris ;
- sur 24 ou plus pour une image en couleurs.

Selon le nombre de bits sur lesquels on choisit de coder un pixel, on obtiendra un choix plus ou moins grand de couleurs pour rendre l'image. On désigne couramment par profondeur du pixel (*bit depth*) le nombre de bits requis pour un pixel.

Le **mode bitonal** code chaque pixel sur un bit, lui donnant ainsi la possibilité de ne prendre que deux valeurs, le noir ou le blanc. Elle est très économique en terme de mémoire. Cependant, si le mode bitonal s'applique facilement aux textes récents et très contrastés, il conviendra de prendre des précautions pour le traitement de documents faiblement contrastés, dont les fonds sont colorés, dont l'encre est pâlie ou de densité variable. L'adoption de ce mode fait perdre le rendu du support papier dont on ne pourra plus évaluer le caractère ancien. Le bitonal nécessite un examen préalable des documents (état du papier, typographie). Il faut par exemple s'assurer par des essais que les taches de roussissures ou d'humidité du papier ne seront pas interprétées par le scanner comme des points à traduire en noir, ce qui peut rendre le document illisible.

Le mode **niveaux de gris** restitue mieux les nuances colorimétriques. Il est plus goumand en mémoire. Plus le nombre de bits utilisés pour coder un pixel est grand, plus large est la palette de gris. Ainsi un codage sur 8 bits (un octet) donne 256 gris différents. Cette technique permet, à l'inverse du bitonal, de préserver correctement l'aspect ancien des documents, ce qui a décidé certains établissements (Bibliothèque du Congrès, Conservatoire des Arts et Métiers) à l'adopter pour numériser des fonds patrimoniaux. On peut l'utiliser pour les gravures.

Le **mode couleurs** suit le même principe que le précédent à la différence que chaque pixel correspond au codage des trois couleurs primaires rouge, vert et bleu (RVB). Chacune de ces couleurs est codée sur un nombre donné de bits. Un codage minimum de 4 bits par couleur donne 4 096 couleurs. En général, on code sur 8 bits (soit 24 bits par pixel) pour obtenir 16 777 216 couleurs. Le principal inconvénient de ce mode réside dans le poids des fichiers qui sont 24 fois plus lourds qu'en mode bitonal. En revanche, elle est recommandée pour des documents présentant un nombre important de couleurs, pour des documents mélangeant textes et illustrations en couleurs.

Il est impossible de donner à une image bitmap une dimension de façon absolue car le pixel n'est pas caractérisé par une taille fixe. Cette dernière varie en effet en fonction de la résolution du périphérique de visualisation ou d'impression. La dimension d'une image bitmap peut donc se définir de la façon suivante :

Nombre de pixels en largeur x nombre de pixels en hauteur = nombre de pixels de l'image

Une image de 700 x 500 pixels en noir et blanc pèse 42,72 Ko.

Une image de 2000 x 3000 pixels en couleur, soit 144 millions de bits, 18 millions d'octets (18 Mo), soit 12 disquettes. On peut sauvegarder 36 images de ce type sur un CD-R.

## ▲ Les avantages et inconvénients du mode image

Le mode image produit un fac-similé électronique du document d'origine. De ce fait, il permet d'étudier la typographie, l'écriture, les illustrations. Relativement simple à réaliser, il nécessite cependant une description bibliographique du document, ce qui inclut une indexation indispensable pour un accès aisé au document.

Le mode image présente deux inconvénients majeurs. Il exclut toute recherche à l'intérieur d'un texte. Par ailleurs, il génère des fichiers importants qui occupent de la place sur les supports de stockage.

## ▲ Le mode texte

Le mode texte offre la possibilité de recherche plein texte. Il est possible d'intervenir directement sur le document électronique. En revanche, le mode texte ne permet pas de conserver la présentation initiale du document original.

Pour obtenir du mode texte, deux solutions sont possibles. La première consiste à saisir manuellement les textes. Cette technique est coûteuse en temps et en main d'oeuvre mais c'est la seule envisageable pour la numérisation de textes manuscrits ou de textes dans des polices anciennes. Elle est très fiable lorsqu'il y a double saisie.

L'autre technique consiste à utiliser un logiciel de reconnaissance optique de caractères (ROC ou OCR, Optical character recognition). A partir d'un document numérisé en mode image, le logiciel convertit la configuration des points en signes typographiques dont il déduit la valeur dans une table de codification comme l'ISO 8859-1 (ISO Latin-1) ou comme l'ISO 10646 (UNICODE).

Les outils de reconnaissance optique ne sont pas encore totalement fiables. Même si certains logiciels annoncent un degré de fiabilité de 99 %, cela signifie que l'on peut trouver jusqu'à 10 mots erronés par page. Les logiciels de reconnaissance haut de gamme peuvent atteindre un taux d'exactitude de 99,95 %, ce qui ne représente pas plus d'une erreur pour 2 000 signes. Les fabricants de logiciels de ROC améliorent régulièrement leurs produits, notamment en ayant recours à des dictionnaires ou des outils d'analyse grammaticale. Ce type d'approche est peu adapté aux documents comportant de nombreux noms propres, des formes grammaticales anciennes ou écrits en plusieurs langues.

## ▲ Le mode vectoriel

Enfin, il existe un troisième mode dit vectoriel. C'est une technique de représentation de forme par des équations mathématiques. Il est principalement utilisé dans le domaine du dessin assisté par ordinateur.

Le passage d'un plan sur support papier à un plan vectoriel par le biais de la numérisation est une opération longue et coûteuse qui nécessite dans la quasi totalité des cas le recours à un opérateur spécialisé qui valide la conversion.

Il existe un format de représentation vectorielle des textes : c'est le format PDF d'Adobe. Il présente deux avantages :

- un poids moyen des fichiers faible par rapport à l'équivalent en mode image ;
- la possibilité de préserver la forme du document quel que soit le système de restitution.

C'est un format de présentation recommandé pour la publication de documents. Il peut indexer le texte et permet une navigation entre les chapitres lorsqu'ils ont été définis.

## ▲ Bibliographie

Anne R. Kenney, Stephen Chapman, *Digital Imaging for Libraries and Archives*, Ithaca (New-York), Cornell University Library, 1996.

Judith A. Zidar, "Optical scanning and text recognition", *Encyclopedia of Library and Information sciences*, ed. Allen Kent, New York, Marcel Dekker, 1995, vol.56 (suppl.19), p. 283-306.

## ▲ Sites à consulter

La couleur - principes, espaces colorimétriques, calibration de la chaîne graphique  
<http://www.erag.ch/MAC/Base.htm>

La couleur en informatique  
<http://www.fundp.ac.be/institution/autser/cc/formations/couleur/couleur.html>

Susan Haigh, *La Reconnaissance optique de caractères en tant que technologie de numérisation*.  
<http://nlc-bnc.ca/pubs/netnotes/fnotes14.htm>

Terry Kuny, *Introduction aux techniques et aux problèmes de la numérisation*  
<http://www.nlc-bnc.ca/pubs/netnotes/fnotes37.htm>

Preparing Quality Images for Computer Networks  
<http://www-personal.umich.edu/~jweise/quality/NetworkedImagesTOC.HTML>

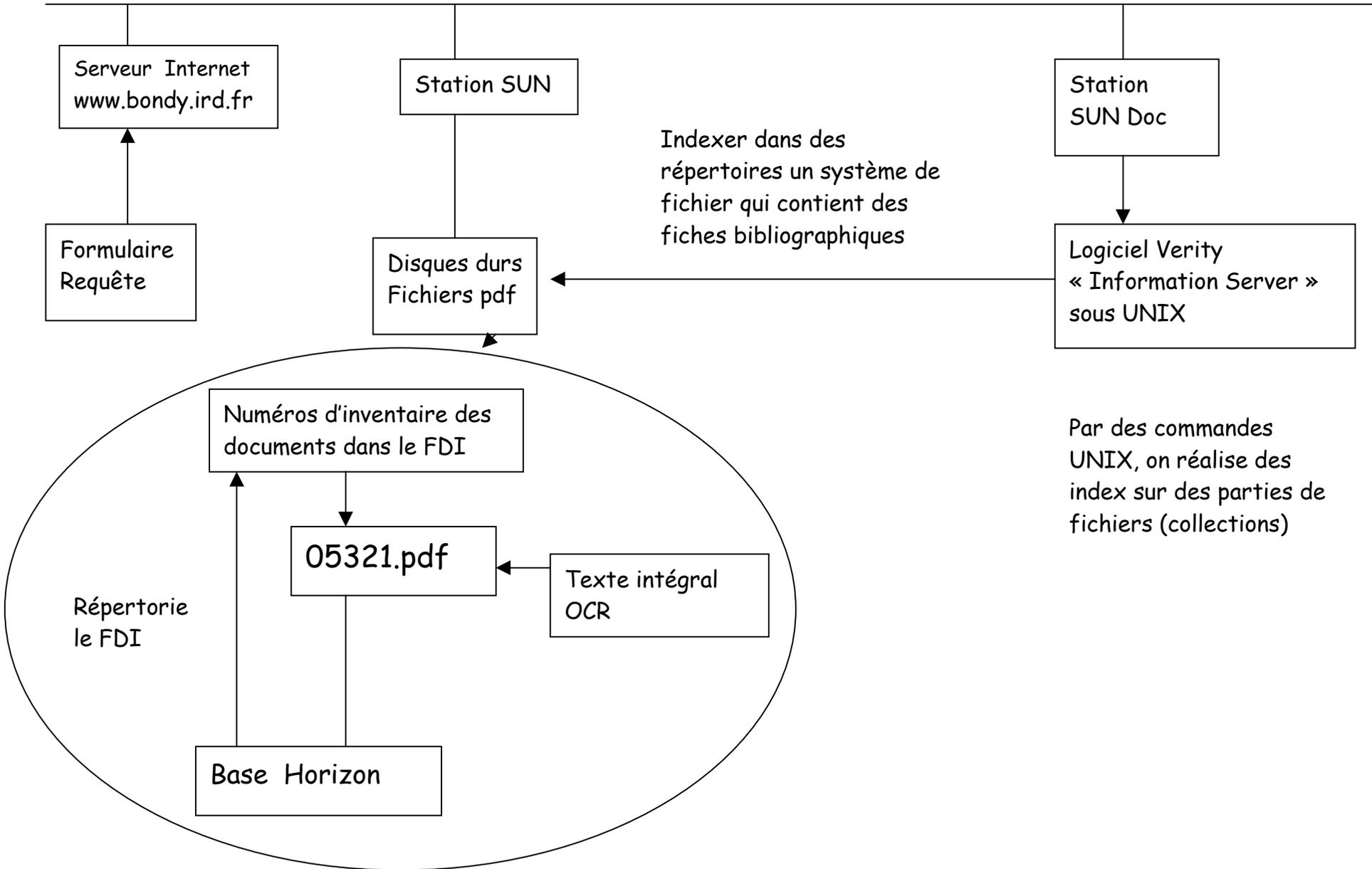
Introduction to Imaging  
Guide clair et concis édité par le Getty, accompagné d'un glossaire : les notions essentielles pour la réalisation d'une banque d'images.  
<http://www.getty.edu/gri/standard/introimages>

Digital Photography Solutions  
Site sur la photographie numérique  
<http://www.image-acquire.com/>

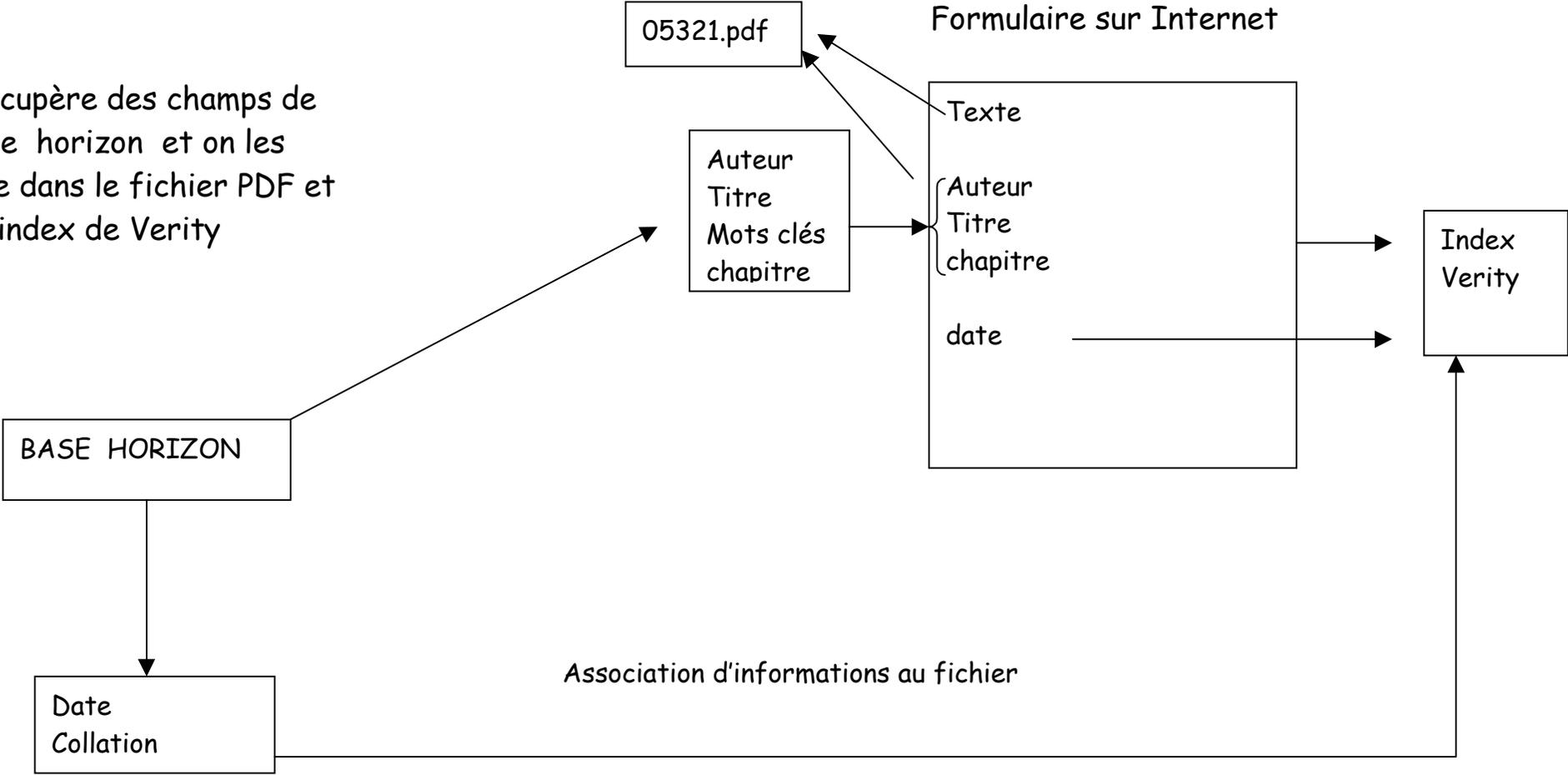
Article de Pier Luigi ROSSI sur  
la bibliothèque électronique  
« Pleins Textes »

Schéma sur le fonctionnement de  
la bibliothèque électronique  
« Pleins Textes »

**Pleins textes 1**



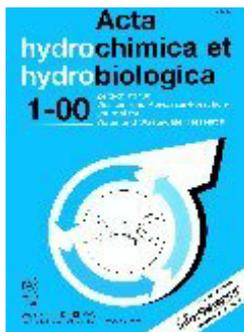
On récupère des champs de la base horizon et on les injecte dans le fichier PDF et dans l'index de Verity



# ANNEXE 11

## Charte graphique de l'IRD

Modèle de page HTML choisi lors de  
l'état des lieux de l'existant



Online ISSN: 1521-401X Print ISSN: 0323-4320

**[Acta hydrochimica et hydrobiologica](#)**

**Volume 28, Issue 5, 2000.**

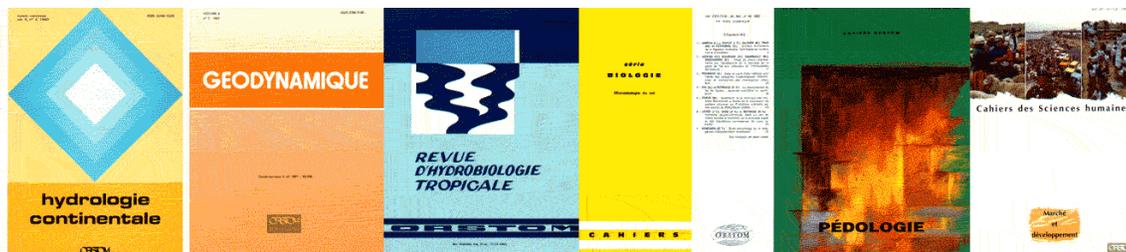
© 2000 WILEY-VCH Verlag GmbH, Weinheim, Fed. Rep. of Germany

### Research Paper

- 241-249 **Trophieindikation in Fließgewässern mit Hilfe des TIM (Trophie-Index Makrophyten) - Erprobung eines neu entwickelten Index im Inninger Bach**  
S. Schneider, Th. Krumpholz, A. Melzer  
[Abstract](#) [PDF Full Text](#) (Size: 1347K)  
**Published Online:** 1 Dec 2000  
**DOI** 10.1002/1521-401X(200005)28:5<241::AID-AHEH241>3.0.CO;2-Z
- 250-255 **Abschätzung der Eignung einer molekularbiologischen Methode zur Charakterisierung der Grundwasserbiozönose**  
B. Kuhlmann, B. Kilb, G. Preuß  
[Abstract](#) [PDF Full Text](#) (Size: 2658K)  
**Published Online:** 1 Dec 2000  
**DOI** 10.1002/1521-401X(200005)28:5<250::AID-AHEH250>3.0.CO;2-Z
- 256-261 **Effect of Preozonation on the Formation of Chlorinated Disinfection By-products for River Ruhr**  
Y.-W. Ko, G. Abbt-Braun, F. H. Frimmel  
[Abstract](#) [PDF Full Text](#) (Size: 522K)  
**Published Online:** 1 Dec 2000  
**DOI** 10.1002/1521-401X(200005)28:5<256::AID-AHEH256>3.0.CO;2-B
- 262-271 **Langzeitprognosen zum Sickerwasseraustrag von Schwermetallen aus Deponien von Müllverbrennungsschlacken**  
G. Hirschmann, U. Förster  
[Abstract](#) [PDF Full Text](#) (Size: 2418K)  
**Published Online:** 1 Dec 2000  
**DOI** 10.1002/1521-401X(200005)28:5<262::AID-AHEH262>3.0.CO;2-N
- 272-276 **Sorption Preconcentration and Determination of Cobalt in Industrial Solutions by Diffuse Reflection Spectroscopy Method**  
O. P. Kalyakina, O. N. Kononova, S. V. Kachin, A. G. Kholmogorov  
[Abstract](#) [PDF Full Text](#) (Size: 232K)  
**Published Online:** 1 Dec 2000  
**DOI** 10.1002/1521-401X(200005)28:5<272::AID-AHEH272>3.0.CO;2-J

Quelques exemples de pages HTML  
réalisées pour le site Internet :  
<http://www.bondy.ird.fr/tdp>

# Sommaires des revues de l'IRD



## ⇒ [Cahiers des Sciences Humaines](#)

Les Cahiers des Sciences Humaines présentent des articles consacrés à différentes disciplines des sciences sociales, telles que l'anthropologie, la sociologie, la démographie, l'économie, l'archéologie, la géographie, la linguistique, etc.

## ⇒ [Cahiers ORSTOM, série Biologie](#)

Les Cahiers ORSTOM, série Biologie présentent des articles consacrés à diverses branches de la biologie végétale et animale (agronomie, biologie et amélioration des plantes utiles, entomologie agricole, microbiologie du sol).

## ⇒ [Cahiers ORSTOM, série Pédologie](#)

Les Cahiers ORSTOM, série Pédologie présentent des articles consacrés à l'étude des sols : morphologie, caractérisations physico-chimique et minéralogique, classification, relation entre sols et géomorphologie, problèmes liés aux sels, à l'eau, à l'érosion, à la fertilité.

## ⇒ [Revue de Nématologie](#)

La Revue de Nématologie présente des articles consacrés à l'étude des nématodes parasites de plantes ou d'insectes, des nématodes libres du sol et des nématodes d'eaux douces ou marins.



Les Cahiers des Sciences Humaines présentent des articles consacrés à différentes disciplines des sciences sociales, telles que l'anthropologie, la sociologie, la démographie, l'économie, l'archéologie, la géographie, la linguistique, etc.

L'[index des auteurs](#) est également disponible.

## Cahiers ORSTOM, série Sciences Humaines

**1963**

⇒ Volume 1, Numéro [1](#), [2](#), [3](#), [4](#) ([Tome 1](#) et [Tome 2](#))

**1965**

⇒ Volume 2, Numéro [1](#), [2](#), [3](#), [4](#)

**1966**

⇒ Volume 3, Numéro [1](#), [2](#), [3](#), [4](#)

**1967**

⇒ Volume 4, Numéro [1](#), [2](#), [3](#) - [4](#)

**1968**

⇒ Volume 5, Numéro [1](#), [2](#), [3](#), [4](#)

**1969**

⇒ Volume 6, Numéro [1](#), [2](#), [3](#), [4](#)

**1970**

⇒ Volume 7, Numéro [1](#), [2](#), [3](#), [4](#)

**1971**

⇒ Volume 8, Numéro [1](#), [2](#), [3](#), [4](#), [Numéro Spécial](#)

**1972**

⇒ Volume 9, Numéro [1](#), [2](#), [3](#), [4](#)

**1973**

⇒ Volume 10, Numéro [1](#), [2](#) - [3](#), [4](#)

**1974**

⇒ Volume 11, Numéro [1](#), [2](#), [3](#) - [4](#)

**1975**

⇒ Volume 12, Numéro [1](#), [2](#), [3](#), [4](#)

**1976**

⇒ Volume 13, Numéro [1](#), [2](#), [3](#), [4](#)

**1977**

⇒ Volume 14, Numéro [1](#), [2](#), [3](#), [4](#)

**1978**

⇒ Volume 15, Numéro [1](#), [2](#), [3](#), [4](#)

**1979**

⇒ Volume 16, Numéro [1](#) - [2](#), [3](#), [4](#)

**1980**

⇒ Volume 17, Numéro [1](#) - [2](#), [3](#) - [4](#)

**1981 - 1982**

⇒ Volume 18, Numéro [1](#), [2](#), [3](#), [4](#)

**1983**

⇒ Volume 19, Numéro [1](#), [2](#), [3](#), [4](#)

**1984**

⇒ Volume 20, Numéro [1](#), [2](#), [3-4](#)

**1985**

⇒ Volume 21, Numéro [1](#), [2-3](#), [4](#)

## Cahiers des Sciences Humaines

**1986**

⇒ Volume 22, Numéro [1](#), [2](#), [3-4](#)

**1987**

⇒ Volume 23, Numéro [1](#), [2](#), [3-4](#)

**1988**

⇒ Volume 24, Numéro [1](#), [2](#), [3](#), [4](#)

**1989**

⇒ Volume 25, Numéro [1-2](#), [3](#), [4](#)

**1990**

⇒ Volume 26, Numéro [1-2](#), [3](#), [4](#)

**1991**

⇒ Volume 27, Numéro [1-2](#), [3-4](#)

**1992**

⇒ Volume 28, Numéro [1](#), [2](#), [3](#), [4](#)

**1993**

⇒ Numéro Hors-série "[Trente ans](#)" (1963 - 1992)

⇒ Volume 29, Numéro [1](#), [2-3](#), [4](#)

**1994**

⇒ Volume 30, Numéro [1-2](#), [3](#), [4](#)

**1995**

⇒ Volume 31, Numéro [1](#), [2](#), [3](#), [4](#)

**1996**

⇒ Volume 32, Numéro [1](#), [2](#), [3](#), [4](#)

20/09/2000.

**Tél :**01-48-02-56-96

**Fax :**01-48-47-30-88

✉ [pleins\\_textes@bondy.ird.fr](mailto:pleins_textes@bondy.ird.fr)



 **Recherche**

 **L'IRD dans le monde**

 **Pratique**



## Cahiers ORSTOM, série Sciences Humaines

1963 - Volume 1, Numéro 1

133 p.

**Facteurs d'intégration et de désintégration du travailleur gabonais à son entreprise**

Biffot, Laurent

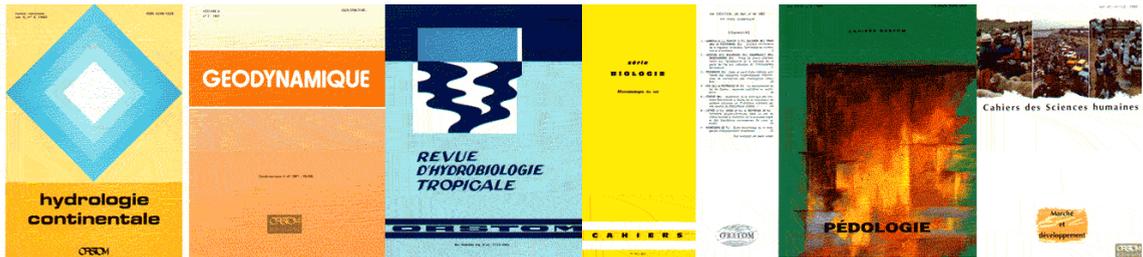
[Document au format PDF](#)

03/10/2000.

Tél : 01-48-02-56-96

Fax : 01-48-47-30-88

 [pleins\\_textes@bondy.ird.fr](mailto:pleins_textes@bondy.ird.fr)



**A** (26) **B** (67) **C** (49) **D** (45) **E** (11) **F** (21) **G** (46) **H** (26) **I** (3) **J** (9) **K** (11) **L** (51) **M** (39)  
**N** (4) **O** (9) **P** (39) **Q** (3) **R** (29) **S** (27) **T** (16) **U** (0) **V** (19) **W** (9) **X** (0) **Y** (1) **Z** (3)

**Z :**

### Zangato, E.

Zangato, E.. Etude du mégalithisme en République centrafricaine : nouvelles découvertes de monuments à chambre dans le secteur de Ndio. Cahiers des Sciences Humaines, 1996, Vol. 32, Num. 2, p. 361-377

[Document au format PDF](#)

### Zimmermann, M.

Pettang, C.; Vermande, P.; Zimmermann, M.. L'impact du secteur informel dans la production de l'habitat au Cameroun. Cahiers des Sciences Humaines, 1995, Vol. 31, Num. 4, p. 883-903

[Document au format PDF](#)

### Zola Baez, Manuel G.

Zola Baez, Manuel G.; Marchal, Jean-Yves (pref.). Aménagement préhispanique des zones inondées du Veracruz (Mexique). Cahiers des Sciences Humaines, 1986, Vol. 22, Num. 1, p. 83-95

[Document au format PDF](#)

