MINISTERE DE L'EDUCATION NATIONALE
CONSERVATOIRE NATIONAL DES ARTS ET METIERS
INSTITUT FRANCAIS D'INGENIERIE

CYCLE DE FORMATION DE CHEF DE PROJET
EN INGENIERIE INFORMATIQUE
(formation post-ingénieur)

MEMOIRE DE FIN D'ETUDES

présenté par

Patricio SOTO

PROPOSITION D'UN SYSTEME INFORMATIQUE OUVERT POUR LA GESTION TRANSDISCIPLINAIRE DES DONNEES DANS LE CADRE DE L'INSTITUT FRANCAIS DE L'ENVIRONNEMENT

SOPHIA ANTIPOLIS / Année universitaire 1990 - 1991

Décembre 1992

RESUME

PROPOSITION D'UN SYSTEME INFORMATIQUE OUVERT POUR LA GESTION TRANSDISCIPLINAIRE DES DONNEES DANS LE CADRE DE L'INSTITUT FRANCAIS DE L'ENVIRONNEMENT

Cette étude concerne le dimensionnement des systèmes informatiques de l'Institut Français de l'Environnement, dont le rôle est d'assurer une validation et une structuration des données recueillies, pour les diffuser auprès des utilisateurs concernés. L'élaboration du Plan National de l'Environnement (février 1989) a montré que le point faible du dispositif actuel est bien le manque de système d'information coordonné, d'expertise et d'organisation.

Les objectifs de l'étude peuvent être résumés de la manière suivante :

* décrire la structure générale de l'Institut et ses relations extérieures,
* décrire des configurations informatiques répondant aux besoins de l'Institut en présentant un schéma d'organisation du système informatique,
* recenser les schémas de réseaux pour mettre en relation l'Institut et le monde extérieur : les organismes externes et le public,
* évaluer les volumes informatiques et de réseaux,
* esquisser une évaluation financière du système de gestion et de traitement de l'information.

Les solutions techniques proposées dans ce rapport sont basées sur les techniques et moyens disponibles, immédiatement ou à brève échéance, qui permettent la mise en place du système d'information de l'Institut.

L'étude consiste à proposer des solutions techniques d'organisation informatique correspondant aux fonctions dévolues à l'Institut et à la structure de ses relations extérieures.

Pour développer et coordonner de manière efficace les politiques sur l'Environnement à l'échelle régionale, nationale, communautaire ou mondiale, il est nécessaire de disposer de systèmes d'information couvrant les aspects scientifiques et techniques des phénomènes.

L'informatique est l'un des moyens privilégiés qui permet d'apporter au système d'information un accroissement des performances de traitement, de gestion, d'organisation, de mémorisation, de recherche d'information et de restitution de celles-ci sous différentes formes (écran, fichier, papier). L'évaluation des systèmes d'information et de communication inclut, bien entendu, l'évaluation de l'efficience de ces systèmes, et par conséquent, des outils qui le servent. L'ordinateur (maxi, mini ou micro) en constitue l'une des composantes essentielles.

L'autre composante qui prend en compte la conception et le développement d'un système d'information intégré pour l'environnement est un cas spécifique d'ingénierie ; elle implique, à côté des moyens informatiques, une organisation homogène et unitaire des applications ; elle s'insère dans le plan directeur informatique et plus généralement dans le plan directeur de chaque entreprise concernée.

Mots-clés : Ecologie - Environnement - Système d'Information - Systèmique - Génie Logiciel - Base de Données - Modélisation - Réseau -

Key-words : Ecology - Environment - Information System - Systemics - Software engineering - Modelling - Data Base - Network -

REMERCIEMENTS

Je tiens à exprimer mes plus sincères remerciements aux personnes qui ont contribué, par leur aide ou par leurs conseils, à ce que ce travail aboutisse.

Tout d'abord, je remercie Jacques VARET, Directeur et Jean-Louis WEBER Chargé des Relations Internationales de l'Institut Français de l'Environnement, Marcel B. BOUCHE, Directeur de Recherche à l'INRA, Gilbert LONG, Directeur de Recherche au CNRS qui ont servi à définir le cadre général de l'étude.

Je tiens à exprimer toute ma reconnaissance aux chercheurs et ingénieurs des différents organismes de recherche à Montpellier qui ont apporté leurs données et leurs connaissances théoriques et pratiques dans leurs disciplines permettant la validation du schéma conceptuel global d'"ECORDRE" et la création d'une véritable synergie de coopération interdisciplinaire et interinstitutionnelle. Je tiens particulièrement à nommer : Jean Marc BARBIER du Laboratoire d'Etude Comparées des Systèmes Agraire (INRA), Alain DERVIEUX et Alain TAMISIER du groupe d'ornithologie du CEFE (CNRS), Alain DUCREUX du CIRAD-SAR, Philippe GRANVAL de l'ONC, Jean-Pierre LAPORTE de la Station de Economie et de Sociologie Rurales (INRA), Georges MANIERE du Génie Rural et Machinisme Agricole (ENSA), Patrick SILAN du Laboratoire de parasitologie comparée (Université Montpellier II).

J'exprime toute ma reconnaissance pour leur appui logistique au niveau institutionnel à : Daniel BOULET Président du Centre INRA de Montpellier, Gilles CARAUX Responsable Informatique du Centre INRA/ENSA de Montpellier, Catherine CHRISTOPHE Responsable de l'Unité Bases de Données à l'INRA, Alfred CONESA Président d'AGROPOLIS, Jacques MARZIN Directeur de la Direction Informatique de l'INRA, Marc MEUNIER Responsable National de la Formation en Informatique à l'INRA, Jean PICHOT Directeur scientifique du CIRAD-SAR.

Je remercie ma collègue de travail Régine ALIAGA pour son dévouement dans le travail de mise à jour permanente de la base de données ECORDRE. Mes remerciements vont aussi à ma femme Rosa pour son travail d'analyse et programmation dans l'équipe de projets que je dirige.

Je voudrais remercier ici toutes les personnes de l'atelier de dessins du CEFE-CNRS Montpellier pour leur extraordinaire travail d'artiste qui ont permis l'élaboration de la quasi totalité des figures présentes dans ce rapport. Mes remerciements vont aussi aux ingénieurs informaticiens du CNUSC, pour leur assistance permanente.

TABLE DES MATIERES

CHAPITRE I. ***Etude du système informatique de l'Institut Français de l'ENvironnement (IFEN)***

1.1. INTRODUCTION 2
1.2. CONTEXTE DE L'ETUDE 4
1.2.1. Présentation 4
1.2.2. Objectif de l'étude 4
1.2.3. Champs de l'étude 5
1.2.4. Modalités de l'étude 5

1.3. DESCRIPTION DE LA STRUCTURE GENERALE DE L'INSTITUT ET SES RELATIONS EXTERIEURES 7
1.3.1. Structure générale et finalités 7
1.3.2. Un Centre d'observation et d'information 9
1.3.3. Un Centre d'expertise et un outil d'aide à la décision 10
1.3.4. L'IFEN correspondant de l'Agence Européenne de l'Environnement et des organisations internationales 11
1.3.5. Les partenaires de l'IFEN 12
1.3.6. Données gérées par l'Institut dans un cadre de partenariat 13

1.4. DECRIRE DES CONFIGURATIONS INFORMATIQUES REPONDANT AUX BESOINS DE L'INSTITUT 14
1.4.1. Rappels des fonctions 14
1.4.2. Banque de données 15
1.4.3. Collecte et mise à disposition de l'information 15
1.4.4. Unité d'études 16
1.4.5. Service d'archives 17
1.4.6. Schéma d'organisation du système informatique 17
1.4.7. Organisation et responsabilité 19
1.4.8. Propositions de configuration informatique 20
1.4.9. Configuration de banque de données 26
1.4.10. Configuration de production d'analyses 30
1.4.11. Configuration de collecte, de validation et d'interrogation 31
1.4.12. Configuration d'études 34

1.5. RECENSER LES SCHEMAS DE RESEAUX POUR METTRE EN RELATION L'INSTITUT ET LE MONDE EXTERIEUR 36
1.5.1. Considérations techniques : réseaux locaux 36
1.5.2. Considérations techniques : réseaux distants 37
1.5.3. Solution possible 38

1.6. EVALUATION DES VOLUMES INFORMATIQUES ET DE RESEAUX 39
1.6.1. Banque de données 39
1.6.2. Collecte et mise à disposition de l'information 41
1.6.3. Unité d'études 46
1.6.4. Service d'archives 48

1.7. ESTIMATION D'UNE ENVELOPPE DES COUTS 50
1.7.1. Moyens de recueil et validation des données 50
1.7.2. Moyens de mise en forme et entrée dans la banque 51
1.7.3. Moyens d'entretien de la banque de données 52
1.7.4. Moyens de l'unité d'études 54
1.7.5. Moyens de production et diffusion 55
1.7.6. Récapitulatif 56

1.8. CONCLUSIONS ET PERSPECTIVES 57

CHAPITRE II. ***Quelques concepts et outils nécessaires pour la gestion d'un système information operationnel au service de l'environnement et de l'intégration transdisciplinaire***

2.1. LES BASES DE DONNEES REPARTIES : UNE SOLUTION A LA GESTION INTEGREE ET TRANSDISCIPLINAIRE DES DONNEES DE L'IFEN 59
2.1.1. Présentation 59
2.1.2. Les bases de données relationnelles 60
2.1.3. Limites des SGBD relationnels 61
2.1.4. Les SGBD avancés 64
2.1.5. Pourquoi une gestion orientée bases de données réparties ? 65
2.1.6. Les bases de données réparties 67
2.1.7. Objectifs des SGBD répartis 69
2.1.8. Architectures de référence 72
2.1.9. Conclusion 76

2.2. L'ARCHITECTURE CLIENT/SERVEUR : SOLUTION AU TRAITEMENT COOPERATIF ET A LA GESTION DU SYSTEME D'INFORMATION DE L'IFEN 78
2.2.1. Présentation 78
2.2.2. Failles et limites 78
2.2.3. Transparence 80
2.2.4. Efficacité et accessibilité 81
2.2.5. Le serveur de données : étape décisive dans le traitement de l'information 81
2.2.6. Traitement différé contre temps réel 83
2.2.7. La transaction : élément important de l'architecture Client/Serveur 84
2.2.8. La communication multiserveur 85
2.2.9. Le two-phase commit 87
2.2.10. Une question de réseaux 88
2.2.11. Cahier de charge de l'architecture Client/Serveur 88
2.2.12. Quelques produits du marché côté client 89
2.2.13. Plusieurs niveaux d'utilisation 92
2.2.14. Conclusion 93

2.3. LE RESEAU NATIONAL DE TELECOMMUNICATIONS POUR LA RECHERCHE : EPINE DORSALE DE L'ARCHITECTURE CLIENT/SERVEUR DE L'IFEN 95
2.3.1. Introduction 95
2.3.2. Le contexte 96
2.3.3. Un réseau fédérateur et performant 96
2.3.4. Des réseaux de campus au réseau national 97
2.3.5. L'architecture du réseau national de télécommunications pour la recherche 100
2.3.6. Les services disponibles sur ce réseau 101
2.3.7. Le cahier des charges de RENATER 102
2.3.8. Les perspectives d'évolution 103
2.3.9. Des enjeux pour la recherche française 104
2.3.10. Le réseau national de télécommunications pour la recherche et la construction de l'Europe 106

2.4. NECESSITE DE DEFINIR LA QUALITE DES OUTILS DU GENIE LOGICIEL POUR REPONDRE AUX BESOINS DE DEVELOPPEMENT D'APPLICATIONS A L'IFEN . 107
2.4.1. Introduction sur le génie logiciel 107
2.4.2. Répondre aux besoins applicatifs 108
2.4.3. Répondre aux besoins de l'IFEN et ses partenaires en matière de maintenance 112
2.4.4. Répondre aux exigences en matière de qualité et de standardisation 114
2.4.5. Répondre aux besoins en matière de développement d'applications dans le cadre d'une architecture Client/Serveur 115

2.5. L'HYPERTEXTE EL LA PROGRAMMATION ORIENTEE OBJETS : DES OUTILS NECESSAIRES POUR LA CREATION DES SYSTEMES D'EXPLICITATION ET D'AIDE A LA DECISION 119
2.5.1. L'Hypertexte 119
2.5.2. La Programmation Orientée Objets 120

2.6. CONCEPTS ET OUTILS POUR LES APPLICATIONS GEOGRAPHIQUES 122
2.6.1. Introduction 122
2.6.2. L'informatique géographique 122
2.6.3. Exemples d'applications géographiques 124
2.6.4. Les différentes approches proposées 125
2.6.5. Evolution de l'offre matérielle et logicielle entre 1987 et 1992 128
2.6.6. Conclusion sur les approches 128

CHAPITRE III. ***Fondement sémantico-conceptuel des données écologiques et environnementales***

3.1. INTRODUCTION GENERALE 131

3.2. L'ECOLOGIE ET LES DISCIPLINES CONNEXES 133
3.2.1. Introduction 133
3.2.2. L'écologie 134
3.2.3. Ecologie et sciences humaines 135
3.2.4. L'écologie et l'interdisciplinarité 136
3.2.5. Ecologie et réfutabilité 137

3.3. L'INFORMATION ECOLOGIQUE 138
3.3.1. Introduction 138
3.3.2. Démarches analytiques, synthétiques et globales 141
3.3.3. Le champ écologique 142
3.3.4. Le temps 145
3.3.5. L'espace : hétérogénéité et continuum 145
3.3.6. Champ écologique et choix des données : analyse, synthèse et globalité 147
3.3.7. Données objectives et hypothèses 148

3.4. FORMES ET FONCTIONS 149
3.4.1. Notion de caractéristique 149
3.4.2. Etats et cinèses 149
3.4.3. Relations entre états et cinèses 151

3.5. ACQUISITION D'INFORMATION 152
3.5.1. L'élaboration des connaissances écologiques 152
3.5.2. La donnée : l'élément d'information sur le champ écologique 153
3.5.3. Donnée et liens 155
3.5.4. Notion de prélevat 158
3.5.5. Sélection des données et options 160
3.5.6. Liens des données au champ écologique 163
3.5.7. Collecte des données 165
3.5.8. Options chronologiques sur les prélevats 169
3.5.9. Options spatiales : prélevats euclidiens, cartésiens, informes et ponctuels 170
3.5.10. Associations de prélevats (prélevats associés et l'associateur) 172

3.6. LES ENSEMBLES DE DONNEES 174
3.6.1. Objectifs, ensembles et caractéristiques 174
3.6.2. Options spatio-temporelles sur les ensembles 175
3.6.3. Qualité des données et des ensembles : les erreurs 175

3.7. LA SAISIE DES DONNEES INITIALES 177
3.7.1. La saisie 177
3.7.2. Le cas des données continues 178
3.7.3. Collections et enregistrements (données non saisies) 179

3.8. LES LIENS ENTRE DONNEES 179
3.8.1. L'organisation des données 179
3.8.2. Situations des prélevats dans le champ écologique 180
3.8.3. Champ écologique et informations 182
3.8.4. Le référentiel des données 182
3.8.5. L'acquisition de données : principes d'optimisation 184

3.9. CONCLUSION GENERALE 185

CHAPITRE IV. ***Le Système d'Information ECORDRE : un outil transdisciplinaire au service de l'environnement***

4.1. INTRODUCTION 188

4.2. APPROCHE DES SYSTEMES COMPLEXES 191
4.2.1. L'approche systémique 191
4.2.2. La modélisation des systèmes 193
4.2.3. Les étapes de la modélisation 194
4.2.4. Les méthodes pour l'approche pratique 194

4.3. LE SYSTEME D'INFORMATION ECORDRE 196
4.3.1. De la base de données ECORDRE 196
4.3.2. ... Au Système d'Information ECORDRE 197
4.3.3. Le nouveau cadre du projet ECORDRE 197
4.3.4. Les moyens pour la vulgarisation scientifique 199

4.4. QUELQUES EXEMPLES DES SCHEMAS D'APPLICATIONS 200
4.4.1. Quelques relations de la base de données communes aux différents sous-schémas d'application 200
4.4.2. Etude des dénombrements de canards et de foulques en Camargue 205
4.4.3. Etude du rôle agronomique et écologique des lombriciens 210
4.4.4. Apport du machinisme à la modélisation globale de l'information sur l'exploitation agricole 217
4.4.5. Etude des mécanismes de l'économie viti-vinicole et du marché du vin au niveau national et international 223

CONCLUSION GENERALE 230
BIBLIOGRAPHIE 233
ANNEXES 240

LISTE DES FIGURES

Figure 1-1 : Contexte de l'IFEN 8

Figure 1-2 : Invariants de l'IFEN 14

Figure 1-3 : Fonctins de l'IFEN 19

Figure 1-4 : Principe d'une sous-traitance possible des mises à jour de la banque de données 20

Figure 1-5 : Modèle "essentiel" d'un système informatique 21

Figure 1-6 : Configuration informatique centralisée de la banque de données 28

Figure 1-7 : Configuration informatique décentralisée de la banque de données 29

Figure 1-8 : Schéma de circulation en entrée 39

Figure 1-9 : Fonction de l'unité d'études 46

Figure 1-10 : Constitution de l'archive 48

Figure 2-1 : Une base de données répartie 67

Figure 2-2 : Différentes approches de conception des bases de données réparties 69

Figure 2-3 : Architecture des schémas d'une base de données répartie 73

Figure 2-4 : Architecture fonctionnelle d'un SGBD réparti 75

Figure 2-5 : L'architecture client/serveur permet à plusieurs applications clientes de partager une base de données commune 82

Figure 2-6 : Modèle client/serveur 85

Figure 2-7 : La solution SYBASE 89

Figure 2-8 : Architecture du réseau national de télécommunication pour la recherche 100

Figure 3-1 : Le limites spatiales de la biosphère sont identiques avec celles de l'écosphère 138

Figure 3-2 : Le champ écologique et ses trois référendaires 143

Figure 3-3 : Représentation "esotérique" d'un ensemble, sous-ensembles et liens entre les données 155

Figure 3-4 : Caractéristiques, situation et limites d'un prelevat 160

Figure 3-5 : La situation spatiale des données 164

Figure 3-6 : Intégration et repères spatio-temporels des données agro-écologiques dans l'écosphère 181

Figure 4-1 : L'interdisciplinarité, lors de l'acquisition des données ... 200

Figure 4-2 : Canard colvert et foulque macroule 205

Figure 4-3 : Localisation géographique des sites de recherche en Camargue 206

Figure 4-4 : Rôle des lombriciens 210

Figure 4-5 : Situation spatiale et dimensions d'un plot 211

Figure 4-6 : Le machinisme agricole 218

Figure 4-7 : L'économie viti-vinicole 223

LISTE DES TABLEAUX

Tableau 3-1 : Niveaux de perception écologique 140

Tableau 3-2 : Le champ écologique et sa globalyse 144

CHAPITRE PREMIER

ETUDE DU SYSTEME INFORMATIQUE DE L'INSTITUT FRANCAIS DE L'ENVIRONNEMENT

1.1. INTRODUCTION

La France est naturellement dotée d'une grande richesse et d'une diversité exceptionnelle dans ses espaces, ses paysages, ses espèces animales et végétales, ses eaux, ses sols et son sous-sol. Ces dernières années ont été marquées par une dégradation de l'environnement dans beaucoup de domaines : contamination des nappes phréatiques, accumulation des toxiques dans les sols, pollution de la haute atmosphère, urbanisation du littoral et de zones sensibles, déchets mal contrôlés.

Ce sont des individus qui innovent technologiquement, qui "spontanément" mettent en oeuvre des nouveaux outils dans le cadre de régulations institutionnalisées antérieurement ou modifiées sous l'effet de ces innovations. Ces politiques institutionnelles sont finalement décidées et infléchies par des individus.

L'accès à l'information, 1) rapidement, 2) sous la forme souhaitée par l'utilisateur, donc 3) sans bornage disciplinaire et 4) parfaitement à jour, est donc indispensable à la gestion environnementale de nos milieux dans le long terme en mettant à disposition la connaissance acquise par et pour les acteurs et les régulateurs.

Opérationnellement, il faut que les acteurs concevant des nouveaux produits ou/et mettant en oeuvre de nouvelles techniques, notamment dans l'espace rural ou urbain, puissent accéder à toutes connaissances fondées :

a) sur des données issues du concret (observations et mesures),

b) sur les opinions interprétant ces données selon le point de vue de divers "experts" y compris les régulateurs (aspects réglementaires).

L'état de ces connaissances étant quantitatif, qualitatif ou flou.

Toute décision partielle, par exemple labourer telle parcelle et tel jour, ou générale, par exemple interdire sur le marché les molécules chimiques entraînant une mortalité supérieure à 15% dans un test de toxicologie appliqué à une espèce donnée, sont des choix empiriques pris "au mieux" par un décideur-acteur (ou un petit groupe de travail) aux informations individuelles et à l'expérience personnelle nécessairement limitée.

Les problèmes d'environnement n'ont pas d'autre origine que cette limitation des décideurs-acteurs et les freins à un développement durable résultent de la mise en oeuvre de techniques à partir d'analyses trop sectorielles.

Il faut donc créer des systèmes d'information échangeant en principe le maximum de connaissance humaine, y compris le doute sur la qualité de celle-ci, sans barrière disciplinaire, sans retard de mise à jour et de transmission, sans occultation des doutes des experts et des représentations (modèles).

Dans ce sens, une mobilisation de tous les acteurs : collectivités locales, industriels, associations, chercheurs, et bien sûr pouvoirs publics est indispensable en prenant en compte les objectifs définis par le Plan National pour l'Environnement.

L'Institut Français de l'Environnement constitue un des éléments essentiels de cette nouvelle politique, qui doit amener la France au niveau des pays les plus avancés dans le monde pour la qualité de leur cadre de vie.

La qualité écologique des produits, des procédés et des investissements fait d'ores et déjà partie des facteurs de compétitivité internationale. Cette tendance ira croissant dans l'avenir, et l'Institut Français de l'Environnement participe à la politique de "développement durable" de l'économie qui se met en place au plan national et européen en aidant à la définition de critères appropriés.

Par ailleurs, l'environnement n'est pas seulement une affaire locale et nationale, et nous percevons tous et de manière croissante notre interdépendance régionale et planétaire.

1.2. CONTEXTE DE L'ETUDE

1.2.1. Présentation

Cette étude concerne le dimensionnement des systèmes informatiques de l'Institut Français de l'Environnement, principalement dans le cadre de la gestion, l'organisation et le traitement de l'information et des outils d'aide à la décision. Elle se base sur l'étude préliminaire réalisé par Jacques VARET à la demande du Ministre de l'Environnement le 12 janvier 1990 (un premier rapport fut édité en avril 1990).

La démarche ayant conduit à cette étude découle aussi des réflexion menées à l'occasion de l'élaboration du Plan National de l'Environnement (février 1989), montrant que le point faible du dispositif actuel est bien le manque de système d'information coordonné, d'expertise et d'organisation.

Cette étude fait suite aussi aux entretiens avec J.L. WEBER (SRETIE), J. VARET (BRGM), M.B. BOUCHé (INRA), G. LONG (CNRS) qui ont servi à définir son cadre général, ainsi qu'aux résultats des travaux antérieurs concernant la faisabilité et l'organisation de l'Institut.

1.2.2. Objectif de l'étude

L'examen des documents cités, complété par les informations fournies par les différents acteurs, fait ressortir les objectifs de l'étude :

* décrire la structure générale de l'Institut et ses relations extérieures,
* décrire des configurations informatiques répondant aux besoins de l'Institut en présentant un schéma d'organisation du système informatique,
* recenser les schémas de réseaux pour mettre en relation l'Institut et le monde extérieur : les organismes externes et le public,
* évaluer les volumes informatique et de réseaux,
* esquisser une évaluation financière du système de gestion et traitement de l'information.

1.2.3. Champs de l'étude

Depuis quelques années, différentes tentatives ont été entreprises pour élaborer des comptes de ressources naturelles, de patrimoine bio-écologique. Il s'agit bien de compter, de recenser les flux et stocks de ressources avec les moyens techniques et comptables disponibles (télédétection par satellite, recensement de terrain,...).

Au niveau Européen, le programme CORINE-Land cover a constitué des banques de données géographiques et cartographiques pour l'Europe du Sud, à savoir le Portugal, l'Espagne et la France-Sud (programme en cours). Il s'agit d'élargir ces actions pour passer du spatial à la gestion de toutes données concrètes dans l'espace-temps et renforcer leur cohérence au moyen d'une structure d'intégration associée à des réseaux d'organismes communautaires. Au niveau communautaire, cet organisme serait chargé de la coordination des données des différents pays membres, chaque acteur conservant son système d'information propre.

Pour la France, une telle structure (un point focal) pourrait être constituée par **l'Institut Français de l'Environnement** dont le rôle est d'assurer une validation et une structuration des données recueillies, pour les diffuser auprès des utilisateurs.

1.2.4. Modalités de l'étude

Les solutions techniques proposées dans ce document sont basées sur les techniques et moyens disponibles, immédiatement ou à brève échéance, qui permettent la mise en place du système d'information de l'Institut.

L'étude consiste à proposer des solutions techniques d'organisation informatique correspondant aux fonctions dévolues à l'Institut et à la structure de ses relations extérieures.

Pour développer et coordonner de manière efficace les politiques sur l'Environnement à l'échelle régionale, nationale, communautaire ou mondiale, il est nécessaire de disposer de systèmes d'information couvrant les aspects scientifiques et techniques des phénomènes.

Il est apparu que les données acquises ne beneficient pas des mêmes soins ; seules actuellement certains sont recueillies. Elles présentent un caractère hétérogène dans le

format des données ; ceci introduit des lacunes et des redondances néfastes à la bonne compréhension des phénomènes mis en jeu. D'où la nécessité de leur description à l'aide respectivement de structures, de procédures et de protocoles mettant en jeu une modélisation des données, des traitements et des communications qui doivent être :

- *fidèles*, c'est-à-dire être une image non exacte mais conforme, significative de la réalité ; la réalité évoluant, un système d'information doit être extensible, c'est-à-dire que l'on doit pouvoir à tout moment mettre à jour des informations sans bouleverser toutes les autres informations.
- *cohérents*, c'est-à-dire composé de parties harmonisées entre elles, de façon à être la base de traitements utilisables ; il ne doit pas être source de distorsion,
- *simples*, afin que l'utilisateur puisse l'appréhender, le valider et le maintenir,
- *attractifs*, afin de stimuler la conservation et la mise en commun de ces données,
- *irredondant* dans la mesure où c'est un bon critère de cohérence et de simplicité.

L'informatique est l'un des moyens privilégiés qui permet d'apporter au système d'information un accroissement des performances de traitement, de gestion, d'organisation, de mémorisation, de recherche d'information et de restitution de celles-ci sous différentes formes (écran, fichier, papier). L'évaluation des systèmes d'information et de communication inclut, bien entendu, l'évaluation de l'efficience de ces systèmes, et par conséquent, des outils qui le servent. L'ordinateur (maxi, mini ou micro) en constitue l'une des composantes essentielles.

L'autre composante qui prend en compte la conception et le développement d'un système d'information intégré pour l'environnement est un cas spécifique d'ingénierie ; elle implique, à côté des moyens informatiques, une organisation homogène et unitaire des applications ; elle s'insère dans le plan directeur informatique et plus généralement dans le plan directeur de chaque entreprise concernée. elle suppose :

- un *ensemble de concepts* (modèles) servant de référentiel et permettant de décrire des données, des traitements ou des transferts de données,
- un *langage de spécification* permettant de décrire les modèles obtenus, se réduisant à des documents standards ou à des dossiers types,
- une *démarche* (ou méthode de conception), définissant les étapes du cycle de développement et formalisation du processus de construction d'un système d'information,
- des *outils* appelés systèmes d'aide à la spécification ou programmes d'aide à l'analyse : dictionnaire de données, documentation et éditeurs structurés, outils graphiques, de simulation, de maquettage, de validation, d'interfaçage, de génération de code, ...

1.3. DESCRIPTION DE LA STRUCTURE GENERALE DE L'INSTITUT ET SES RELATIONS EXTERIEURES

1.3.1. Structure générale et finalités

POURQUOI L'INSTITUT FRANCAIS DE L'ENVIRONNEMENT ?

Pour répondre au développement des préoccupations relatives à l'environnement, s'accompagnant d'une demande sociale croissante d'informations objectives et facilement accessibles sur l'état et l'évolution des milieux, du patrimoine naturel et de la qualité du cadre de vie.

Pour que la France dispose d'un dispositif d'information, de documentation et des capacités d'expertise suffisantes, nécessaires pour répondre efficacement aux exigences précédentes dans des conditions assurant toute la neutralité indispensable.

POUR QUI ?

Pour répondre aux responsables publics et aux entreprises qui ont par ailleurs besoin de pouvoir disposer de données synthétiques précises et donc non controversées pour prendre leurs décisions ou évaluer l'efficacité de leur action.

Pour que la France dispose d'un organisme à vocation scientifique et statistique susceptible d'être le correspondant national de l'Agence Européenne de l'Environnement.

QUE FAIRE ?

Permettre une révision en profondeur et une rationalisation du dispositif de traitement, d'harmonisation et de valorisation de données environnementales, ainsi que des méthodes de mesure et d'évaluation ou de surveillance continue des milieux.

Assurer une large diffusion dans le public de l'information et de la documentation ainsi synthétisée et pour animer des structures de formation.

Constituer un pôle de référence et contribuer à la présence française en faveur de l'environnement à l'échelle internationale.

C'est pour ces raisons que l'IFEN constitue en France le "point focal national" de l'Agence Européenne, contribuant ainsi à une meilleure prise en compte de l'environnement dans l'Europe de demain.

OU EST IMPLANTE L'IFEN ?

C'est pour les raisons précédentes, à la fois nationales et internationales, qu'a été proposée dans le cadre du Plan National pour l'Environnement, la création d'un Institut Français de l'Environnement ayant le statut d'un Etablissement Public à caractère administratif placé sous la tutelle du Ministère de l'Environnement (décret constitutif de l'IFEN N° 911177 du 18 novembre 1991).

Créé le 18 novembre 1991, sa mise en place à Orléans en 1992 contribue à une meilleure intégration de la France en Europe et apporte plus de transparence au débat social sur l'environnement.

COMMENT EST STRUCTURE ?

Son organisation est conçue de manière à pouvoir à la fois mener des missions de service public et répondre avec la plus grande souplesse aux demandes qui lui seront adressées. Sa vocation est de travailler avec de nombreux partenaires publics ou privés, français ou étrangers.

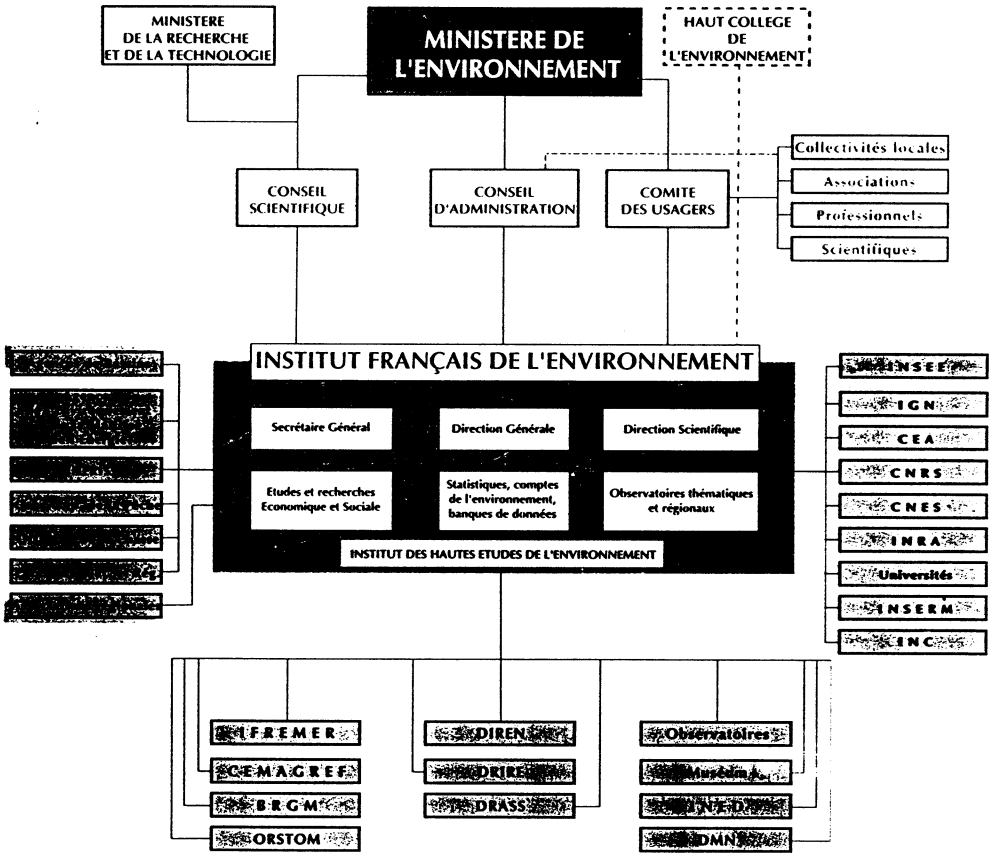


Figure 1-1 ; *Contexte de l'IFEN*

Son conseil d'administration comprend des représentants des administrations, des élus, des scientifiques, des responsables d'entreprises et/ou d'associations.

Il dispose en outre d'un comité scientifique de 15 membres chargés de veiller à la qualité de ses travaux, ainsi que d'un comité d'usagers chargé de l'assister dans la définition de ses programmes et dans la diffusion de ses productions.

1.3.2. Un Centre d'observation et d'information

a) ***Observation, recueil des données***

L'IFEN joue le rôle d'un observatoire de l'environnement. Pour ce faire, il doit animer et coordonner la collecte et le traitement des données en s'appuyant sur un grand nombre de partenaires.

Il assure la cohérence des données et la centralisation des réseaux de surveillance, en faisant appel aux moyens les plus modernes (télédétection, logiciels de gestion de bases de données et de traitement statistique, télétransmission, systèmes d'informations géographiques, intelligence artificielle...).

Toutes les informations (grâce auxquelles l'état actuel et prévisible de l'environnement peut être décrit) devront être recueillies, notamment celles qui portent sur :

- la qualité,
- les pressions subies par l'environnement,
- la sensibilité de l'environnement,
- les réponses des administrations, des entreprises, des associations et des individus.

b) ***Réseau d'observatoires***

Les problèmes d'environnement se posent à diverses échelles. L'IFEN contribue au développement et organise en réseau :

- des observatoires thématiques,
- des observatoires régionaux ou locaux,
- des laboratoires de terrain pour les études environnementales, des prototypes et des réalisations expérimentales, dans le cadre de conventions, de groupements d'intérêt public ou économiques ou d'associations,
- d'observatoires d'intérêt planétaire (Pacifique, Sahara et Sahel, Méditerranée...).

c) *Information du public*

L'IFEN diffuse les informations sur l'environnement auprès du public, des acteurs économiques, des milieux scientifiques, des associations et de l'administration.

Il procède à des enquêtes d'opinion et développe des produits d'information adaptés à la demande de diverses catégories d'usagers.

Les données seront accessibles selon diverses formes : centres de documentation, serveurs télématiques, publications, données numériques, synthèses cartographiques, etc.

1.3.3. Un Centre d'expertise et un outil d'aide à la décision

a) *Statistiques, comptes et indicateurs*

L'IFEN (outil statistique du Ministère de l'Environnement) intègre dans un système cohérent l'ensemble des données qu'il collecte, contrôle et stocke directement ou avec l'aide d'organismes ayant déjà, pour partie, cette vocation.

Il établit des comptes du patrimoine naturel et produit des cartes et autres systèmes d'informations sur l'environnement.

Il élabore des indicateurs sur l'état et l'évolution de l'environnement et les risques naturels et technologiques.

b) *Expertises et études socio-économiques*

L'IFEN réalise des études et des rapports sur des thèmes d'actualité ou de prospective à la demande du Ministère de l'Environnement ou de l'office parlementaire d'évaluation des choix scientifiques et technologiques.

Il effectue notamment des études économiques sur les dommages subis ou la restauration de l'environnement.

Il est aussi habilité à effectuer des recherches, des études et des expertises au profit de toute personne publique ou privée, française ou étrangère, dans le cadre des conventions approuvées par son conseil d'administration.

c) *Evaluation et synthèse*

L'IFEN, outil de synthèse, d'aide à la décision et d'évaluation, établit des modèles, des scénarios et des prévisions à l'amont des études et des expertises.

Il réalise périodiquement un rapport sur l'état de l'environnement qu'il publie après avis du Haut Collège de l'Environnement et contribue à l'évaluation des politiques publiques de l'environnement.

d) *Critères pour les bilans écologiques*

Les bilans écologiques se multiplient dans divers domaines (collectivités locales, entreprises, produits, procédés, labels) et nécessitent la définition de critères de référence.

Par sa compétence diversifiée sur l'ensemble des milieux, des pollutions et des risques, l'IFEN contribue à la définition de l'approche multicritères et à la méthodologie des écobilans.

e) *Formation*

L'Institut Français de l'Environnement engage des actions de sensibilisation et de formation des cadres des administrations, des établissements publics et des entreprises dans tous les domaines de l'environnement. Il contribue à la définition des orientations générales de la politique de formation dans ces domaines. Dans le cadre de l'Institut des Hautes Etudes de l'Environnement, il organise des cycles de formation et d'études pour des responsables de haut niveau de l'administration, du secteur public, des entreprises et des milieux associatifs.

1.3.4. L'IFEN correspondant de l'Agence Européenne de l'Environnement et des organisations internationales

L'Agence Européenne de l'Environnement a pour objet de fournir à la Communauté et à ses Etats membres les données scientifiques et techniques nécessaires pour l'élaboration de ses actions, l'évaluation de leur mise en oeuvre et l'information du public. Elle repose sur un système permanent d'information et d'observation mobilisant deux catégories de partenaires :

- des "points focaux nationaux" responsables chargés de la collecte, de la synthèse et de l'évaluation des données dans chacun des pays membres ;
- des "centres thématiques européens" chargés de la définition et de l'harmonisation des normes de mesures et de traitement des données.

A l'avenir, l'Agence Européenne pourra également se voir confier l'établissement de labels "environnement" et de critères d'attribution de ces labels à des produits, technologies ou programmes, et de promotion de technologies et de procédés respectueux de l'environnement.

Plusieurs pays européens sont dotés d'Agences ou d'Instituts nationaux chargés de ces tâches et qui seront les correspondants naturels de l'Agence Européenne : l'Allemagne, le Danemark et les Pays-Bas, par exemple. Des Instituts régionaux de l'environnement sont opérationnels dans plusieurs régions voisines en Espagne, en Italie et en Allemagne.

L'Institut Français de l'Environnement est le point focal national de l'Agence Européenne et assure l'interface entre les réseaux français, européens et internationaux.

1.3.5. Les partenaires de l'IFEN

L'IFEN mobilise tous les partenaires compétents dans un réseau national d'observation de l'environnement.

Il doit avoir accès à toutes les informations relatives à l'état de l'environnement collectées par les diverses administrations (notamment celle de la Santé, de l'Agriculture, de l'Education Nationale, de l'Industrie, de l'Equipement...) et les établissements publics de l'Etat.

Parmi les principaux producteurs de données, citons :

- l'Agence de l'Environnement et de la Maîtrise de l'Energie, les associations en charge des réseaux de mesure et de l'information sur la pollution de l'air, l'observatoire des déchets et l'inventaire des décharges et des points noirs,
- les six Agences de Bassin, les Ministères de l'Agriculture et de la Santé pour la banque nationale des données sur l'eau,
- l'Institut Géographique National (IGN) pour les données topographiques et hydrographiques,
- la Météorologie Nationale pour les données pluviométriques et thématiques,

- le Bureau de Recherches Géologiques et Minières (BRGM) pour les données sur l'eau souterraine, l'érosion, les risques naturels et les sites d'intérêt géologique,
- l'Institut Français de la Mer (IFREMER) pour les données sur la qualité des eaux littorales et les écosystèmes marins,
- le Centre National d'Etudes Spatiales (CNES) pour l'observation de la terre et la gestion de systèmes globaux de surveillance de l'environnement,
- le Muséum National d'Histoire Naturelle, l'Office National de la Chasse, le Conseil Supérieur de la Pêche et les Parcs nationaux et régionaux, pour les écosystèmes terrestres,
- l'Institut français de recherche scientifique pour le développement en coopération (ORSTOM) pour les connaissances de l'environnement outre-mer,
- l'INSEE pour les données économiques et sociales,
- et aussi l'INED, l'INC, le CEMAGREF, l'INRS, l'INRETS, le CEA, l'INRA, le CNRS, l'INERIS...

En outre, l'Institut développe, dans un cadre contractuel, un partenariat avec les collectivités locales et les régions, les associations, les entreprises de toute nature, pour élargir le réseau de connaissance de l'environnement. Dans le cadre d'accords de coopération bilatéraux et multilatéraux, il participe aux programmes internationaux et étrangers.

1.3.6. Données gérées par l'Institut dans un cadre de partenariat

- L'utilisation de sol et les ressources naturelles,
- l'occupation du territoire et les paysages,
- la faune, la flore et les écosystèmes terrestres et marins,
- la protection du littoral, de la montagne et en général des espaces protégés ou sensibles,
- la qualité de l'air, les émissions de polluants et les ressources aquatiques,
- la gestion des déchets et l'état des sols,
- les émissions sonores,
- l'environnement urbain,
- la dissémination des substances chimiques dangereuses pour l'environnement et leurs effets,
- les risques naturels et technologiques,
- les perceptions et les pratiques socio-économiques de l'environnement.

1.4. DECRIRE DES CONFIGURATIONS INFORMATIQUES REPONDANT AUX BESOINS DE L'INSTITUT

1.4.1. Rappels des fonctions

L'IFEN peut être considéré comme étant composé des invariants suivants :

* une banque de données sur l'Environnement, servant de référentiel de données pour les études d'analyses et synthèses au niveau national ou européen,
* une structure d'études, servant à promouvoir des démarches méthodologiques centrées sur les données de la banque, et fournissant des études conjoncturelles aux instances gouvernementales et communautaires,
* un service d'échange institutionnalisé entre la banque et les grands organismes nationaux, de type tableau de bord dans un sens, et de mises à jour recueillies régulièrement de l'autre,
* une archive historique de données de base, à des fins d'études d'évolution.

Ces invariants revêtissent tous un aspect de "services" important : service de recueil, de gestion, de synthèse et de vulgarisation de l'information. Ils se résument dans le schéma suivant :

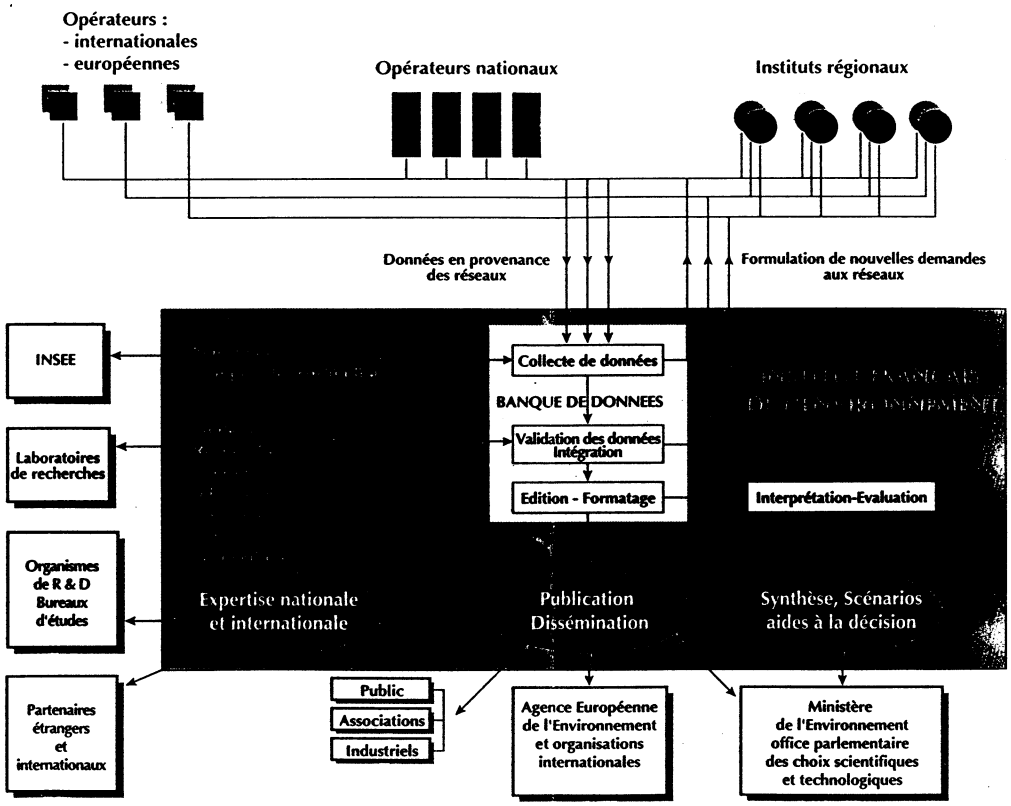


Figure 1-2 : Invariants de l'IFEN

1.4.2 Banque de données

La banque de données de l'Institut constitue l'une de ses principales missions. Elle se compose de plusieurs bases de données indépendantes et réparties ayant chacune trait à un aspect distinct de l'Environnement, par exemple :

* Données expérimentales de type agro-écologiques relatives à différentes caractéristiques du milieu,
* Plan d'occupation territoriale de type CEE-CORINE,
* Données statistiques relatives à un paramètre d'environnement, comme par exemple des données fournies par une agence de bassin.

Elle présente un aspect "services" qui s'articule autour des missions de constitution et de maintenance de données cohérentes sur l'Environnement.

"Offrir un service" signifie garantir :
* une exigence de qualité, concernant les données mises à disposition,
* une exigence de disponibilité et d'accès aux informations,
* une exigence de crédibilité de celles-ci, crédibilité qui passe par leur fiabilité.

1.4.3. Collecte et mise à disposition de l'information

En corollaire au service de constitution et de maintien de données, on trouve les missions de collecte et de mise à disposition de ces informations après validation et intégration suivant un schéma conceptuel global des données.

<u>La collecte</u> : l'Institut reçoit des organismes collecteurs des données de base qui alimentent la mise à jour de sa banque. La valeur ajoutée de l'Institut réside dans la vérification de validité et de cohérence de ces données à une échelle globale, ainsi que dans leur mise au standard commun à l'échelon national et communautaire. La prestation se fait donc sous la forme d'un échange d'informations.

<u>Mise à disposition</u> : il s'agit d'une fourniture d'analyses et de synthèses à partir de ces sources, en termes réguliers ou du moins institutionnalisés, pour des consommateurs de ce type d'informations : directions régionales de l'aménagement, chambres d'agriculture, collectivités locales, associations, etc.

Enfin, dans la mesure où l'Institut joue le rôle de point focal national de l'Agence Européenne de l'Environnement (AEE), la banque de données doit aussi servir de passerelle ou de représentant de son équivalent de l'AEE, ou encore assurer entièrement le service de banque de données au niveau communautaire.

Deux scénarios s'offrent à ce stade :

- soit l'AEE dispose d'une banque de données propre, et l'Institut joue alors le rôle de collecte et de maintien de cohérence des données, qui sont mises à la disposition de la banque de l'AEE,
- soit ce rôle est délégué à l'Institut, qui rassemble alors les informations collectées par les organismes nationaux et des autres pays, pour les mettre à la disposition des organismes utilisateurs membres.

Les données conservées par l'IFEN se présentent sous deux plans : le plan multi-thématique associé à un pays ou une région, et un plan trans-régional concernant un thème d'Environnement donné.

Sur l'ensemble des fournitures de données de sa part, l'Institut pourrait présenter aux utilisateurs des engagements de type "services" :

* service de mise à disposition d'analyses et de synthèses concernant l'état de l'Environnement et son évolution, suivant une méthodologie uniforme et constante,
* service de recueil et de validation des mises à jour en provenance des sources locales,
* service d'accès direct du public aux données d'ordre général (Minitel, bulletin périodique ou autre type de diffusion).

1.4.4. Unité d'études

En amont de ces services, et pour les alimenter en méthodes et valider leurs procédures thématiques, se trouve l'unité d'études de l'Institut.

Cette unité joue un rôle démonstratif et incitatif, en étant une unité d'expérimentation et de démonstration à l'attention des utilisateurs potentiels des données de la banque.

Elle est destinée aussi à servir d'interlocuteur aux autres organismes de recherche, pour maintenir une veille scientifique constante sur l'état des méthodes et des emplois des informations sur l'Environnement. Pour cela, elle entreprend et valide de nouvelles méthodologies, qui pourront ensuite constituer le fond permanent des analyses et des synthèses du service de recueil et de diffusion évoquées précédemment.

Enfin, cette unité d'études pourrait être chargée des études conjoncturelles effectuées à la demande, et qui ne rentrent pas dans le cadre des travaux institutionnels du service de recueil et de diffusion : suivi d'une catastrophe, analyse d'impact, étude exceptionnelle,...

1.4.5. Service d'archives

Pour compléter le rôle de dépositaire et de conservateur des données sur l'Environnement, l'Institut assume aussi un rôle de suivi de l'évolution de cet environnement.

Les archives des données de base, avant leur traitement pour la mise à jour de la banque de données, sont destinées à permettre de reconstituer les états antérieurs et de suivre ainsi la succession d'événements qui ont conduit de ces états vers la situation actuelle.

1.4.6. Schéma d'organisation du système informatique

Une fois définis les invariants de fonctionnement de l'Institut Français de l'Environnement, il s'agit de déduire l'organisation à mettre en place pour assumer ces responsabilités.

a) Flux de données en amont de la banque

La mission principale de l'IFEN consiste à être le dépositaire et le garant de la cohérence des données concernant l'Environnement.

En amont de la banque de données, les données collectées subissent un premier examen sur l'opportunité d'une mise à jour de la banque, vue à la fois sous l'angle de la cohérence d'ensemble et celle de la présence significative de ce thème dans la banque. Cette étape correspond à une **validation lexicale** des unités d'information.

En cas de décision, les données constituent un dossier technique de mise à jour, qui est instruit quant à la validité et la conformité du format du contenu. Cette étape correspond à une **validation syntaxique** des éléments de mise à jour.

Le contenu du dossier sert alors à mettre à jour une copie extraite de la banque de données. Cette opération "blanche" permet de contrôler l'impact des adjonctions sur la cohérence de l'ensemble des données existantes de la banque. Cette étape correspond à une **validation sémantique** des unités d'information.

Après la validation d'ensemble, les données sont introduites dans la base de données correspondante, et l'état précédent est archivé ainsi que les données de base qui ont servi à obtenir cet état.

b) Flux de données en aval de la banque

En aval de l'IFEN se trouvent les services visant à mettre en oeuvre les données conservées pour des applications ayant pour toile de fond l'Environnement, qu'elles soient institutionnelles ou grand-public.

L'Institut offre un accès standard à la banque de données (tableaux de bord, cartes thématiques numériques).

Il propose en outre une gamme de produits standards, c'est-à-dire des synthèses réalisées à la demande sur un thème et une région donnés, mais avec des logiciels reposant sur des algorithmes et méthodes courantes "du catalogue". Les demandes peuvent être soit occasionnelles soit sur "abonnement" périodique ou sur mise à jour de la banque de données. Sauf exception, le délai d'élaboration de ces commandes exclut une livraison de type interrogation interactive.

En plus des produits standards, il pourrait exister des produits "grand-public" à des fins de service public ou de sensibilisation. Ces produits pourraient être accessibles interactivement sur Minitel ou être diffusés par les canaux de distribution classiques (cartographie ou presse).

Vis-à-vis de l'AEE, l'Institut offre un canal d'accès privilégié qui peut être une mise à disposition de ses utilisateurs du catalogue AEE (cas de la passerelle vers l'AEE) ou

encore une mise à son catalogue des données du niveau communautaire (cas de la délégation complète des données de l'AEE vers l'IFEN).

Pour ses propres besoins de recherche et de validation méthodologiques, l'Institut se présente aussi comme consommatrice des données conservées dans sa banque de données. Ce service d'études pourrait en plus offrir des prestations à façon dans des cadres qui correspondent à ses compétences tout en se situant en dehors des services habituels (catastrophes, études conjoncturelles, ...).

1.4.7. Organisation et responsabilités

Les missions principales de l'Institut Français de l'Environnement font apparaître plusieurs niveaux de responsabilité avec pour chacun différents liens de dépendance qui sont décrits dans le schéma ci-dessous.

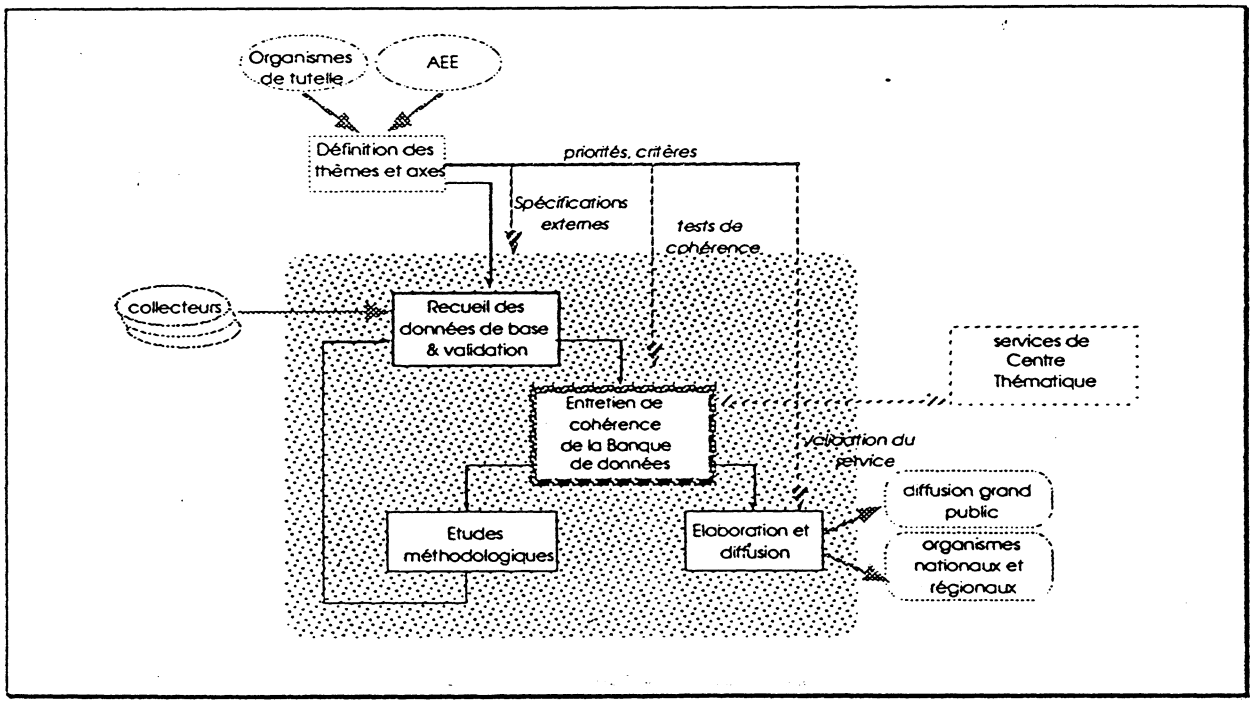


Figure 1-3 : Fonctions de l'IFEN

Les responsabilités recensées pour chaque étape de l'analyse de missions de l'IFEN, telles que décrites par ce schéma pourraient conduire à un découpage d'organisation élaboré suivant le partage fonctionnel suivant :

* décision concernant l'application des priorités et des thèmes désignés par les autorités de tutelle, suivi et validation de leur conformité : confiée à la Direction de l'Institut,
* décision d'opportunité de mise à jour avec des données nouvelles : confiée au responsable de la cellule désignée par l'Institut,
* instruction "syntaxique" et "sémantique" de la mise à jour de la base : peut être confiée à une entité technique spécialisée. L'opportunité d'une sous-traitance de cette activité à une entité externe, rendant compte de la qualité de service au responsable désigné par l'IFEN, est à envisager,
* responsabilité de la réception des demandes, de l'élaboration et distribution des produits standards : confiée à une cellule au sein de l'Institut,
* poursuite des études méthodologiques et validation des algorithmes d'élaboration de produit thématique standard : confiée au Centre de recherche de l'Institut,
* élaboration d'études conjoncturelles et de produits spéciaux : confiées au Centre de recherche de l'Institut,
* pôle d'entrée de l'AEE : confiée à une cellule au sein de l'Institut.

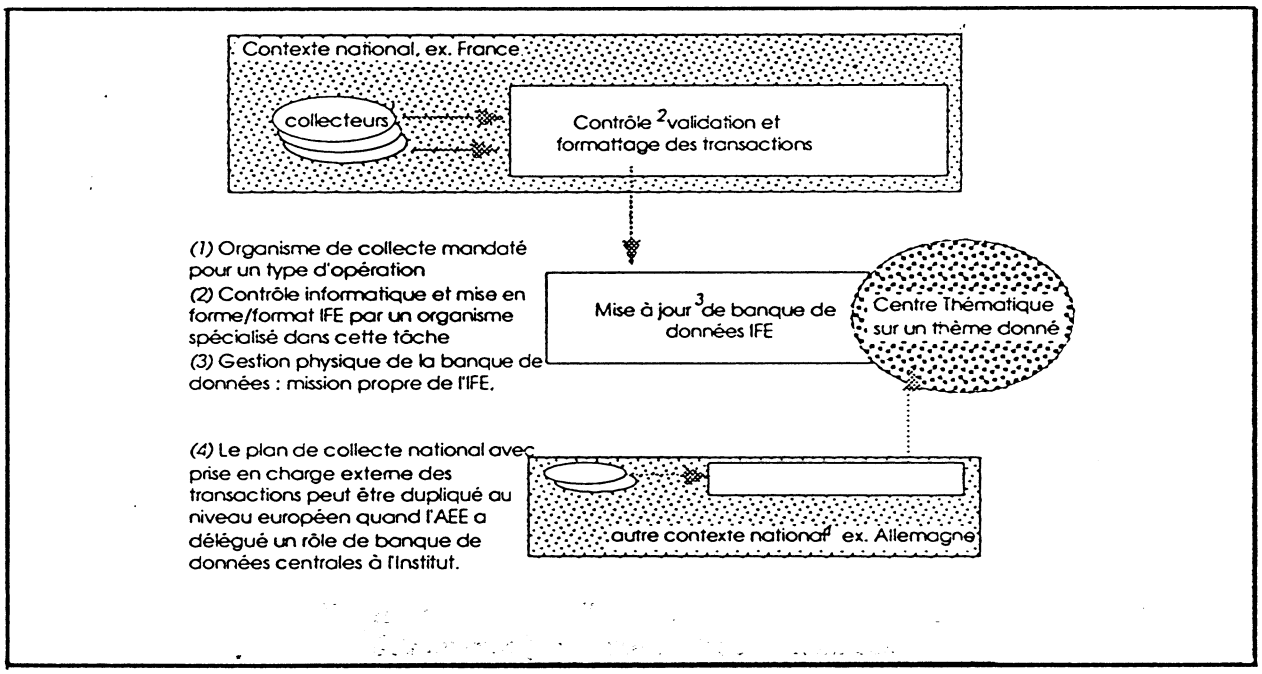


Figure 1-4 : Principe d'une sous-traitance possible des mises à jour de la banque de données.

1.4.8. Propositions de configuration informatique

Avant de procéder à l'étude des configurations informatiques de l'IFEN, il est utile de rappeler certaines considérations générales d'ordre informatique qui vont sous-tendre

l'analyse des systèmes de l'Institut. Il s'agit en particulier du modèle d'analyse essentiel de système, et des tendances aux architectures décentralisées.

Ensuite les différents sous-systèmes informatiques de l'Institut sont examinés.

a) Composition formelle d'un système informatique

Dans la méthode dite d'analyse essentielle de systèmes, on tente d'identifier les composants qui forment la trame essentielle du système informatique étudié, indépendamment du domaine d'activité considéré. Puis on détermine ensuite les aspects qui concourent à la spécificité du problème.

Par exemple, un système d'information peut être organisé informatiquement selon un modèle à trois couches :

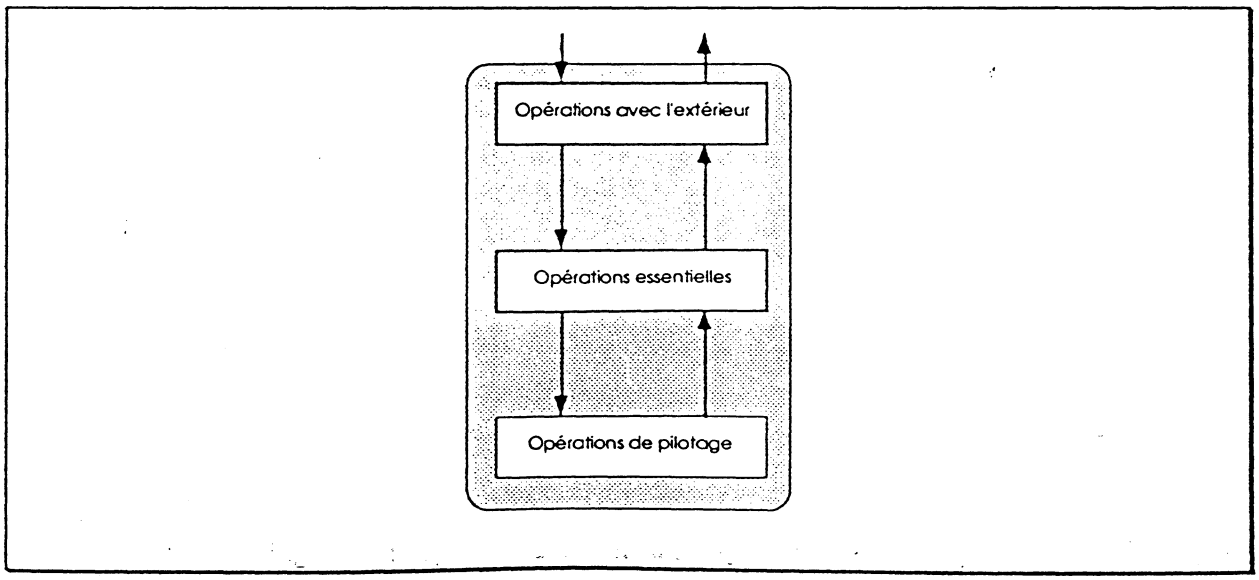


Figure 1-5 : Modèle "essentiel" d'un système informatique

La **couche externe** contient les opérations et données qui permettent au système de dialoguer avec le monde environnant. Elle comprend par exemple les opérations de publicité, de renseignement, de réservation de billet, de prise de commande, de livraison de facturation, de contentieux,... Les opérations de cette couche sont susceptibles d'être gérées par des logiciels transactionnels. Des pannes et incidents localisés dans cette couche ne mettent pas en danger l'organisme. Pour l'Institut, cette couche correspond aux accès extérieurs (collecte, diffusion grand public ou institutionnelle, publications,...).

La **couche dite "essentielle"** rassemble les opérations qui forment la raison d'être de l'organisme, et lui donnent sa compétitivité vis-à-vis du monde extérieur. Pour une usine, il s'agit de la tenue des comptes clients, des prêts ou des fonds en circulation. Pour un hôtel, c'est le remplissage des chambres et la gestion des repas. Pour une société de services, c'est l'état d'avancement des contrats. Pour l'Institut Français de l'Environnement, c'est la validation de la collecte, la mise à jour de la banque de données et l'élaboration des analyses et synthèses. Cette couche repose en général sur un réseau au sein de l'organisme, et une panne sérieuse localisée dans cette couche peut mettre en danger gravement le fonctionnement du système. Les fonctions de cette couche gagnent donc à être protégées informatiquement : redondance, protection d'accès, encryptage des données.

La **couche de pilotage** concerne les données et opérations de caractère stratégique pour l'organisme : indicateurs, tableaux de bord, informations des évolutions externes et internes, suivi d'actions de marketing, gestion de trésorerie, administration... Ces données et traitements ne sont pas volumineux, mais leur perte ou leur dévoilement peuvent mettre en péril le système. Pour l'Institut, cette couche correspond à la fonction de gestion de l'Institut, mais aussi aux structures d'études qui élaborent les futures méthodes d'analyse des données recueillies. Les consultations des données sont transactionnelles, mais en revanche la mise à jour et l'élaboration sont effectuées en temps différé et avec des contrôles soignés. Les autres couches peuvent accéder à celle-ci par des passerelles logiciels bien identifiées et protégées.

b) Facteurs de choix d'architecture centralisée et décentralisée

Plusieurs facteurs techniques soutiennent la controverse sur le choix entre des architectures informatiques centralisées ou décentralisées : la montée en puissance des stations de travail, la diffusion des réseaux locaux, la baisse des prix des logiciels sur les stations par rapport à ceux sur les gros calculateurs, la convivialité graphique des stations, les outils de programmations de 4è génération, la baisse de coût du stockage.

Dans la discussion qui suit, le terme de "station de travail" désigne indifféremment des micro-ordinateurs individuels (principalement de la série 486) et des machines Unix. En effet, la différence entre ces machines existe en termes de marché et non en termes de puissance ou de possibilités.

- Puissance de calcul

Une station de travail de bas de gamme sans espace disque présente une puissance de calcul de 10 à 15 Mips, soit 7 fois la puissance d'un VAX - 780 du début des années 80

pour le prix inférieurs à 100 KF. Les stations de haut de gamme comme les modèles DEC station 50, RS 6000 (IBM) ou SUN Sperc 80 affichent toutes des puissances supérieures à 40 Mips et 15 Mflops. Bien que la majorité de cette puissance soit employée à gérer des écrans haute définition et à 256 couleurs, il reste suffisamment pour des calculs aussi fastidieux que les éléments finis, la synthèse et le traitement d'images.

- Stockage de données

La technologie de stockage magnétique autorise couramment des unités de 300 Mo au format 5,25 pouces, avec des possibilités jusqu'à 1200 Mo avec des temps d'accès moyens de l'ordre de 14 ms et des débits de transfert de 2 à 4 Mo/s.

La technologie magnéto-optique autorise de plus grandes capacités à prix égal (700 Mo à 2500 Mo), mais reste handicapée par un temps d'accès supérieur à 50 ms.

Concrètement, cela signifie qu'il est possible de mettre sur chaque station de travail individuelle la puissance et la capacité de stockage des unités centrales d'un ancien centre d'étude.

Pourtant, si l'application ou l'organisation des données sont telles qu'elles exigent des échanges importants de données entre les espaces de stockage des stations, le temps d'attente qui en résulte rend alors inutile toute cette puissance installée : le réseau local constitue un goulot d'étranglement.

Pour Ethernet, ceci se produit à partir d'un débit continu de 3-Mbps c'est-à-dire environ 48 images SPOT-XS échangées entre les stations en une heure. En général, on peut supposer que les données à traiter résident sur chaque station concernée mais cette situation peut survenir quand un fort courant d'échange temporaire se superpose à un envoi d'images sur restituteur diapo, Vizir ou sur tout autre périphérique de masse.

- Réseaux locaux

Le protocole NFS introduit par SUN est devenu une quasi norme tant le nombre de constructeurs qui le proposent est élevé. Il présente l'avantage de permettre aux machines de voir les espaces disques des autres machines du réseau local comme un seul espace disque, et d'être robuste car il ne présente pas de mémoire d'états.

En revanche, si le réseau est bruité ou en cas de surcharge du réseau, le protocole "bégaie" et les paquets répétés d'informations chargent encore plus le réseau. La gestion du réseau devient alors une affaire de "gourou".

Il est certain que dans le cas d'une architecture centralisée, toutes les données transitant des disques vers la mémoire centrale passent par le bus central de la machine qui est de deux ou trois ordres de grandeur plus rapide que Ethernet. Les erreurs de configuration se voient moins mais tous les utilisateurs se partagent une mémoire centrale coûteuse.

Les solutions à ce problème de bande passante sont soit mieux répartir les données, soit augmenter la bande passante entre deux machines. Cette dernière solution peut se faire soit avec un bus spécialisé ("cluster" de DEC) soit avec des fibres optiques comme le Fiber Data Distribution Interface (solution encore à ses débuts mais prometteuse).

- Possibilités graphiques

L'emploi d'écrans graphiques de haute définition rend obligatoire le travail sur station individuelle. En effet, le rafraîchissement d'un écran couleurs de 1024 *x* 768 implique l'envoi de 2359 ko toutes les 1/30è de seconde. La carte vidéo réside donc obligatoirement sur le bus central et de manière privilégiée.

Il n'est donc pas imaginable de disposer des écrans graphiques de ce type sur le bus, une machine centrale serait partagée entre *n* utilisateurs. Il n'existe pas de bus central pouvant supporter ce débit pour un nombre d'utilisateurs-*n* supérieur à 1.

Ceci signifie que même dans une architecture centralisée, il est nécessaire pour l'Institut de disposer de stations graphiques individuelles indépendantes de la machine centrale.

- Langages et programmes d'application

Compte tenu du nombreux parc de stations installées, les logiciels d'application ou des langages de 4è génération (bases de données, tableurs) sont proposés à des prix imbattables et donnent une souplesse de travail aux utilisateurs que ne peuvent proposer les constructeurs de machines centrales, à cause de leur faible marché.

De plus en plus, les services informatiques classiques des entreprises proposent d'opérer un "downsizing", c'est-à-dire le remplacement d'un mini ordinateur départemental par un parc de stations de puissance supérieure et effectuant la même mission. Un réseau local et un serveur de données constituent alors le lien entre ces "petites" machines individuelles.

- Protections et confidentialités

Le principal problème qui se pose aux architectures décentralisées, en dehors de la bande passante du réseau local, est celui de l'archivage et de la confidentialité d'accès.

L'archivage des données individuellement par station impose des contraintes opérationnelles et psychologiques difficiles à admettre. Il est donc de pratique courante d'opérer la nuit et de manière centrale, en lançant un "démon" vers chaque station pour lui faire émettre les données nouvellement modifiées vers la centrale d'archivage.

Sur une durée de 3 heures au milieu de la nuit, on peut transférer de l'ordre de 4 Go, ce qui semble un volume de données "raisonnable" mais qui dépend en fait du nombre de stations servies par le secteur d'archivage.

Les protections d'accès constituent un autre aspect délicat d'un système ouvert, que ce soit en réseau local ou distant. Dans le cas d'un réseau local, il faut éviter qu'un poste de travail puisse lire ou modifier par inadvertance ou par malveillance des informations stratégiques. Il existe des solutions contre cela, mais chacun sait que sécurité signifie aussi contraintes et procédures lourdes. Le découpage par groupes de travail, aidé par le modèle en couches ci-dessus, contribue à limiter les contraintes à un niveau strictement nécessaire.

c) Conclusions générales

Dans la méthode dite d'analyse essentielle d'un système, on tente d'identifier les composants qui forment une trame essentielle du système informatique étudié, indépendamment du domaine d'activité considéré. Puis on détermine ensuite les points qui forment la spécificité du problème.

Indépendamment de l'étude d'organisation propre à l'Institut, les considérations techniques générales ci-dessus montrent que des contraintes opérationnelles liées à l'état de la technique informatique dictent une certaine logique de configuration de système.

Les propos généraux ci-dessus conduisent à favoriser une informatique décentralisée calquée sur le modèle par couches (externe, essentielle et de pilotage) qui permet de garantir un découplage fonctionnel entre groupes de travail. Au sein de chaque couche et selon la fonction considérée, il peut subsister une centralisation locale (par exemple serveur de donnée) ou une décentralisation (postes de traitement). Ces entités s'échangent des informations au moyen d'un ou de plusieurs réseaux locaux ou distants dont la protection se fait de manière de plus en plus stricte à mesure qu'on remonte vers les niveaux de pilotage.

Un autre facteur à prendre en compte dans les choix de configuration informatique pour l'Institut, non technique cette fois-ci, réside dans le support des industriels français oeuvrant dans l'intégration de systèmes d'information sur l'Environnement. En effet, si ponctuellement telle ou telle technique matérielle semble bien maîtrisée par les industriels français, la vision "système" (qui est bien maîtrisée par les sociétés de logiciels) reste à acquérir ou à démontrer pour préparer les industriels français à ce marché naissant des années '90.

1.4.9. Configuration de banque de données

Dans le modèle par couches de l'analyse essentielle de système, la banque de données de l'IFEN appartient aux fonctions qui constituent la raison d'être de cet organisme, de même que la préparation d'analyses et de synthèses concernant l'environnement. Elle peut être considérée à cet égard comme une fonction autonome pourvue de facilités informatiques propres.

Son rôle est de conserver en dépôt des données expérimentales, statistiques, temporelles et spatiales sur l'environnement, organisées par thèmes. Elle doit en assurer le maintien de la cohérence et l'intégrité. Elle est en relation directe avec le service d'analyse (qui appartient à la même couche) et assure aussi le dépôt des analyses et conclusions.

Vis-à-vis de la couche "externe", elle prend en entrée des paramètres d'environnement recueillis et validés et fournit en sortie soit des réponses à des interrogations directes de la banque soit des analyses et des synthèses élaborées.

Compte tenu du rôle de dépositaire des données sur l'Environnement qui est dévolu à l'IFEN, l'étude d'organisation a montré qu'il est judicieux d'organiser la banque de données en des bases de données séparées dans des systèmes de gestion répartis et interconnectés.

La séparation en bases de données thématiques distinctes optimise les performances d'interrogation, et la procédure de mise à jour de ces données garantit leur cohérence au sein de l'Institut.

La controverse "centralisé-décentralisé" se traduit alors par le choix entre la concentration de toutes les bases de données de la banque sur une seule machine, ou la

répartition par machine autonome de chaque base de données, avec une possibilité intermédiaire de "fédération" de plusieurs machines autour d'un calculateur principal.

a) Considérations techniques particulières

L'estimation du volume des données à entretenir par l'IFEN montre que la principale caractéristique de ces données est moins leur volume total que leur diversité : caractéristiques diverses, séries spatio-temporelles, données d'occupation territoriale à différentes échelles, ...

Le problème ainsi posé de l'accès à ces données diversifiées est dans une certaine mesure analogue à celui posé par l'Annuaire Electronique français, qui doit gérer un grand nombre de petits articles concernant chacun un abonné.

D'autre part, en dehors des bases de données opérationnelles, l'IFEN doit disposer aussi d'espaces de stockage auxiliaires servant de bases secondaires : études de validation des données recueillies en amont, extractions partielles en vue d'études conjoncturelles, maquettes pour l'Unité d'études, besoins isolés,... Il s'agit alors de garantir la cohérence entre les données de ces différentes bases secondaires.

Cette contrainte générale de cohérence, associée à la variété des données thématiques et statistiques à gérer, pose un problème de gestion de configuration. Une solution possible serait fournie par les futures bases de données "à objets" qui permettent d'attacher des procédures de contrôle ou de signification inter-bases à chaque modification de données. De telles bases gèrent non seulement des données mais aussi des procédures et sont bien adaptées aux données complexes.

b) Solutions possibles : architecture centralisée

Dans une optique centralisée classique, qui est celle adoptée pour l'Annuaire Electronique, le problème principal qui se pose est la performance d'une unité centrale unique confrontée à des interrogations qui se traduisent par des nombreux accès à la base, chacun concernant des enregistrements de petite taille.

Une solution possible consiste à coupler à cette machine centrale non pas des disques magnétiques mais des périphériques intelligents multiprocesseurs qui effectuent ce travail directement sur les données lues sur les disques. Ces périphériques constituent des "machines bases de données".

Un exemple de matériel qui répond aux besoins exprimés ci-dessus est celui proposé par la société Copernique pour l'Annuaire Electronique avec des "machines bases de données" DIRAM.

Ces machines sont composées de disques magnétiques rapides à double accès, couplés à des contrôleurs intelligents à haut débit qui présentent une architecture optimisée pour des interrogations variées et portant sur des quantités de données relativement petites. Elles présentent à l'heure actuelle des coupleurs pour DPS-6, tandis que la génération qui sera introduite fin 1992 pourra se connecter par bus SCSI-2 à tout type de machine UNIX.

Cette architecture par machines bases de données se prête bien à un partitionnement de la banque de données de l'IFEN en des bases distinctes mais globalement cohérentes. Elle satisfait ainsi à la fois aux besoins opérationnels et aux besoins d'études.

La machine centrale hôte est alors reliée à un frontal de communication qui prend en compte toutes les demandes (interrogations, mises à jour) provenant de l'extérieur ainsi que de toutes les autres fonctions de l'Institut, y compris celle de collecte et de validation des mises à jour.

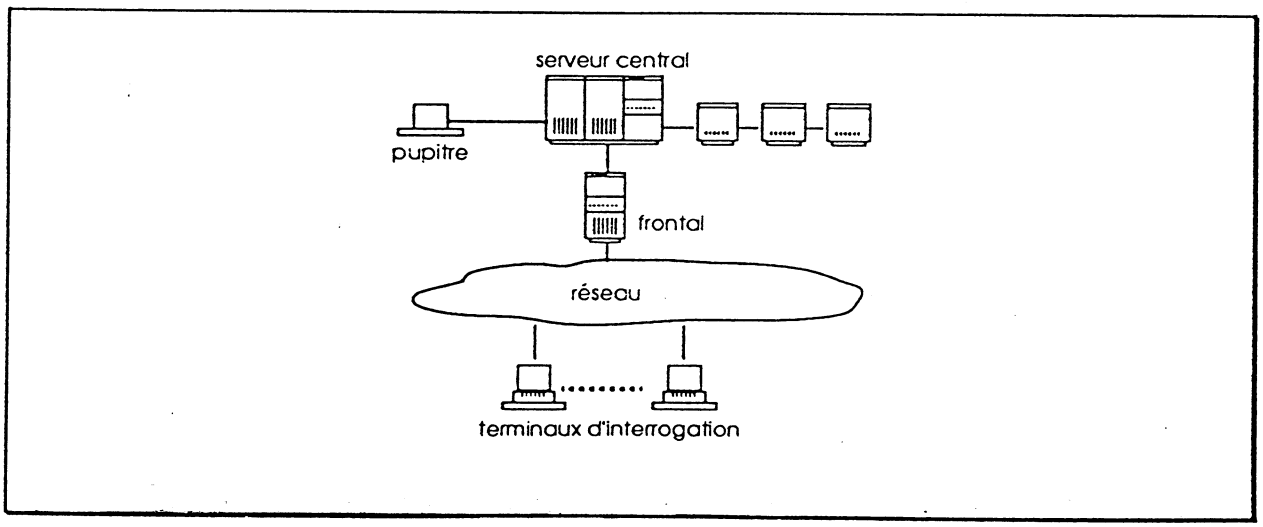


Figure 1-6 : Configuration informatique centralisée de la banque de données.

c) Solutions possibles : architecture décentralisée

En partant de l'idée que chacune des bases de données constituant la banque d'information de l'IFEN forme un tout homogène, il est possible d'imaginer une organisation informatique décentralisée autour de chacun des thèmes des bases de données : l'occupation territoriale, l'eau, le littoral, la forêt,...

Parmi ces bases de données, il en existe une qui peut servir de point d'entrée privilégié aux interrogations plurithématiques et qui est alors chargée de traduire ces interrogations en requêtes pour les autres.

Cette constatation amène à proposer une architecture de type serveur-client avec des stations de travail, qui a l'avantage d'une plus grande souplesse de mise en place, de robustesse de fonctionnement et de latitude d'évolution par ajout de nouvelles bases, mais qui exerce de fortes contraintes de débit sur le ou les réseaux locaux.

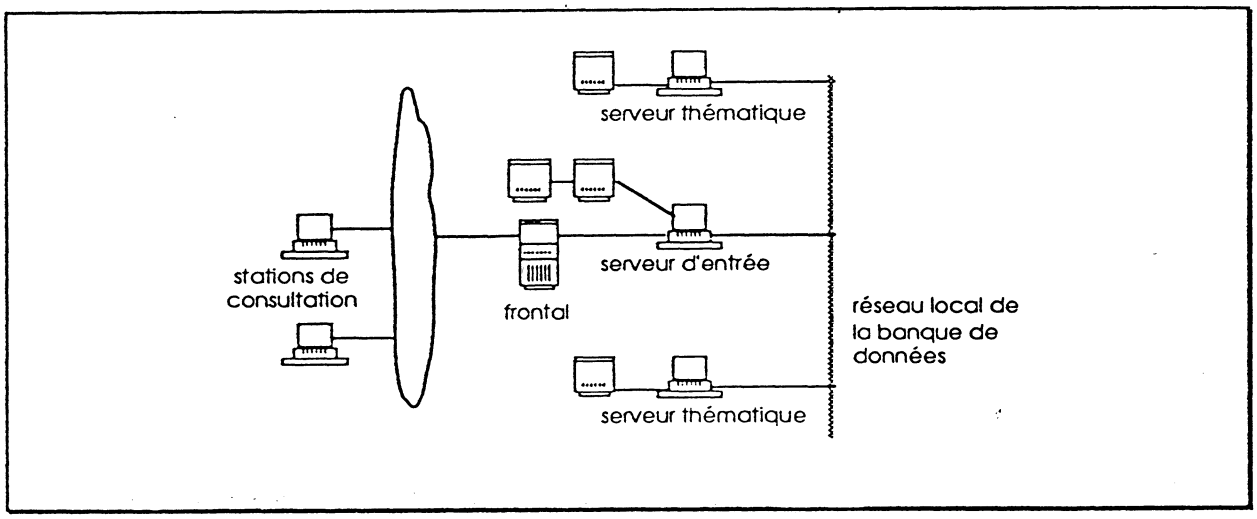


Figure 1-7 : Configuration informatique décentralisée de la banque de données.

Dans cette optique, les bases de données peuvent constituer chacune un serveur autonome gérant un ou plusieurs thèmes d'environnement. L'utilisateur (qu'il soit extérieur ou appartenant à l'Institut) ne voit qu'un seul serveur sur le plan logique car le protocole d'accès masque cette gestion de détail aux logiciels d'interrogation.

Une telle architecture est d'ores et déjà réalisable avec les logiciels existants de base de données comme ORACLE ou INGRES, moyennant le développement de certains protocoles de couche applicative. Dans le futur proche, les logiciels de base de données dites "orientées objet" proposeront de semblables protocoles en standard. Par exemple, le logiciel O2 commercialisé depuis mars 1991 par le GIE Altaïr est en cours de validation à l'IGN et à l'Université de Paris VII-Jussieu.

Pour gérer les données archivées, dans le cas où le volume s'avère être important, on peut imaginer que leurs serveurs puissent être des machines puissantes disposées en réseau, optimisées pour la gestion de ce type de données. On retrouve ainsi en quelque sorte les machines hôtes précédentes, mais sans leur accorder une place privilégiée dans l'architecture.

1.4.10. Configuration de production d'analyses

De même que pour la banque de données, les fonctions de production d'analyses et de synthèses concernant l'Environnement constituent l'essence de l'Institut. Leur équipement informatique demande donc une certaine protection et optimisation pour que cette production se fasse dans les meilleures conditions de qualité.

Il ne s'agit pas ici d'un Unité d'études mais d'une unité de production de synthèses et d'indicateurs sur l'Environnement. Elle utilise des logiciels et des méthodes éprouvées pour produire, soit de manière régulière soit à la demande, des informations issues du croisement de différents thèmes ou de l'examen des séries spatio-temporelles.

Les équipements informatiques de cette fonction accèdent à la banque de données au même titre que les utilisateurs externes, c'est-à-dire au moyen des points d'entrée télématiques de la banque. Leur nature et leur architecture sont donc indépendantes des choix effectués pour le matériel de la banque de données.

Les moyens de production d'analyses et de synthèses de l'Institut devront reposer sur deux types de services : l'un est la production de cartes thématiques, le second est l'élaboration d'études et de documents de synthèse.

Dans le premier cas, les outils requis sont les systèmes d'information géographique qui offrent en standard de nombreuses fonctions de gestion et d'assistance à la production de cartes thématiques. Dans le second cas, il s'agit d'outils classiques de *bureautique* destinés à extraire des données, à élaborer des observations synthétiques (tableau de bord), à les mettre sous forme facilement compréhensible (graphiques).

Les moyens pour y parvenir sont donc les mêmes que pour une bureautique intégrée: moyens d'interrogation, logiciels de décision dédiés, tableurs et traitements de texte généraux, outils de PAO et de présentation,...

En considérant la large diffusion des matériels UNIX dans le monde de la recherche, des possibilités de messagerie électronique, des échanges de logiciels, il semble donc qu'une configuration décentralisée de ce type serait adaptée aux besoins et aux évolutions de ce segment fonctionnel de l'Institut.

La convivialité et l'ouverture sur d'autres moyens bureautiques deviennent des critères de sélection des machines. Les critères de choix pour les postes de travail sont :

- performances de traitement,
- performances de réseau local,
- performances d'affichage,
- ouverture aux standards existants (NFS, FDDI, X-WINDOW/MOTIF, ABI,...),
- large bibliothèque d'applications.

Pour ce type de besoin, les ordinateurs individuels de type Macintosh semblent tout indiqués, de par leur grande facilité graphique (images, texte, présentation, PAO) et l'intégration sans bavures et donc sans manipulations complexes de leurs périphériques comme les imprimantes, les télécopies ou leurs réseaux locaux avec les moyens de communication comme le RNIS.

Une alternative intéressante actuellement consistirait à utiliser un matériel à base de compatibles PC de la série 486 sous l'environnement Windows. En effet, cet environnement offre dans l'immédiat et encore plus à terme, le meilleur compromis entre convivialité et faible coût, avec une grand diffusion des matériels.

Le double emploi de Unix avec cohabitation avec les systèmes d'exploitation bureautiques comme Windows sous DOS 5 ou Macintosh est garant d'un bon compromis entre un univers Unix et un univers graphiques Mac ou PC-Windows.

Les logiciels techniques de Unix manipulent en effet des données codées sur 8 bits et sont donc limités à la manipulation des signes alphabétiques occidentaux, tandis que ceux du monde graphique Mac ou PC gèrent des informations codées sur 16 ou 32 bits et sont donc adaptés à la manipulation des entités graphiques et cartographiques, des déformations de caractères, des images en vraies couleurs ou encore des signes des civilisations orientales (arabe, chinois, japonais).

1.4.11. Configuration de collecte, de validation et d'interrogation

Les opérations de collecte de données, de validation avant mise à jour de la base, ainsi que les fonctions de distribution des informations produites par l'Institut font partie de la couche externe.

Les moyens informatiques de cette couche sont particulièrement exposés aux erreurs de manipulation, aux tentatives de fraude, aux pannes. Ils doivent cependant être très simples d'emploi, ce qui exclut toute protection élaborée. Une solution classique consiste

alors à fragmenter ces moyens en zones étanches, afin d'isoler facilement des pannes ou des tentatives d'intrusion.

a) Collecte et validation

Selon les thèmes des données à valider, concernant l'une ou l'autre des bases, les opérations de validation peuvent être différentes et le volume des informations auxiliaires à mettre en oeuvre peut varier. Il est donc important de permettre aux stations de travail de pouvoir se consacrer indifféremment à l'une des bases. Pour cela, l'architecture décentralisée présente un avantage opérationnel important.

Dans une telle configuration de collecte et de validation, le modèle serveur-client reste encore plus d'actualité. La différence notable provient sans doute de la facilité de programmation qui complète la convivialité comme critère de sélection des machines ou des programmes d'application. Dans une optique d'échange de données, la compatibilité ou l'existence de passerelles dans les trois environnements Unix, DOS-Windows et Macintosh constitue un critère de choix à retenir.

b) Information spatio-temporelle et collecte

Dans de nombreux cas, les données collectées sur l'environnement ne devront pas seulement posséder une qualité technique ou statistique satisfaisante, mais elles devront encore être parfaitement définies par rapport à un localisant géographique particulier.

Suivant les situations rencontrées, la forme (coordonnées géographiques, référence à un domaine géographique particulier, ...) mais aussi la précision des informations de localisation peuvent être très variables.

Très souvent, la nature de l'information sur la situation géographique associée à la description d'un paramètre d'environnement est liée directement à l'utilisation prévisible de cette donnée sur l'environnement. On distingue donc plusieurs niveaux d'utilisation pouvant se résumer brièvement en trois grandes catégories :

* utilisation régionale, locale et sectorielle des informations sur l'environnement,
* utilisation nationale, comme dans le cas de l'inventaire de l'occupation des sols,
* utilisation européenne.

L'inventaire de l'occupation du sol, en cours de réalisation dans le cadre du programme CORINE, est représentatif du niveau national. Il va contribuer à fournir à la

fois la carte de base de description du territoire (échelle 1:100000e) et le répertoire statistique permettant l'extrapolation des observations de terrain.

L'utilisation régionale et locale des données spatio-temporelles sur l'environnement va de plus en plus être liée à la nécessité d'introduire cette information dans les schémas de planification et d'aménagement de territoire.

Enfin, les bases de données de la Communauté Européenne vont viser à couvrir l'ensemble du territoire de la Communauté par des informations fiables, car collectées selon des méthodologies standardisées.

c) Distribution et interrogation

Plusieurs possibilités techniques sont disponibles pour les fonctions de diffusion de synthèse produites par l'Institut, dépendant du caractère grand public ou institutionnel du service.

Dans une diffusion institutionnelle, on peut envisager des mises à distribution de cartes thématiques et de tables d'indicateurs d'Environnement. Il s'agit le plus souvent de fichiers textes (tableaux d'indicateurs), de fichiers graphiques prêts à l'emploi (cartographies pour la presse) ou de fichiers d'information géographique détaillés dans un format SIG à définir. Ces fichiers sont plus ou moins volumineux, selon leur échelle de précision et leurs destinataires visés, et sont destinés à être téléchargés pour un usage différé.

Dans une distribution grand public, on peut envisager la production des écrans pour des serveurs vidéotex à consultation immédiate, et à terme des fichiers numériques accessibles par Numéris.

Dans les deux cas, les informations sont détachées de l'ensemble de la banque des données de l'Institut vers un serveur spécialisé. Il y a donc peu de risques de malveillance ou de contamination du reste de la configuration.

d) Information géographique et distribution

L'émergence des concepts industriels d'éco-produit et d'éco-procédé, avec en parallèle la prise de conscience croissante de critères d'environnement à différents niveaux

d'échelle dans les plans d'aménagement, sont autant de signes qui vont dans le sens d'un besoin sans cesse mieux affirmé en matière d'information sur l'environnement.

L'impact de politiques ne peut plus être maintenant jugé uniquement sur des critiques techniques, leurs effets dans le temps et dans l'espace doivent être également évalués.

Cette situation explique la place de plus en plus grande donnée à l'information traitée, analysée et présentée sous forme graphique et cartographique. C'est ainsi par exemple que les organismes chargés de gérer et de surveiller les milieux naturels s'équipent de systèmes d'information géographique.

Les échanges entre systèmes prennent également de plus en plus d'ampleur. Les organismes comme le CNIG (Conseil National de l'Information Géographique) se préoccupent de proposer des moyens techniques de transfert de données, et d'en déduire des normes.

L'objectif poursuivi est double. Il s'agit :

1) d'optimiser l'efficacité et la valeur des échanges de données géographiques sous forme numérique,

2) de réduire au strict minimum le coût des échanges nécessaires entre systèmes.

Dans l'hypothèse où les travaux du groupe de normalisation EDIGEO, travaillant dans le cadre de EDIFrance, vont aboutir en France à l'adoption d'une norme d'échanges des informations géographiques, l'IFEN sera confronté à l'utilisation de cette norme et à son insertion au sein des systèmes informatiques prévus dans cet Institut.

1.4.12. Configuration d'études

Pour déterminer les algorithmes pour valider les données collectées, pour produire des synthèses à partir des informations de la banque de données, pour analyser les tendances concernant l'Environnement, l'Institut a besoin d'une structure qui appartient à la couche de pilotage.

Cette unité d'études n'a pas besoin d'être en contact direct avec les moyens de production de l'Institut, et peut travailler sur des échantillons de données. Les moyens

informatiques de cette couche, s'il sont en contact avec l'extérieur, le sont plutôt sur des thèmes scientifiques et sur des réseaux de recherche comme Renater, Bitnet, Usenet, Telnet ou uucp.

La protection de ces équipements ne demande pas un contrôle aussi rigoureux que pour des données de gestion ou administratives, qui appartiennent à la même couche. Exceptionnellement, on peut imaginer cependant qu'une étude "confidentiel défense" puisse être demandée à l'Institut. Dans ce cas, on retrouve le même besoin de protection que la plupart des organismes avancés comme les constructeurs d'avions, d'automobiles, d'ordinateurs confère à leurs centres d'études.

1.5. RECENSER LES SCHEMAS DE RESEAUX POUR METTRE EN RELATION L'INSTITUT ET LE MONDE EXTERIEUR

Pour constituer la banque de données de l'IFEN, pour en extraire les informations à des buts d'étude, pour les distribuer aux consommateurs de résultats, il est nécessaire d'offrir un moyen de transport : les réseaux distants.

Pour la circulation à l'intérieur de l'IFEN des informations qui sont constituées à la fois des données techniques et des données de fonctionnement (gestion, approbation de documents, échanges de coordination interne, documentation,...) un autre type de réseau est nécessaire : les réseaux locaux.

1.5.1. Considérations techniques : réseaux locaux

De plus en plus, le câblage des bâtiments avec des techniques modernes prend en compte les besoins de réseaux locaux d'entreprise, et essaie de proposer une solution globale et pérenne. Malheureusement, le consensus est loin d'être établi et la technique est en perpétuelle évolution.

En effet, l'augmentation de performances (jusqu'à 10 ou 16 Mbps) des *tranceivers* se traduit par une tendance à employer les paires torsadées blindées indifféremment pour les réseaux informatiques (Ethernet 10base-T) et pour le téléphone. D'autre part, ce dernier se différencie de moins en moins avec le numérique, avec l'arrivée du service Numéris de France-Télécom.

En haut de la gamme des performances, des passerelles à fibres optiques (FDDI : Fiber Distributed Data Interface) sont proposées pour relier des réseaux locaux à paires torsadées qui seraient alors limités à la desserte des ailes d'étage (en mai 1991, une norme FDDI sur paires torsadées a été introduite, et témoigne du dynamisme de ce secteur de l'industrie du réseau local à haut débit).

Une configuration possible et probable pour les bâtiments de l'IFEN serait alors composée de paires torsadées blindées connectées, soit à l'autocommutateur téléphonique et RNIS (service Numéris), soit à un réseau Ethernet d'aile ou d'étage. Ces différents réseaux locaux sont alors mis en contact par des passerelles FDDI.

Les composants de cette configuration sont déjà disponibles dans le commerce, intégrés pour les matériels et logiciels Unix, PC ou Macintosh :

- PhoneNet de Farallon (234 Kbps sur paire torsadées),
- EtherPort de Kinetics (10 Mbps sur paire torsadées),
- logiciel Internet Router de Apple (passerelle),
- boîtier Transtalk de P-Ingénierie pour connexion AppleTalk au Numéris,
- FDDI proposés directement par les constructeurs comme SUN ou DEC, ou encore par de sociétés spécialisées comme Ungermann-Bass,
- ...

1.5.2. Considérations techniques : réseaux distants

Pour des échanges de données entre bâtiments, et à plus forte raison à distance, le monopole des PTT et les règles économiques impliquent le passage par des réseaux distants qui n'obéissent pas aux mêmes contraintes techniques que pour des réseaux locaux (foudres, accidents, attentats,...).

Selon le volume, la dispersion géographique et la fréquence des échanges, les moyens offerts varient depuis la ligne spécialisée haut-débit au télex 110 bps, en passant par les techniques comme Numéris, X25 et la ligne asynchrone 1200 bps.

La ligne synchrone à haut-débit (service TRANSFIX de 9600-19200 bps de France Télécom) est avantageuse pour de gros volumes de données fréquemment transférés.

La ligne moyen débit Numéris (64 Kbps) sur réseau commuté sera à terme avantageuse dans les échanges européens ou intra-OCDE, dans les pays qui offrent ce service. A titre d'exemple, la France qui est pourtant en pointe dans ce domaine, ne propose encore l'abonnement RNIS standard que dans certaines régions. Dans les autres pays, ce service accessible au coup par coup selon que le central téléphonique régional est numérique ou non.

L'accès PAD-X25 (service appelé "Transpac" en France) est avantageux pour des échanges intermittents et à faible volume. En France, on dispose d'un service Minitel qui peut constituer un point d'accès grand public privilégié pour l'IFEN, dans la mesure où l'Institut a l'intention d'offrir de tels services.

Les liaisons asynchrones de 1200 bps à 4800 bps (sur ligne commutée) prennent peu à peu un caractère international grâce au marché américain qui privilégie des modems asynchrones de plus en plus performants (jusqu'à 9600 bps) avec des algorithmes de traitement du signal et de compression de données plus ou moins normalisés de type MNP-3 (Microcom Networking Protocol) à MNP-5. Pour des liaisons intercontinentales bruitées, cette solution est souvent la seule possible, en équipant les deux extrémités de la ligne avec le même modem.

La liaison télex est mentionnée à titre anecdotique. Elle est valable mondialement mais son faible débit et son codage sur 5 bits n'autorise que des transferts de données *très synthétiques.*

Le nouveau réseau national de télécommunication pour la recherche (RENATER) dont la réalisation et l'exploitation sont confiées à France Télécom permettant d'interconnecter les établissements de recherche doit être sérieusement envisagé par l'Institut. Le service ainsi offert doit permettre l'usage des protocoles IP-DOD et X25, les débits disponibles étant compris entre 64Kb et 2Mb. Les premières expériences pilotes ont eu lieu entre RERIF (Ile-de-France), REMIP(Toulouse) et VIKMAN (Caen). Cette phase pilote est aujourd'hui achevée et un plan de déploiement de Renater est établi et doit permettre la couverture de l'ensemble du territoire fin '92.

1.5.3. Solution possible

Compte tenu du texte technique exposé ci-dessus et de la mission principalement française et européenne de l'IFEN, on peut imaginer une configuration de réseau suivant pour l'Institut.

Le réseau local de l'Institut sera constitué d'Ethernet en paires torsadées blindées. La préférence qui lui est donnée sur Tokem Ring n'a pas de valeur technique absolue mais se situe plutôt dans la logique des équipements du monde UNIX et des stations de travail. Entre les étages du bâtiment, des passerelles FDDI relieront chaque réseau local au reste. Entre bâtiment distants, des passerelles Numéris assureront ce même service, à une vitesse inférieure.

Le réseau distant de l'Institut sera essentiellement basé sur Renater et Numéris pour des débits moyens et sur des accès PAD-X25 pour des faibles débits (consultation grand-public). Pour ces derniers accès, l'Institut disposera soit d'un serveur Minitel en propre, soit d'une liaison Numéris pour télécharger les écrans vers un serveur "36-15" du commerce, soit d'une liaison Renater vers un serveur de la recherche possédant les services 3614, 3615 et 3616.

1.6. EVALUATION DES VOLUMES INFORMATIQUES ET DE RESEAUX

1.6.1. Banque de données

La mission principale de l'IFEN consiste à être le dépositaire et le garant de la cohérence des données concernant l'Environnement :

* données spatio-temporelles de base utiles aux études d'environnement issues des projets avec des organismes nationaux et régionaux,

* Plans d'Occupation Territoriale (POT), cartographies au 1:100000è issues du projet communautaire CORINE-Land Cover,

* cartes thématiques de données de base utiles aux études d'environnement,

* et statistiques diverses sur l'Environnement en provenance des sources locales (air, pollution, ...) soit sur une cartographie à la même échelle, soit au niveau d'une région ou d'un bassin couvrant plusieurs pays.

a) Schéma de circulation en entrée

Le schéma de circulation de ces données en entrée de la banque de données est le suivant :

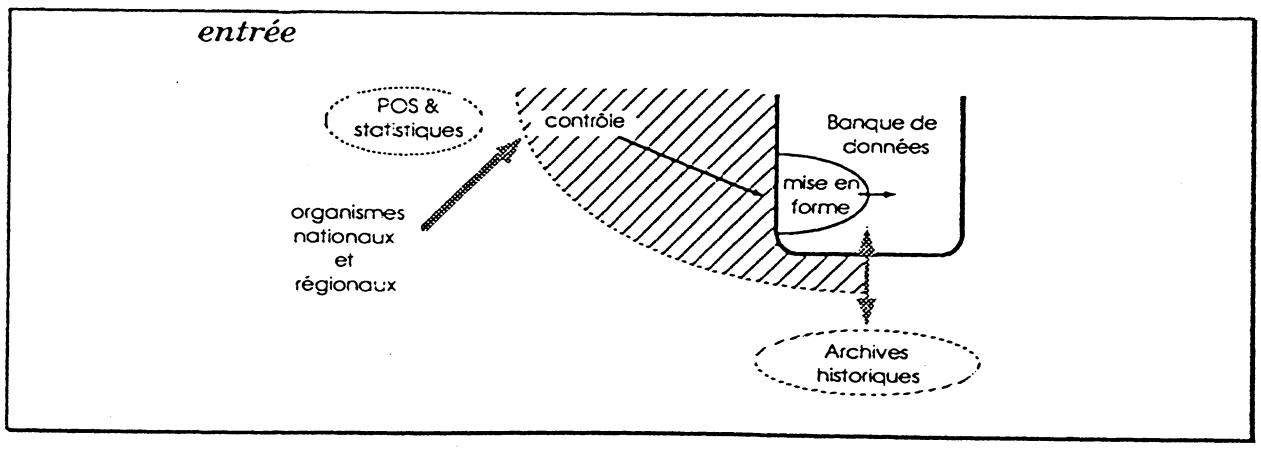


Figure 1-8

Les données sont recueillies auprès des organismes collecteurs avec lesquels l'Institut entretient des relations d'échange institutionnalisées.

Elles sont contrôlées et validées par les services thématiques de l'Institut, en vue de la cohérence de l'ensemble des données de la banque. Eventuellement, les données manquantes sont demandées aux fournisseurs ou complétées par exemple par extrapolation des séries temporelles ou spatiales. Il s'agit donc d'une validation *sémantique* du contenu. Ces données validées sont ensuite transmises au service de banque de données proprement dit.

Ce service met en forme les données pour correspondre au mécanisme de stockage interne à la banque de données (mise en forme *syntaxique*), puis les intègre dans la base de données.

b) Dimensionnement informatique

D'après les chiffres disponibles, le volume brut de l'information spatiale n'est pas très élevé en soi : environ 11 000 000 éléments de POT pour la France, chiffre auquel il faut ajouter les statistiques concernant l'Environnement.

Cependant, le point à étudier en détail est l'organisation de l'information spatio-temporelle dans la base de données en vue d'une interrogation rapide et non structurée pour les analyses et synthèses thématiques.

En effet, les techniques de base de données actuelles sont soit de type *hiérarchique* soit de type *relationnel.*

- l'organisation hiérarchique des données favorise la vitesse d'accès, mais ne permet qu'une seule structure d'interrogation prédéfinie (la hiérarchie arborescente).

- les techniques relationnelles apportent une souplesse d'interrogation car les relations sont établies dans la machine au moment de l'accès au lieu d'être prédéfinies. En contrepartie, elles consomment énormément de ressources machine quand la structure des données est complexe, comme c'est souvent le cas en information géographique. Ceci les rend inadaptées aux performances exigées.

Une troisième technique de base de données commence à se diffuser et va sans doute être largement disponible dans la décennie '90, il s'agit des bases de données à objets. Dans ces outils, les relations ne sont pas établies à l'avance comme dans la technique hiérarchique, elles ne sont pas non plus recréées à chaque interrogation comme dans la

technique relationnelle. Elles suivent un graphe d'héritage à partir d'un certain nombre de propriétés de base (les parents) pour construire rapidement les propriétés nécessaires.

Un GIP créé par l'INRIA et l'Université d'Orsay produit actuellement un tel prototype, qui est en cours d'évaluation pour la gestion des cartes de zones de circulation aérienne. Le problème est similaire à celui de l'Information Géographique et les résultats semblent prometteurs.

A l'horizon de l'Institut et à partir de ces observations, il est clair qu'une étude de définition précise de la base serait opportune tant dans la structuration logique que dans l'organisation physique et l'architecture de la base.

c) Matériels et logiciels

Les matériels et les logiciels de la base de données sont à déterminer en fonction des hypothèses de volume de stockage et d'interrogation.

En tout état de cause, les outils standards disponibles ou attendus pour la décennie '90 permettent de répondre aux besoins discernables.

1.6.2. Collecte et mise à disposition de l'information

En amont et en aval du gisement d'informations que constitue la base de données de l'Institut, on trouve les services qui mettent en valeur ces données sur l'Environnement.

Le schéma de circulation de ces données en entrée et autour de la banque de données est le suivant :

a) Schéma d'entrée

Les données de base en entrée (données expérimentales, statistiques, POT, ...) proviennent des organismes qui sont aussi consommateurs de produits mis en forme par l'Institut.

Les données ainsi collectées sont de nature multithématiques :

- les données expérimentales proviennent des organismes nationaux ou régionaux chargés de mettre des dispositifs expérimentaux sur le terrain dans le cadre agronomique, écologique et environnemental.
- les données de télédétection spatiale alimentent les Plans d'Occupation Territoriale,
- les données statistiques proviennent des Agences de Bassin ou d'organismes nationaux chargés de la surveillance d'un paramètre d'Environnement comme par exemple les biotopes.

Les données de base arrivent de manière asynchrone ou avec une périodicité quelconque, soit sous forme de fichier soit par un réseau téléinformatique.

Ces données constituent un "dossier de mise à jour" qui sera instruit par un service de l'Institut pour contrôle et validation, avant d'être transmises à la base de données proprement dite.

Les critères de validation du dossier portent sur :

* *la nature des données* : validation de l'autorité à mettre à jour l'information en question (une commune ne peut demander la mise à jour d'informations concernant une autre commune),
* *le format des données* : tous les formats ne sont pas forcément manipulables par l'Institut. Les formats reconnus font partie des négociations avec les organismes interlocuteurs,
* *la cohérence sémantique des données* : un code de population, isolé au milieu d'un fichier de modèle numérique de terrain, provient sans doute d'une erreur,
* *la complétude* : une mise à jour incomplète, sauf exception, n'apporte pas de valeur ajoutée à la banque de données,
* *l'homogénéité* : dans le même ordre d'idées, une parcelle contenant une précision de 1:10000è est inutile dans un ensemble au 1/50000è,
* ...

Les expérimentations nécessaires aux validations de ces dossiers portent sur une image réduite de la base de données, de taille significative par rapport à ces besoins thématiques.

b) Schéma de diffusion standard

Du point de vue accès extérieurs, les bases de données qui composent la banque de données de l'IFEN doivent être vues comme un système d'information fournissant sur demande des données élaborées ou des cartes thématiques et numériques.

Le thème considéré est bien sûr fonction de la base, mais la définition spatio-temporelle ou géographique des données qui seront fournies à la suite d'une interrogation dépendra beaucoup des règles de gestion qui seront attachées à ces données. Il s'agit là d'un point important qui devra être étudié avec soin lors d'une étude de définition.

Les choix de gestion correspondants devront aussi tenir compte des contraintes externes appliquées à ces ensembles de données. Il s'agira :

* des contraintes d'organisation communiquées par les instituts nationaux et internationaux chargés de la normalisation en matière de données descriptives et de cartographie numérique,
* des contraintes de nature davantage thématiques imposées notamment à travers des programmes de la CEE.

c) Schéma de diffusion institutionnelle

Les organismes institutionnels qui sont consommateurs de produits mis en forme par l'Institut sont par exemple les organismes régionaux, les instituts de recherche spécialisés (agronomie, géologie, écologie, ...).

Les produits que l'Institut leur fournit périodiquement sont des synthèses, des cartes thématiques, des indicateurs de l'Environnement, des tendances,...

Ces organismes peuvent aussi accéder aux données d'ordre général mises à la disposition du grand public, et par convention pour des études particulières aux données de base.

d) Schéma de diffusion grand public

Parmi les consommateurs grand-public des produits de l'Institut, l'Education Nationale jouit d'une position privilégiée de par son rôle illustrateur.

Ensuite viennent les médias, qui consomment des synthèses, des statistiques, des cartes thématiques d'ordre général.

Enfin, il reste un travail de fond de promotion et de sensibilisation du public à l'Environnement. Il se concrétise par des affiches, des expositions, un service télématique Minitel, des parrainages de colloques, etc.

e) Relations avec l'AEE

Plusieurs scénarios peuvent être envisagés pour permettre à l'Institut de remplir son rôle de point focal national de l'AEE, en ce qui concerne la diffusion des données.

Dans l'hypothèse où l'AEE délègue à l'Institut sa banque de données totalement ou sur la base d'une coopération de type Centre Thématique, la diffusion des données sur ce thème concernant le reste de la communauté est traitée comme pour les données nationales.

Dans l'hypothèse où l'AEE délègue à l'Institut simplement un rôle de représentation de sa propre banque de données, les requêtes venant de France et concernant des pays membres de la communauté peuvent être :

* soit redirigées informatiquement vers la banque de l'AEE, de manière transparente à l'utilisateur, avec traitement immédiat des erreurs ou absences. Il s'agit d'une fonction de *passerelle*,
* soit enregistrées, assistées éventuellement par un catalogue interactif et graphique, puis retransmises sous forme de commande de produits à l'AEE, les erreurs ou absences étant signalées dans un deuxième temps. Il s'agit alors d'une fonction de *point de vente*.

f) Dimensionnement informatique de l'instruction des dossiers

Le volume des informations nécessaires pour l'instruction des dossiers de mise à jour n'est pas très élevé. Cependant, le caractère systématique de ces opérations demande un niveau de performances élevé pour éviter un effet de goulot d'étranglement à la mise à jour.

Concrètement, par exemple, des consoles graphiques couleur de haute définition seront nécessaires pour visualiser rapidement des images satellites ou les cartographies numériques accompagnant les dossiers. De même, les interrogations et calculs liés aux

vérifications de qualité des données demandent un temps de réponse réduit aux interrogations de *l'image* de la base de données.

Ceci suppose donc le recours à des systèmes de gestion de bases de données et à des systèmes d'information géographiques très puissants, et configurés spécialement en fonction de ce travail de mise à jour.

g) Dimensionnement informatique de la diffusion institutionnelle

La diffusion institutionnelle est une tâche de fond qui ne réclame pas un haut degré d'interactivité, bien que sans doute une certaine puissance de calcul soit nécessaire.

Il s'agit d'une production périodique qui se traduit par exemple par la mise à disposition d'un certain nombre de points d'entrée télé-informatiques et de comptes utilisateurs sur la configuration informatique de l'Institut, dotés de volumes de stockage conséquents contenant les produits de synthèse fournis à ces organismes.

Une solution complémentaire probable consiste à des envois par bande magnétique de données volumineuses comme par exemple des retours d'informations de base d'Environnement, traitées et mises en état par l'Institut.

L'emploi des médias optiques pourrait être envisagé, sous réserve d'une plus grande normalisation de ces supports dans une échéance de temps compatible avec la mise en place de l'Institut, les rendant aptes à un échange entre de multiples organismes.

h) Dimensionnement informatique de la diffusion grand-public

La diffusion grand-public demande un autre type de moyens informatiques car les volumes concernés ne sont pas énormes, mais sont en général accompagnés de commentaires explicatifs et didactiques.

Le premier moyen de diffusion est le Minitel, pour la sensibilisation et la diffusion de masse.

Un autre moyen sera disponible très vite pour le transfert des fichiers de volume moyens : il s'agit du Numéris (64 kbits/s). Ce moyen permet de diffuser des cartes thématiques annotées et commentées sous la forme d'un fichier avec un format d'usage répandu comme le format PICT (pour Macintosh, employé par l'AFP pour ses dépêches).

D'autre part, pour la réalisation de documents éducatifs, des moyens de publication assistée par ordinateur sont nécessaires. Il ne s'agit pas de réaliser des maquettes directement par l'Institut, mais de disposer d'un moyen d'échange commode avec les sociétés prestataires de ces services.

i) Matériel et logiciel pour l'instruction des dossiers

Les matériels et les logiciels du service d'instruction des dossiers sont typiquement des postes de type ordinateurs individuels ou stations de travail, reliés en réseau local à la banque de données de l'IFEN.

j) Matériel et logiciel pour la relation avec l'AEE

Les matériels et les logiciels pour ce service de relation avec l'AEE dépendant fortement des scénarios de fonctionnement envisagés, et de la configuration informatique de l'AEE.

L'ensemble des besoins peut être satisfait avec les offres standards du marché.

1.6.3. Unité d'études

L'unité d'études joue un rôle exemplaire et démonstratif vis-à-vis des utilisateurs de la banque de données, en étant une consommatrice exemplaire des données de la banque à titre expérimental.

a) Fonction de l'unité d'études

La position de l'unité d'études vis-à-vis de la banque de données est représentée par le schéma suivant :

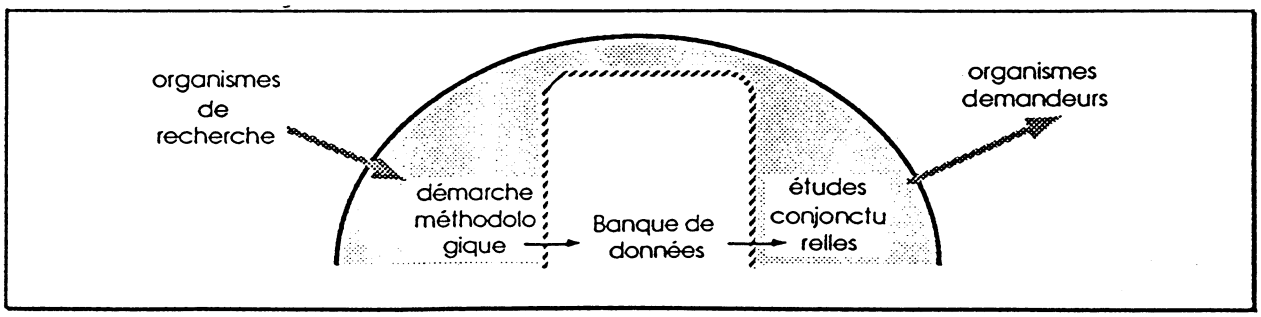


Figure 1-9 :

L'unité d'études joue le rôle d'interlocuteur représentant l'Institut face aux organismes de recherche nationaux et internationaux, publics et privés, travaillant sur le thème de l'Environnement.

Elle assure la veille scientifique et explore les voies de recherche qui permettent d'améliorer le service offert par l'Institut, soit par la mise en valeur des informations de la banque de données, soit par la garantie de leur qualité.

Les méthodes et traitements validés par l'unité d'études serviront ensuite de base pour élaborer les produits standards de l'Institut.

En parallèle, l'unité d'études doit contribuer à étendre l'efficacité et à accroître le niveau de performance des contrôles effectués avant la mise à jour de la banque de données.

En plus de ce travail méthodologique, l'unité d'étude prendra aussi en charge l'élaboration des "produits spéciaux" ou des études conjoncturelles demandées par certains organismes "clients" de l'Institut.

Ce rôle entre dans une des missions de l'unité, d'être un "utilisateur éclairé" de la banque de données, et d'expérimenter des produits qui pourront à terme figurer en standard au catalogue du service de diffusion.

b) Dimensionnement informatique

Les moyens informatique de l'Unité d'études sont classiques d'un tel organisme : architecture ouverte, réseau local, stations de travail graphiques élaborées, restituteurs et digitaliseurs, moyens de communication téléinformatique avec la communauté scientifique (Bitnet, UseNet,...).

Il est recommandé, pour des raisons de sécurité informatique, de séparer physiquement le réseau local de l'unité d'études du reste de la configuration de l'Institut : banque de données, moyens de diffusion institutionnelle et grand-public.

1.6.4. Service d'archives

Pour compléter le rôle de dépositaire et de conservateur des données sur l'Environnement, l'Institut assume aussi un rôle de suivi de l'évolution de cet environnement. Le service d'archive lui permet de remplir ce rôle.

a) Composition de l'archive

L'étude de l'évolution de l'Environnement requiert deux démarches. Il s'agit de constater une évolution, puis de l'expliquer.

Pour constater une évolution, on compare des états successifs des données contenues dans la base.

Pour expliquer une évolution, on peut avoir recours aux données de base (données expérimentales, POT, statistiques, images, satellites,...) qui ont servi à élaborer la base.

De plus, même pour constater l'évolution de l'Environnement, il est souvent nécessaire de repartir des données de base afin de leur appliquer le même traitement, pour obtenir ainsi des séries spatio-temporelles comparables. Par exemple, pour comparer une évolution du parcellaire, deux images satellites doivent être corrigées géométriquement pour les rendre superposables.

Il faut donc conserver en archive les données de base aussi bien que les produits élaborés de la banque de données.

b) Constitution de l'archive

L'archivage des données de la banque s'effectue au moment de l'acceptation d'un dossier de mise à jour, selon le circuit suivant :

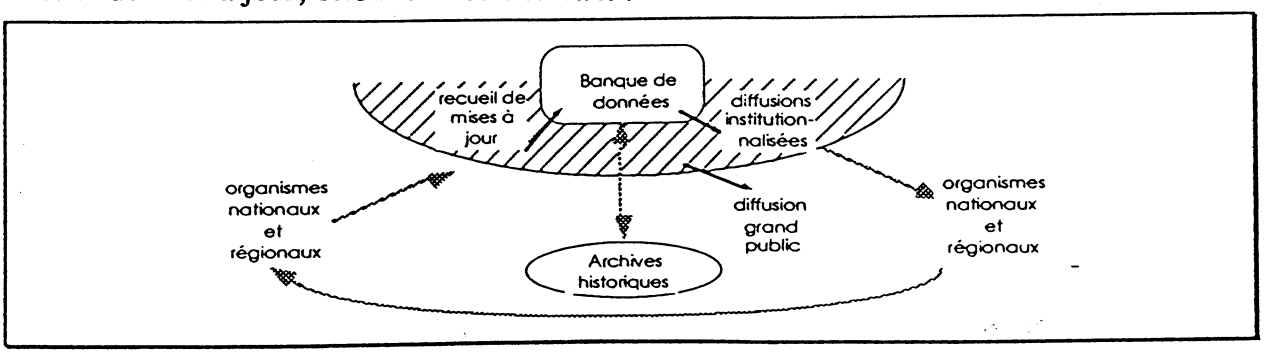


Figure 1-10 :

c) Accès aux archives

L'accès aux données archivées se fait sur la base de requêtes différées, en dehors des formules de fourniture de produits standards de l'Institut. En effet, la demande de ces services est relativement rare, car l'Institut fournira déjà en tant que produits standards des synthèses d'évolution qui répondent à ces préoccupations.

d) Dimensionnement informatique

Le volume des archives est variable, selon qu'elles contiennent, en plus des données statistiques traitées, des données de base comme les images satellites, les données automatiques des centrales d'acquisition, etc. qui sont volumineuses.

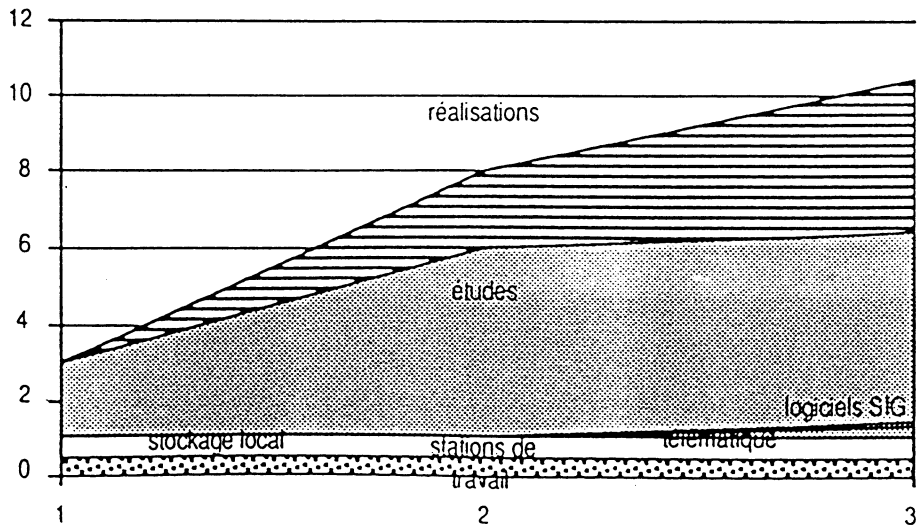
Il existe sur le marché des moyens de stockage sur disques optiques numériques susceptibles de répondre à ces exigences. On peut aussi envisager à terme les nouveaux supports comme les bandes DAT à enregistrement hélicoïdal, sous réserve que la durée de conservation du support vinyle soit prouvée.

1.7. ESTIMATION D'UNE ENVELOPPE DES COUTS

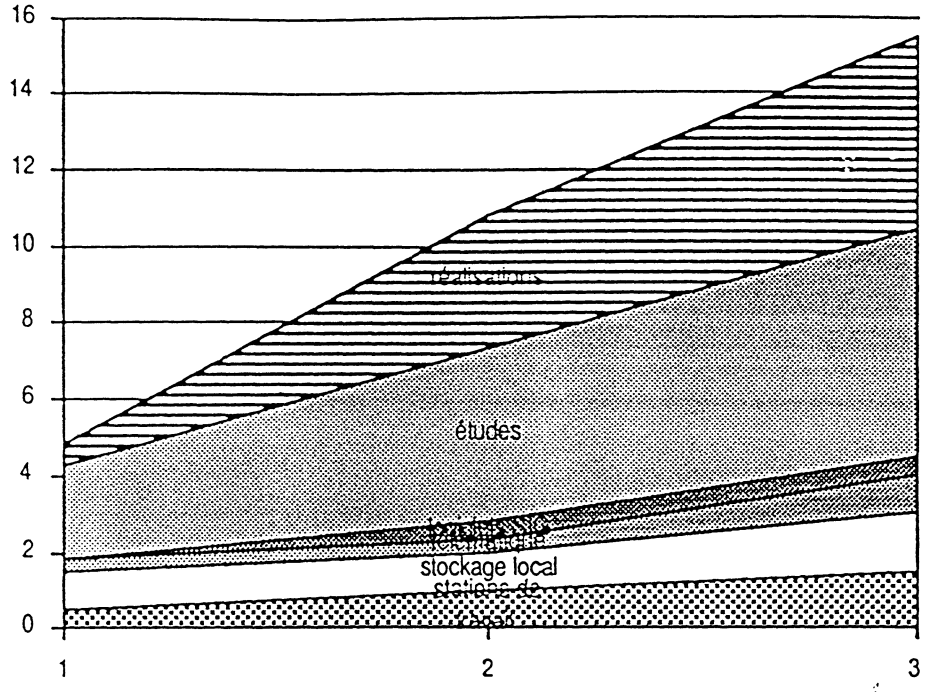
Dans la perspective d'une constitution des moyens informatiques de l'IFEN, basée sur des techniques permises actuellement ou dans un avenir proche, les estimations suivantes peuvent être avancées à titre indicatif.

1.7.1. Moyens de recueil et validation des données

recueil et validation	cumul hyp. basse (MF)			cumul hyp. haute (MF)		
	année 1	année 2	année 3	année 1	année 2	année 3
stations de travail	0,5	0,5	0,5	0,5	1	1,5
stockage local	0,5	0,5	0,5	1	1	1,5
télématique	0	0	0,3	0,3	0,3	1
logiciels SIG	0	0	0,1	0	0,5	0,5
études	2	5	5	2,5	4,5	6
réalisations	0	2	4	0,5	3,5	5
total	3	8	10,4	4,8	10,8	15,5



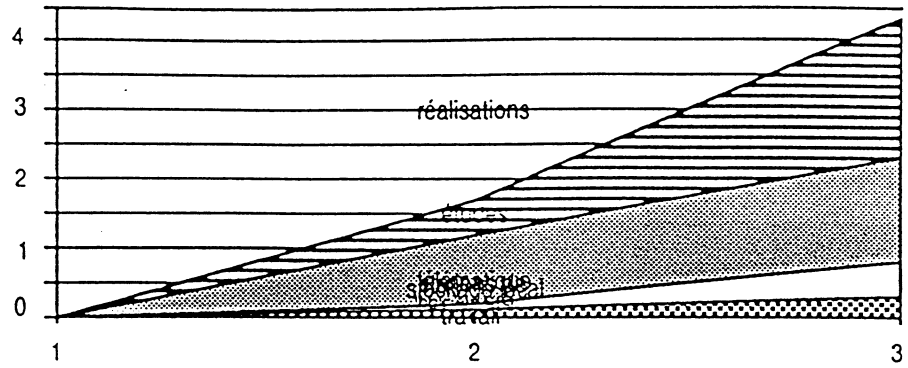
Evolution configuration IFEN, recueil de données, en MF (hypothèse basse)



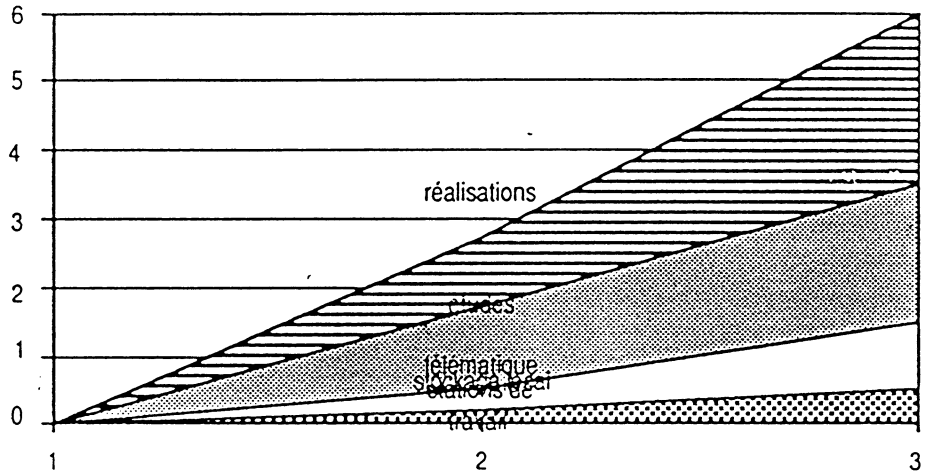
Evolution configuration IFEN, recueil de données, en MF (hypothèse haute)

1.7.2. Moyens de mise en forme et entrée dans la banque

mise en forme	cumul hyp. basse (MF)			cumul hyp. haute (MF)		
	année 1	année 2	année 3	année 1	année 2	année 3
stations de travail	0	0,1	0,3	0	0,2	0,5
stockage local	0	0,1	0,5	0	0,3	1
télématique	0	0	0	0	0	10
logiciels SIG	0	0	0	0	0	10
études	0	1	1,5	0	1,2	2
réalisations	0	0,5	2	0	1	2,5
total	0	1,7	4,3	0	2,7	26



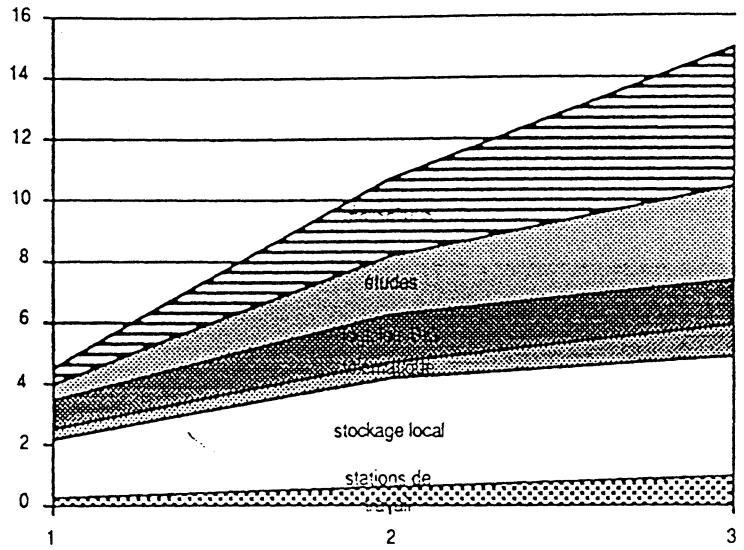
Evolution configuration IFEN, entrée de données, en MF (hypothèse basse)



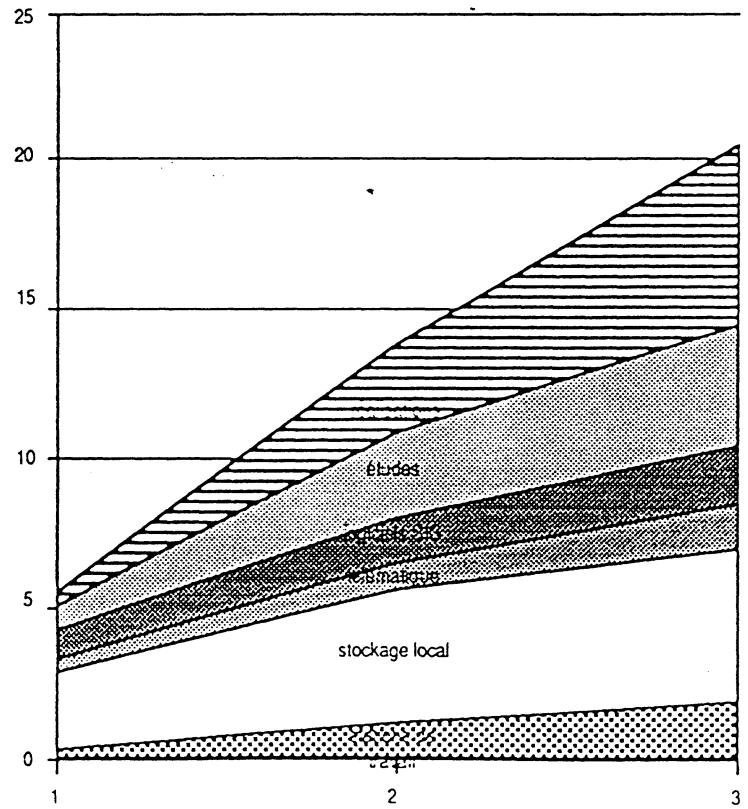
Evolution configuration IFEN, entrée de données, en MF (hypothèse haute)

1.7.3. Moyens d'entretien de la banque de données

entretien banque	cumul hyp. basse (MF)			cumul hyp. haute (MF)		
	année 1	année 2	année 3	année 1	année 2	année 3
stations de travail	0,2	0,6	0,9	0,4	1,2	2
stockage local	2	3,5	4	2,5	4,5	5
télématique	0,3	0,6	1	0,4	0,9	1,5
logiciels SIG	1	1,5	1,5	1	1,5	2
études	0,5	2	3	0,8	2,8	4
réalisations	0,5	2,5	4,5	0,5	3	6
total	4,5	10,7	14,9	5,6	13,9	20,5



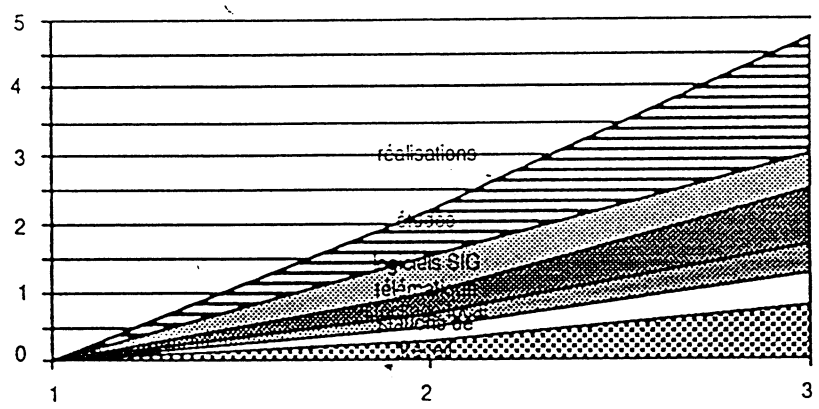
Evolution configuration IFEN, banque de données, en MF (hypothèse basse)



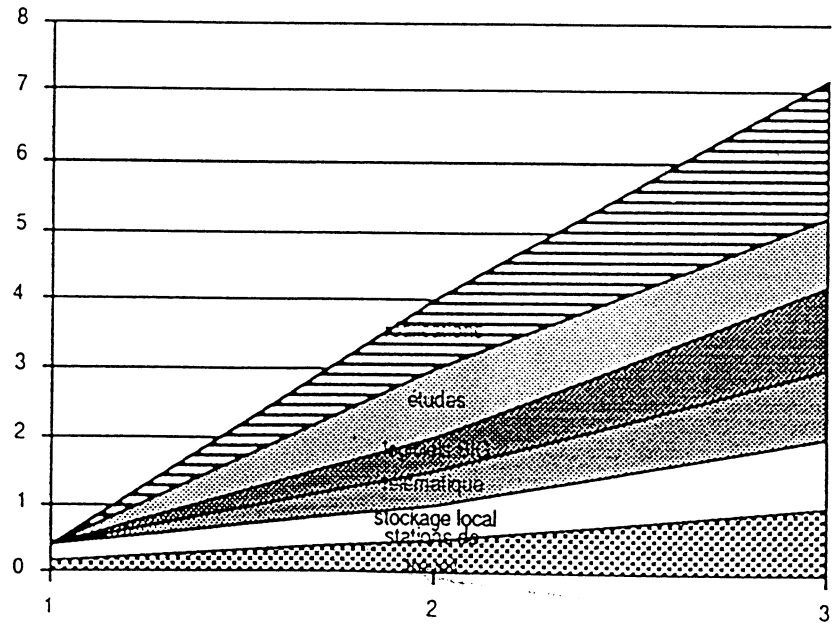
Evolution configuration IFEN, banque de données, en MF (hypothèse haute)

1.7.4. Moyens de l'unité d'études

unité recherche	cumul hyp. basse (MF)			cumul hyp. haute (MF)		
	année 1	année 2	année 3	année 1	année 2	année 3
stations de travail	0	0,3	0,8	0,2	0,5	1
stockage local	0	0,2	0,5	0,2	0,5	1
télématique	0	0,2	0,4	0	0,5	1
logiciels SIG	0	0,3	0,8	0	0,5	1,2
études	0	0,5	0,5	0	1	1
réalisations	0	0,7	1,7	0	1	2
total	0	2,2	4,7	0,4	4	7,2



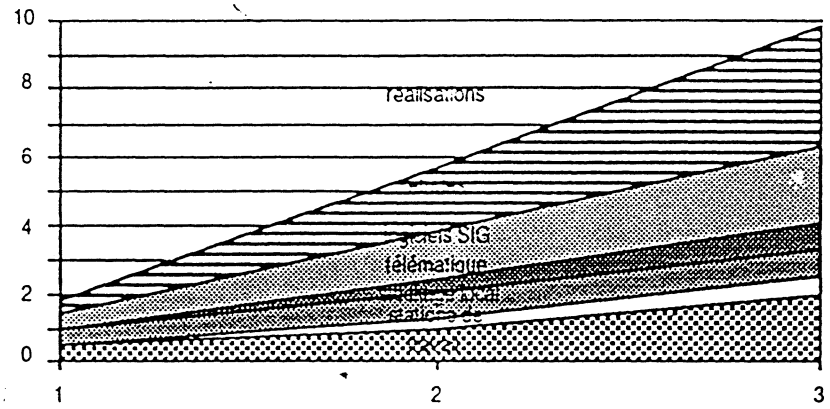
Evolution configuration IFEN, unité d'étude, en MF (hypothèse basse)



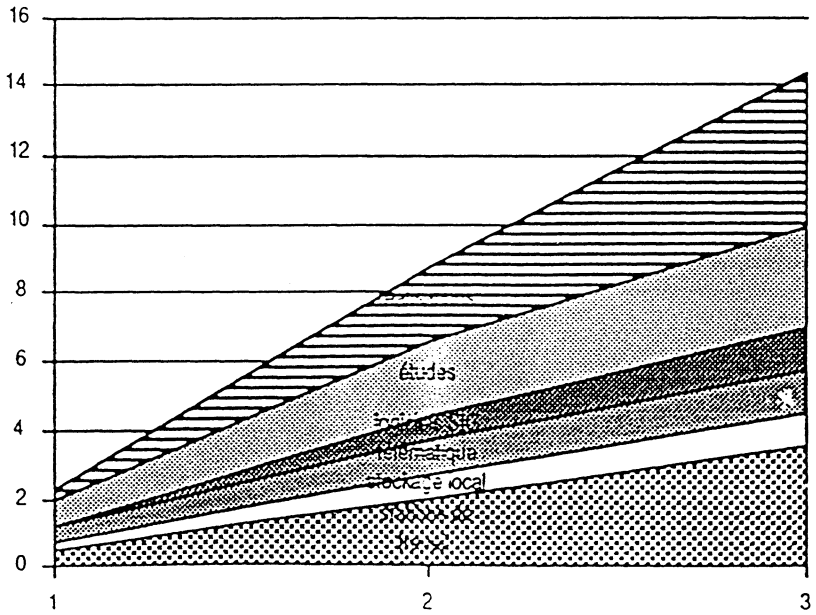
Evolution configuration IFEN,.unité d'étude, en MF (hypothèse haute)

1.7.5. Moyens de production et diffusion

production, diffusion	cumul hyp. basse (MF)			cumul hyp. haute (MF)		
	année 1	année 2	année 3	année 1	année 2	année 3
stations de travail	0,5	1	2	0,5	2	3,5
stockage local	0	0,3	0,5	0,2	0,7	1
télématique	0,5	0,8	0,8	0,5	1	1,2
logiciels SIG	0	0,3	0,8	0	0,7	1,2
études	0,5	1,5	2,2	0,8	2,2	3
réalisations	0,3	1,8	3,5	0,3	2,1	4,5
total	1,8	5,7	9,8	2,3	8,7	14,4



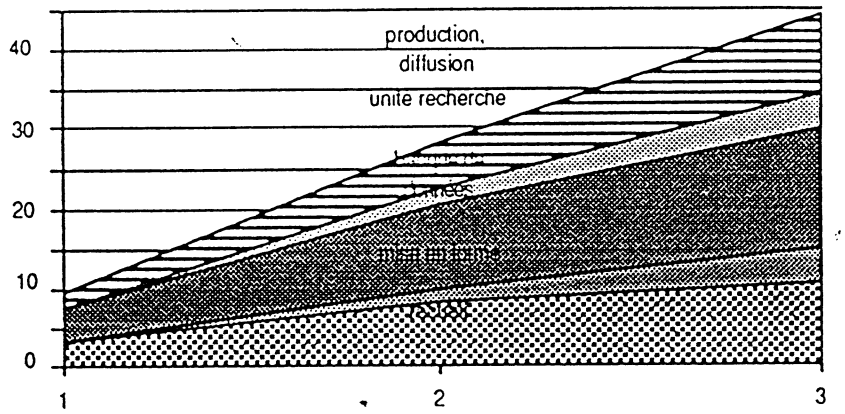
Evolution configuration IFEN, production et diffusion, en MF (hypothèse basse)



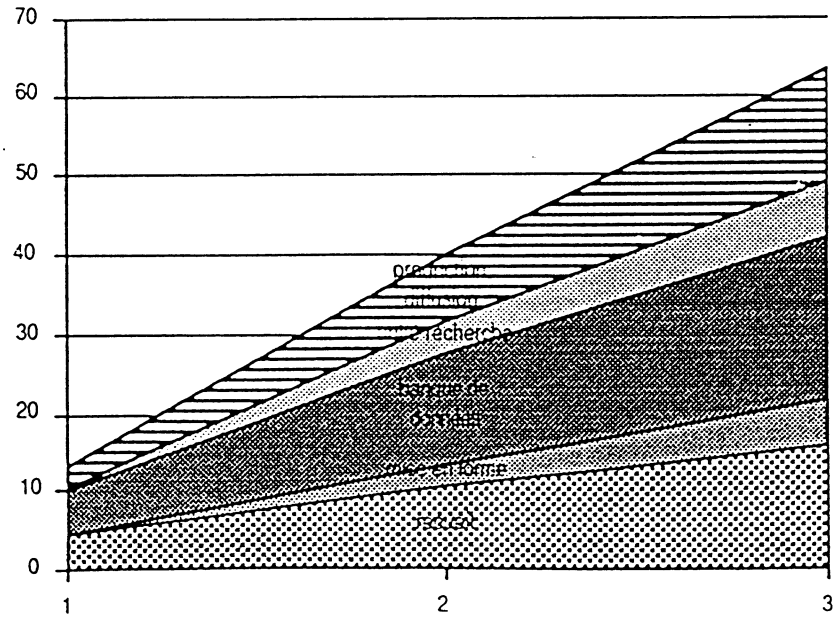
Evolution configuration IFEN, production et diffusion, en MF (hypothèse haute)

1.7.6. Récapitulatif

Récapitulatif	cumul hyp. basse (MF)			cumul hyp. haute (MF)		
	année 1	année 2	année 3	année 1	année 2	année 3
recueil	3	8	10,4	4,8	10,8	15,5
mise en forme	0	1,7	4,3	0	2,7	6
banque de données	4,5	10,7	14,9	5,6	13,9	20,5
unité recherche	0	2,2	4,7	0,4	4	7,2
production, diffusion	1,8	5,7	9,8	2,3	8,7	14,4
total	9,3	28,3	44,1	13,1	40,1	63,6



Evolution configuration IFEN globale, en MF (hypothèse basse)



Evolution configuration IFEN globale, en MF (hypothèse haute)

1.8. CONCLUSIONS ET PERSPECTIVES

Sur la base de l'étude de faisabilité et d'organisation ci-dessus, il s'avère que l'évolution de la technique informatique présente actuellement et à terme toutes les ressources informatiques nécessaires à l'équipement et au fonctionnement de l'Institut Français de l'Environnement.

Le schéma d'organisation tel qu'il a été identifié, permet de constater une très nette connotation "service" de l'Institut :

* service de collecte et de maintien de données sur l'Environnement,
* service de mise à disposition de données sur l'Environnement,
* service de promotion de l'Environnement,
* service d'études et de validation méthodologiques,
* service de gestion des données d'archives sur l'Environnement.

Ces services relèvent tous de la mission de l'Institut, qui est de constituer et maintenir un compte des ressources d'Environnement au niveau national et communautaire.

Toutefois, on constate que le service de mise à disposition de données (la banque de données) relève uniquement d'une prestation technique et non thématique, et son exploitation peut donc être sous-traitée à un organisme externe à l'Institut, soit à un serveur spécialisé analogue aux serveurs Minitel, soit à une société informatique intervenant dans le secteur de l'Environnement. La garantie de qualité de service serait assurée par l'Institut (en particulier l'unité d'études) qui en serait un consommateur exemplaire.

Cette organisation présente l'avantage de mieux centrer l'Institut sur sa mission à caractère thématique, et de lui permettre de se positionner d'emblée en tant que point focal national vis-à-vis de l'Agence Européenne de l'Environnement.

CHAPITRE DEUXIEME

QUELQUES CONCEPTS ET OUTILS NECESSAIRES POUR LA GESTION D'UN SYSTEME INFORMATIQUE OPERATIONNEL AU SERVICE DE L'ENVIRONNEMENT ET DE L'INTEGRATION TRANSDISCIPLINAIRE

2.1. LES BASES DE DONNEES REPARTIES : UNE SOLUTION A LA GESTION INTEGREE ET TRANSDISCIPLINAIRE DES DONNEES DE L'IFEN

2.1.1. Présentation

Un **Système de Gestion de Bases de Données** (SGBD) est aujourd'hui un logiciel de base essentiel dans un système informatique de gestion. Intuitivement, il permet à des utilisateurs concurrents de manipuler (insérer, modifier et rechercher) efficacement des données contenues dans une base de données. Depuis l'apparition des premiers SGBD vers 1962, d'importants résultats théoriques et pratiques ont ponctué l'histoire de la recherche sur le bases de données. La mise en oeuvre de ces résultats a permis de faciliter l'administration, la gestion et la manipulation de données, et d'accroître ainsi la productivité des utilisateurs des bases de données (administrateurs, programmeurs d'applications et utilisateurs finals).

L'histoire des SGBD peut être résumée en distinguant trois générations. La **première génération** de SGBD s'appuie sur les modèles de données **hiérarchiques et réseaux**. Elle vise à étendre un système de fichiers (l'ancêtre des SGBD) par des possibilités de liaisons interfichiers matérialisées par des pointeurs. Les bases de données sont représentées au niveau des types d'articles par une hiérarchie ou un graphe. Un SGBD de première génération fournit une faible indépendance physique, compliquant ainsi l'administration et la manipulation des données. En particulier, son langage de manipulation de données **navigationnel** impose au programmeur de spécifier les chemins d'accès aux données en naviguant dans le graphe de la base.

La **deuxième génération** de SGBD est née vers 1970 avec l'apparition du modèle de données **relationnel**. Une dizaine d'années d'efforts de recherche et de développement furent nécessaires pour aboutir à la commercialisation des premiers SGBD **relationnels**. Depuis, la technologie des **bases de données relationnelles** a fait l'objet de progrès remarquables en termes de facilité d'usage et de performances. Aujourd'hui, tout SGBD relationnel offre un ensemble intégré d'outils basé sur un Langage de 4e Génération (L4G) afin d'accroître la productivité des utilisateurs. Aussi, certains systèmes supportent efficacement les applications transactionnelles, très exigeantes en performances. En conséquence, le marché actuel de bases de données est couvert en grande majorité par les SGBD relationnels.

Les premières et secondes générations de SGBD ont été conçues pour les applications de gestion classique. Ces SGBD sont généralement implantés comme une couche externe d'un système opératoire installé sur un ordinateur général (gros-ordinateur, mini-ordinateur ou micro-ordinateur). Deux tendances complémentaires récentes soulignent les limites de ces systèmes. La première est la manifestation de besoins pressants en gestion de données de la part d'applications nouvelles, telles que la bureautique, la conception assistée par ordinateur (CAO) ou le génie logiciel, qui se singularisent par rapport aux applications traditionnelles de gestion. La seconde tendance est la diversité croissante des environnements informatiques, essentiellement décentralisés, qui favorisent la répartition et l'hétérogénéité des données.

On parle souvent de **troisième génération** de SGBD pour désigner des systèmes qui supportent bien ces applications nouvelles ou exploitent des environnements opérationnels complexes (répartis, parallèles ou hétérogènes). Cette nouvelle génération est initiée dans quelques laboratoires de recherche et de développement en France comme à l'étranger (surtout aux USA). Déjà, quelques produits issus de ces travaux apparaissent sur un marché encore mal connu mais prometteur.

2.1.2. Les bases de données relationnelles

Le modèle relationnel, inventé par E.F. Codd [Codd 1970], peut être introduit informellement par les trois propriétés suivantes :

a) Les structures de données sont simples et se construisent à partir de la théorie des ensembles. Ce sont des tables à deux dimensions dont chaque élément appartient à un ensemble de valeurs appelé **domaine**. Puisqu'elle définit une relation entre des domaines, une telle table est appelée **relation**. Une colonne d'une relation, appelée **attribut**, est définie sur un même domaine. Une ligne d'une relation, appelée **n-uplet** ou **tuple**, relie logiquement les éléments d'information fournis par les valeurs d'attributs. Toute l'information de la base de données doit être représentée explicitement par des valeurs dans des tables. En particulier, il n'y a pas de pointeurs interrelations visibles par l'utilisateur.

b) Un ensemble d'opérateurs appliqués à des relations, constituant **l'algèbre relationnelle**, permet la définition, la recherche et la mise à jour des données. Chaque opérateur prend une ou deux relations en argument et produit une relation. L'algèbre relationnelle comprend les opérateurs classiques de la théorie des ensembles (produit

Cartésien, union, intersection, différence) et des opérateurs permettant de composer des sous-ensembles d'une ou deux relations (projection, restriction, jointure et division).

c) Un ensemble de **contraintes d'intégrité sémantiques** définit les états cohérents de la base de données. Il existe au minimum deux contraintes d'intégrité **structurelles**, c'est-à-dire, inhérentes au modèle relationnel. La première est **l'unicité de clé**, qui spécifie qu'un ou plusieurs attributs constituent la clé unique d'une relation. La seconde est la **contrainte de référence** qui permet de garantir que la valeur d'une donnée dans une relation est présente dans une autre relation.

L'avantage majeur du modèle relationnel est sa faculté à assurer l'indépendance complète entre les descriptions de données logiques (en termes relationnels) et physiques (en termes de fichiers et liaisons inter-fichiers). Cette indépendance physique a permis le développement de langages de définition et de manipulation de données de haut niveau, appelés **langages de requêtes**, au-dessus de l'algèbre relationnelle. Ces langages sont typiquement **assertionnels**, c'est-à-dire, basés sur la logique des prédicats, et libèrent le programmeur de la spécification des chemins d'accès aux données. En conséquence, l'optimisation des requêtes de manipulation de données peut être entièrement automatisée. L'existence du langage de requêtes standard SQL [ANSI 1986] contribue d'ailleurs fortement à la promotion du modèle relationnel. SQL fournit une interface uniforme aux administrateurs, programmeurs d'applications, et utilisateurs finals, pour la définition, le contrôle et la manipulation de données. Un autre avantage de cette indépendance physique est de fournir une base solide pour l'aide à la conception de schémas conceptuels et internes.

La plupart des SGBD relationnels commercialisés supportent le langage SQL. De plus, ils offrent généralement un ensemble intégré d'outils L4G (générateur d'applications, gérant de menus et de fenêtres, aide à la conception de schémas, etc.) qui facilitent le développement des applications bases de données et améliorent ainsi la productivité des utilisateurs. La réalisation de ces outils a été facilitée par la simplicité et la puissance du modèle relationnel.

2.1.3. Limites des SGBD relationnels

Pour des raisons historiques, la plupart des SGBD relationnels actuellement sur le marché ont été conçus pour les applications de gestion et développés sur des ordinateurs classiques. Il est donc naturel que certains nouveaux besoins exprimés par les applications

ne puissent être bien satisfaits. Les limites majeures des SGBD relationnels actuels concernent le support d'objets complexes et dynamiques, la gestion des connaissances et la gestion de données réparties.

1) Support d'objets complexes et dynamiques

Un objet **complexe** désigne informellement un objet à structure complexe ou un objet atomique potentiellement grand. Par exemple, un texte, le schéma d'un composant VLSI, une image ou un programme sont des objets complexes. Le support et la manipulation efficaces des objets complexes sont indispensables aux nouvelles applications des bases de données (bureautique, CAO, etc.). Le support d'objets complexes par les SGBD relationnels actuels est difficile pour les raisons suivantes :

a) Les types de données supportés sont en général limités à quelques domaines alphanumériques (par exemple, entiers, réels, chaînes de caractères et dates). Il est impossible à l'utilisateur d'étendre les types de base avec des types spécifiques (par exemple, montant en Francs) ou génériques (par exemple, liste). Une solution est de modéliser les objets atomiques de type riche comme des chaînes de caractères et de les manipuler à l'extérieur du SGBD grâce à un langage de programmation (qui permet de spécifier les types nécessaires). Cette solution souffre de problèmes importants : redondance du code qui interprète les objets atomiques dans les programmes d'application, dépendance entre programmes et données, et inefficacité de la manipulation d'objets atomiques de grande taille.

b) Les structures de données fournies par le modèle relationnel sont trop simples pour modéliser des objets à structure complexe comme une hiérarchie ou un graphe. Un **objet complexe** (par exemple, un objet hiérarchique de CAO) peut être facilement décomposé en relations. Cependant, cette solution pose deux problèmes sérieux. D'abord, l'information sémantique véhiculée par la structure complexe d'un objet est dispersée sous forme de valeurs dans différentes relations, imposant à l'utilisateur d'adapter sa vision à celle du SGBD. Ensuite, la composition de données d'un même objet complexe peut exiger un grand nombre de jointures (une opération coûteuse en temps d'exécution).

c) Un langage de requête ne permet que difficilement les manipulations sophistiquées d'objets complexes. La solution pratique consiste à intégrer le langage de requête dans un langage de programmation (par exemple, SQL dans PASCAL, PL1 ou COBOL). Cette intégration pose le problème délicat de l'interface entre deux mondes qui diffèrent par leurs

types et leurs modes d'opérations (le langage de requête est ensembliste alors que le langage de programmation est procédural).

Une application a un aspect statique, représenté par les données, et un aspect dynamique, représenté par les traitements sur ces données. Les objets manipulés par un SGBD relationnel sont essentiellement statiques, leur dynamique (ou comportement) étant décrite séparément par les requêtes et les programmes d'application qui les manipulent. Cette approche impose au programmeur de connaître la structure (relationnelle) des objets. En d'autres termes, un SGBD relationnel ne supporte pas les **objets dynamiques**, incorporant à la fois une partie données et une partie traitements.

2) Gestion des connaissances

Une relation est une collection de tuples représentant des **faits**. Un SGBD relationnel permet donc essentiellement la manipulation de faits. Les **connaissances** sont des assertions générales et abstraites sur les faits. Par exemple, l'assertion "Socrate est un homme" est un fait alors que l'assertion "les hommes sont mortels" est une connaissance. Les connaissances permettent de raisonner. Dans le monde informatique, ceci peut se traduire par deux fonctionnalités :

a) l'inférence de nouveaux faits à partir de faits connus ; par exemple, "Socrate est mortel" est un fait nouveau que l'on peut déduire de la base ;

b) la production de réponses "intelligentes" ; par exemple, on peut répondre à la question "qui est mortel ?" de façon concise par "tous les hommes" (plutôt que de les énumérer tous).

Le support des connaissances autour de la base de données est un besoin important des applications de l'intelligence artificielle, comme les systèmes experts, qui peuvent opérer sur de grandes bases de faits.

Un SGBD relationnel supporte une forme limitée de connaissances : les contraintes d'intégrité. Cependant, les connaissances plus générales doivent être intégrées dans les programmes d'applications. Cette approche soulève trois problèmes importants. D'abord les connaissances sont codées dans les programmes d'application souvent avec duplication. Ensuite, il est difficile voire impossible de partager les connaissances entre les utilisateurs. Enfin, le raisonnement sur de grands volumes de faits intervient à l'extérieur du SGBD et peut être très inefficace.

3) Gestion de données réparties

Un SGBD désigne un logiciel classiquement implanté sur un ordinateur central. Cependant, le monde informatique évolue rapidement vers des environnements de travail plus variés et mieux adaptés à notre monde complexe. En particulier, les systèmes répartis offrent des alternatives intéressantes à la gestion de données.

Un **système réparti** est constitué d'un ensemble de machines interconnectées par un réseau de communication et utilisées dans un but global : la gestion de l'entreprise (au sens large). Dans un environnement décentralisé, un besoin essentiel est l'administration et la manipulation de données **réparties**, c'est-à-dire situées sur différents ordinateurs, et éventuellement gérées par des systèmes **hétérogènes**. Idéalement, la gestion de telles données doit assurer la transparence de la répartition et de l'hétérogénéité aux utilisateurs. Ceci exige alors des extensions considérables aux fonctions de gestion de données centralisées.

2.1.4. Les SGBD avancés

Les SGBD avancés ont pour ambition de lever les limites des SGBD relationnels actuels. Ils suscitent aujourd'hui de nombreux travaux de recherche et de développement. Cependant, quatre approches importantes peuvent être isolées : les **bases de données orientées objets**, les **bases de données déductives**, les **bases de données relationnelles étendues** et les **bases de données réparties**.

Les bases de données orientées objet [Bancilhon 1988] ont pour objectif essentiel le support d'objets complexes et dynamiques tel que nous l'avons introduit précédemment. Cette approche consiste à intégrer les techniques des langages de programmation orientés objets (comme C++) et des bases de données afin d'offrir un langage unique et général pour programmer les applications traditionnelles et récentes des bases de données.

Les bases de données déductives [Gardarin 1987] visent à gérer les connaissances autour de la base de données qui reste souvent relationnelle. Elles permettent essentiellement la déduction d'informations nouvelles à partir de données stockées dans la base. Cette approche peut être perçue comme la combinaison des techniques de la programmation logique (comme PROLOG) et des bases de données relationnelles. La logique du premier ordre lui fournit une base théorique solide.

Les bases de données relationnelles étendues [Gardarin 1989a, Stonebraker 1986] cherchent à atteindre les objectifs des bases de données orientées objets et des bases de données déductives en restant dans un contexte relationnel. Les extensions majeures souhaitables à un SGBD relationnel sont les suivantes : le support de types abstrait de données pour objets dynamiques [Cheiney 1989], le support d'objets complexes [Valduriez 1987] et le support de vues déductives [Madelaine 1986]. Ces nouvelles fonctions peuvent être intégrées uniformément dans un langage de requêtes, par exemple, une extension de SQL [Gardarin 1989c].

Les bases de données réparties [Ozsu 1989] permettent la gestion de données logiquement corrélées et physiquement stockées dans un réseau d'ordinateurs. Cette approche intègre les techniques des systèmes répartis et des bases de données. En général, elle s'appuie sur le modèle relationnel pour étendre l'indépendance physique en particulier à la répartition des données et à l'hétérogénéité des systèmes. Les récents travaux autour de cette approche concernent d'une part, la coopération entre systèmes hétérogènes et d'autre part, l'implantation de bases de données sur des ordinateurs parallèles.

L'approche bases de données orientées objets est singulière dans le sens où elle s'oppose à l'approche relationnelle. Au contraire, les trois autres approches s'appuient sur le modèle relationnel pour en conserver ses avantages.

2.1.5. Pourquoi une gestion orientée bases de données réparties ?

Depuis les débuts de l'informatique, l'approche dominante à la gestion de données dans l'entreprise a été **centralisée** ; les données sont gérées par un seul ordinateur contrôlé par un service informatique omnipotent. Jusqu'au début des années 1980, le coût élevé des ordinateurs faisait de cette approche la seule solution économiquement viable. Ce contexte a par ailleurs favorisé la notion de base de données qui tend à centraliser la définition et l'administration des données. Plus récemment, les développements remarquables de domaines aussi divers que les réseaux de communication, les mini et micro-ordinateurs, et les bases de données relationnelles ont permis à l'approche **base de données répartie** de devenir une solution alternative à la centralisation. Intuitivement, une base de données répartie est une collection de données logiquement corrélées et physiquement réparties sur plusieurs machines interconnectées par un réseau de communication. Un programme d'application qui manipule une base de données répartie peut alors accéder à des données résidant sur plusieurs machines, sans que le programmeur ait à connaître la localisation de ces données.

Une gestion orientée base de données répartie apporte actuellement des avantages appréciables en fonctionnalités et performances. D'abord, chaque groupe de personnes (ou chaque personne) disposant d'un ordinateur peut exercer un contrôle plus direct sur les données gérées par cet ordinateur. Ceci se traduit par une meilleure intégrité des données et une plus grande efficacité des traitements de données locales. Ensuite, par comparaison avec l'approche centralisée où les données doivent être transférées depuis chaque groupe vers l'ordinateur central afin d'être traitées, le traitement local des données par l'ordinateur du groupe réduit les coûts de communication. Enfin, les performances et la fiabilité de la gestion de données peuvent être augmentées en exploitant les capacités de traitement parallèle et de redondance offertes par plusieurs machines. Par ailleurs, l'approche bases de données réparties reste la solution naturelle à la gestion de données d'une grande entreprise géographiquement répartie.

Une base de données répartie permet d'intégrer et de partager des données gérées par des ordinateurs différents. Le partage de données réparties repose sur l'interconnexion des ordinateurs par un réseau local ou général (national par exemple), et un système spécifique pour gérer les données réparties. En particulier, ce système doit fournir un haut degré d'indépendance par rapport à l'environnement réparti qui peut être très complexe. La technologie des bases de données relationnelles offre une grande indépendance physique des données, en rendant transparent aux programmes d'application tout changement des structures de données physiques. Pour cette raison essentielle, la plupart des travaux visant à construire des bases de données réparties s'appuient sur le modèle relationnel. Ainsi, la plupart des SGBD relationnels actuellement sur le marché offrent des facilités de gestion de bases de données réparties [Gardarin 1989a, Valduriez 1989a].

Un SGBD réparti augmente les fonctions d'un SGBD afin de permettre la définition et la manipulation de bases de données réparties. Sa construction pose des problèmes difficiles qui résultent de la combinaison de deux technologies importantes et complexes, les SGBD et les systèmes répartis. Ces problèmes ont été à l'origine de nombreux travaux de recherche fructueux qui ont produit de nouveaux principes et de nouvelles solutions, faisant des bases de données réparties un domaine important de l'informatique. Cette technologie est en pleine expansion et son rôle dans la gestion de données sera de plus en plus important.

2.1.6. Les bases de données réparties (distributed database, ddb)

Une base de données répartie est un ensemble de bases de données coopérantes qui résident sur différentes machines, que nous appelons **sites**, interconnectées par un réseau de communication. Idéalement, un utilisateur sur n'importe quel site peut accéder et manipuler des données sur n'importe quel site.

La différence essentielle avec plusieurs bases de données distantes est que l'utilisateur n'a pas à être informé de la répartition des données sur plusieurs sites et qu'il perçoit la base de données répartie comme une base non-répartie. La répartition des données est alors **transparente** à l'utilisateur. La gestion d'une base de données répartie nécessite sur chaque site l'installation des composants systèmes suivants : un système de communication de données (responsable de l'interface avec le réseau), un SGBD local, et un gérant de données réparties (voir Figure 2-1).

Le système qui gère une base de données répartie est un SGBD réparti, c'est-à-dire un ensemble de logiciels systèmes constitués des gérants de données réparties, des gérants de communication et des SGBD locaux résidant sur chaque site de la base de données répartie.

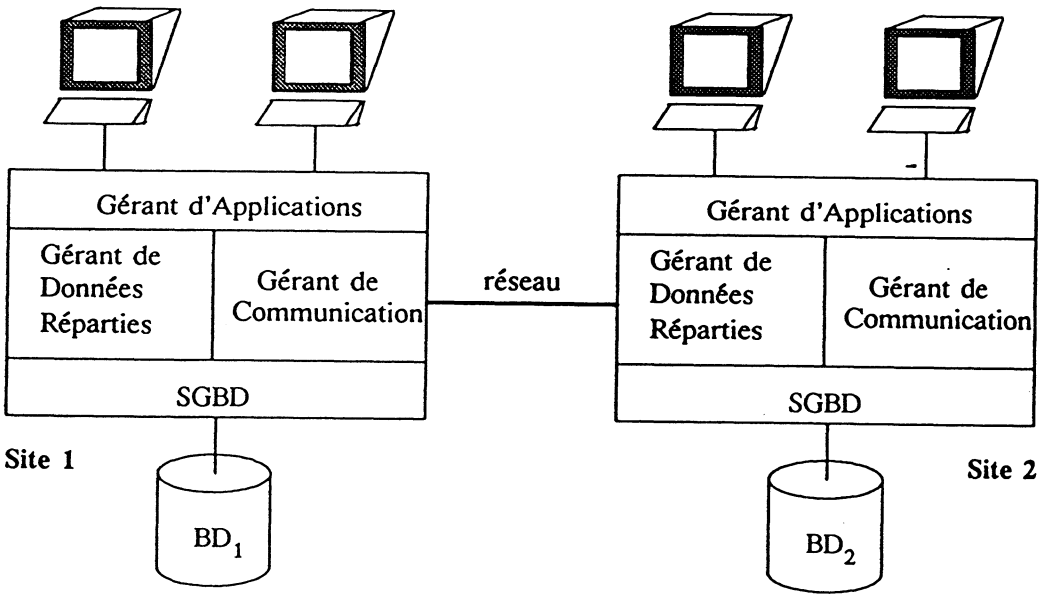


FIGURE 2-1 : *Une Base de Données Répartie*

Les fonctions spécifiques d'un SGBD réparti sont, en dehors de celles des SGBD locaux et du gérant de communication :

1) la gestion d'un dictionnaire de données global qui maintient l'information relative aux données réparties,

2) la définition de données réparties,

3) le contrôle sémantique des données réparties,

4) l'évaluation de requêtes réparties, émises par les utilisateurs et invoquant la base de données répartie,

5) la gestion de transactions réparties, programmes atomiques à l'intérieur desquels sont émises les requêtes réparties.

Une propriété importante d'une base de données répartie est l'homogénéité ou l'hétérogénéité (voir figure 2-2).

Une **base de données répartie homogène** est composée de bases de données locales gérées par le même SGBD.

Une base de données répartie homogène résulte en général d'une approche de conception descendante où l'on conçoit une base de donnée répartie à partir d'une base de données centralisée. Elle favorise la croissance incrémentale de l'entreprise, en facilitant l'ajout d'un nouveau site dans le réseau. Elle permet aussi d'augmenter les performances de la gestion des données, en exploitant les capacités de traitement parallèle fournies par des sites multiples.

La caractéristique essentielle d'une **base de données répartie hétérogène** est d'intégrer des bases de données locales gérées par des SGBD différents. Deux niveaux d'hétérogénéité peuvent être distingués :

1) les bases de données ont le même modèle (par exemple, relationnel) mais sont gérées par des SGBD différents (par exemple : DB2 et ORACLE),

2) les bases de données ont un modèle différent (par exemple, ORACLE et IMS).

Cette approche est bien plus complexe que l'approche homogène. Cependant, elle favorise une approche de conception ascendante qui permet l'intégration de bases de données existantes et indépendantes sans avoir à créer une base de données répartie complètement nouvelle. En plus des fonctions précédemment indiquées, le gérant de données réparties hétérogènes doit fournir des interfaces pour faire communiquer les différents SGBD.

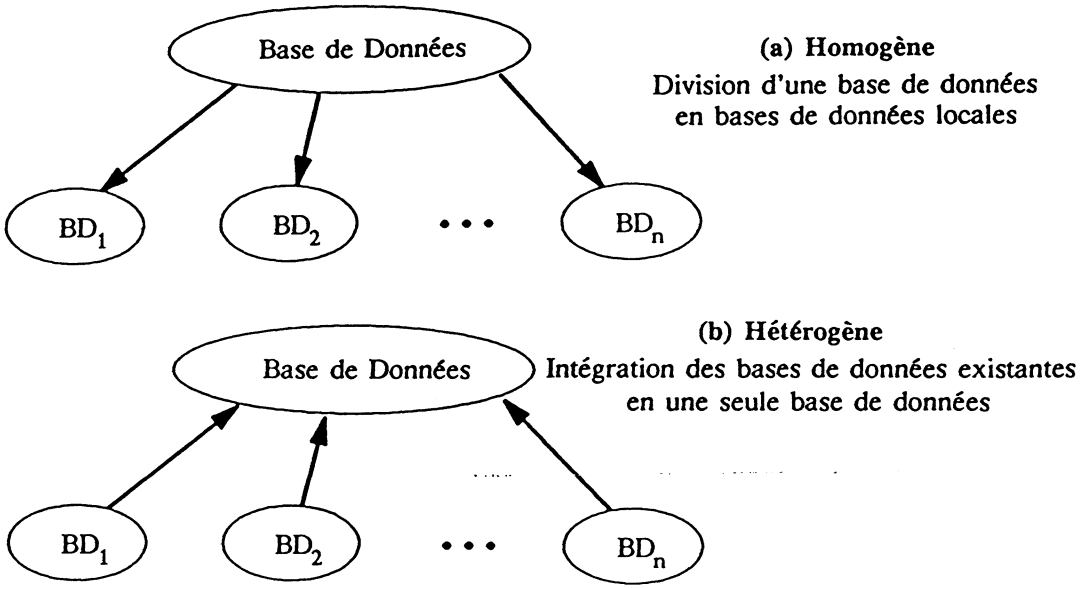


FIGURE 2-2 : *Différentes Approches de Conception des Bases de Données Réparties*

2.1.7. Objectifs des SGBD répartis

Une base de données répartie peut fournir différents niveaux d'indépendance. Cependant, chaque niveau d'indépendance contribue au même objectif général : rendre l'utilisation d'une base de données répartie pas plus difficile que celle d'une base de données centralisée. Il n'existe pas de système qui atteigne tous les objectifs suivants. En fonction du marché d'applications visé, un SGBD réparti n'atteindra qu'un sous-ensemble de ces objectifs. Par exemple, un SGBD réparti homogène ne fournit pas l'indépendance des SGBD.

1) Indépendance à la localisation

L'indépendance à la localisation des données est certainement le premier objectif d'une base de données répartie. Elle cache le fait que les données ne résident pas nécessairement sur le site utilisateur. L'utilisateur de la base de données répartie peut ignorer la localisation des données. En particulier, les requêtes invoquant des relations stockées à des sites différents n'ont pas à préciser la localisation des relations. La transparence à la localisation fournit l'indépendance physique par rapport à l'environnement réparti. L'information concernant la localisation des données est maintenue dans le dictionnaire de données et consultée par le SGBD réparti pour déterminer la localisation des relations impliquées dans les requêtes des utilisateurs. L'avantage est que la réorganisation physique de la base de données, qui peut conduire à transférer des relations d'un site à un

autre, est sans impact sur les programmes d'application qui accèdent à la base de données répartie. Cependant, les requêtes qui accèdent à des relations impliquées par la réorganisation doivent en général être recompilées.

2) Indépendance à la fragmentation

La façon la plus simple de gérer un objet conceptuel (une relation) dans une base de données répartie est de le stocker sur un seul site. Cependant, les accès aux données sont généralement définis sur des sous-relations plutôt que des relations entières. Pour des raisons d'efficacité, il est alors désirable de diviser une relation en plus petites sous-relations appelées **fragments**, pouvant être stockées à des sites différents. La fragmentation augmente les performances d'une base de données répartie car elle permet de favoriser les accès locaux. Considérons par exemple la relation LIEUX d'une base de données répartie entre Montpellier, Paris et New York. Le placement optimal de cette relation consiste probablement à la fragmenter en une relation de lieux Américains et une relation de lieux Européens, la première localisée à New York et la seconde à Montpellier ou Paris, afin que les données se trouvent là où elles sont le plus souvent accédées.

L'indépendance à la fragmentation des données cache à l'utilisateur le fait que les données sont fragmentées. L'utilisateur ne manipule que des relations conceptuelles non réparties. L'information concernant la fragmentation est maintenue dans le dictionnaire des données et utilisée par le SGBD réparti pour traduire automatiquement les requêtes sur les relations conceptuelles en requêtes sur des fragments.

3) Indépendance à la duplication

L'unique solution à la fiabilité des données est la redondance sous certaines formes. Dans une base de données répartie, la duplication de données peut servir à augmenter la fiabilité et la disponibilité des données ainsi que les performances d'accès. En général, l'unité de duplication est le fragment, ou la relation si la fragmentation n'est pas supportée. Un fragment est dupliqué lorsqu'il est matérialisé par deux ou plusieurs copies, chacune résidant sur un site différent. Une telle duplication, contrôlée par le système, fournit une grande disponibilité des données. Si un site devient non opérationnel, à la suite d'une panne par exemple, la copie du fragment est toujours accessible sur un autre site. La duplication permet aussi d'améliorer les performances d'accès en augmentant la localité des références. Lorsque le coût de communication est un facteur dominant, la duplication d'un fragment sur le site où il est le plus souvent accédé favorise les accès locaux et évite l'accès au réseau.

Les avantages apportés par la duplication sont à comparer avec la complexité et les coûts supplémentaires de maintenance des copies qui doivent, en théorie, rester identiques à

tout moment. La mise à jour d'un fragment doit être répercutée automatiquement sur toutes ses copies. Le problème est compliqué par la présence de pannes de site. Lorsqu'un site redevient opérationnel après une panne, tous ses fragments dupliqués doivent retrouver le même état que leurs autres copies, ces dernières ayant pu être modifiées lorsque le site était en panne. Le compromis recherché entre performance d'accès en consultation et en mise à jour des données rend difficile le choix du niveau de duplication. Celui-ci est très dépendant de la charge de travail demandée par les applications.

L'indépendance à la duplication des données rend la duplication invisible à l'utilisateur, qui ne voit que des relations non dupliquées (de même que non fragmentées). L'information concernant la duplication est aussi maintenue dans le dictionnaire de données. Elle est exploitée par le SGBD réparti lors de la traduction des requêtes utilisateurs en requêtes optimisées sur des fragments, avec le souci de maintenir la cohérence des copies.

4) Indépendance aux SGBD

L'indépendance aux SGBD permet de cacher le fait que les SGBD locaux peuvent être différents. Ces SGBD peuvent différer parce que leur modèle de données (relationnel, réseau, etc.) et le langage associé sont différents, ou plus simplement parce qu'ils ont été conçus par des groupes différents. L'indépendance aux SGBD est l'objectif particulier des bases de données réparties hétérogènes. Cet objectif reste très difficile à atteindre totalement. Pour cette raison, peu de SGBD réparties hétérogènes ont été développés. Le support total de l'indépendance aux SGBD exige un mécanisme de traduction du modèle de données et du langage de chaque SGBD local dans ceux de chaque autre SGBD local. La solution la plus courante consiste à employer un <u>modèle pivot</u> (et son langage associé), et un traducteur entre ce modèle pivot et le modèle de chaque SGBD différent. Pour des raisons de simplicité, le modèle relationnel est généralement choisi comme modèle pivot. En pratique, des solutions partielles sont proposées comme, par exemple, l'indépendance aux SGBD uniquement relationnels.

5) Autonomie des sites

L'autonomie des sites est un objectif récent qui permet à chaque site opérationnel de contrôler et de manipuler ses données locales indépendamment des autres sites. Pour ce faire, chaque site doit disposer localement de toute l'information qui se trouve dans le dictionnaire de données afin d'être autonome, sans avoir à compter sur un dictionnaire de données global centralisé (géré sur un autre site) et sur un administrateur de base de données globale.

L'avantage immédiat de l'autonomie de site est que l'administration de la base répartie peut être complètement décentralisée. Chaque base de donnée locale peut être indépendamment contrôlée par un administrateur local. Cependant, la coopération entre sites doit être coordonnée par les administrateurs locaux grâce à des commandes spécifiques fournies par le SGBD réparti. De plus, le fonctionnement local d'un site autonome n'est aucunement affecté par la panne des autres sites. Le support de l'autonomie de site va de la non-autonomie (c'est-à-dire contrôle centralisé) jusqu'à l'autonomie totale (c'est-à-dire contrôle décentralisé). En pratique, des solutions intermédiaires sont souvent employées.

6) Extensibilité

L'extensibilité d'une base de données répartie est sa capacité d'augmentation incrémentale, par introduction de nouveaux sites dans le réseau, avec un impact minimal sur les bases de données locales et les programmes d'application existants. Cet objectif est relativement facile à atteindre lorsque la base de données répartie supporte l'autonomie des sites, car l'ajout d'une nouvelle base locale n'est pas une opération globale. L'extensibilité est particulièrement utile pour augmenter les capacités de stockage et de traitement de données de l'entreprise, ce qui est une tendance naturelle.

7) Performances

Comme pour toute technologie, l'obtention de bonnes performances lors de la manipulation de données réparties est un objectif indispensable à la viabilité de l'approche base de données répartie. Il peut être atteint grâce à la fragmentation et à la duplication de données qui favorisent les traitements parallèles.

2.1.8. Architectures de référence

Comme les bases de données centralisées, les bases de données réparties peuvent être implantées de nombreuses façons selon leurs objectifs et leurs choix de conception.

1) Architecture des schémas

L'architecture des schémas, illustrée figure 2-3, propose une organisation idéale d'une base de données répartie supportant tous les niveaux d'indépendance possibles. Cette architecture est décrite en termes de niveaux de schémas (représentés par des boîtes) et transformations de niveaux de schémas (représentées par des lignes reliant les boîtes). Suivant les niveaux d'indépendance supportés, l'architecture d'une base de données répartie peut ne comporter qu'un sous-ensemble de ces niveaux de schémas. De plus, comme la conversion d'un niveau de schéma à un autre peut induire un coût de traitement important,

l'objectif de performances peut demander de combiner certains niveaux de schéma. L'architecture de référence est une extension simple de l'architecture à trois niveaux proposée par le groupe de travail ANSI/X3/SPARC pour bases de données centralisées [Tsichritzis 1978]. Ainsi, l'indépendance des données est facilement garantie. Toutes les informations concernant les schémas et les transformations de schémas sont maintenues dans le dictionnaire de données.

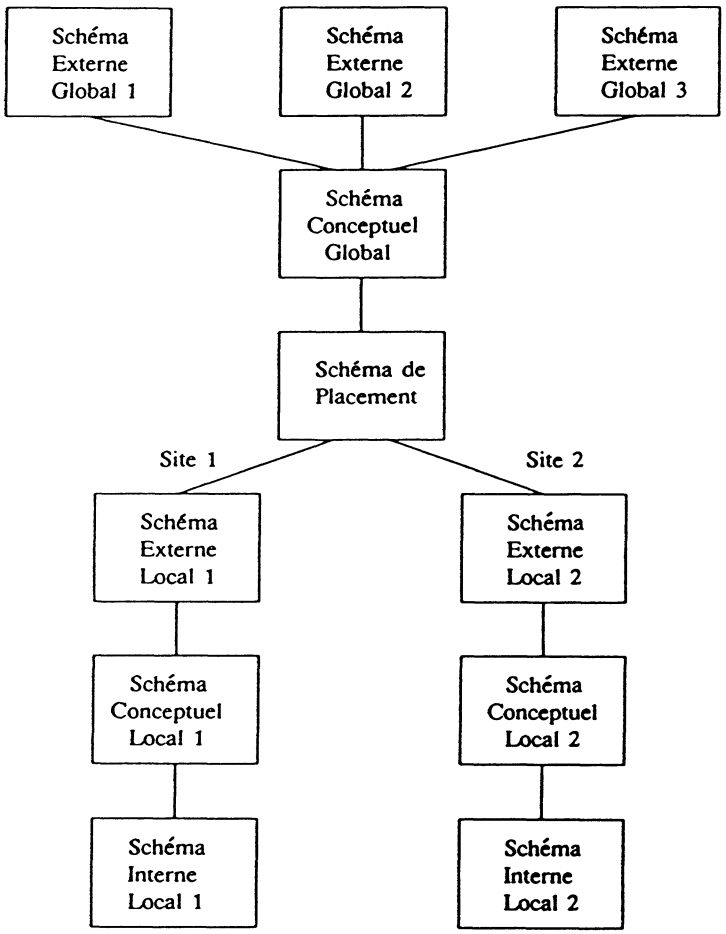


FIGURE 2-3 : *Architecture des Schémas d'une Base de Données Répartie*

Les niveaux de schéma qui décrivent la base de données répartie indépendamment de toute base locale sont appelés **globaux**.

Une requête utilisateur (de recherche ou de mise à jour) sur une base de données répartie est appelée requête répartie.

Trois niveaux de schéma global peuvent être distingués : le schéma conceptuel global, les schémas externes globaux et le schéma de placement. Le schéma conceptuel global définit toutes les relations appartenant à la base de données répartie et fournit l'indépendance physique des données par rapport à l'environnement réparti. L'indépendance

logique des données est assurée par les schémas externes globaux, chacun supportant la vue d'une classe d'utilisateurs particulière. Les règles de définition des vues servent à transformer les données du niveau schéma externe global en données du niveau schéma conceptuel global. Le schéma de placement précise la façon dont les relations sont placées sur les différents sites dans le réseau. Il contient donc toute les informations relatives à la localisation, la fragmentation et la duplication des données, en particulier les règles de transformation des données du niveau schéma conceptuel global en données localisées, fragmentées et dupliquées. Le schéma de placement assure donc l'indépendance à la localisation, à la fragmentation et à la duplication.

Les niveaux de schéma qui décrivent une base de données locale sont appelés **locaux**. Par opposition au schéma global qui est unique, il y a autant de schémas locaux que de sites dans le réseau. Une requête exprimée sur un schéma local est une **requête locale**.

Comme dans l'architecture de référence ANSI/X3/SPARC, il peut exister trois niveaux de schémas locaux. Le schéma local interne et le schéma local conceptuel sont identiques à ceux d'une base de données centralisée. Seul le schéma local externe à un rôle différent. Il représente les fragments tels qu'ils sont décrits dans le schéma de placement en objets externes, considérés comme des vues de la base locale. Cette transformation est dépendante du SGBD local. Le schéma local externe assure donc l'indépendance des SGBD, et doit être présent si l'environnement est hétérogène. Il est aussi utile lorsque la base de données répartie n'inclut qu'un sous-ensemble (défini par des vues) de la base locale.

2) Architecture fonctionnelle

La figure 2-4 illustre l'architecture fonctionnelle simplifiée d'une base de données répartie. Les fonctions sont groupées dans deux composants systèmes différents, le **gérant d'applications** (ou client) et le **gérant de données** (ou serveur), présents sur chaque site. En général, une requête d'un utilisateur invoque le gérant d'applications du site client et les gérants de données d'un ou plusieurs sites serveurs.

Les fonctions du gérant d'applications sont implantées par le gérant de données réparties. Quatre fonctions essentielles sont à distinguer : l'interface utilisateur, le contrôle sémantique des données, l'évaluation de requêtes réparties, et la gestion de transactions. L'interface utilisateur reçoit une requête de l'utilisateur, vérifie que les données spécifiées dans la requête sont conformes au schéma externe, et, après exécution de la requête, restitue les résultats à l'utilisateur (données, messages d'erreur, etc.). Le contrôle sémantique des données transforme la requête exprimée sur des vues, définies dans le

schéma externe global, en une requête exprimée sur des relations conceptuelles globales, définies dans le schéma conceptuel global, puis ajoute les contrôles d'autorisations et les contrôles d'intégrité sémantique.

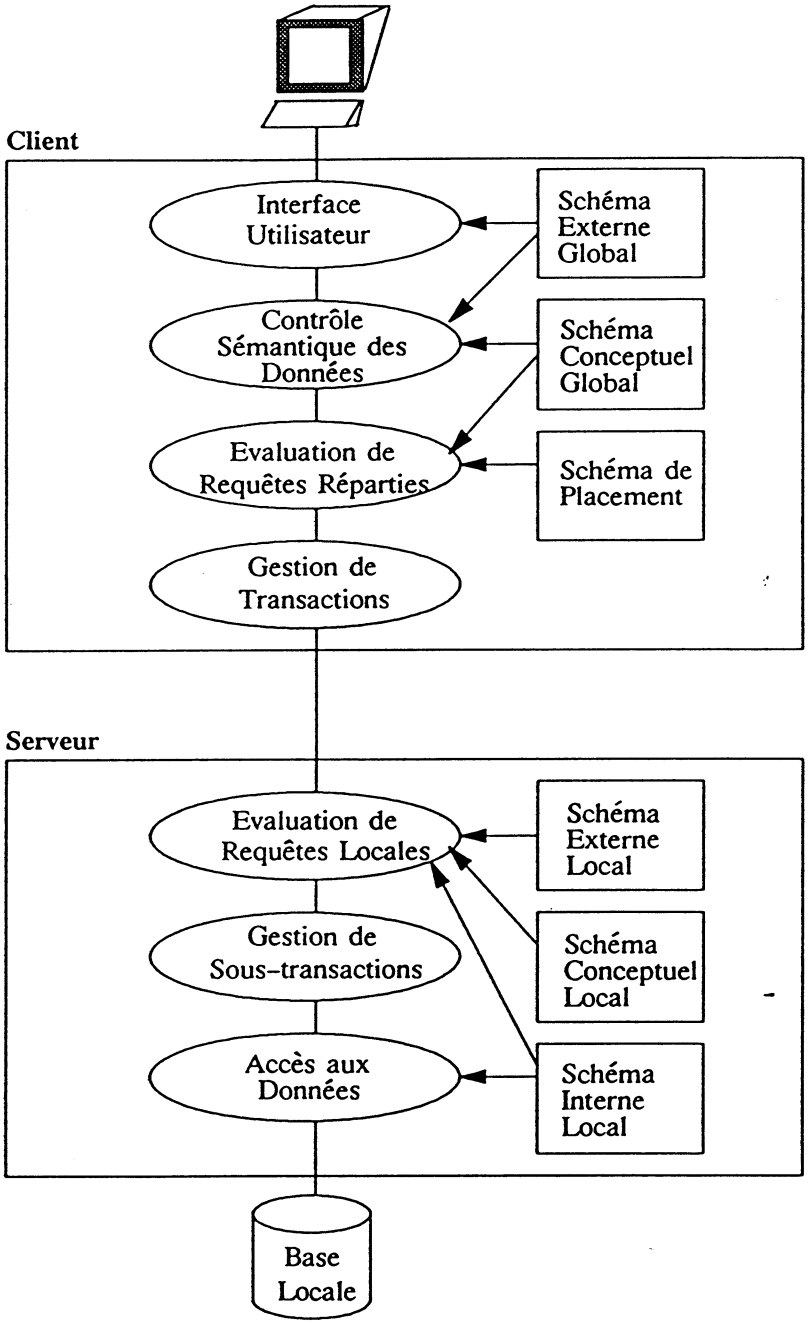


FIGURE 2-4 : *Architecture Fonctionnelle d'un SGBD réparti*

L'évaluation de requêtes réparties transforme une requête exprimée sur des relations conceptuelles en un **plan d'exécution réparti.**

L'optimisation de requêtes réparties consiste à choisir la meilleure localisation des copies de fragments, définies dans le schéma de placement, et à ordonner les requêtes locales de manière à minimiser une fonction de coût.

La gestion de transactions (réparties) coordonne l'exécution d'une requête répartie, qui constitue une transaction, en communiquant avec les gérants de données des sites sur lesquels les requêtes locales s'exécutent comme des sous-transactions. Cette fonction assure le contrôle de concurrence réparti et la validation de transactions.

Les fonctions du gérant de données sont implantées dans le gérant de données réparties et le SGBD local. Trois fonctions sont essentielles : l'évaluation de requêtes locales, la gestion de sous-transactions et l'accès aux données. L'évaluation de requêtes locales traduit d'abord la requête exprimée conformément au schéma de placement en une requête exprimée conformément au schéma externe local. Dans le cas d'une base hétérogène, cette traduction nécessite la conversion de la requête exprimée dans le modèle et le langage pivot en une requête exprimée dans le modèle et le langage de la base locale. La requête peut alors être transformée en une requête optimisée exprimée suivant le schéma interne local, après un passage intermédiaire optionnel par le niveau de schéma conceptuel local.

La gestion de sous-transactions coopère avec le gérant d'applications, qui a initié la transaction, et les autres gérants de données afin de synchroniser l'exécution des sous-transactions de la même transaction. L'accès aux données comporte les fonctions normales de gestion de données d'un SGBD (par exemple, les méthodes d'accès et la gestion de mémoire).

2.1.9. Conclusion

Les bases de données réparties sont maintenant un domaine important de la gestion d'informations. L'objectif général d'une base de données répartie est de permettre à l'utilisateur de manipuler un ensemble de bases de données gérées sur des sites différents comme une base de données centralisée. L'indépendance à l'environnement réparti est obtenue en assurant la transparence de la localisation, de la fragmentation et de la duplication des données. D'autres objectifs importants sont l'autonomie des sites, l'extensibilité du SGBD réparti, la grande disponibilité des données, les performances et l'indépendance à l'hétérogénéité des bases de données locales. Ce dernier objectif est le plus

ambitieux et le plus difficile à atteindre complètement. En pratique, les SGBD répartis existants ne visent qu'un sous-ensemble de ces objectifs.

Le modèle de données relationnel et les langages associés (SQL) sont le support naturel des bases de données réparties. D'abord, une base de données répartie exige divers niveaux d'indépendance des données qui peuvent être fournis plus simplement par le modèle relationnel que par des modèles complexes. Ensuite, les langages relationnels sont assertionnels et ensemblistes, ce qui facilite l'optimisation de requêtes réparties. En conséquence, tous les SGBD répartis existant aujourd'hui sont relationnels.

2.2. L'ARCHITECTURE CLIENT/SERVEUR : SOLUTION AU TRAITEMENT COOPERATIF ET A LA GESTION DU SYSTEME D'INFORMATION DE L'IFEN

2.2.1. Présentation

Allier les avantages respectifs des gros systèmes et de la micro-informatique en passant par les stations de travail, tel est en quelque sorte le défi de l'architecture client/serveur. Derrière le phénomène de mode se cache un bouleversement : la manière d'envisager la micro-informatique devient de moins en moins individuelle et de plus en plus intégrée et coopérative.

L'architecture client/serveur n'est-elle qu'une mode ou un réel progrès ? Progrès, certainement. Mode, également. Car si son intérêt est indéniable pour certaines applications, elle ne touche actuellement qu'une infime partie des utilisateurs de la micro-informatique et son implantation ne peut être que lente. Son rôle sera cependant décisif dans le domaine stratégique de l'accès à l'information. Elle redonne un sens nouveau aux réseaux et désenclave la micro-informatique.

"Le meilleur des deux mondes." Cette expression revient fréquemment à propos de l'architecture client/serveur, qui est indissociable des notions de réseau et de système d'exploitation multitâche. "Elle s'oppose aux concepts qui prédominaient auparavant, l'informatique centralisée d'une part, et le partage de ressources dans les réseaux locaux d'autre part", résume Hugo Lunardelli, chef de produit Lan Manager et SQL Server chez Microsoft. L'expression n'est pas très bien choisie car elle peut rappeler, à tort, celle de "maître/esclave". Le "client" est bien actif, et pas seulement au moment de la demande. L'objectif est d'utiliser la puissance disponible sur l'ensemble des machines.

En théorie, tout est simple et logique. Si ce type de configuration de réseau ne s'est pas développé plus rapidement, c'est surtout faute d'outils, mais également faute de besoins bien cernés. Une fois touchées du doigt les limites des réseaux traditionnels, on prend conscience de l'intérêt de l'architecture client/serveur.

2.2.2. Failles et limites

Avec un **serveur de réseau classique**, lorsque l'un des utilisateurs cherche une information contenue dans la base de données, il charge d'abord sur son propre poste

l'ensemble de la base, quelle que soit sa taille. Des centaines, des milliers de kilo-octets vont donc transiter sur le réseau, pour venir s'installer provisoirement sur le poste, qui assurera ensuite lui-même la recherche. Les inconvénients sont multiples. Le trafic généré sur le réseau est important. Si plusieurs personnes le sollicitent de la sorte au même moment, celui-ci sera rapidement "à genoux". La sécurité n'est pas assurée dans de bonnes conditions car la promenade sur les câbles de grosses masses d'informations est toujours dangereuse.

Sur le serveur lui-même, il est difficile de mettre toutes les protections nécessaires. Durant le travail effectué sur la base, celle-ci est interdite à tout autre utilisateur, puisqu'elle a quitté le serveur. Une copie reste bien sûr présente, mais elle est verrouillée pour des raisons évidentes. Si un second utilisateur la modifiait, deux versions différentes d'une même base se trouveraient en conflit. Dernier inconvénient, la recherche est assurée par l'un des postes du réseau, dont la puissance est très moyenne : vitesse processeur limitée, taille mémoire mesurée. Ce deuxième paramètre ne fait que multiplier les allers retours entre le serveur et le poste afin de faire transiter l'ensemble de la base par parties, nouveau facteur de ralentissement.

Sur un **gros système**, le processus est totalement inverse puisque tout se passe sur le serveur. L'inconvénient : une interface trop souvent rébarbative et l'impossibilité de réaliser des traitements en local. Le serveur supporte donc toute la charge de travail.

Au sein d'une **architecture client/serveur**, tout change. L'utilisateur qui recherche une information pose sa question au serveur. Ce dernier se charge seul de la recherche et ne renvoie que la réponse. Tous les inconvénients cités plus haut sont caducs. Le trafic sur le réseau est réduit au minimum. Quelques dizaines d'octets pour la requête et seulement la partie de la base concernée pour la réponse, dont la taille varie bien entendu selon la question. L'intégrité de la base est assurée. Elle ne quitte jamais le disque dur du serveur, ce qui facilite les protections et assure qu'elle n'existera jamais qu'en une seule et même version. Son accès n'est interdit que pendant le temps de la recherche, qui est lui-même réduit grâce à la puissance du serveur.

Plusieurs utilisateurs pourront interroger la même base dans une période très courte, sans devoir attendre que l'accès à celle-ci soit libéré. Le temps de réponse lui-même sera amélioré. Et, à condition qu'ils disposent des outils adéquats, ils ne se seront aperçus de rien, si ce n'est de cette vitesse accrue. Dans le même temps, ils conservent la facilité d'emploi de la micro-informatique, sa souplesse et sa puissance pour traiter l'information en local, ne serait-ce que pour réaliser des graphiques, insérer les éléments recueillis au sein d'un texte... Tout serait-il donc parfait ? En théorie, certainement. La réalité est plus complexe. Les outils manquent encore.

Des travaux sont en cours pour une totale normalisation. Les plus optimistes prévoient l'aboutissement de ces discussions pour dans deux ou trois ans...

L'utilisation directe du langage SQL est possible, mais elle n'est pas accessible à un non-informaticien, ce qui va à l'encontre de l'évolution de la micro-informatique, qui devient de moins en moins complexe à utiliser. Il est donc indispensable de passer par des logiciels standards, connus et bien maîtrisés par leurs utilisateurs, capables de traduire par eux-mêmes les interrogations en langage SQL. Ce dernier possède l'avantage d'exister sur toutes sortes de systèmes. Par lui, la micro-informatique sort de son "ghetto", ou de sa "cage dorée", pour communiquer enfin avec la "grande" informatique.

2.2.3. Transparence...

L'architecture client/serveur ouvre la porte de **l'architecture répartie**, où l'on va chercher des informations dans différents systèmes. Le serveur, en tant qu'unité logique, peut être constitué de plusieurs unités physiques, sans que l'utilisateur ne s'aperçoive de rien. Lorsqu'il pose une question, si la réponse n'est pas contenue dans la base du serveur micro par exemple, elle sera automatiquement routée vers un autre ordinateur pour y chercher les informations demandées. Si nos interlocuteurs ont toujours insisté sur cette transparence, ils se sont faits également plus discrets sur les conséquences pour l'ingénieur système responsable de la bonne marche de l'ensemble. "Les outils arrivent, note Hugo Lunardelli. Cela va introduire des bouleversements dans les méthodologies de développement."

Toutes les entreprises agissant sur ce marché se rejoignent dans leur présentation de l'architecture client/serveur. Certains insistent d'abord sur le gain de performance, d'autres sur l'ouverture vers différents systèmes, mais leur analyse globale est similaire. C'est sur l'utilisation des produits micro que les avis se modulent, surtout en fonction de l'offre de chacun. Michel Doric, responsable commercial SQLSystem à Ise-Cégos, insiste sur SQL Windows, un générateur d'applications grâce auquel l'utilisateur aura une application réalisée avec l'interface graphique Windows. Ce produit est vendu indépendamment de SQL Base et peut travailler avec Oracle, DB2 et sera prochainement compatible avec SQL Server.

2.2.4. ... Efficacité et accessibilité

Avec la généralisation de Windows sur les postes de travail, la présence d'OS/2 sur le serveur et l'arrivée de produits facilitant la communication entre les logiciels standards et la base de données, l'architecture client/serveur va devenir plus efficace et plus accessible. Pour Pierre Métivier, elle justifie OS/2. C'est l'un des éléments nécessaires à la généralisation de réseaux performants, ouverts à d'autres systèmes. En perdant son indépendance, la micro-informatique offre ses interfaces, sa facilité d'emploi et sa productivité à l'informatique répartie (ou centralisée).

2.2.5. Le serveur de données : étape décisive dans le traitement de l'information

Le **serveur de données** permet d'effectuer une partie du traitement de l'information sur le serveur même, avant toute consultation ou manipulation. Par exemple, au lieu de charger dans son intégralité la table "Matériel_Agricole" sur son micro, puis d'effectuer un tri ou une recherche multicritère (opérations qui peuvent prendre un temps considérable lorsque le fichier est de grande taille), l'utilisateur demande au serveur de données d'opérer directement la sélection. Cette "page" de données lui est ensuite envoyée par le serveur, qui ne verrouille l'accès qu'aux lignes correspondantes dans la base et non l'intégralité de la table. Cette manière de fonctionner offre plusieurs avantages : le trafic sur le réseau est optimisé et les accès concurrents à une même base sont facilités.

Dans un tel système, il y a répartition des tâches entre le serveur et le client en fonction de leurs capacités propres. La gestion de grandes bases de données, quel que soit leur modèle :

- relationnel, hiérarchique, ...
- en accès multi-utilisateur

suppose le **"fine tuning"**, c'est-à-dire le "réglage fin" d'un certain nombre de paramètres concernant la structure des données et leur accès.

Ces réglages sont fort consommateurs de temps CPU et suffisamment délicats à réaliser pour être réservés aux spécialistes de l'administration des données. Ils comprennent, entre autres : la reconstruction des index, la tenue d'une table système indiquant les accès qui ont été faits aux bases, la définition d'un cache pour y placer les données auxquelles il est le plus souvent fait appel, l'analyse de la charge des bases, la gestion des données sur des emplacements physiques distincts et, dans certains SGBD,

l'optimisation des chemins de requêtes. Il faut aussi, à ce niveau, définir l'allocation des ressources matérielles, comme l'espace disque alloué aux bases et leur emplacement physique.

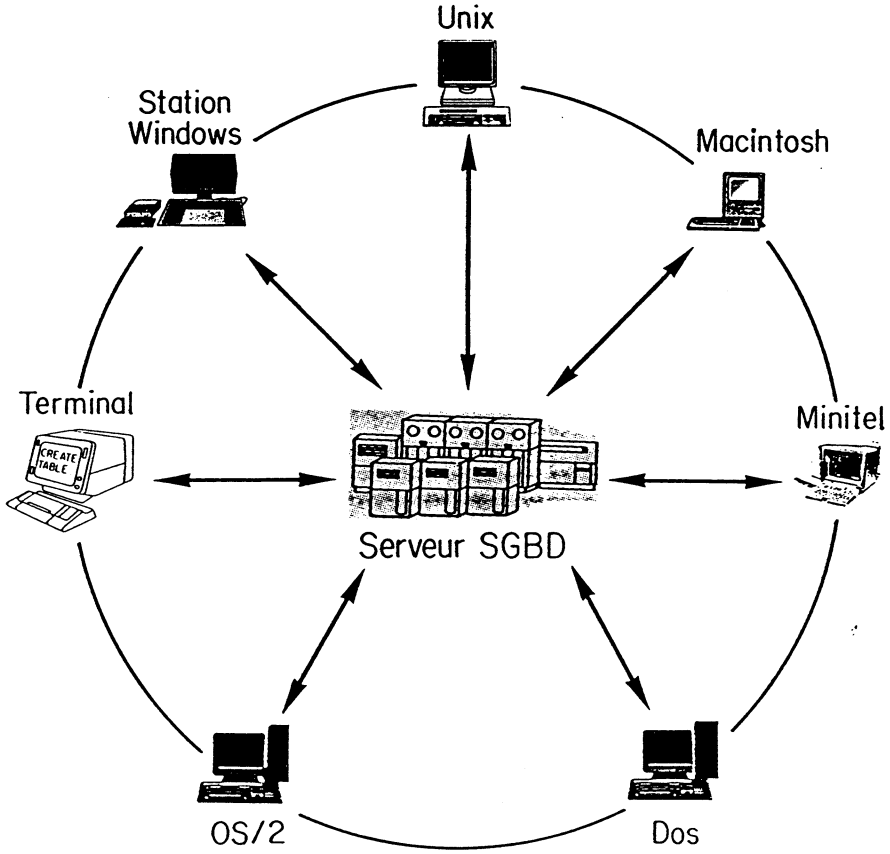


Figure 2-5 : *L'architecture client/serveur permet à plusieurs applications clientes de partager une base de données commune*

Orientés dès leur origine vers le service multi-utilisateur, les grands SGBD utilisés comme serveurs de données permettent également une meilleure gestion de la sécurité. Les privilèges d'accès peuvent être définis à tous les niveaux : sur la base dans son intégralité, sur certaines tables voire sur des enregistrements particuliers, avec faculté ou non pour l'utilisateur de donner tout ou une partie de ses privilèges à d'autres utilisateurs.

Toutes ces opérations, sans nouveautés pour les administrateurs de bases de données centralisées, prennent une importance plus grande lorsque le client n'est plus un simple terminal mais un ordinateur capable de dialoguer d'égal à égal (**"peer to peer"**, pour reprendre la terminologie IBM) avec le serveur. Alors que le terminal se contente de consulter les informations et demande à l'unité centrale d'effectuer les traitements pour lui, le poste micro bénéficie de capacités suffisantes pour faire en local un grand nombre d'opérations. Les deux extrêmes de l'éventail des tâches sont faciles à répartir : le **fine**

tuning est évidemment réservé au poste serveur, tandis que les outils d'aide à la décision (graphes, simulation...) appartiennent au client. Reste la gestion des transactions, de l'intégrité et des données remote ou locales.

2.2.6. Traitement différé contre temps réel

80% des bases de données sont encore gérées aujourd'hui par des **systèmes "batch"**. Ces systèmes (IMS et autres SGBD IBM, par exemple) permettent de rentabiliser le site central en effectuant une partie des traitements durant la nuit, de manière automatisée. Le principe d'un gestionnaire de données batch part du postulat suivant : nombre de traitements ne requièrent pas une exécution en temps réel. Très souvent, ce sont ceux qui consomment le plus de temps CPU.

Par exemple, dans une banque, la récapitulation des ordres de virement, des débits et des crédits représente chaque jour une masse extrêmement volumineuse de transactions à réaliser. Un simple bordereau de remise de chèque peut mettre en relation plusieurs comptes, dans plusieurs banques différentes, au milieu de giga-octets de données réparties sur divers serveurs. Les différentes opérations de ce type sont donc stockées dans des fichiers de requêtes temporaires (ou notées sur des formulaires papier). En fin de journée, le site central consolide ces différentes transactions et génère le programme de mise à jour du site central, qui s'effectuera durant la nuit.

Dans d'autres cas, le traitement batch s'appliquera à la génération d'états sur une partie de la base, comme la liste des clients ayant eu un découvert durant les mois de janvier, février et mars et dont le compte est ouvert dans la succursale X...

Ces systèmes peuvent aussi fonctionner en **temps réel** pour tout ce qui concerne la consultation. Ainsi, lorsque le client demande une copie des dernières opérations effectuées sur son compte, le résultat de la requête est immédiat. De même, lorsqu'il consulte ses relevés sur le minitel, il obtient l'état de ses comptes sans attendre le retour du courrier. Malgré cette possibilité, les systèmes batch restent très lourds et peu adaptés à la demande des utilisateurs en matière d'information. S'ils se justifient dans des secteurs informatiques très consommateurs de ressources, comme les banques, EDF..., ils ne peuvent faire la base d'une architecture client/serveur orientée vers une gestion décentralisée de l'information.

En **mode transactionnel**, chaque demande du client est traitée en temps réel par le serveur. Il y a deux acceptions du mot transactionnel : la première concerne ce fonctionnement en temps réel, l'autre la manière dont sont traitées les demandes des clients.

L'architecture client/serveur visant à donner à l'utilisateur une plus grande interactivité avec le site central, il est nécessaire de permettre l'accès en temps réel à l'information. La délocalisation d'une partie des traitements sur le poste micro permet, de surcroît, de décharger le site central de certaines opérations.

Prenons l'exemple d'une agence de voyages qui réserve des billets d'avion pour ses clients. Chaque réservation met en jeu les données suivantes : horaires, lieux de départ et de destination, compagnie, type d'avion, places disponibles, classe de transport et réductions éventuelles. Tous ces éléments sont utilisés pour donner une place au client et valider sa réservation. Etant donné le nombre d'aéroports, de compagnies et d'horaires possibles, chaque requête est en théorie très complexe. Pour faciliter la gestion de ces informations, les vols sont référencés par un numéro unique dont la liste est publiée tous les six mois. L'agence de voyages stocke cette liste sur son poste local, étant donné la périodicité des mises à jour, ce qui lui permet de simplifier sa requête. Le client indique ses desiderata en matière d'horaires et de lieux ; le poste local détermine alors le numéro de vol correspondant. Il fournit ce numéro au site central, qui envoie les informations concernant les places disponibles dans les différentes catégories. Il ne reste plus qu'à choisir et valider.

Cet exemple montre combien l'architecture client/serveur facilite le travail des entreprises, où le nombre de données à manipuler est important. Il met en relief la répartition des données entre le serveur et le client. Un tel système requiert évidemment des temps de réponse réduits au minimum, le client de l'agence de voyages n'étant généralement pas disposé à attendre longtemps sa réservation.

2.2.7. La transaction : élément important de l'architecture client/serveur

La réception d'une requête par le serveur peut être traitée de deux manières : soit en interprétant chaque instruction de la requête, soit en considérant la globalité de la requête, nommée dans ce cas **transaction**. La transaction offre deux avantages : une plus grande rapidité d'exécution, (les appels aux différentes ressources étant groupés) et une meilleure sécurité.

Si une requête est traitée comme une transaction, c'est-à-dire un ensemble indissociable, le serveur lit la séquence d'instructions comme un petit programme complet et l'exécute de la même manière qu'un programme compilé, d'où des temps de réponse bien meilleurs. De plus, les différentes tables concernées sont toutes verrouillées simultanément, ce qui évite de créditer un compte sans débiter l'autre ou inversement. La

validation n'est effectuée que si toutes les instructions ont pu être exécutées correctement. On évite ainsi des situations peu agréables.

La transaction s'impose plus encore lorsque plusieurs serveurs sont concernés par une même requête (base de données répartie). Les risques d'interruption entre les différentes instructions sont plus grands et la validation doit intervenir sur tous les serveurs ou sur aucun (cf. **"Le two-phase commit"**). Or la possibilité de communiquer avec plusieurs serveurs, simultanément ou non, est un des apports de l'architecture client/serveur à la gestion du système d'information de l'entreprise.

2.2.8. La communication multiserveur

Dans les environnements distribués actuels, il est nécessaire d'avoir accès de manière transparente aux données, quel que soit leur emplacement physique ou logique. Le poste client doit pouvoir consulter et manipuler l'information issue de différents SGBD et bases de données. Cet accès doit être capable de prendre en compte l'hétérogénéité des modèles de gestion des données (relationnel, hiérarchique...), des SGBD, des systèmes d'exploitation, du matériel ou du réseau (cf. **"Une question de réseaux"**). Il doit être possible de créer des transactions distribuées qui utilisent plusieurs ordinateurs ou serveurs, accédant aux données et aux services de chacun.

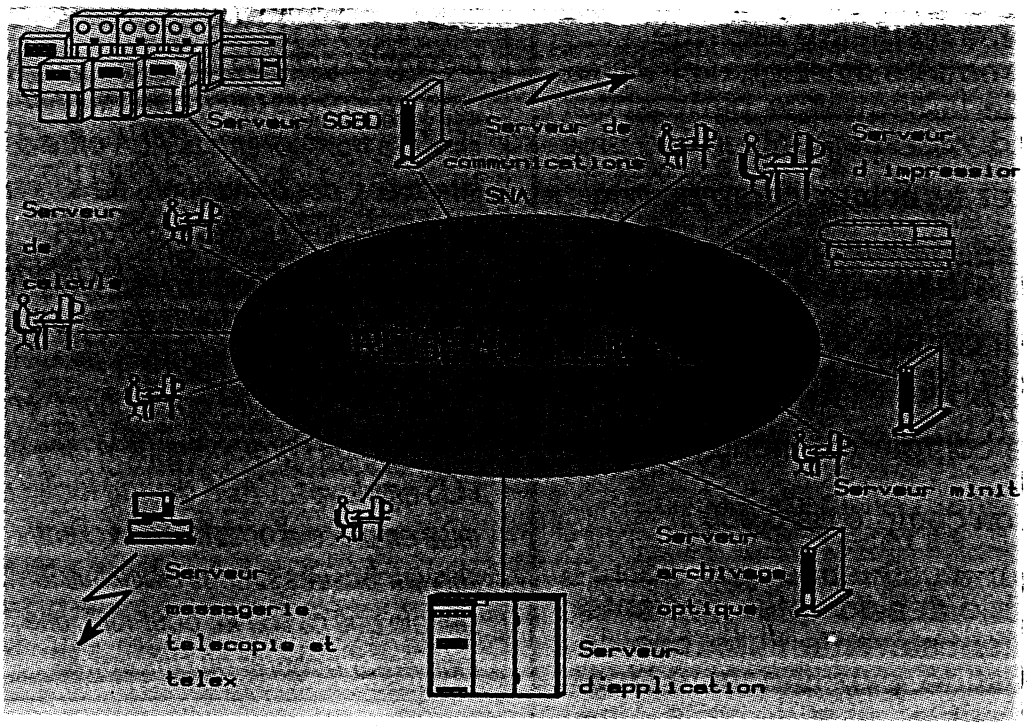


Figure 2-6 : *Modèle client/serveur*

Chez un grossiste, le service facturation accède aux données du service stocks et envoie des messages au service comptable, chaque service possédant son propre serveur. Ce type de transactions distribuées doit être capable de gérer plusieurs ordinateurs, applications et données dans un environnement intégré. Le but de l'architecture client/serveur est de permettre l'accès aux informations sans qu'il soit nécessaire de se préoccuper de leur emplacement et de leur système de gestion. De ce fait, une architecture complète ne doit pas seulement gérer l'accès aux données mais doit également permettre de créer des transactions distribuées sur plusieurs serveurs.

Aujourd'hui, la solution généralement retenue par les éditeurs est de décliner leur SGBD sur un plus grand nombre de plates-formes matérielles et logiciels.

La méthode adoptée par tous ces grands éditeurs est l'utilisation d'API, pour **Application Program Interface**. Les API sont des routines d'accès au serveur que l'on insère dans le programme client, permettant aux concepteurs de systèmes de bâtir des transactions distribuées d'une manière standardisée.

Les postes clients peuvent aussi utiliser un langage comme SQL, suffisamment standardisé pour permettre l'émission de requêtes vers de nombreux sites, à condition toutefois qu'ils soient basés sur le modèle relationnel.

SQL est un standard largement utilisé pour accéder aux bases de données. Néanmoins, chaque implantation diffère des autres et peut contenir des extensions incompatibles avec le SQL d'une autre. La gestion des erreurs diffère également. Enfin, 80% des SGBD dans le monde sont encore bâtis sur le modèle hiérarchique, et SQL n'est d'aucune utilité dans ce domaine. C'est pourtant ce langage et les concepts qui l'accompagnent qui ont contribué à l'essor du modèle client/serveur. SQL est avant tout un langage de communication avec les bases de données. Il est indépendant du système interne de gestion des bases et peut être mis en oeuvre (bien que difficilement) par des non-professionnels de l'informatique.

Par exemple, en donnant aux postes clients, travaillant sous dBase ou autres logiciels micro, la possibilité d'envoyer une requête SQL à un ou plusieurs serveurs pour récupérer des données et les traiter directement dans l'application cliente, on a permis à l'architecture client/serveur de faire ses premiers pas.

Depuis, les solutions de ce type se sont améliorées et les interfaces SQL prennent également en compte d'autres éléments, comme la gestion des protocoles de communication dans DAL, l'implantation SQL d'Apple. DAL est une couche logiciel qui se place sur le

serveur et transforme les requêtes arrivant en une transaction compréhensible par chacun de ces SGBD, qui possèdent des versions différentes de SQL. DAL comporte aussi une couche cliente chargée de la gestion des requêtes à laquelle se rapportent les applications générant des procédures DAL, moins complexes que les requêtes SQL mais gérant les protocoles de communication. Avec cette solution, le monde Macintosh prend encore quelques longueurs d'avance sur les compatibles PC dans le domaine de la communication hétérogène.

2.2.9. Le two-phase commit

Le two-phase commit a pour but de permettre des modifications sur des bases réparties. Par exemple, si les informations concernant les comptes épargnes sont stockées sur un serveur et les informations concernant les comptes courants le sont sur un autre, tout transfert de fonds implique une modification des données sur les deux serveurs. Pour assurer l'intégrité de ce type d'opération, les serveurs utilisent un protocole nommé two-phase commit.

Lorsqu'il faut faire une modification répartie, l'application cliente émet la requête vers un des serveurs concernés considéré comme le commit server. Le commit server prend la décision de faire un commit (validation) ou un rollback (annulation) de la modification.

Phase 1

La première phase du two-phase commit commence par une sous-transaction sur chacun des serveurs concernés par la modification distribuée. Lorsque chaque serveur a terminé sa part de la transaction, il envoie un message au commit server pour lui dire qu'il est prêt à valider cette partie de transaction.

Phase 2

Lorsque tous les serveurs ont indiqué qu'ils étaient d'accord pour valider, le commit server émet un message de confirmation et enregistre la transaction comme ayant été validée en appelant un utilitaire spécial de service du commit. Une fois que cela est fait, la transaction est validée.

Si un des serveur a un problème avant d'envoyer le message, le commit server annule toute l'opération et le notifie aux autres serveurs pour qu'ils annulent leur portion de transaction.

Si un serveur a un problème après avoir envoyé le message, mais avant de valider sa sous-transaction, il saura qu'il n'a pas fait le nécessaire lorsqu'il redémarrera. Il demandera

donc au commit server de lui dire s'il doit valider ou annuler sa partie de transaction distribuée.

2.2.10. Une question de réseaux

Le réseau est l'épine dorsale de l'architecture client/serveur. De ses performances dépendent en grande partie celles de la configuration globale. Il est en effet inutile de disposer d'un puissant serveur de données couplé à un client efficace si l'information circule trop lentement sur le réseau. Il est donc important de choisir un support de communication adapté aux performances des postes serveurs et clients.

Les protocoles jouent également un rôle dans l'optimisation des transmissions. Ils ne doivent pas diminuer outre mesure la capacité du support physique et être suffisamment ouverts pour supporter des conversions rapides en cas de sites hétérogènes. Il vaut mieux , dans la plupart des cas, se baser sur le protocole du serveur que sur celui implanté chez les clients. Le principal problème, en effet, est de faire communiquer des réseaux utilisant des protocoles différents.

2.2.11. Le cahier de charge de l'architecture client/serveur

Les protocoles d'un système distribué ou architecture client/serveur qui doivent être prévus dans le cahier de charges sont :

* Autonomie locale
* L'architecture du système ne doit pas reposer sur un site central unique
* Les opérations doivent s'effectuer en continu
* Indépendance physique des données
* Indépendance logique des données
* Traitement des requêtes réparties
* Two-phase commit et détection des étreintes fatales (deadlocks)
* Indépendance du matériel
* Indépendance du système d'exploitation ; les mêmes SGBD sur différents systèmes d'exploitation
* Indépendance du réseau ; les mêmes SGBD sur différents réseaux
* Indépendance du SGBD ; différentes bases de données dans un même environnement ou dans plusieurs différents

2.2.12. *Quelques produits du marché côté client*

L'univers SQL Server

Le SGBD de Sybase, SQL Server, s'est vu mis en vedette par son portage sous OS/2 et adopté par Microsoft. Il repose entièrement sur le concept d'architecture client/serveur et va même plus loin, en proposant de véritables architectures réparties à base de two-phase commit. SQL Server comporte différents éléments logiciels visant à permettre la communication hétérogène entre clients et serveurs. C'est un SGBD de la nouvelle génération, qui sait gérer des données multimédias et offre des API performantes et très complètes.

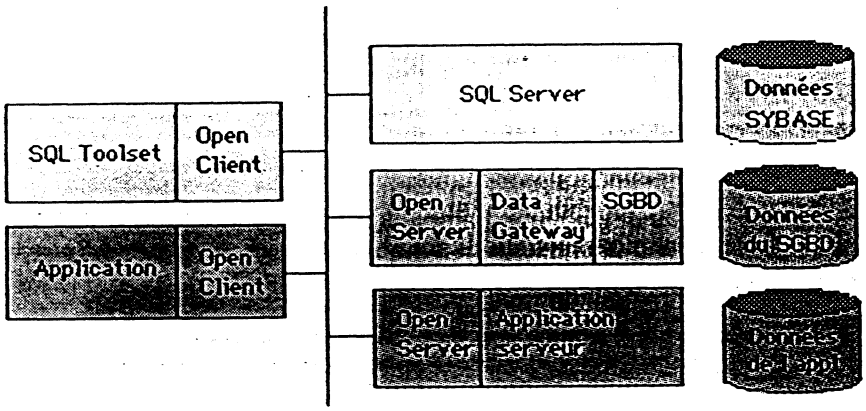


Figure 2-7 : *La solution SYBASE*

L'ensemble client/serveur de Sybase, est composé d'éléments modulaires. Les DB-Libraries servent au développement sur le poste client et une "Net-Libraries" est chargée de gérer la conversion des protocoles réseau. Open Server est une couche serveur qui autorise les connexions multiples, gère la sécurité, permet l'appel remote aux procédures, offre la connexion logique multiple sur un réseau unique et donne accès aux protocoles réseau. Puis Server-Library est une interface pour la création d'applications serveur particulièrement axée sur le **monitoring** des accès multiples en temps réel. Enfin, Data Stream Library est un ensemble de règles utilisé pour le transfert des requêtes et de leurs résultats entre le(s) serveur(s) et le(s) client(s).

Malgré son portage sous OS/2, SQL Server est avant tout un produit dédié à la mini-informatique. De ce fait, il souffre du peu d'ergonomie de ses outils de développement et de présentation. Reste qu'actuellement, dans sa nouvelle version sous Unix, SQL Server est une des réalisations les plus abouties en matière d'architecture client/serveur.

Focus, un cas particulier

Le L4G/SGBD, distribué par Focus, édité par Information Builders, est un des rares systèmes orientés vers le traitement coopératif entre SGBD non relationnels. Bien qu'il ne prétende pas au titre d'architecture client/serveur, il se comporte, dans une large mesure, comme tel. Focus est un L4G multi-environnement qui peut, entre autre, permettre des développements autour de sa propre base, ainsi que sur DB2, SQL/DS, IMS, IDMS et VSAM. Dans sa nouvelle version micro (DOS et OS/2), ce produit permet d'accéder à la fois à des données Focus et non Focus du serveur comme si elles étaient résidentes sur le disque local, prenant en charge toutes les conversions de données, transferts ou extractions de fichiers. Bien que Focus ne soit pas un modèle d'ergonomie (Information Builders annonce cependant que son interface a été considérablement remaniée), sa puissance et sa présence sur de nombreux environnements en font un L4G très prisé.

SQL Windows : fenêtre sur serveur

Conçu à l'origine pour présenter les concepts de SQL System, l'architecture client/serveur de Gupta Technologies, SQL Windows, rencontre un succès grandissant dans de nombreux autres environnements. SQL Windows est un langage de 4e génération performant basé sur SQL, qui s'appuie sur la version monoposte de SQL Base et peut donc travailler en local. Son interface Windows, très soignée, est apte à séduire le plus récalcitrant des utilisateurs de consoles en mode texte. L'utilisateur appréciera la présence d'un générateur automatique de formats, Express SQL, qui permet de créer une application en quelques clics de souris.

En lui-même, SQL Windows est un excellent SGBD/générateur d'applications sur PC. Mais ce n'est pas sa qualité première. Son principal intérêt réside dans ses possibilités de communication avancée avec différents serveurs en environnement hétérogène.

L'ensemble SQL System comporte en effet un module client/serveur, composé de deux couches logiciels à placer sur le client et sur le serveur. Basé sur le principe des API, ce module gère, outre la connexion avec SQL Base multi-utilisateur sur Sun (Unix) ou sous OS/2, différentes passerelles vers DB2, Oracle, SQL Server. L'utilisateur travaillant avec SQL Windows n'a pas à se soucier de l'emplacement des données ni de la manière d'y accéder. La connexion est totalement transparente.

La 4e dimension du monde Macintosh

Le Macintosh, pour sa survie, a dû s'intéresser aux problèmes de connexion hétérogène bien avant le PC. De plus, son interface se prête parfaitement à la productivité personnelle de l'utilisateur. Il n'est donc pas étonnant de constater que les logiciels Macintosh ont pris de nombreuses longueurs d'avance sur leurs homologues PC.

4e Dimension est un fervent partisan de l'architecture client/serveur, il est un SGBD performant basé sur une utilisation intensive du graphisme. Il est possible de travailler avec ce produit sans écrire une seule ligne de programmation. Le langage lui-même est complet, français et proche d'une syntaxe naturelle. Cette recherche de la simplicité se retrouve dans le module de communication. Tous les add-ins de 4e Dimension peuvent travailler avec les données provenant du site central de manière interactive. L'utilisateur peut se créer un environnement de travail adapté à ses besoins, le must pour une application cliente.

Des tableurs face aux SGBD

En matière de connexion hétérogène, Excel fait l'objet d'attentions particulières pour de nombreux développeurs. Alors que Microsoft ajoute à son catalogue Q+E, un module d'accès à l'information sur sites SQL distants, de nombreux autres éditeurs proposent des solutions performantes.

1-2-3, grâce à Datalens (norme SQL intégrée dans le tableur) a pris de l'avance sur son principal concurrent en accédant à Ingres, Oracle, DB2, SQL Server et bien d'autres SGBD SQL. Mais Excel reste dans la course, grâce à l'offre d'éditeurs indépendants de Microsoft.

Ainsi, SQL Vision, de Gupta Technologies, est une commande en langage Excel ou 1-2-3 permettant d'accéder à SQL System et de profiter de toutes ses fonctions, en particulier en matière de connexion avec d'autres sites centraux. Excel possède également des possibilités de travail interactif avec Oracle, grâce au logiciel Second Wind. Ce programme permet de construire des applications utilisant Oracle à partir du langage Excel, plus simple que le C requis d'ordinaire pour ce faire. Bien qu'Excel et 1-2-3 occupent le devant de la scène, il ne faut pas oublier Wingz, conçu dès l'origine pour être client du SGBD Infomix, et ouvert depuis à d'autres environnements. Wingz possède un module optionnel qui ajoute à son langage Hyperscrip toutes les commandes nécessaires au travail interactif avec le serveur.

Ces tableurs, dont les qualités ne sont plus à décrire, sont des outils idéaux en matière de traitement de l'information provenant de sites distants. Leur langage puissant permet l'élaboration d'applications en architecture client/serveur, mais n'offre pas, néanmoins, toutes les possibilités d'un SGBD. Cependant, dans de nombreux cas, ils sauront s'imposer, même face aux ténors des SGBD micros, du fait de leur excellente adaptation, aux besoins de l'utilisateur.

2.2.13. Plusieurs niveaux d'utilisation

Le logiciel client a plusieurs possibilités de dialogue avec le serveur. La plus simple consiste à sélectionner des données pour les consulter et les intégrer dans une simulation, un tableau statistique ou un état récapitulatif. Ces opérations ne sont guère complexes. L'utilisateur de tableur va appeler une macro-commande, celui qui possède un SGBD, une procédure. Les données arrivent dans le fichier, il ne reste plus qu'à les utiliser : calculs, graphes, présentation...

Dans ce domaine, le tableur constitue un client parfait, le tableur organise l'information qu'il reçoit et offre tous les éléments nécessaires au traitement de celle-ci, qu'il s'agisse de totaliser des ventes par secteurs, de grapher des résultats trimestriels ou d'éditer un rapport d'activité.

Les choses se compliquent lorsqu'il y a manipulation de l'information en interactivité avec le serveur. En effet, tableurs et SGBD possèdent de nombreuses différences de structure qui peuvent empêcher certaines actions : le tableur s'étend horizontalement et verticalement, tandis qu'une table de données croît uniquement verticalement. De plus, le SGBD type ses données de manière plus rigide que le tableur. Dans une feuille de calcul, chaque cellule peut avoir son propre type. Dans une table, il y a un type précis par champ.

Enfin, la plupart des SGBD centraux utilisent des index sophistiqués nécessitant l'envoi d'informations très structurées. Chaque enregistrement (**row**) doit posséder une clé unique qui peut porter sur plusieurs champs. Si le typage est approximatif, ou si la séquence de données envoyée est incomplète par rapport à cette clé, la saisie ne sera pas validée. Or le tableur lui-même est incapable d'effectuer ces vérifications en local. L'erreur n'est détectée qu'une fois la transaction effectuée, d'où une perte de temps notable.

En fait, le rôle du client peut être considéré selon deux angles différents, selon la répartition des tâches qui est envisagée. Lorsque les informations sont gérées sur un ou plusieurs sites centraux, les postes micros souhaitent évidemment y accéder et l'utiliser directement dans leurs applications bureautiques courantes. Dans ce cas, le tableur client prend en charge tout ce qui concerne la productivité individuelle et s'adapte à l'architecture client/serveur.

Lorsqu'il est nécessaire de mettre en place des architectures plus complexes, soit pour soulager le site central (**downsizing**), soit pour permettre à plusieurs applications de traiter des données interactivement en liaison avec le serveur (comptabilité, trésorerie, service du personnel travaillent sur les mêmes fichiers mais y accèdent de manière différente) le SGBD micro reprend ses droits, offrant plus de possibilités que le tableur en matière de gestion locale des données et de développement.

2.2.14. Conclusion

Les SGBD micros ne se prêtent pas encore tous parfaitement au rôle de client. Les qualités requises pour être un "bon" client sont une interface soignée, un mode d'utilisation accessible directement sans programmation et la présence d'outils de productivité personnelle (éditeur d'états complexes, fonctions de mise en pages et de présentation...). De plus, si l'on utilise un SGBD micro, c'est pour assurer une cohérence plus grande entre le serveur et le client. Or certains produits ont des structures qui vont totalement à l'encontre de celles des SGBD sur sites centraux : absence partielle ou totale du relationnel, peu de possibilités de typage des données, indigence de l'interface de requêtes...

Ces logiciels, du reste boudés de plus en plus par les utilisateurs, seront progressivement supplantés par des concurrents mieux adaptés aux nouveaux modèles informatiques, comme l'architecture client/serveur.

Après avoir voulu concurrencer la "grosse informatique", le micro revient à ce pour quoi il a été conçu : assister l'utilisateur. Sa puissance sans cesse croissante devrait, dans les années à venir, être consacrée au confort personnel et non à rivaliser avec les minis et mainframes (gros ordinateurs). Les frères ennemis, enfin réconciliés et complémentaires, grâce à l'architecture client/serveur ? Ce n'est pas une vue de l'esprit, mais la nouvelle manière d'envisager le système d'information de l'entreprise.

Pour arriver à ce résultat, les logiciels clients devront évoluer. De nouveau concepts, tels l'EIS (Executive Information System) qui traite l'information provenant du site central d'une manière adaptée à chaque utilisateur et automatiquement, grâce à l'Intelligence Artificielle, voient le jour. Les schémas informatiques tendent à se rapprocher du comportement humain.

En outre, il semble que le concept d'architecture client/serveur soit celui des années 1990, même si beaucoup pensent encore qu'il s'agit plus d'un lifting marketing de produits existants plutôt que de solutions réellement nouvelles.

En fait, l'architecture client/serveur est avant tout une manière d'organiser le système d'information de l'entreprise. On ne peut pas donner un label "client/serveur" à certains produits plutôt qu'à d'autres. Il est possible, en théorie, de mettre en place une telle architecture autour de n'importe quel type de serveur, quels que soint son modèle, son type de fonctionnement, sa plate-forme matérielle, son système d'exploitation ou son origine. C'est là la grande force de ce concept : il respecte l'existant et vise à proposer des solutions plutôt que des outils.

2.3. LE RESEAU NATIONAL DE TELECOMMUNICATIONS POUR LA RECHERCHE : EPINE DORSALE DE L'ARCHITECTURE CLIENT/SERVEUR DE L'IFEN

2.3.1. Introduction

Le 5 Février 1991, un protocole d'accord pour la mise en oeuvre du Réseau National de Télécommunications pour la Recherche a été signé entre le Ministère des Postes, des Télécommunications et de l'Espace, le Ministère de la Recherche et de la Technologie et le Ministère de l'Education Nationale. Cet accord a pour objectif de mettre en oeuvre un Réseau National de Télécommunication pour la Recherche (RENATER) dont la réalisation et l'exploitation sont confiées à France Télécom.

L'un des objectifs du réseau national est de fédérer les structures existantes ou en cours de réalisation tels les réseaux régionaux. Ce rôle fédérateur sur l'ensemble du territoire doit permettre de fournir des services de capacité supérieure, et de bénéficier d'une offre globale de services de la part de l'opérateur public France Télécom. Les échanges nationaux se feront au travers de l'infrastructure nationale et des diverses composantes régionales.

Un comité de pilotage comprenant des représentants du MEN-PTE-MRT a été institué en mai 1990 et conduit l'ensemble des actions liées à l'élaboration de RENATER. Ce comité de pilotage s'appuie sur un comité technique dont le rôle a été de préparer un cahier des charges après analyse des besoins. Ce comité technique a lui-même été décomposé en groupes de travail faisant appel si nécessaire à des spécialistes. Parmi les groupes de travail mis en place on peut citer :

* Groupe X25
* Groupe IP
* Groupe DECNET
* Groupe SNA
* Groupe sécurité
* Groupe liaisons internationales
* Groupe expériences pilotes

Le réseau national de télécommunications pour la recherche doit permettre d'interconnecter les campus et les établissements de recherche pour offrir une infrastructure efficace et évolutive. Ce réseau ne veut pas être un simple remaniement des structures

actuelles. Il se fixe pour objectif de faciliter le travail quotidien des ingénieurs, chercheurs et universitaires, en permettant aux communautés d'utilisateurs d'accroître leur capacité de communication en se déchargeant des contraintes liées à l'exploitation et la maintenance des infrastructures.

2.3.2. Le contexte

Le monde de la Recherche connaît en France un développement rapide et diversifié de ses besoins en communication, en raison de l'arrivée massive de nouveaux outils informatiques puissants dans les laboratoires.

Par ailleurs, les équipements spécialisés, de plus en plus coûteux, utilisés par la communauté scientifique (super-ordinateurs, grandes bases de données...) nécessitent d'être partagés entre tous les utilisateurs. Ils doivent donc être accessibles au moyen de réseaux étendus et performants.

Enfin, les grands projets scientifiques et techniques se mènent aujourd'hui à l'échelle européenne ou mondiale : la navette Hermès, dans le domaine spatial, la recherche médicale sur le SIDA ou les recherches sur l'environnement appellent des structures de collecte et de comparaison des données recueillies dans les laboratoires du monde entier.

Déjà, des réseaux de campus ont vu le jour à l'initiative des utilisateurs, suivis de réseaux locaux et de réseaux disciplinaires.

2.3.3. Un réseau fédérateur et performant

Un Réseau ouvert : le Réseau National de Télécommunications pour la Recherche est naturellement tourné vers l'ensemble des opérateurs de la recherche publique (organismes de recherche et universités).

Il est par ailleurs conçu dans un esprit d'ouverture pour permettre, en conformité avec la Loi sur la Réglementation des Télécommunications, un raccordement de ce réseau aux centres de recherche et de développement des industriels. Ce réseau a vocation à s'intégrer aux réseaux internationaux existants ou à venir.

Un Réseau fédérateur : le Réseau National de Télécommunications pour la Recherche fédérera les réseaux existant sur l'ensemble du territoire et permettra la fourniture de services de capacité supérieure.

Un Réseau structurant : en intensifiant les échanges au sein de la communauté scientifique entre établissements de recherche et universités, le Réseau National de Télécommunications pour la Recherche favorisera le développement de nouvelles méthodes de travail coopératives et intégrées, susceptibles de renforcer l'interdisciplinarité et transdisciplinarité de la recherche française.

En retenant l'opérateur public France Télécom pour la réalisation et l'exploitation de ce réseau, les ministères et les utilisateurs bénéficieront d'une offre globale de services sur l'ensemble du territoire et, par ailleurs, des possibilités d'interconnexions internationales grâce aux étroites relations de coopération entretenues par France Télécom avec ses homologues étrangers. La recherche française se dote ainsi d'un outil primordial pour maintenir son rôle essentiel dans le développement de la connaissance scientifique.

2.3.4. Des réseaux de campus au réseau national

Les organismes de recherche et d'enseignement utilisent depuis dix ans des réseaux de communications internes. La mise en commun de l'exploitation de moyens coûteux (expériences sur les grands appareils scientifiques, superordinateurs, grandes bases de données) et l'extension à l'échelle nationale puis internationale des collaborations sur des actions scientifiques ou techniques entre équipes de divers organismes, ont rendu nécessaire d'interconnecter ces réseaux.

1) Les réseaux de campus :

L'apparition des mini-ordinateurs, adaptés aux besoins et à la taille d'une équipe de recherche, puis, plus récemment, le développement explosif des stations de travail et des micro-ordinateurs, ont profondément modifié les méthodes de travail du chercheur et de l'ingénieur. Leur outil quotidien, permanent, interactif, est désormais une station de travail scientifique, connectée à celle de leurs collègues, ou avec des ordinateurs centraux. Gros calculs, interrogation de grandes bases de données, dépouillement des données issues des expériences, simulation numérique et modélisation : ces besoins conduisent à un développement spectaculaire des communications informatiques.

Les grands centres de recherche et les Universités ont été les premiers à mettre en place des réseaux de campus internes à leurs sites. Les plus grands relient plus d'un millier d'ordinateurs et de stations de travail. Aujourd'hui, tous les sites du monde de la recherche sont équipés de tels réseaux, qui se développent au même rythme que les équipements qu'ils relient (actuellement proche d'un facteur 10 tous les 5 ans).

2) Les réseaux régionaux :

A l'initiative des communautés locales d'utilisateurs, des réseaux régionaux se développent depuis deux ans. Les premiers sont déjà en cours de réalisation par France Télécom : Rhône-Alpes ("GRENET"), Provence, Alpes, Côte d'Azur ("R3T2"), Basse Normandie ("VIKMAN"), Toulouse. Les études avancent rapidement pour les autres. Ces réseaux bénéficient de l'aide des régions.

3) Les réseaux de dimension nationale et internationale existant :

L'organisation et la répartition des activités à l'intérieur des collaborations scientifiques et techniques impliquent des échanges d'informations constants à l'échelle française et européenne au moins. Le courrier électronique est l'outil essentiel de coordination. L'utilisation à distance d'équipements lourds, de grands appareils expérimentaux, de grandes bases de données, de superordinateurs, génère en permanence des flux de données informatiques croissants à travers la France et l'Europe.

a) Les réseaux des organismes de recherche

Ces besoins ont conduit à la mise en place, par les grands organismes ou par les équipes de recherche elles-mêmes, de réseaux supportant des services nationaux tels que le réseau des Universités ("EDUNET"), le réseau de la Physique Nucléaire du CNRS, celui du CNES, et celui du CNET...

b) Les services spécifiques

D'autres services spécifiques existent. Ils sont, en fait, des sous-ensembles de réseaux plus vastes, mis en place par des communautés d'utilisateurs de dimension européenne ou mondiale : service EARN/BINET de messagerie et de transfert de fichiers ; services EASINET reliant des superordinateurs IBM ; service FNET/EUNET reliant des ordinateurs et stations de travail Unix ; service HEPNET de la Physique des Hautes Energies ; service SPAN des sciences de l'espace, et d'autres encore...

4) Les insuffisances du système actuel :

Si elle illustre bien l'importance des besoins de communication dans tous les secteurs du développement technologique, de la recherche et de l'éducation, cette floraison de services pose plusieurs problèmes : performances limitées, duplication des efforts et des moyens, donc coûts élevés par rapport aux performances, peu ou pas d'intercommunication. Cela freine les actions interdisciplinaires et les transferts entre recherche et technologie.

Par ailleurs, ces services ne s'appuient pas sur une infrastructure assez puissante. Ils ne permettent pas le traitement des besoins des communications liés aux développements actuels de la simulation numérique, de l'imagerie, ni ceux liés à la nouvelle génération de très grands équipements scientifiques (LEP, COLUMBUS...).

5) Le réseau national :

Le Réseau National de Télécommunications pour la Recherche, à travers ses diverses composantes régionales, permettra d'interconnecter les campus et les établissements de recherche. Il offrira ainsi aux réseaux existants une infrastructure efficace, et qui leur permettra d'évoluer.

Ce réseau permettra d'améliorer les communications liées à la recherche et au développement technologique. Il préfigurera les moyens de communication de demain pour l'industrie et les entreprises.

Il reposera sur les infrastructures de télécommunication et les techniques de transport de données informatiques les plus avancées. Ceci lui permettra d'intégrer les flux de communication de divers types d'utilisateurs.

Fournissant un service cohérent et homogène, le réseau sera susceptible d'évoluer plus facilement vers des performances supérieures.

Il permettra l'accès aux autres réseaux de la recherche fonctionnant dans les pays d'Europe, et aux Etats-Unis, depuis quelques années : DFN en Allemagne, JANET en Angleterre, GARR en Italie, NSFNET aux Etats-Unis.

2.3.5. l'architecture du réseau national de télécommunications pour la recherche

Le réseau national reposera sur une architecture de réseau structurée en trois niveaux:

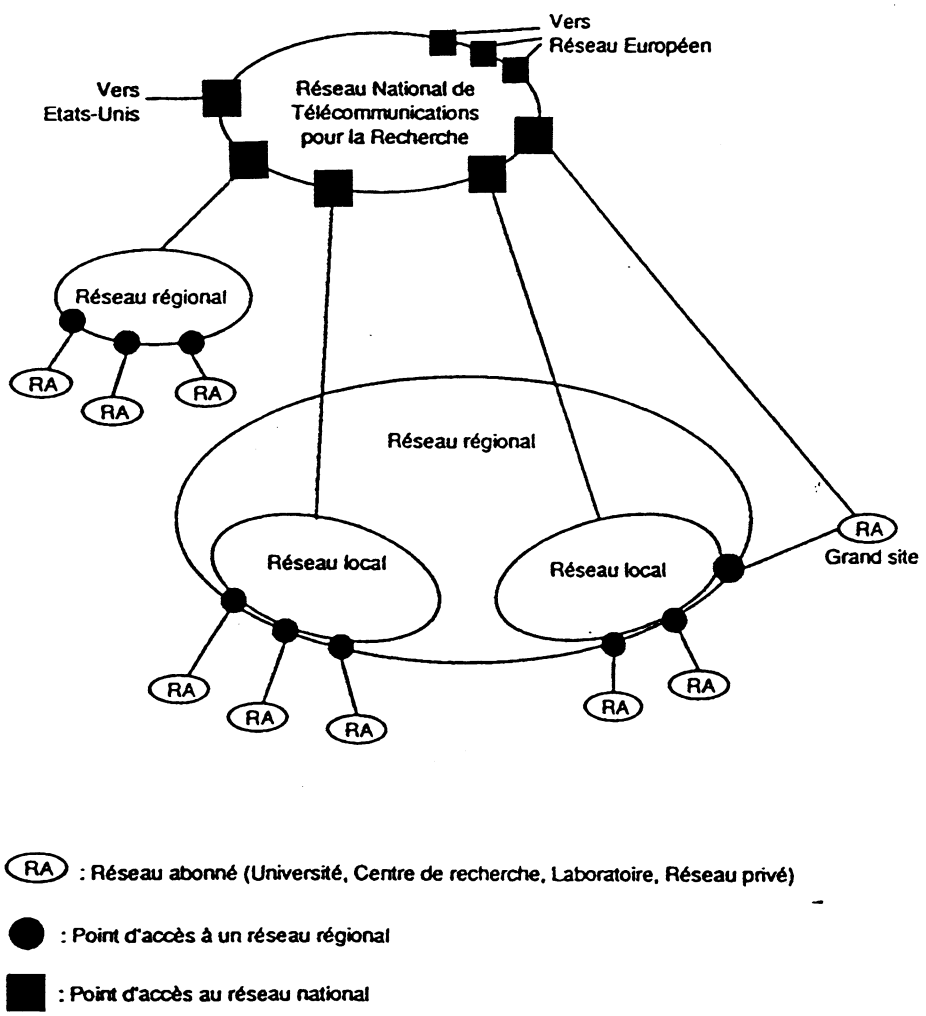


Figure 2-8 : *Architecture du réseau national de télécommunication pour la recherche*

- **1er niveau :** les 200 réseaux de campus universitaires et des établissements de recherche (15000 ordinateurs et stations de travail scientifiques y sont déjà raccordés), dont le rôle est d'assurer les communications locales ;

- **2e niveau :** les réseaux régionaux ou métropolitains en cours de réalisation et auxquels se raccordent les réseaux de campus ou centres de recherche ;

- **3e niveau :** le Réseau National, qui fédérera les réseaux des deux premiers niveaux et leur offrira l'accès aux réseaux de la recherche à l'étranger. Quelques très grands campus, ainsi que des ressources communes d'intérêt national (très grands équipements de recherche, centres de documentation, bases de données), y seront directement connectés. Le Réseau National de Télécommunications pour la Recherche constituera ainsi une gigantesque banque de données et un moyen de calcul unique proposés à la recherche en France.

2.3.6. Les services disponibles sur ce réseau

L'installation, le fonctionnement, l'évolution à moyen et long terme du Réseau National devront atteindre les objectifs suivants :

- offrir aux réseaux préexistants des services de communication ;
- assurer une capacité de communication satisfaisante avec les autre grands réseaux et sites, européens et américains ;
- obtenir dès le démarrage, et préserver sur le long terme, une haute qualité de service, dans un contexte de croissance rapide du trafic ;
- encourager les nouveaux services issus de l'essor des communications à haut débit d'images et des secteurs multimédia.

Le Réseau National sera un outil de travail quotidien pour les ingénieurs, les chercheurs et les universitaires. Il permettra de faciliter les coopérations ainsi que les transferts de connaissances entre les personnes et les laboratoires.

Dès sa mise en exploitation, les principaux services apportés par ce réseau seront :

- le courrier électronique, national et international ; le réseau offrira l'accès aux services de messagerie existants, ainsi qu'aux services de sécurité et de protection des données ;
- l'accès interactif à des ressources d'intérêt général situées sur d'autres sites que celui de l'utilisateur : grandes bases de données scientifiques et techniques, médicales, grandes puissances de calcul, etc... ; les possibilités de consultation fournies par le réseau inciteront à l'extension des bases existantes et au développement de nouvelles bases spécialisées, environnementales, juridiques, bibliographiques, patrimoine, ... ;
- les transferts de fichiers, de documents graphiques, etc... ;

- les passerelles et connexions vers d'autres réseaux de la recherche et vers les réseaux publics ;
- les services de surveillance du trafic et des incidents ;
- les services d'information et d'aide aux utilisateurs.

Le réseau permettra aux communautés d'utilisateurs d'améliorer leur communication en les déchargeant des contraintes liées à l'exploitation et à la maintenance d'une infrastructure spécifique ; il leur permettra de consacrer leurs efforts aux applications spécifiques à leurs activités.

Par la suite, les capacités du réseau devront suivre le rythme de l'extension des possibilités des stations de travail et du développement de nouvelles applications, parmi lesquelles figureront :

- Le transfert d'informations en mode multimédia ;
- les communications de groupe : visioconférence, télé-enseignement ;
- les nouvelles applications basées sur le développement du travail coopératif et intégré, la puissance de calcul et les données étant réparties sur le réseau.

La mise en place de ce réseau ouvre à la recherche et à l'industrie informatique françaises, ainsi qu'à France Télécom, tout un domaine d'études, de réalisation et commercialisation nouveau dans ces directions, et dans bien d'autres.

2.3.7. Le cahier des charges de Renater :

RENATER ayant pour vocation d'être un réseau ouvert, offrant une large interconnectivité, **les protocoles** de communications prévus dans le cahier des charges sont :

* Protocole X25 (évolution prévue vers le niveau 88)
* Protocole IP (évolution prévue vers IP/ISO)
* Protocole DECNET phase IV (évolution prévue vers la phase V)

Pour chacun de ces protocoles un service de niveau 3 (selon le modèle ISO) est demandé à l'opérateur.

Pour les liaisons n'entrant pas dans ce cadre (SNA, Hyper Canal, etc.) ou ayant des contraintes particulières de sécurité, un service indépendant du type de protocole est demandé :

* Service de liaisons permanentes point à point (L3P).

Dans ce cas, les informations sont véhiculées en mode transparent de bout en bout au moyen de l'infrastructure du réseau de la Recherche.

Les classes de débits demandées sont :
* un service opérationnel jusqu'à 2 MB
* un service expérimental à 34 MB

Ces classes de débits sont celles envisagées au démarrage du réseau. Après une période de 18 mois des accès opérationnels à 34 MB doivent être fournis.

A titre indicatif voici **le calendrier** de mise en oeuvre envisagé :

Fin 91 : prototype sur 6 sites à quelques MB,
début 92 : interconnexion de réseaux régionaux, prototype à 34 MB,
fin 92 : déploiement du réseau et qualité de service garantie,
début 93 : généralisation du service 34 MB.

Ceci est un calendrier prévisionnel soumis à d'éventuelles modifications.

2.3.8. Les perspectives d'évolution

La cible initiale du réseau est constituée de 15000 ordinateurs et stations de travail actuellement utilisés dans le monde de la recherche en France.

La croissance des échanges s'annonce dès aujourd'hui vertigineuse.

D'une part, l'augmentation de la puissance des ordinateurs et des stations de travail va continuer à son rythme actuel entraînant une croissance analogue dans le volume des informations à échanger. Un super ordinateur réalisait, il y a cinq ans, 2 milliards d'opérations par seconde ; la génération actuelle permet 20 milliards d'opérations par seconde.

D'autre part, on peut penser qu'avec la banalisation de l'informatique liée à la baisse des coûts, de nouvelles catégories d'utilisateurs souhaiteront bénéficier de ce service.

Pour maintenir un service de qualité, cette augmentation doit être anticipée et l'évolution du réseau préparée, ce qui suppose un travail technique considérable, se situant dans le domaine de la technologie des réseaux informatiques et des télécommunications.

Une politique d'actions pilotes sera développée : elles définissent les études à engager et les expérimentations permettant de valider les futures fonctionnalités du service.

Par cette approche, qui associera, en partenariat, France Télécom aux utilisateurs et aux industriels, la communauté scientifique dans son ensemble pourra bénéficier des derniers développements technologiques.

2.3.9. Des enjeux pour la recherche française

La mise en place du réseau national de télécommunications pour la recherche répond à trois enjeux.

1) Un enjeu d'organisation :

Outil de communication, le réseau national permettra à l'ensemble de la communauté scientifique d'accéder aux ressources de calcul, aux banques de données, aux bibliothèques informatisées, dont elle a besoin en temps réel dans son travail quotidien. En rapprochant les hommes il contribuera à tisser entre eux de nouveaux liens, fédérant une communauté encore trop dispersée. L'étude et la mise en place, dans le courant de l'année 1991, d'une structure représentative de l'ensemble des utilisateurs parachevera cet effort de structuration.

Le raccordement au réseau des différents laboratoires permettra de développer le travail coopératif entre équipes et influera sur les méthodes de travail, en offrant la possibilité à chaque instant de communiquer ou de modifier à distance un projet en cours d'exécution.

Dédié à la recherche et au développement technologique, le réseau contribuera au rapprochement déjà engagé entre la recherche publique exécutée dans les établissements publics ou les universités et la recherche privée conduite dans les laboratoires des entreprises.

2) Un enjeu scientifique :

L'accès à distance, rendu possible par le réseau aux ressources de calcul que représentent les centres équipés de super calculateurs, encouragera le recours à la modélisation numérique dans toute une série de disciplines permettant de faire l'économie d'expérimentations souvent plus coûteuses. Déjà fortement développée en physique des plasmas, en mécanique des fluides ou en astronomie, cette démarche devrait se renforcer dans les sciences de la vie, et notamment en climatologie.

La possibilité accrue de transfert de fichiers, de dialogue interactif entre stations de travail, et le développement des techniques de visualisation permettront d'agir en direct et plus rapidement sur le déroulement de programmes d'observation ou de manipulation, en disposant d'informations sous une forme plus synthétiques (images, tableaux, graphiques) qui assurent une meilleure représentation, donc une meilleure exploitation des résultats.

Ainsi, dans le domaine de l'observation spatiale, les données transmises par les sondes et les satellites seront-elles plus rapidement interprétables par les équipes dans les laboratoires.

L'accès rapide à des banques de données ou des bibliothèques d'informations permettra une meilleure diffusion et une meilleure exploitation du travail scientifique. Dans certaines disciplines comme la biologie ou les sciences de la vie, la consultation en temps réel de banques de données résultant du travail des équipes est indispensable.

3) Un enjeu technologique :

La demande spécifique de la communauté scientifique en matière de réseau suppose pour être satisfaite de faire appel aux technologies les plus récentes en matière de télécommunications, du triple point de vue des protocoles, des normes et des débits que celui-ci doit offrir. C'est donc l'occasion, pour l'opérateur public France Télécom, de faire évoluer son offre technologique dans le haut de sa gamme, le situant au mieux sur un marché très concurrentiel au niveau mondial.

La mise en place du réseau, ainsi que les services que celui-ci va offrir, contribueront à accélérer la mutation technologique en cours vers une **informatique mieux répartie**, par le biais de stations de travail de plus en plus puissantes et du développement de grosses machines conçues pour des applications spécifiques (calculateurs vectoriels, machines à architecture massivement parallèle).

2.3.10. Le réseau national de télécommunications pour la recherche et la construction de l'Europe

Grâce à ce nouveau réseau, la France se situera, à la fin de la décennie, au premier rang des pays équipés d'un réseau tout optique, supportant les services les plus modernes.

Pour parvenir à cet objectif, l'équipement en fibre optique du réseau interurbain national de France Télécom a été décidé, avec comme objectif qu'il soit entièrement optique à la fin du siècle.

La technologie optique permettra en effet à un réseau de supporter des services numériques à très haut débit. Le réseau de télécommunications pour la recherche préfigure ce type de services sur les infrastructures de France Télécom.

Mais le réseau de télécommunications pour la recherche a aussi pour but de constituer le maillon français d'un réseau européen harmonisé de la recherche. Il a donc vocation à être ouvert sur d'autres réseaux comparables qui peuvent exister au niveau international.

La création de tels réseaux harmonisés à l'échelle européenne a été proposée par la France lorsqu'elle assurait la Présidence du Conseil des ministres européens, à la fin de l'année 1989.

C'est en cohérence avec cette proposition, et dans la perspective de la concrétiser rapidement, que la France s'attache maintenant à créer le réseau de télécommunications pour la recherche.

La France agit dans le même sens en participant, au niveau européen, à l'interconnexion des infrastructures RNIS des différents pays. Ceci a été fait, par exemple, en décembre 1990 avec l'Allemagne entre les deux opérateurs nationaux de télécommunications, France Télécom et Deutsche Telekom.

2.4. NECESSITE DE DEFINIR LA QUALITE DES OUTILS DU GENIE LOGICIEL POUR REPONDRE AUX BESOINS DE DEVELOPPEMENT D'APPLICATIONS A L'IFEN

2.4.1. Introduction sur le génie logiciel

Lorsqu'on développe des applications informatiques importantes, on rencontre souvent d'énormes problèmes. Certains d'entre eux se retrouvent dans tout projet relevant des sciences de l'ingénieur : gestion de projet, contrôle des coûts et des délais, gestion du personnel, choix des outils, contrôle de la qualité..., ce qui explique l'introduction du terme "génie logiciel" pour décrire ce type d'activité.

Le génie logiciel est une discipline qui repose principalement sur l'informatique, il fait aussi appel à des disciplines de base comme la logique, les mathématiques, la psychologie et à d'autres disciplines comme l'ergonomie et la gestion.

Un ingénieur informaticien doit être capable d'évaluer l'efficacité de ses outils informatiques et de les mettre en oeuvre dans un projet donné. Il doit aussi comprendre les problèmes qui se posent aux utilisateurs dans leur relation avec un système dont ils ne connaissent pas l'organisation interne. Enfin, il doit traiter les problèmes de gestion de projet associés à la production de logiciel.

Grâce à la diminution des coûts du matériel informatique, due à l'évolution de la technologie de semi-conducteurs, les ordinateurs ont pu être associés à la réalisation d'un nombre de plus en plus important de produits. De plus, des applications évoluées de l'informatique, telles que les bases de données avancées, l'intelligence artificielle, les systèmes d'aide à la décision, sont devenues économiquement viables.

Malheureusement, les coûts des logiciels n'ont pas diminué autant que ceux du matériel. En fait, ils ont même accusé une forte croissance. Dans beaucoup de systèmes, le coût du logiciel représente aujourd'hui 80% ou plus du coût total. La pratique du génie logiciel a pour objectif de diminuer sensiblement les coûts (et les délais) de développement des applications informatiques. Une amélioration, même faible, de la productivité dans ce domaine peut donc entraîner une diminution significative des coûts de développement.

En outre, le terme "logiciel" ne désigne pas que les programmes associés à telle application ou tel produit. Il rentre dans une approche systémique globale du système

d'information de l'entreprise, dans lequel est inclus le système informatique où nous pouvons définir : le cycle de vie du logiciel (analyse et définition des besoins, conception, réalisation, tests, exploitation et maintenance), les méthode techniques et outils de conception (programmation structurée, modèle de données, modèle de traitement) et l'évolution vers l'intégration des systèmes (architecture client/serveur).

Par ailleurs, il ne faut pas oublier que le génie logiciel étant une activité cognitive, il est normal que les aspects relationnels avec les utilisateurs et les équipes de développement doivent être pris en compte pour éviter de fixer des objectifs irréalisables.

2.4.2. Répondre aux besoins applicatifs

1) La démarche proposée :

L'expérience des travaux réalisés en commun avec différentes Unités de Recherches et Laboratoires dans le cadre de la conception et la construction du système d'informations ECORDRE, nous autorise à présenter une démarche pour la mise en oeuvre rapide d'une solution opérationnelle au niveau du développement des applications informatiques.

Cette démarche est fondée sur les principes suivants :

a) Pérenniser les investissements à forte valeur ajoutée

L'investissement consenti pour la mise en place de méthodes doit porter ses fruits et aboutir à des réalisations concrètes.

La démarche et les outils à mettre en oeuvre doivent s'inscrire donc dans un cadre scientifique et méthodologique et ne doivent pas impacter toute décision de l'Entreprise visant à garder ou à modifier la définition de son organisation.

Par ailleurs les méthodes en vigueur chez les différents partenaires de l'IFEN (et l'existant) doivent être prises en compte, sans aucune pression pour être modifiées ou remplacées.

b) Préparer l'avenir

La qualité de ses applications informatiques est la seule composante totalement stratégique pour l'IFEN et ses partenaires. Les méthodes de développement et les outils associés, adoptés depuis de nombreuses années par des utilisateurs dans le cadre écologique, agronomique et environnemental doivent être étudiés et pris en compte pour le maintien d'un bon rapport stratégique avec l'environnement global de l'Entreprise.

Les méthodes de développement doivent être avant tout fondées sur une architecture ouverte, dans laquelle des outils complémentaires et communicants doivent permettre une couverture aussi complète que possible du cycle de vie des projets.

La stratégie informatique doit être considérée comme un moyen, un support, à la stratégie globale définie par l'IFEN dans le cadre du traitement de l'information environnementale.

Les outils sont des moyens tactiques qui doivent permettre à l'Entreprise d'utiliser au maximum les éléments conceptuels globaux, au service de sa stratégie.

Les outils proposés doivent permettre d'avoir à la disposition de l'IFEN la meilleure technologie pour progressivement adopter les moyens correspondant à l'architecture client/serveur.

c) Alléger tout de suite la charge de réalisation

Tout d'abord l'IFEN doit mettre en oeuvre des outils du génie logiciel pour répondre tout de suite à des contraintes applicatives au niveau informatique.

Ces outils doivent permettre de très rapidement démarrer la réalisation des projets les plus prioritaires avec des temps d'apprentissage et de mise en oeuvre très courts grâce à un plan de formation et d'assistance technique adapté aux besoins.

Ces outils doivent être d'une grande simplicité de mise en oeuvre et d'une grande efficacité permettant aux différentes équipes travaillant dans des projets au niveau environnementales d'atteindre très vite un niveau de productivité élevé, en contribuant ainsi à alléger la charge de développement des applications.

Les outils de génie logiciel à utiliser doivent offrir les meilleures garanties au niveau de la qualité de leurs références (part du marché mondial) et de la nature des projets réalisés (références dans d'autres entreprises).

2) Répondre aux besoins de développement

a) Diminuer le temps de développement des systèmes

L'utilisation d'un Atelier de Génie Logiciel (AGL) avec une approche non procédurale de haut niveau doit permettre de simplifier considérablement le développement des applications, et prouver sa capacité à livrer des systèmes complets en moitié moins de temps qu'avec des techniques conventionnelles.

Ce niveau de productivité est possible par l'intégration dans un AGL de la technologie adressant la totalité du cycle de vie des applications, et en libérant le développeur de toutes les contraintes liées à l'environnement technique.

b) Apporter une vraie solution de génie logiciel

Les composants d'un AGL doivent aller de l'Analyse des applications, (en utilisant les outils du choix de l'utilisateur) en passant par le prototypage, la génération du code, jusqu'aux tests unitaires et tests d'intégration, permettant d'arriver à une solution aussi complète que possible en matière de développement d'applications.

En aval, lorsque le code a été généré et testé, l'AGL doit proposer une solution permettant de résoudre les problèmes de gestion de source et de load-modules et répondre aux besoins d'automatisation de la mise en production des applications.

Cet ensemble d'outils qui va de l'analyse jusqu'à la maintenance, constitue une solution permettant une couverture aussi complète que possible du cycle de vie des systèmes d'information, en assurant la continuité et le contrôle qualité à tous les niveaux.

c) Accroître la qualité des systèmes

En plus de l'apport qualitatif induit par les outils d'analyse, principalement situé au niveau de la cohérence de l'ensemble, les facilités de prototypage proposées par un AGL doivent permettre aux analystes de présenter rapidement à l'utilisateur final un prototype complet du futur système.

Ceci doit être réalisé sans génération de code, donc sans compilation ni link-edit.

De ce fait, toute remarque ou suggestion émanant de l'utilisateur final peut être prise en compte immédiatement. Toute erreur détectée à ce niveau devra être corrigée sans aucun effort, car très tôt dans le cycle de vie ; la qualité de la communication entre professionnels de l'informatique et utilisateurs finals doit être considérablement améliorée.

Enfin, l'AGL à utiliser doit générer automatiquement des procédures de contrôle d'erreur et les aides en ligne (HELP-HOLD)... Ces fonctionnalités sont habituellement négligées par les éditeurs car elles sont coûteuses et difficiles à réaliser ; l'AGL à utiliser devra permettre de les mettre en oeuvre simplement, augmentant en cela la satisfaction des développeurs et des utilisateurs finals.

d) Valoriser l'utilisation des SGBD et de la technologie SQL

L'AGL à utiliser doit faciliter la reprise de l'existant. En effet, l'AGL doit être interfacé avec les catalogues des SGBD (au moins du SGBD principal qui doit gérer la Banque de Données) et récupère dynamiquement les définitions de données, telles qu'elles sont décrites par les administrateurs des bases de données.

L'AGL doit générer automatiquement les ordres d'accès aux bases de données et offrir aux développeurs la possibilité de les visualiser, et éventuellement de les personnaliser.

La génération automatique des accès non procédurale et paramétrable aux bases de données, doit très largement contribuer au succès de l'utilisation d'un AGL à l'IFEN et chez ses partenaires, et à éliminer les contraintes associées à la codification d'ordres complexes et aide à la standardisation des applications.

e) Libérer la créativité, accroître les fonctionnalités des systèmes

L'adoption d'un AGL comme outil standard pour le développement des applications par l'IFEN est possible actuellement grâce à l'offre de divers éditeurs proposant des systèmes ouverts permettant la personnalisation des applications, où des modules (des routines) existants peuvent être insérés au niveau des outils de "design".

Ces modules sont placés aux endroits corrects de la structure lors de la génération du code par le générateur d'application. Ces modules spécifiques sont considérés comme des spécifications liées au design de l'application : ils font partie d'une logique.

L'AGL doit permettre d'écrire des modules particuliers, que ceux-ci soit attachés à un programme, ou au contraire réutilisables par tout système d'information. Permettant de libérer le développeur des contraintes techniques, celui-ci peut se concentrer sur ce qui est vraiment important : la logique et la qualité de son application.

La simplicité d'utilisation doit être un atout important d'un AGL, il doit posséder une grande richesse de fonctionnalités autorisant la construction des applications très élaborées qui auraient souvent dû être simplifiées compte tenu de leur coût de réalisation avec des techniques habituelles.

2.4.3. Répondre aux besoins de l'IFEN et de ses partenaires en matière de maintenance

1) Répondre aux besoins de maintenance de l'Entreprise

La principale spécificité des Entreprises est la vitesse à laquelle leurs applications doivent se développer et évoluer. La réactivité et la productivité des équipes doivent alors être optimum.

a) <u>Libérer des ressources pour le développement</u>

Ceci ne peut être réalisé que si l'on sait dégager des ressources mobilisées sur la maintenance, en réduisant le coût de cette tâche.

La productivité moyenne en maintenance d'applications "online" produites avec un AGL doit être de l'ordre de 3 pour 1, par rapport aux techniques habituelles.

La maintenance des applications étant réalisée à partir des spécifications de haut niveau, et non plus au niveau du code, le développeur peut très rapidement prendre en compte une demande de modifications et, dès lors, contribuer à réduire le retard de maintenance (ou tout du moins à ne pas l'aggraver !).

b) Gérer la mesure d'impact des changements et les planifier

L'AGL à utiliser devra générer de nombreux états détaillant l'emploi des données élémentaires, tant au niveau des applications, des états et des écrans, qu'au niveau des fichiers ou des bases de données.

Une demande de modification pourra alors être anticipée et planifiée puisque ses conséquences en seront connues a priori.

La fonction maintenance peut dès lors être gérée, et pas seulement subie.

c) Bénéficier d'un code de grande qualité

Le code source produit doit être de très grande qualité, toujours structuré de la même manière, très documenté, et tous les paragraphes doivent être clairement décrits et commentés.

La qualité du code produit par un AGL doit éliminer tous risques de non pérennité de la solution, ceci constitue souvent l'un des aspects les plus appréciés par les professionnels de la programmation.

d) Réduire les coûts de maintenance sur l'existant

L'AGL à utiliser doit gérer des programmes modulaires et autonomes, permettant la construction des nouveaux modules, des extensions ou des sous-systèmes sur des application existantes.

De fait, la plupart des professionnels de la programmation considèrent qu'il est beaucoup plus économique de réécrire des modules ou des applications avec un AGL lorsqu'ils requièrent des adaptations fréquentes. Ils reconnaissent tous que l'investissement préalablement consenti pour la réécriture est largement compensé par les gains ultérieurs obtenus en maintenance et en évolution d'applications avec ce type d'outil.

e) Considérer la maintenance comme du développement

La maintenance n'est plus à considérer comme la correction d'anomalies dans les applications. Avec un AGL on peut participer au contrôle qualité, en évitant de générer des informations non conformes aux spécifications validées.

La maintenance est bien plus le processus qui consiste à faire évoluer les applications en fonction des modifications imposées par les utilisateurs, que la correction d'erreurs de codification.

La plate-forme de développement d'un AGL doit permettre de reporter les modifications demandées au niveau le plus approprié, soit au niveau analyse, soit au niveau du prototypage jusqu'aux tests d'intégration, en fonction du type d'intervention demandée.

2.4.4. Répondre aux exigences en matière de qualité et de standardisation

1) Répondre aux contraintes de qualité et de standards

a) <u>Proposer des méthodes standardisées pour le développement et la maintenance</u>

L'AGL à utiliser doit proposer une démarche standard permettant de normaliser les étapes, pour construire et utiliser une application.

Il doit compter avec une approche guidée et paramétrable, présentée sous forme de panels, permettant aux chefs de projets de mieux maîtriser le contrôle d'avancement d'un projet, et d'identifier précisément la charge du travail résiduelle.

b) <u>Intégrer l'existant</u>

L'intégration avec l'existant doit être réalisée via les zones de transfert, ou de tout autre processus en service sur le site. Il ne doit y avoir aucune caractéristique limitative dans le source généré par l'AGL.

Les programmes source des applications doivent être répertoriés, soit sous la forme originelle des spécifications de l'AGL, soit sous la forme du source généré ne demandant aucun runtime ou module de l'AGL pour s'exécuter.

Il doit requérir très peu de temps pour obtenir la productivité recherchée en faisant appel aux compétences existantes des équipes en matière de programmation.

Les outils habituels de stockage et de gestion de bibliothèques doivent être utilisés sans aucun problème, de même que les différents outils de sécurité et de contrôle en service sur le site.

c) Garantie d'obtenir des applications efficaces en production

Les programmes générés par l'AGL doivent être réputés fonctionner aussi bien (voire mieux) que les programmes écrits à la main. Leur qualité doit être équivalente à celle des programmes produits par un bon spécialiste en programmation structurée.

Au fil des années, les applications développées à la main se dégradent, à cause des innombrables "verrues" rajoutées à l'occasion des opérations de maintenance.

L'AGL à utiliser qui régénère les applications doit fournir de nombreuses options afin d'obtenir la meilleure efficacité possible, permettant d'éliminer le problème de rajout constant de modifications sur des systèmes existants.

d) Respecter les procédures de sécurité

L'AGL doit être totalement assujetti aux procédures de sécurité mises en oeuvre sur le site.

La partie réservée à l'administration des données sous l'AGL doit fournir des contrôles plus restrictifs si l'utilisateur le souhaite, et affiner ainsi le système de contrôle et de gestion des données.

e) Supporter les standards de l'entreprise au niveau analyse/conception

L'architecture ouverte de l'AGL à utiliser doit autoriser une grande liberté dans le choix des stations d'analyse et de conception, permettant de respecter les choix de l'Entreprise et de ses partenaires dans ce domaine.

2.4.5. Répondre aux besoins en matière de développement d'applications dans le cadre d'une architecture client/serveur

Cette architecture doit constituer l'axe stratégique des années '90 pour l'IFEN et ses partenaires.

L'AGL à utiliser doit intégrer dans ses solutions les fondements de cette architecture que sont CUA, le traitement coopératif, le support des protocoles évolués (LU6.2), le Repository, et sur différents plate-formes : MS/DOS, OS2, UNIX, MVS/ESA,...

1) Le développement sur le serveur

a) Rappel des besoins

L'AGL devra être capable de générer une transaction aveugle ; c'est-à-dire :

- Orientée gestion de messages IN et OUT.
- Détachée de la gestion des terminaux.
- Pas de gestion de données de type "Transaction Conversationnelle".
- Pas de gestion des attributs de champs et messages d'erreur.
- Pas de limite de la longueur des messages
- Maintien de fonctionnalités, associées non plus à un champ d'écran mais à un élément de structure de message.
- Exits CUSTOM-CODE juste après l'acquisition du message IN et avant l'envoi du message OUT.

b) Architecture fonctionnelle du SERVEUR

Le SERVEUR d'informations développé dans la philosophie Batch de l'AGL doit comporter deux couches fondamentales :

- Le gestionnaire des conversations.
- Les routines de service.

Le gestionnaire des conversations

. Contrôle du protocole de conversation
  * Séquence des opérations RECEIVE, SEND.

. Acquisition des demandes de service (message)
  * Réception
  * Validation

. Activation de la routine de service

* Mise en forme des éléments de la demande de service.
* Appel dynamique de la procédure de service.
* Mise en forme des éléments de la réponse.

. Restitution de la réponse de service

* Mise en place des informations de type préfixe de message.
* Envoi du message.

. Prise en charge de toutes les anomalies de fonctionnement des procédures de service.

* Restitution d'une réponse compatible avec celle attendue par le CLIENT,

décrivant succinctement la raison de non exécution de sa demande de service.

Les routines de services

Elles contiennent les diverses règles de gestion attachées à un service donné (Exp : Accès à un contrat) et sont indépendantes des informations de contrôle contenues dans les messages.

Elles sont capables d'assurer la navigation et/ou l'accès aux données de façon transparente à la couche de gestion des conversations, et ce, afin de rendre plus aisée tout processus de modification du protocole de communication mis en oeuvre.

2) Le développement sur le poste client

a) <u>Description générale</u>

Les années 90 vont se caractériser par l'utilisation de plus en plus fréquente d'environnements graphiques sophistiqués tels que Présentation Manager, Windows, Nextstep, Motif, Macintoch, etc...

Les grands consommateurs en informatique classique vont devoir fournir à leurs utilisateurs des applications s'intégrant bien dans ces environnements.

Une partie importante de ces applications sera développée par les équipes de développement de l'utilisateur.

b) <u>Face à ces besoins la solution à proposer doit permettre de :</u>

Simplifier considérablement le développement d'applications dans ces nouveaux environnements graphiques.

Construire une application graphique d'une manière simple et intuitive en respectant les connaissances de l'utilisateur, en particulier en matière de langage de troisième génération.

Supporter en plus du développement d'applications graphiques, l'accès aux données (DB) et aux couches de communication (DC).

Développer les applications dans une architecture moderne de type client/serveur.

2.5. L'HYPERTEXTE ET LA PROGRAMMATION ORIENTEE OBJETS : DES OUTILS NECESSAIRES POUR LA CREATION DES SYSTEMES D'EXPLICITATION ET D'AIDE A LA DECISION

Nous présentons conjointement l'importance des logiciels hypertexte et la programmation orientée objets qui en est la logique naturelle.

2.5.1. L'hypertexte

L'hypertexte est une technique d'organisation et de gestion de l'information liée au développement récent des interfaces utilisateurs graphiques (par exemple l'environnement Windows), il s'agit en effet de rendre l'information aussi accessible que possible.

L'intérêt des techniques d'hypertexte en environnement est de donner la possibilité aux utilisateurs d'un système d'explicitation ou d'aide à la décision de se déplacer rapidement dans celui-ci en s'affranchissant de la nature séquentielle des opérations propres aux applications classiques en arborescences. (A la différence des menus classiques, c'est l'utilisation qui guide la structure du "menu").

Chaque utilisateurs d'un hypertexte se déplace dans le document composé de texte et de dessins selon ses propres désirs, et accède beaucoup plus rapidement à l'information qu'il recherche. Ceci est indispensable quand on a conscience de l'extrême diversité et complexité des connaissances environnementales. On peut organiser cette information sous forme conversationnelle, l'utilisateur étant mis à contribution éventuellement pour introduire des données complémentaires et déclencher au choix le fonctionnement de "modèles" décrits dans l'hypertexte (procédure, heuristique) avec ses algorithmes. Toute une flexibilité est ainsi organisée à partir de connaissances antérieures décrites initialement en modèles, gérée en base de données, ou introduites manuellement à la demande avec l'aide de menus conversationnels. Il y a possibilité de hiérarchiser l'accès pour atteindre les "niveaux de perception" souhaités en explorant si nécessaire un approfondissement des zones sensibles.

La technique hypertexte permet dans une application d'utiliser des zones sensibles ou des "boutons" qui, lorsque l'on "clique" dessus, renvoient à d'autres informations.

Lorsque par exemple un document hypertexte est affiché dans une fenêtre, on distingue très facilement les zones sensibles des autres parties du document car, d'une part, le curseur de la souris change d'aspect et prend une forme qui dépend du bouton sur lequel

il se trouve et, d'autre part, chaque type de bouton peut se voir attribuer un style propre (gras, italique, souligné).

On a donc à notre disposition au moins quatre types de boutons pour structurer un hypertexte. Ces boutons produisent chacun un effet particulier.

Les boutons de remplacement : ils peuvent être symbolisés par un viseur. Le fait de "cliquer" sur un bouton de remplacement provoque l'effacement du bouton puis l'affichage d'une information de remplacement associée à ce bouton. Cette information supplémentaire peut contenir du texte, des graphiques mais aussi de nouveaux boutons.

Les boutons de référence : ils servent à relier deux points d'un même hypertexte ou de deux hypertextes différents. Le bouton de référence est le point d'origine et il doit être associé à un point de référence qui est le point d'arrivée. Le fait de "cliquer" sur le bouton de référence affiche l'information qui commence au point de référence. Il est possible de revenir au point d'origine en "cliquant" sur une icône de retour.

Les boutons de note : ils servent principalement à annoter un document. Lorsque l'on "clique" sur un bouton de note, une nouvelle fenêtre contenant une autre information est affichée tant qu'on laisse enfoncé le bouton de la souris. En pratique, il s'agit de commentaires utiles à la compréhension.

Les boutons de commande : c'est l'ouverture vers le monde extérieur ou vers des nouvelles applications. Le fait de "cliquer" sur un bouton de commande provoque, soit l'exécution d'une nouvelle application (permettant même le partage dynamique des données entre applications), soit la commande des périphériques comme les lecteurs de vidéodisque ou les magnétoscopes pilotables (on peut créer des hypertextes composés de texte, de graphiques et d'images vidéo).

2.5.2. La programmation orientée objets

La programmation orientée objets est une nouvelle méthode de programmation. Elle tend à se rapprocher de notre manière naturelle d'appréhender le monde. Etape importante dans l'évolution des techniques de programmation, les langages orientés objets (dits LOO) sont plus structurés, plus modulaires et réalisent une bien meilleure abstraction de données que ceux de la génération précédente (les langages structurés).

Les trois principales caractéristiques propres aux LOO sont :

L'encapsulation : elle consiste à combiner une structure de données (un enregistrement) et un ensemble de procédures et de fonctions habilitées à la manipuler. Il en résulte un nouveau type de donnée que l'on appelle un **objet**. Il faut remarquer que ce nom se rapporte effectivement à un nouveau concept puisqu'il s'agit en fait de bien plus qu'un nouveau mode de gestion des données, d'un usage dynamique de celles-ci.

L'héritage : en procédant tout d'abord à la définition d'un objet (ou d'une classe d'objets), on peut ensuite créer, à partir de cette définition, toute une hiérarchie d'objets. Chaque descendant hérite des propriétés de son ancêtre (accès au code et aux données) ; il a ensuite des caractéristiques propres.

Le polymorphisme : qualité que peut avoir un identificateur d'être associé à une hiérarchie d'objets sous plus d'une forme, l'action effective liée à cet identificateur différant suivant l'objet choisi. L'identificateur "change de forme" en fonction de l'objet où il s'applique.

La programmation orientée objets est la réponse à la complexité croissante des applications. L'héritage et l'encapsulation sont des moyens très efficaces de maîtriser cette complexité. L'orientation objets, plus encore que la programmation structurée, impose un ordre rationnel dans les structures de données et d'actions. Il faut ajouter à cela la possibilité d'enrichir et de faire évoluer ces objets dans des situations qui ne peuvent être connues que plus tard ; il s'agit là d'une propriété essentielle : la science et encore moins la complexité de notre environnement ne peuvent être décrites d'emblée.

2.6. CONCEPTS ET OUTILS POUR LES APPLICATIONS GEOGRAPHIQUES

2.6.1. Introduction

L'information géographiques et les applications qui la manipulent sont assez spécifiques ; aussi les approches classiques des bases de données (notamment la technologie relationnelle) s'appliquent-elle relativement mal. Ce domaine est actuellement en pleine expansion, avec la création de plusieurs grands projets de recherche tant aux Etats-Unis (création du National Center for Geographic Information ans Analysis) qu'en Europe (par exemple le réseau CASSINI en France).

L'information géographique est également de plus en plus précise et de plus en plus répandue sur support numérique. Les applications informatiques utilisant des informations géographiques sont extrêmement nombreuses, depuis la cartographie numérique, la télédétection, la planification urbaine, la gestion des différents réseaux nationaux (par exemple le projet de gestion des lignes téléphoniques de Télécom). Dans toutes ces applications, la gestion de l'information est cruciale et c'est pourquoi les bases de données apparaissent maintenant comme le noyau des systèmes d'informations géographiques (SIG).

2.6.2. L'informatique géographique

La géographie est la science qui décrit la réalité du terrain. Son objectif est de fournir des informations utilisables au niveau de l'homme. Ces informations géographiques ne sont que des descriptions approximatives de la réalité géographique. Elles se manifestent sous des formes diverses : descriptions littéraires, mesures de terrains, photographies, cartes, ...

L'information géographique peut aussi être le résultat d'un traitement. Il a pour but de répondre à des questions sur l'état du terrain et fait appel à des notions de topologie [David 88] :

- quelle est l'altitude du Mont-Blanc ?
- la RN14 coupe-t-elle le département de la Somme ?
- quel est le plus court chemin ferroviaire de Paris à Grenoble ?

Quelle que soit sa source, une caractéristique importante de l'information géographique est la persistance des sous-informations élémentaires invariables qui vont permettre son calcul : la réalité géographique est en effet peu modifiable. Même dans le cas d'une modification, on souhaite presque toujours conserver ses anciennes versions.

Une application informatique peut être qualifiée de géographique lorsqu'elle manipule des informations et qu'elle leur applique des traitements de nature géographique. Les informations géographiques sont traduites dans l'application informatisée par des données :

- les mesures de terrains deviennent une liste de réels,
- les descriptions littéraires sont des chaînes de caractères ou des documents structurés,
- les photographies et les cartes peuvent être représentées par des images numériques ou des graphiques.

Une application géographique manipule donc des données de nature multimédia :

- des données alphanumériques,
- des images sous format raster (matrice de pixels),
- des graphiques, c'est-à-dire des ensembles de figures constituées à la base par certains types de dessins élémentaires (en général points, lignes, segments et polygones),
- des textes sous forme de documents volumineux structurés suivant la décomposition classique en chapitres, sections et paragraphes.

Pour des raisons d'efficacité, on ne digitalise pas une carte ou une image à chaque exécution d'une application : il est donc nécessaire de gérer la persistance de certaines données.

La modélisation géographique se traduit par l'utilisation de structures de données destinées à saisir la sémantique de l'application. Les traitements géographiques sont de nature très variable, ils correspondent, par exemple, à des algorithmes répondant aux questions précédemment citées. Parmi celles-ci, on peut mettre en valeur la notion de requêtes spatiales [Laurini 88] [Orenstein 88] [Scholl 89]. Il s'agit de répondre à des questions sur la géométrie des objets :

R1 - calcul de distances ou de surfaces,
R2 - recherche des objets contenant un point donné,
R3 - recherche des objets contenus dans une région donnée.

Ce type de requêtes intervient pratiquement dans toutes les applications géographiques. Ce sont des commandes fréquentes et interactives qui conditionnent l'efficacité des applications.

2.6.3. Exemples d'applications géographiques

On peu citer ici quelques applications de nature géographique, développées de façon industrielle, qui permettent de mieux appréhender les fonctionnalités et les besoins de telles applications :

- **la photo-interprétation :** le travail du photo-interprète consiste à étudier des photographies (généralement aériennes) prises au cours d'une reconnaissance du terrain. Il cherche alors à y reconnaître certains objets pour en déduire des éléments d'information. Un poste de photo-interprétation informatisé doit mettre à la disposition du photo-interprète toutes les informations géographiques qui sont rattachées soit au lieu de la prise de vue, soit aux objets à reconnaître ;

- **la cartographie :** la cartographie informatique a pour but de reconstituer des cartes à partir d'informations géographiques préalablement enregistrées. Cette reconstitution peut être plus ou moins complexe car elle dépend de la nature des informations qu'elle utilise et du type de cartes demandées par l'utilisateur ;

- **la télédétection :** la télédétection consiste à gérer et à exploiter des images satellites. L'image satellite brute (telle qu'elle est reçue depuis l'espace) n'est pas utilisable. Elle doit être traitée pour corriger les déformations subies à la fois à cause de l'imprécision de son capteur et à cause des phénomènes géographiques comme la rotation de la terre. Pour pouvoir effectuer ces traitements, on doit faire intervenir de nombreuses informations géographiques concernant cette image : altitude et coordonnées du capteur, grille de déformation, etc.

Quelle que soit l'application, on se heurte aux mêmes problèmes de gestion des données géographiques :

- **persistance de données multimédia :** certaines données doivent survivre à la fin d'une application, notamment les images, les cartes, les graphiques ou les textes ;

- **présence d'un gros volume de données :** le volume de données varie selon les applications. Cependant il reste très important. Il faut environ 160 gigaoctets pour enregistrer la cartographie de la France à l'échelle de 1/50 000 ;

- **importance des relations entre les données :** la modélisation des applications en terme de structures de données n'est pas simple. Elle nécessite de nombreuses possibilités :
  * la définition d'objets complexes. Par exemple, une région contient une liste de départements, chaque département est lui-même composé d'une préfecture et d'une liste de circonscriptions ;
  * la gestion d'historique. On désire presque toujours garder une trace des modifications appliquées aux objets géographiques ;
  * la gestion de relations spatiales entre les objets. Tous les objets sont liés à leur position dans l'espace et l'accès à ces objets s'effectue presque toujours suivant des requêtes spatiales (distance, objets contenant un point et objets contenus dans une région) ;

- **partage de ces données dans une communauté d'utilisateurs :** les données géographiques sont communes à différentes applications et sont donc partagées entre ces applications ;

- **fonctionnement en mode dégradé et reprise après panne :** la plupart du temps, les applications géographiques fonctionnent sur un réseau local disposant d'un serveur de données et de postes de travail spécialisés qui coopèrent. Un des postes de travail peut être momentanément en panne ; les autres applications du réseau doivent être en mesure de fonctionner partiellement en préservant l'intégrité et la cohérence des données. Lors de la reprise du poste arrêté, ce dernier doit avoir perdu un minimum du travail effectué avant la panne et le réseau ne doit pas être trop perturbé.

2.6.4. Les différentes approches proposées

Il existe plusieurs domaines de recherche sur la gestion des données géographiques :

- un domaine issu des besoins informatiques du milieu géographique : il a conduit à l'apparition des **systèmes d'information géographique** (SIG),

- un domaine issu de la recherche en base de données qui s'est spécialisé dans les applications géographiques. Il a conduit à l'apparition d'une nouvelle génération de bases de données appelées **bases de données géographiques.**

a) Les SIG

Le but des SIG est de faciliter le travail des utilisateurs pour la recherche, la gestion et la planification géographique. En phase d'analyse, un SIG met en oeuvre une suite d'actions : sélection, observation et étude des données. D'autres phases de fonctionnement sont aussi importantes : saisie des données, restitution des résultats. Le logiciel commercialisé ARC/INFO [Schaller 87] est un bon exemple de tels systèmes. Calkins et Tomlinson [Calkins 77] donnent une présentation synthétique des premiers SIG. Ils sont définis comme des systèmes d'information devant accomplir cinq tâches :

* saisie et formatage des données,
* organisation et gestion des données,
* accès rapide aux données,
* analyse et traitement sur les données,
* sortie et restitution des données et des résultats.

Smith [Smith 87] montre l'insuffisance d'une telle définition et décrit les futurs SIG comme des outils géographiques munis des fonctionnalités suivantes :

* la gestion de données géographiques dans une base de données,
* la possibilité d'interrogation de ces données suivant des critères géographiques (requêtes spatiales),
* l'exécution efficace de ces interrogations pour pouvoir travailler intéractivement,
* une souplesse d'adaptation du système aux opérations de maintenance et à la modification des schémas des applications,
* des possibilités d'apprentissage sur les objets spatiaux, leur signification et leur implantation durant la vie du système.

Smith [Smith 87] montre qu'un SIG répondant à ces fonctionnalités assure obligatoirement les tâches décrites dans [Calkins 77].

b) Les SGBD géographiques

Dans le début des années 80, les chercheurs se sont intéressés à l'extension des SGBD pour de nouvelles applications dans de nombreux domaines (intelligence artificielle, génie logiciel, bureautique, CAO). Parmi les applications géographiques, les systèmes de

traitement d'image ont été les premiers demandeurs de fonctionnalités base de données, mais ils demeurent assez éloignés des fonctionnalités géographiques.

Le concept de base de données spatiales est relativement récent. Il est issu de la cartographie et des fonctions de traitement de l'espace. Il repose sur une démarche qui prend sa source dans l'étude des besoins des SIG et aboutit à la définition d'un SGBD complet : interfaces spécifiques à la géographie, modélisation de cartes et méthodes d'implantation pour les opérateurs spatiaux (notamment pour les requêtes spatiales).

Les objets géographiques sont décrits par trois composants :

* un ensemble de propriétés générales sur leur nature pouvant s'exprimer simplement avec des données alphanumériques,
* une représentation graphique,
* une représentation spatiale : localisation, surface, espace occupé.

Les deux derniers points constituent de nouveaux types de données encore peu pris en compte par les bases de données.

Les objets géographiques complexes (par exemple des cartes) sont modélisés par des ensembles de sous-objets géographiques. Leur représentation graphique n'est pas stockée, elle est obtenue par union des représentations graphiques des sous-objets. De la même façon, leur représentation spatiale peut être le résultat d'un calcul sur les représentations spatiales de leurs sous-objets.

Les modèles de données de bases de données spatiales peuvent être considérés comme des extensions du modèle relationnel. David [David 88] et Güting [Güting 88] proposent des extensions du modèle relationnel et définissent de nouveaux domaines qui ont pour valeur des représentations spatiales. Scholl [Scholl 89] définit un modèle relationnel étendu pour la gestion d'objets géographiques complexes et leur utilisation à l'aide d'une algèbre de régions.

Les méthodes mises en oeuvre dans les bases de données spatiales rejoignent celles des SIG. Elles ont pour but d'implanter efficacement les requêtes spatiales (algorithme optimisé de produit spatial [Scholl 89] [Güting 88], ou jointure de Peano [Laurini 88]).

2.6.5. Evolution de l'offre matérielle et logicielle entre 1987 et 1992

Si l'on essaye de mesurer le chemin parcouru depuis cinq ans en matière de SIG (en schématisant un peu le propos), on peut noter les évolutions suivantes.

- En matière d'organisation des données graphiques, le passage s'est effectué de logiciels de DAO à fichiers graphiques et sans structuration des données, à des systèmes capables de gérer la topologie et de manipuler de véritables bases de données graphiques. Ces outils sont capables d'assurer une continuité géographique de manière transparente.

- Des interfaces ont été établies entre logiciels graphiques et SGBD relationnels, entre objets graphiques et enregistrement de tables, à tel point qu'on considère pratiquement aujourd'hui qu'un SIG doit être constitué d'un logiciel graphique et d'un SGBD.

- Les matériels ont vu leur puissance augmenter, ce qui a rendu envisageable l'exploitation de certains types de données (raster par exemple) et la mise en oeuvre de certaines architectures matérielles (stations de travail). Les solutions totalement centralisées sont remises en question.

- L'adoption de standards a remis en cause les plates-formes matérielles propriétaires au profit de solutions permettant ouverture et facilités de communication entre applications.

- Les échanges de données sont possibles grâce au développement d'outils d'interfaçage et à la généralisation de certains formats d'échange de données (DXF notamment).

- Les interfaces utilisateurs se sont améliorées grâce à l'introduction des interfaces graphiques ; ceci a parfois remis en cause le pouvoir des informaticiens (fondé sur la difficulté de mise en oeuvre de certains outils), en mettant l'accent sur la convivialité et la facilité d'utilisation.

2.6.6. Conclusion sur les approches

Il ressort de l'étude des applications géographiques des problèmes de modélisation, de cohérence, de partage, de volume et d'efficacité dans la gestion des données.

Les SIG n'apportent pas de solutions. Ils se heurtent aux mêmes problèmes de modélisation, de cohérence et de partage. Par contre, ils mettent en oeuvre des algorithmes efficaces dans le traitement des requêtes spatiales.

Les bases de données spatiales apportent une solution aux problèmes des SIG. Leurs fonctionnalités sont intéressantes mais ces systèmes (encore au stade de la recherche) n'intègrent pas de gestion de données multimédia.

Une autre voie nouvellement explorée est celle des SGBD extensibles et des SGBD orienté-objet. Ceux ci par leur capacité à modéliser les objets complexes, à intégrer les opérations (au moyen de types abstraits de données) semblent avoir le potentiel pour réaliser de vraies bases de données spatiales. On peut citer un certain nombre de projets dans ce sens, PROBE [Manola 87], DASDBS Geo-kernel [Wolf 89], POSTGRES [Stonebraker 86 et 91], le modèle ESTRELLA et le système OMEGA (groupe MATRA).

CHAPITRE TROISIEME

FONDEMENT SEMANTICO-CONCEPTUEL DES DONNEES ECOLOGIQUES ET ENVIRONNEMENTALES

3.1. INTRODUCTION GENERALE

Tous les systèmes écologiques sont aujourd'hui sous influence humaine de sorte que les caractéristiques des actes humains (agronomiques, aménagements hydrauliques, mouvements d'éléments, y compris polluants, ...) sont partie intégrante de la description de ces systèmes.

Le développement d'une gestion informatisée et uniformisée des données écologiques est une nécessité actuelle qui exige toutefois une clarification conceptuelle. Si chaque discipline, concourant à l'étude des systèmes écologiques, a ses caractéristiques propres (caractéristiques décrivant les éléments chimiques, les horizons de sols, les débits de cours d'eau, les travaux aratoires, les récoltes céréalières, les états forestiers, ...), il est indispensable d'avoir un minimum de langage commun pour comprendre la situation des caractéristiques dans les systèmes écologiques et les étapes d'acquisition, gestion et traitement des données.

Les systèmes d'information en environnement sont actuellement des moyens techniques permettant la gestion de l'information dont l'homme dispose sur son environnement. Il s'agit essentiellement de moyens informatiques de saisie de données, de gestion-conservation, de traitement et d'édition de l'information.

Selon les objectifs et les moyens disponibles ces systèmes peuvent être très divers quant à leur objet (nature des données prises en compte), à leurs objectifs (répondre à une demande finalisée précise ou générale), à leurs coûts financiers et à la puissance et convivialité des machines et logiciels mis en oeuvre.

Nous proposons ici cette base sémantico-conceptuelle, nécessaire pour la conception et la réalisation de bases de données, pour la description de toute donnée issue de l'écosphère dans le cadre d'une approche systémique, ceci permettant les échanges transdisciplinaires et ouverts des sciences liées à l'écologie.

Cette base sémantico-conceptuelle a été peu à peu développée à partir d'une pratique de l'écologie visant à rendre celle-ci à la fois rigoureuse et opérationnelle (Bouché, 1990). Mon travail principal a été le développement conceptuel et informatique de cette nouvelle méthodologie d'aborder l'écologie appliquée à l'environnement.

Au cours de ce développement bien des concepts ont été éprouvés dans la pratique et une terminologie rigoureuse a dû être peu à peu dégagée afin de rendre possible de façon efficace l'intégration des connaissances.

Le présent chapitre est un effort de réflexion destiné à faciliter la mise en oeuvre pratique des indispensables bases de données avancées, à la fois puissantes, efficaces et adéquates pour gérer la complexité des informations issus des (agro-) écosystème.

Ce sont ces développements appliqués qui nous ont démontré, d'abord au niveau circonscrit de la problématique d'un laboratoire, puis en les testant sur des applications tout à fait interdisciplinaires, l'intérêt stratégique actuel des Bases de Données Relationnelles qui comme leurs noms l'indiquent, permettent de relier les données entre elles quelle soit la discipline qui les a produites.

Ces applications, aujourd'hui indispensables à une véritable intégration éco-environnementale, ont nécessité une conceptualisation écologique rigoureuse qui ensuite nous a permis de gérer des données issues d'équipes diverses et allant de l'ornithologie au poisson, des lombriciens au machinisme, de la végétation au traitement pesticide et au type de sol.

En permettant une gestion ordonnée de l'information exhaustive, non bornée par un objectif sectoriel, ces bases rendent possibles des échanges qui sont optimisés au niveau des bases elles-même. Elles constituent des modules qui, fédérés et couplés à des réseaux, permettront une mobilisation quasi-instantanée des données objectives quelque soit leur origine disciplinaire ou géographique.

Les progrès spectaculaires de l'informatique nous donnent dès aujourd'hui, ou dans le très proche futur, des outils d'Intelligence Artificielle permettant une véritable maîtrise de la complexité des (agro-) écosystèmes et de la grande diversité de nos relations avec ceux-ci. Malheureusement, il ne suffit pas de disposer de machines sophistiquées et de logiciels performants si on ne ré-évalue pas les pratiques nécessairement sectorielles dans un cadre global en s'appuyant sur ces fondements sémantico-conceptuels univoques, transdisciplinaires et éprouvés par la pratique.

Le présent chapitre vise à réduire le divorce qui existe entre l'inadéquation des pratiques sectorielles et l'existence d'outils d'intégration indispensables mais encore mal employés. Il expose ces fondements sémantico-conceptuels, préalables indispensables à l'usage de ces outils.

3.2. L'ECOLOGIE ET LES DISCIPLINES CONNEXES

3.2.1. Introduction

Comme toute science, l'écologie a un objet d'étude général et des domaines particuliers impliquant des méthodes, des procédés, des techniques, des observations de faits, une reconstitution synthétique, etc. Ceci implique un langage associé à ces opérations, les mots recouvrant des concepts traduits par une définition plus ou moins précise.

La difficulté provient du fait que les langages utilisés, sont très souvent incompatibles les uns avec les autres. Dès que l'on veut échanger, on constate que les termes utilisés sont employés avec des sens différents et encore plus avec des précisions très variables.

Les défauts sémantiques proviennent aussi d'une nécessité sociale entre chercheurs s'unissant pour tendre à une cohérence ; dans un premier temps, cela conduit à masquer, sous des vocables flous, l'hétérogénéité initiale des concepts pour ne pas soulever à tout instant trop de questions. Enfin, à quoi sert-il de définir avec rigueur un élément quelconque dont on sait que les techniques d'études ou les moyens financiers ou humains disponibles ne permettent pas la mesure ou l'observation précise ? Les difficultés techniques initiales ont, d'une certaine façon, "justifié" un certain laxisme lexical. Mais cette situation se retourne, faute de concepts précis, les échanges sont mauvais, car chacun est attaché à son langage, justifié ou non.

Les échanges entre écologues, ayant subi des influences différentes, sont en conséquence difficiles ; ce qui est rigoureux pour l'un, est flou pour l'autre, ou encore le même terme peut avoir deux significations différentes (car adoptées à partir de disciplines différentes), enfin le même phénomène perçu par divers écologues praticiens ou fondamentalistes sera décrit et formulé mathématiquement de façon indépendante.

Correcte, oui, mais par rapport à quoi ? Les avis sont partagés, divers, et chaque scientifique a une opinion à la fois nuancée et particulière. Il sera volontiers exigeant pour un mot dont il a l'usage et inattentif au sens d'un autre terme, important mais externe à sa pratique.

Nous essayerons d'approfondir et d'éclaircir une série de concepts généraux permettant une contribution méthodologique aux démarches actuelles pour le développement d'une écologie opérationnelle, ouverte et transdisciplinaire.

3.2.2. L'écologie

< <Nous entendons par ECOLOGIE la science globale des relations des organismes avec leur monde extérieur environnant dans lequel nous incluons, au sens large, toutes les conditions d'existence> > (HAECKEL, 1866b. trad. BOUCHé, 1980a).

Cette définition mérite une lecture attentive pour analyser chaque propriété de la définition et en déduire ce qui est dans et hors le domaine de cette discipline.

1) L'écologie est une science. HAECKEL utilise dans ses nombreux travaux ce mot (Wissenschaft) au sens précis, moderne, de science exacte. L'écologie devra donc dans ses démarches suivre les méthodes des sciences exactes, *a contrario* elle ne saurait couvrir des domaines techniques, philosophiques ou politiques, même si elle peut contribuer à l'élaboration de (ou influencer) ces domaines.

2) Cette science a trait aux organismes pris individuellement ou dans leur ensemble. Elle se réfère donc, entre autres, à un individu ou à la somme de tous les individus. Par contre, elle ne concerne pas directement les propriétés internes des organismes (régulation physiologique interne, histologie, etc.). Si l'écologie peut évidemment utiliser les connaissances acquises sur ces propriétés, celles-ci ne sont pas son objet d'étude.

3) L'écologie étudie les relations des organismes vivants avec leur monde extérieur environnant (= leur milieu). Il y a symétrie. L'écologie porte autant sur l'étude des organismes que sur le milieu (environnement) de ces organismes, c'est-à-dire les autres organismes, mais aussi les éléments physico-chimiques non vivants : l'écologie n'est donc pas qu'une science biologique mais a trait aussi à d'autres objets analysés par certaines disciplines (minéralogie, climatologie, pédologie, géologie, hydrologie, etc.).

4) L'écologie étudie les relations des organismes vivants avec leur milieu... y compris toutes les conditions d'existence. Toute étude écologique devra donc restituer directement, ou se référer explicitement et de façon démonstrative au milieu permettant toutes les conditions d'existence de l'organisme (ou des organismes en cause).

5) L'écologie, étant une science globale porte d'abord sur l'ensemble des conditions d'existence. L'écologie ne peut évidemment pas se refuser les informations apportées par des données analytiques mais celles-ci n'ont de valeur que si elles restent (ou peuvent être scientifiquement restituées) dans un cadre général et cohérent. Toute démarche analytique

en écologie se fait pour une conclusion synthétique. Une approche exclusivement réductionniste ne peut pas être écologique.

6) Organisme(s) et milieu sont envisagés globalement dans leurs relations (à actions réciproques). Nous sommes en présence d'un ensemble d'éléments dépendant les uns des autres et formant un tout organisé c'est-à-dire un système. Ce dernier point n'est pas, pour des raisons historiques, explicite dans la définition mais découle de ses propriétés puisque au moins deux compartiments (organisme(s) + milieu(x)) et leurs relations en découlent explicitement.

En conclusion, l'écologie, de par sa définition originale, est une science visant à une compréhension générale, intégrée, "totale" des relations de tout organisme ou groupe d'organismes avec l'ensemble des éléments environnant ceux-ci. L'écologie étudiera autant ces éléments environnants que les organismes et de telle sorte que ses résultats, ses démarches s'appliquent à des relations ayant prévalu et prévalant dans l'édification de ces rapports.

3.2.3. Ecologie et sciences humaines

L'écologie porte sur les **systèmes écologiques**, c'est-à-dire des systèmes d'échanges entre les différentes parties biologique, organique et inorganique de l'écorce terrestre, ces systèmes suivent exclusivement les lois biophysicochimiques dans leurs régulations internes et dépendent de variables externes. Parmi ces variables externes, il y a les actions humaines ou actes.

On pourrait considérer l'homme comme un être vivant parmi les autres répondant aux lois biophysicochimiques et inclure ainsi les activités humaines dans les systèmes écologiques, mais celui-ci a développé au cours de son évolution une aptitude intellectuelle tout à fait inaccoutumée non observable dans la totalité des systèmes biophysicochimiques.

La rationalité des décisions humaines relève des études conduites par les sciences humaines (économie, psychologie, politique, sociologie, ...) même si l'homme comme tout élément de système écologique dépend de son environnement global, les motivations de ces actes ne sont pas *ipso facto* la résultante d'une interaction simple homme-environnement à la différence des autres événements biophysicochimiques de **l'écosphère** c'est-à-dire de la partie de la sphère terrestre où vivent les organismes vivants, envisagée notamment du point de vue des relations entre ces organismes et leurs milieux [GILLARD 1969]. D'une façon

générale, l'homme s'est largement affranchi des systèmes écologiques et réciproquement agit sur ceux-ci d'une façon souvent indépendante de leur nature intime. Par exemple, la mise en production d'un nouveau pesticide peut dépendre des liquidités financières de l'entreprise, elle-même influencée par la cotation d'une valeur mobilière dépendante de la tenue de la bourse. De même, pour le choix du matériel de traitement, le choix des méthodes culturales conduisant au traitement pesticide qui finalement agit seul concrètement sur le système écologique..., etc. L'étude de ces processus de décision humaine relève de recherches socio-économiques et non de celles des relations des organismes avec leur milieu : elle relève des sciences humaines et non de l'écologie *stricto sensu*.

L'écologie étudiant les relations des organismes avec leur monde extérieur environnant "au sens le plus large" ne peut exclure les multiples actions humaines dans cet environnement (par exemple, l'agriculture). On sait qu'elles sont omniprésentes dans toute l'écosphère, à commencer par les transferts d'éléments artificiels créés par l'homme (par exemple DDT) ou largement produits sous son contrôle (CO_2, césium radioactif, ...). Aucun système écologique ne fonctionne sans les conséquences d'actes humains.

3.2.4. L'écologie et l'interdisciplinarité

L'écologie, comme toute science exacte, a son propre domaine d'étude et ses démarches et simultanément interfère avec de nombreuses autres divisions des sciences exactes (et humaines).

Il en ressort que l'écologie hérite des concepts, donc du vocabulaire, de formulations synthétiques et d'un renouvellement des concepts dans chaque discipline concourant à l'écologie. A cela s'ajoute la diversité des sources (physiques, chimiques, biologiques, naturalistes et techniques).

On constate en fait que, pour des raisons d'efficacité évidentes, les diverses disciplines n'abordaient qu'une partie "réduite" de l'écologie ; ce réductionnisme historique pèse toutefois très lourd sur les pratiques et concepts dont l'écologie a hérités.

D'un point de vue opérationnel, il faut surmonter la contradiction de la nécessaire collaboration entre experts et du caractère réduit et spécialisé de la perception de chaque intervenant apportant des données objectives, seules bases sérieuses de réfutation.

Il est facile de concevoir que chaque discipline soit conduite au cours de ses progrès à élaborer son vocabulaire, ses formules, sa présentation particulière des connaissances ; ceci est normal. Il n'en reste pas moins que l'écologie se doit d'avoir une présentation unitaire avec un vocabulaire univoque et rigoureux.

Le choix de ce vocabulaire ne peut et ne doit pas se faire en ignorant celui des disciplines connexes car comme toute science, l'écologie n'est qu'une contribution particulière aux connaissances et son aptitude à transmettre dépend de ces choix.

Il en découle que :

1) Dans toute la mesure du possible les termes retenus doivent être ceux qui sont consacrés par la discipline connexe.

2) Toutefois, des difficultés existent lorsque le même terme est utilisé avec des sens différents, parfois partiellement différents dans deux disciplines adjacentes et sécantes à l'écologie.

3) Les néologismes sont évidemment indispensables pour désigner des objets ou concepts nouveaux. Mais la tentation du néologisme pour exclure des malentendus est toutefois lourde de conséquences car elle conduit des fois à des non-entendus devant ce terme nouveau, incompris.

3.2.5. Ecologie et réfutabilité

L'objet d'étude de l'écologie, n'est perçu que par des faits, leur interprétation implique un choix d'hypothèses conduisant à un modèle physique, mathématique et théorique. L'étendue de l'écologie fait que cette approche du global se fait à partir de points de vue, de séries de faits, toujours partiels et incomplets. L'aptitude d'une théorie est dans sa généralisation. C'est la généralité qui permet d'inférer au global des propriétés observées sur des cas particuliers. Comme l'écologie, sa théorie ne peut être que globale, son élaboration ne pourra s'effectuer qu'après une acquisition de données et une élaboration des informations sur les systèmes écologiques.

En l'état actuel toutefois, les interprétations écologiques restreintes tendent à contribuer à cette théorie générale et il est essentiel de rendre réfutables celles-ci et de pratiquer une intégration et un contrôle permanent des connaissances. Une interprétation sera réfutable lorsqu'elle permettra de prédire des résultats vérifiables.

3.3. L'INFORMATION ECOLOGIQUE

3.3.1. Introduction

Toute connaissance sur les < <**relations des organismes avec leur monde extérieur environnant**> > constitue une information écologique. Acquérir de telles informations, les ordonner, les interpréter constitue la tâche primordiale de l'écologue.

En raison de l'extraordinaire complexité de l'écosphère et de son étendue, le travail journalier de l'écologue, ou du technicien dans la **biosphère** c'est-à-dire l'ensemble des organismes vivants animaux et végétaux qui vivent à la surface du globe [CDAF 1842] et/ou dans la **toposphère**, c'est-à-dire l'ensemble abiotique (lithosphère + hydrosphère + atmosphère) [SCHWERDFEGER 1975], ne peut être appréhendé à la fois en totalité et en détail (Figure 1).

L'un des problèmes majeurs concrets des démarches écologiques est donc de permettre à la fois les études les plus fines et précises nécessaires pour comprendre des systèmes écologiques partiels, et simultanément d'inscrire concrètement ces observations particulières dans les ensembles généraux pour obtenir à la fois une perception globale (par exemple, le bilan des échanges du CO_2 dans l'écosphère, dans le cadre du cycle du carbone) et une connaissance des mécanismes y contribuant (par exemple, la respiration/photosynthèse de telle espèce végétale). Ce problème majeur ne sera pas résolu sans une rigueur dans l'élaboration des connaissances.

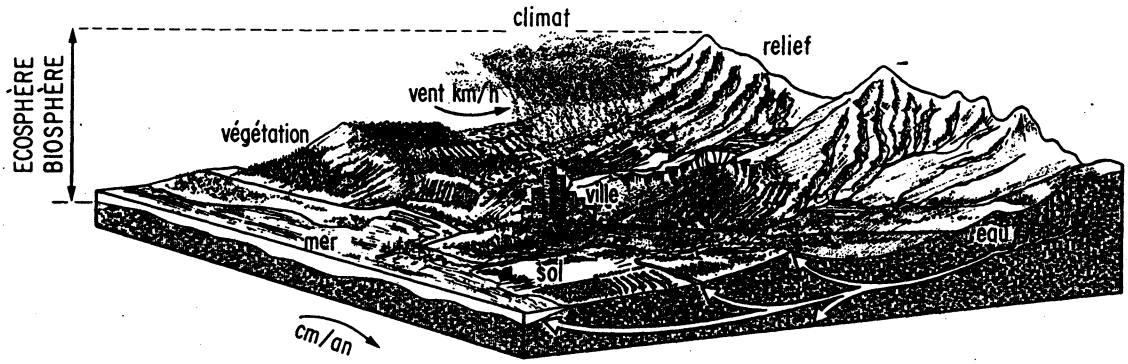


Figure 3-1 : *Les limites spatiales de la biosphère sont identiques avec celles de l'écosphère. La biosphère est la somme des êtres vivants. Elle n'est pas un système. Mais les organismes constituent le compartiment biotique en échange permanent avec le compartiment abiotique (lithosphère + hydrosphère + atmosphère). L'ensemble constitue un système : l'écosphère.*

Le domaine de l'écologie est très étendu. Science globale recouvrant tous les lieux où des organismes vivent dans leur milieu, elle porte sur toute l'écosphère. Certaines études écologiques peuvent toutefois être circonscrites à l'étude des relations des caractéristiques du milieu prévalant autour d'un individu, par exemple d'une plante. Les écologues peuvent ainsi considérer des espaces allant de la dizaine de milliers de kilomètres carrés au micromètre carré ! Les observations écologiques portent de même sur des durées temporelles différentes : un débit respiratoire est observable en quelques secondes, un déplacement d'une cohorte d'animaux sur de grandes périodes de temps, allant d'une saison à plusieurs dizaines d'années, voire, par une documentation historique ou paléogéographique, une reconstitution paléo-écologique peut porter sur des dizaines ou des centaines de millions d'années, notamment pour tenter de reconstituer l'évolution qui a abouti aux relations écologiques actuelles.

Une telle variété d'échelles de temps et d'espace ne va pas sans difficultés. Une des difficultés apparente majeure de la démarche écologique réside dans ces changements d'échelle : une approche globale de l'écologie oblige à adopter pour chaque étude un domaine de l'échelle spatio-temporelle. Cette difficulté peut aujourd'hui être surmontée, mais exige pour cela une élaboration et gestion sérieuse des données.

Toutefois, il ne s'agit pas d'un objet inaccessible, si la gestion des connaissances, fondée objectivement sur les informations écologiques est effectuée rigoureusement, ce qui devient chaque jour plus efficace par les progrès et l'avènement d'outils notamment informatiques. La gestion de l'information écologique doit être faite de telle sorte que les données acquises à une échelle spatio-temporelle soient utilisables à d'autres niveaux de perception de façon méthodologiquement critique.

Ajoutons enfin que les caractéristiques écologiques observées peuvent être perçues, et de fait acquises, à des niveaux de perception très différents. Par exemple et de façon toujours plus réductionniste, on peut enregistrer un climat régional à l'intérieur de celui-ci, une prairie et dans celle-ci un faciès de végétaux associés et pour celui-ci noter le profil d'un sol où l'on observe un horizon dont on peut décrire l'azote total, ou seulement son état ammoniacal NH_4^+ et dans ce dernier la fraction de l'isotope ^{15}N.

Plus difficiles peut-être seront les études portant sur des caractéristiques concrètes rapportées à un objet spatio-temporel abstrait, tel le débit d'une substance dans une portion d'écosystème exprimé en $q/s/m^2$ (q = quantité, s = seconde, m^2 = mètre carré) où la seconde est une unité de temps arbitraire quelconque se référant en fait à un phénomène instantané et où le m^2 est également un arbitraire quelconque (dans la nature, tous les

mètres carrés sont fonctionnellement différents et hétérogènes, dans leur surface, etc.). Il s'agit en fait d'une quantité se rapportant à un point, débitée de façon instantanée, qui est mesurable et exprimée par nécessité technique dans des unités spatio-temporelles conventionnelles.

Les changements d'échelles se font avec une relative aisance lorsque l'on réfère les données à des échelles métriques d'espaces et de temps. Ces changements d'échelle sont quelque peu délicats lorsque ces références sont non métriques, nous sommes amenés alors à parler de station, région, ... pour indiquer ces changements.

Ainsi la notion de niveau de perception, développée notamment par les phyto-écologues [LONG, 1974], permet une première hiérarchie qualitative des objets d'étude dans l'espace (Tableau 1-1) temps négligé.

Niveaux de perception écologique	*Unités de végétation concrètes*	*Variables écologiques prépondérantes* (non exhaustif)
ZONE ÉCOLOGIQUE	GRANDS TYPES DE VÉGÉTATION (ex. zone des forêts sclérophylles sempervirentes...)	latitude : positions relatives des masses continentales et marines et des grands massifs montagneux ; climat général
RÉGION ÉCOLOGIQUE	TYPES DE VÉGÉTATION (ex. maquis, garrigues...)	relief général ; climat régional ; unités géomorphologiques majeures ; homme
SECTEUR ÉCOLOGIQUE	GRANDES FORMATIONS VÉGÉTALES (ex. divers maquis ou garrigues...)	positions topographiques ; climats locaux ; types de formations superficielles ; conditions d'exploitation des ressources
STATION ÉCOLOGIQUE	FORMATIONS VÉGÉTALES en pays *faiblement* influencé par l'homme et/ou FORMES DE VÉGÉTATION en pays *fortement* influencé par l'homme	micro-relief ; variables édaphiques ; action de l'homme et des animaux
ELÉMENT DE STATION ÉCOLOGIQUE	portions de formations végétales ou de formes de végétation	variables micro-climatiques et micro-édaphiques ; variables biotiques (y compris action des micro-organismes)

Tableau 3-1 : *Niveaux de perception écologique* (d'après LONG 1974, simplifié).

3.3.2. Démarches analytiques, synthétiques et globales

Le niveau d'observation situe dans le spatio-temporel toute caractéristique entre le plus global et le plus particulier. En fait, toute caractéristique peut être décomposée en parties plus élémentaires et réciproquement entre dans la composition d'un système plus large.

Pour l'interprétation écologique, le niveau d'observation n'a *a priori* aucune importance, si ce n'est d'un point de vue d'efficacité du choix de la composition optimale à étudier pour un objectif de recherche précis. Il faut toutefois bien distinguer les démarches d'analyse (recherches de caractéristiques plus élémentaires, à partir d'un caractère plus général) de la synthèse (reconstruction à partir de caractères détaillés d'un ensemble synthétique) et de la globalité (qui d'emblée observe une caractéristique à un niveau de composition donnée). Une synthèse ne peut contenir que la résultante des analyses et des hypothèses la sous-tendant : elle ne correspond pas *ipso facto* à la globalité qu'elle tend à représenter.

Confronter le résultat d'une synthèse à une information globale constitue un important moyen de vérifier la qualité des choix d'analyse et des choix de traitement effectués dans la démarche d'analyse-synthèse.

En pratique, pour toute démarche intellectuelle relative à l'élaboration des connaissances écologiques, on réfère le travail *ipso facto* à deux niveaux d'observation ; par exemple les plantes de la prairie, le taux de carbone du sol, la croissance printanière, le débit de la rivière, etc. Il y a un niveau N global (prairie, sol, printemps, rivière, dans les exemples ci-dessus) auquel se rapporte certaines caractéristiques analytiques (réduites ou analytiques, en ce sens qu'il ne s'agit que de quelques caractéristiques parmi beaucoup d'autres propres aux niveaux N). Ces caractéristiques analytiques (respectivement chaque plante, le taux de carbone, un état de croissance, le débit) décrivent une partie de la composition du niveau N et appartienne à un niveau inférieur (plus analytique) égal à N-1.

Opérationnellement, on envisage donc toujours au moins deux niveaux d'observation, l'un par rapport à l'autre. On peut évidemment changer de paire de niveaux d'observation et considérer ensuite soit un niveau supérieur (plus global) égal à N+1, telle la croissance annuelle (incluant N, la croissance printanière) soit un niveau inférieur égal à N-2 (telle un organe d'une plante de la prairie). Mais du point de vue opérationnel, nous faisons en ce sens une option qui ne peut porter que sur une paire de niveaux d'observation ou sur un seul niveau à chaque opération.

3.3.3. Le champ écologique

Toute information écologique se situe à l'intérieur d'un hyper-espace ayant, entre autres, les trois dimensions orthogonales de l'espace habituel (= euclidien) et le temps. Elle est elle-même une indication particulière (information particulière : un chou ; 15 % d'azote ; 25 millibars) résultant d'une analyse l'isolant d'un ensemble plus vaste d'autres informations du même espace-temps ; elle est un élément de la composition de la portion d'espace-temps prise en compte.

Les informations issues de la composition peuvent être recomposées en ensembles moins analytiques ou au contraire précisées par des analyses plus fines. La composition peut enfin s'exprimer selon une infinité de niveaux d'observation analytiques allant du global au plus particulier selon des modalités techniques définissables. La composition est l'ensemble des caractéristiques que l'on peut attribuer à un objet d'étude (= de niveau d'observation quelconque) quelle que soit l'analyse. La somme des informations possibles constitue donc un hyper-espace très complexe et souvent hiérarchisé. Une portion d'espace peut comporter une branche d'arbre dont on analyse la morphologie foliaire dans laquelle on choisit par sélection hiérarchique dans une série juvénile, un pétiole, pour en extraire les protéines puis on sépare l'acide aminé cystéine dont on isole le carbone pour en caractériser l'isotope ^{14}C... Remarquons que cette démarche constitue une cascade de choix... On n'a pas étudié les autres branches, ni les autres isotopes du carbone, ni les autres éléments, ni les autres acides aminés... Il y a à la fois une hiérarchie analytique et une sélection entre caractéristiques étudiées ou non étudiées sur la composition d'un objet d'étude.

On appellera référendaire écologique chacune des trois composantes majeures du champ écologique : l'espace, le temps, la composition [BOUCHé 1990].

Le référendaire étant ici pris au double sens de portion du champ écologique et de nécessaire référence. Ces trois référendaires imposent leur maîtrise dans l'organisation de l'écologie opérationnelle ; leur distinction claire s'impose principalement dans le référentiel des données, c'est-à-dire dans le repère spatio-temporel-compositionnel de chaque donnée acquise dans l'écosphère et ses liens aux données ayant la même position dans l'espace-temps écologique.

Repérer une information écologique est essentiel pour l'interprétation. Celle-ci se fera toujours en cherchant à acquérir, ordonner ou regrouper des données par rapport au champ écologique, c'est-à-dire en suivant les ordres d'analyse très divers de la composition d'un objet repéré par les positions spatiales (3 axes euclidiens) et le repère temporel.

Toute information n'est réciproquement valide que si elle est clairement située dans le champ écologique par la référence explicite ou implicite à ses trois référendaires. Le niveau de perception n'étant qu'une façon qualitative de désigner la portion du champ écologique que l'on considère en particulier pour une analyse plus détaillée.

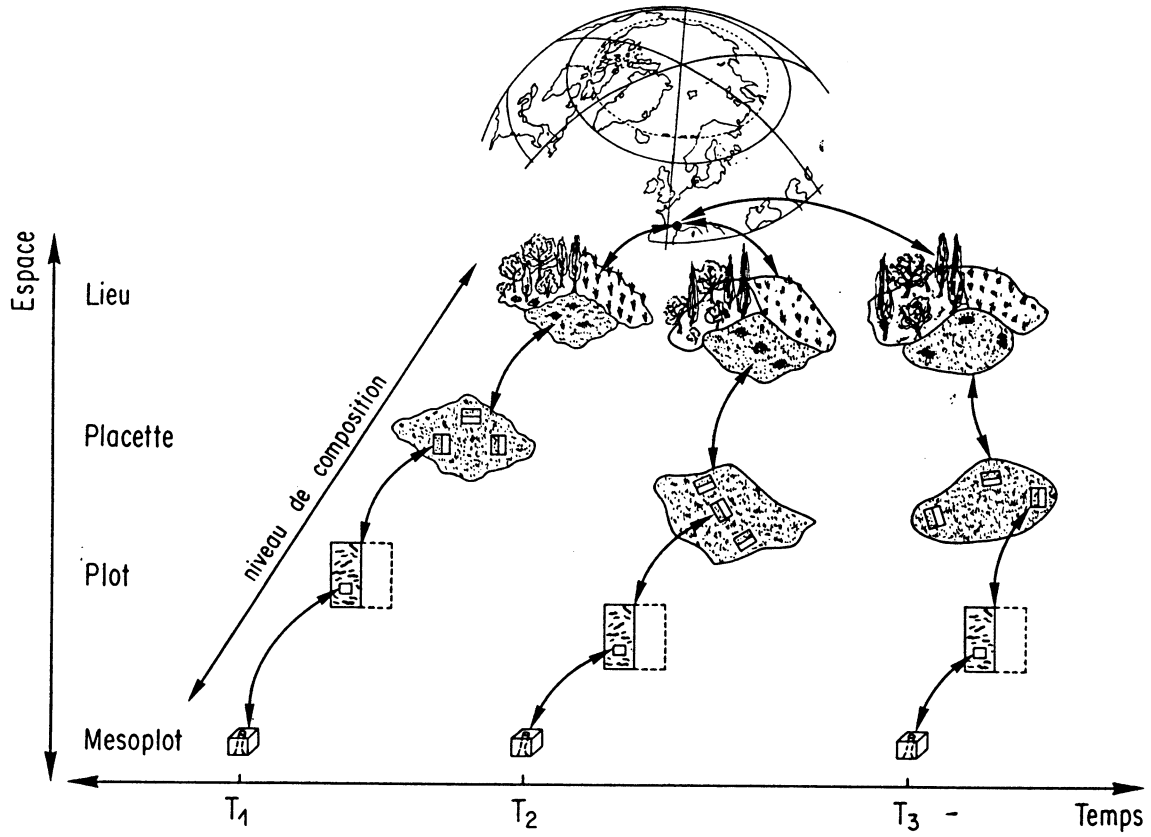


Figure 3-2 : *Le champ écologique et ses trois référendaires : l'espace, le temps et la composition.*

On peut donc considérer depuis le champ écologique le plus général défini simultanément par ses référendaires euclidiens, temporels et de composition, des niveaux d'observation dans un ensemble allant du plus global au plus analytique (par réduction euclidienne ou/et temporelle ou/et compositionnelle) ; il n'y a scientifiquement pas de limites inférieure ou supérieure à ces études : il y a toutefois limite du domaine écologique, du champ écologique supérieur et inférieur. L'art de l'écologie n'est ni dans l'étude de l'infiniment petit, ni dans celle de l'infiniment grand, mais bien au contraire dans l'approfondissement des moyens d'analyse-synthèse et dans la confrontation des résultats au global, la synthèse étant différente de la globalité.

Les observations concrètement effectuées (= données acquises) résultent d'options sélectives. L'écologue limite ses études à certaines parties du champ écologique ; réciproquement, l'information acquise ne porte que sur des fractions du champ écologique se situant entre le "plus" global et les diverses focalisations analytiques possibles (dans l'espace, dans le temps et en ce qui concerne la composition).

Cette fraction perçue, connue, du champ écologique est sa globalyse c'est-à-dire la somme de données objectives disponibles sur un système. La globalyse joue un rôle essentiel dans toutes les approches concrètes de l'écologue, dans toute l'écologie opérationnelle, puisqu'elle représente la (seule) partie effectivement perçue et observée du champ écologique : elle en est l'information objective disponible [BOUCHé 1990].

Le système écologique est l'objet, la globalyse sa partie objectivement perçue, le modèle sa représentation transposée.

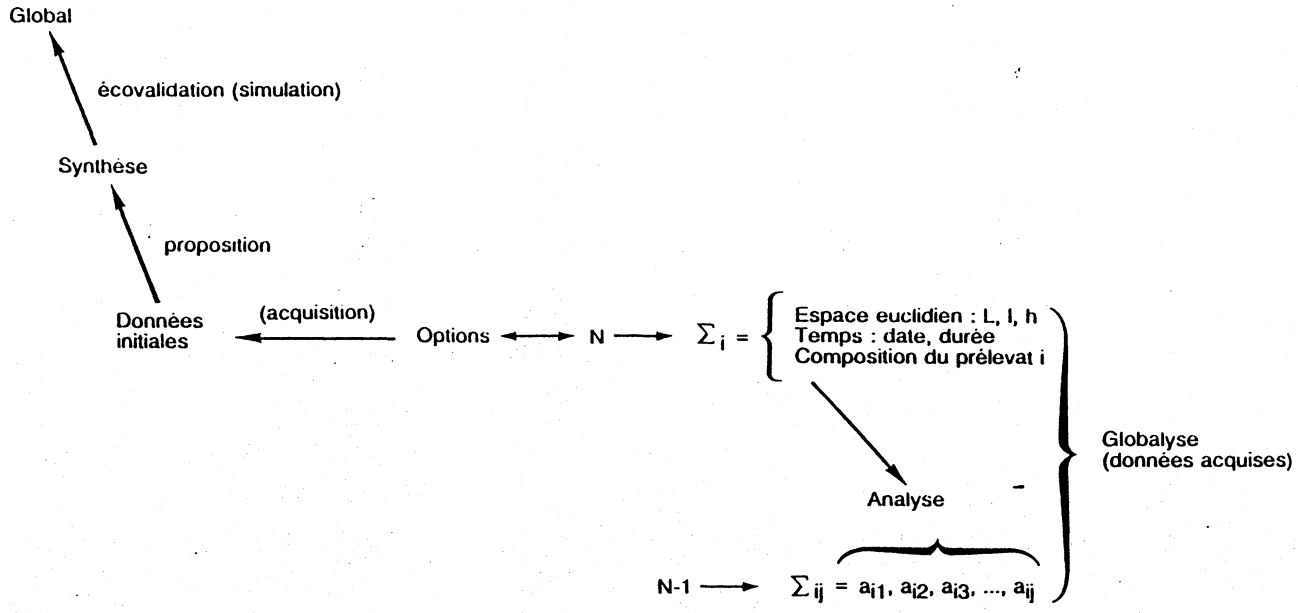


Tableau 3-2 : *Le champ écologique et sa globalyse : A l'intérieur d'un niveau de perception (station, secteur, région), repéré par de coordonnées géographiques (latitude, longitude, altitude), il faut opérationnellement choisir un niveau d'observation N (composition) contenant un nombre i de caractéristiques observables dans l'espace (L=longueur, l=largeur, h=hauteur) et dans le temps (Date, durée). Ce niveau étant susceptible de descriptions plus fines au niveau d'observation N-1 par analyse. Les choix effectués lors de l'acquisition résultent d'hypothèses dénommées options, ces options portent sur la pertinence des caractéristiques retenues pour représenter au mieux la portion du champ écologique sélectionnée dans la globalyse. L'accumulation des données permet d'interpréter synthétiquement la globalyse puis confronter cette synthèse aux données globales que nous avons du champ écologique.*

3.3.4. Le temps

Tout objet d'étude écologique est issu du passé. A un instant (pris au sens instantané) les données observables (comparables à une image photographique) résultent de phénomènes ayant un passé plus ou moins lointain. L'algue qui pousse sur le sol de la prairie s'est développée depuis sa spore il y a quelques heures, tandis que le puceron sur la plante a quelques dizaines de jours, la vache quelques années, le chêne du bouquet d'arbres quelques dizaines d'années et la pierre sur le sol s'est modifiée en des dizaines de milliers voire de millions d'années quoiqu'elle participe avec les sables et les argiles à un ensemble-sol en perpétuel remaniement. Chaque organisme est lui-même un individu particulier issu de recombinaisons parentales sur des millions d'années.

Toute interprétation écologique se place dans l'espace-temps, toute interprétation fait intervenir le temps, l'évolution des faits, des choses et des êtres vivants, des informations, des réalisations au cours du temps : l'enregistrement du temps est un lien obligatoire, un référendaire de toute donnée écologiquement référée. Il faut distinguer la date comme "instantanée" dans l'échelle du temps (même si cette date n'est pas calendaire, mais exprimée en secondes par exemple) et la durée comme étant le temps entre deux dates.

Du point de vue du fonctionnement des systèmes écologiques, le temps peut ne pas être perçu comme homogène au sens de l'écoulement de celui-ci dans une montre. Les choix de dates ou/et durées peuvent être arrêtés par rapport à des critères phénoménologiques, tels les stades phénologiques d'un arbre fruitier. Toutefois ceux-ci sont toujours en outre repérables dans le temps universel.

3.3.5. L'espace : hétérogénéité et continuum

D'une façon générale le champ écologique, qu'il soit vu globalement (l'écosphère) ou sur l'un de ses détails, est en perpétuelle transformation (changement de formes des états) sous l'effet des cinèses (flux, débits) qui échangent matière et énergie. D'une façon générale tout ensemble peut être analysé pour décrire sa diversité morphologique et fonctionnelle. Deux ensembles distincts dans l'espace, qui peuvent paraître semblables à certains égards, vont apparaître différents si on les analyse différemment soit plus en détail, soit plus globalement, soit avec d'autres analyses, soit à d'autres moments. Dans l'absolu, seule l'hétérogénéité existe, à tous les niveaux d'observation.

Les trois référendaires du champ écologique sont en pratique hétérogènes. Toute portion d'espace comporte une hétérogénéité interne, tout niveau de composition peut être décomposé en éléments plus "fins" et plus précis. Quant au temps, s'il semble s'écouler de façon continue, régulier du point de vue chronologique, du point de vue de l'ensemble des systèmes biophysicochimiques il est en fait profondément hétérogène. Sa signification est notamment totalement différente en fonction des rythmes "jour-nuit" (nycthéméraux), saisonniers (annuels) et pluriannuels (cycle de 11 ans, etc.), ces cycles étant dus notamment à la relation complexe de notre planète avec sa principale source énergétique (le soleil), animatrice essentielle des systèmes écologiques.

On devra toujours acquérir des données, puis organiser et interpréter celles-ci pour en étudier l'hétérogénéité.

Certains ensembles de données, par rapport à d'autres, sont moins hétérogènes. Cette identité relative très localisée dans la globalyse du champ écologique (c'est-à-dire somme des données analytiques acquises dans le champ écologique) permet alors de parler d'ensembles homogènes. L'homogénéité est un cas particulier de l'hétérogénéité et n'est usuellement qu'une hypothèse commode en vue d'une approche synthétique. En toute rigueur, on ne fait qu'étudier et comparer des hétérogénéités plus ou moins grandes, l'homogénéité n'étant qu'un cas limite de celles-ci.

Dans la pratique de l'élaboration écologique toutefois, il est fréquent d'opposer plusieurs ensembles dont on souhaite que chacun d'entre eux soit (le moins hétérogène) le plus homogène possible. On parle alors par simplification, et sur la foi d'une supposition empirique, d'ensembles homogènes (champs cultivés homogènes, prise de sol homogène, paysage homogène), voire pour certains travaux, on homogénéisera (par exemple, par brassage de sol, en ne semant qu'une variété de blé, en éclaircissant une forêt, etc.). L'homogénéité est donc avant tout un concept portant sur des suppositions, pratiquement commodes, pour exprimer une opinion générale abstraite. Inversement l'hétérogénéité, plus ou moins grande, est mesurable dans les ensembles (considérés initialement, ou non, comme homogènes) par des expressions statistiques.

Il n'est pas rare que l'étude de l'hétérogénéité à l'intérieur d'un ensemble (par la comparaison de sous-ensembles) permette d'observer un changement graduel. Par exemple, la flore d'un territoire finit par changer totalement sur un autre mais entre les territoires il y a graduellement disparition des espèces du premier et apparition d'espèces du second. Il y a au niveau de perception choisi et pour le type d'analyse impossibilité de situer dans le champ écologique un changement brutal : il y a continuum [GODRON 1984]. Discontinuité

et continuum constituent un autre doublet conceptuel à la base de multiples "difficultés" d'interprétation. Pour reconnaître, classer, l'information écologique, les écologues sont conduits en effet à définir des classes dont souvent (mais pas toujours) les limites sont arbitraires ou dépendent d'un niveau de perception dans le champ écologique si mal défini que les controverses, souvent par défaut d'information, peuvent exister.

3.3.6. Champ écologique et choix des données : analyse, synthèse et globalité

La démarche analytique (analyse) est assez généralisée pour ne pas poser de problèmes ici (si ce n'est celui du choix du niveau des observations retenues pour arrêter l'analyse). Tout dépend des objectifs et moyens disponibles ; une analyse plus détaillée entraînera toujours pour corollaire un accroissement des données, des coûts d'acquisition et parfois des difficultés de vraie synthèse. On peut donc choisir, selon les moyens disponibles, soit de concentrer l'effort pour analyser à un niveau d'observation détaillé sur un nombre de points d'étude restreint, soit accroître le nombre de points d'étude pour une acquisition de données plus globales.

La synthèse offre par contre plus de difficultés car elle est souvent confondue avec globalité et avec généralisation. La synthèse est la reconstruction à partir des éléments analysés d'un ensemble plus complexe ; c'est une recomposition d'un ensemble préalablement décomposé. Pour que cette recomposition aboutisse à la composition initiale, il faut et il suffit que l'on possède toute l'information sur le contenu des sous-ensembles de l'ensemble décomposé et sur l'organisation de ces sous-ensembles.

La globalité est une perception qui d'emblée reste à un niveau d'observation sans décomposition et recomposition. Comparer les résultats d'une étude globale (pour un niveau d'observation) avec les résultats obtenus après analyse puis synthèse (décomposition et recomposition) jusqu'à ce même niveau est un test de la qualité de ces analyses et synthèses.

La généralisation est une démarche purement subjective d'extrapolation de données éparses à des ensembles plus vastes. La généralisation n'est pas en soi fausse... mais n'est souvent pas réfutable directement à partir de ses éléments constitutifs.

Au plan technique, analyse, synthèse et globalité portent sur des données et sur leur organisation et sont réfutables tandis que la généralisation est une approximation subjective généralement non construite pour être testée.

3.3.7. Données objectives et hypothèses

En pratique, les connaissances écologiques sont fondées sur des données. Ces données ne représentent qu'une infime partie de l'information potentiellement présente dans le champ écologique. Cette fraction résulte des hypothèses, nécessairement réductionnistes d'options. Ces options limitent le travail à une problématique (et non à d'autres qui sont écartées), puis à un objectif (et non à d'autres), puis... Cela réduit l'information acquise à partir d'options qui sous-entendent hypothétiquement que ces choix sont les moins mauvais possibles pour représenter le réel. Si ces options "réduisent" les données disponibles, elles n'en modifient pas la valeur intrinsèque, les données reflétant directement notre perception, même limitée, de l'objet d'étude (= données objectives).

Mais ces données initiales réduites et analytiques ne permettent pas isolément de conclusions générales de type synthétique. Pour les synthèses, une autre famille d'hypothèses les propositions tendent à transformer les Données Initiales Contrôlées (DICs) en Données Secondaire Elaborées (DSEs) qui en résument certaines propriétés tout en posant l'hypothèse que ces DICs sont suffisantes pour que les propriétés issues des DSEs soient celles du système réel. Les propositions sont une famille d'hypothèses (choix des données préalablement retenues pour constituer des ensembles, prédicats de fonctions mathématiques, suppositions, assemblages des éléments d'un modèle, extrapolations,...) dont dépend intrinsèquement la qualité des DSEs, mais nullement les DICs.

Les données initiales sont donc directement issues du champ écologique après un certain nombre d'hypothèses (options) réduisant considérablement notre connaissance par rapport à l'ensemble des données qui décrivent totalement ce qui est étudié par l'écologue (= l'écosphère). Les propositions sont des hypothèses qui d'une certaine façon tendent à suppléer à la carence des données initiales par rapport à la réalité. Tout l'art des choix d'acquisition est de prendre des options aux faibles conséquences sur notre perception du réel et l'art des interprétations est de retenir des propositions qui permettent opérationnellement une connaissance que l'on peut confronter à la réalité de l'écosphère. La pertinence plus ou moins grande des options (choix réductionniste des données objectives) et des propositions (hypothèses généralisantes de synthèse) doit pouvoir être appréciée face au réel dans le cadre des interprétations (= processus de validation).

3.4. FORMES ET FONCTIONS

3.4.1. Notion de caractéristique

Une caractéristique est un trait qualificatif, quantitatif ou flou relatif à un organisme, à un système vivant simple ou complexe ou à son milieu. L'acidité du sol, le nombre d'animaux d'une espèce, le nombre d'espèces, la couleur de la feuille, le taux respiratoire sont des caractéristiques.

Le mot caractère, plus court, a été écarté en raison de son usage en informatique, une caractéristique étant décrite par plusieurs caractères (nécessitant chacun par exemple 8 bits).

Le mot descripteur est, par ailleurs, réservé aux opérations d'interprétations, un descripteur pouvant notamment être une combinaison quelconque de caractéristiques ou une classe des valeurs de l'une d'entre elles.

Descripteurs et caractéristiques sont des variables pouvant prendre plusieurs valeurs qualitatives, quantitatives ou floues. Les caractéristiques peuvent se rapporter à des états ou des cinèses.

3.4.2. Etats et cinèses

L'observation de caractéristiques peut conduire à deux types de données initiales très différentes :

1) des qualités et des quantités présentes à un instant t (nombre de pierres, couleur de la feuille, poids du bétail, concentration en zinc d'un tronc d'arbre, etc.) ; il s'agit de caractéristiques liées à des formes ou structures, ou états. Nous les appellerons caractéristiques d'état. La durée est négligée : la caractéristique est alors instantanée et se rapporte à une date.

2) des débits et flux liés à l'une des caractéristiques, exprimées en fonction du temps dq/dt, par exemple nombre d'individus morts par seconde, quantité de CO_2 absorbée à l'heure. La donnée est liée à une durée (deux repères dans l'axe du temps). La valeur est fonction d'un temps. Nous appellerons cinèses les caractéristiques décrivant ces débits ou flux.

L'observation écologique porte sur des états ou/et des déplacements (mouvements) d'objets, c'est-à-dire leur cinèse. Ces deux types d'observations sont de nature différente. Elles peuvent éventuellement être simultanées, mais doivent être très fondamentalement et très clairement distinguées, les méthodes mises en oeuvre et la signification des observations faites pour une cinèse ou un état étant totalement différentes.

L'accumulation de données sur des états à l'instant t est de la morphologie. L'acquisition de données sur des cinèses pendant dt est de la cinétique écologique [BOUCHé 1990].

Le mot état est pris ici au sens strict "configuration des corps et des objets" mais aussi au sens plus large "toutes propriétés physiques de ces configurations" prises instantanément (t = 0) par exemple : masse ou volume d'un objet. Il s'agit de la somme des diverses propriétés que l'on sait enregistrer d'un objet observé, dans la pratique, dans un temps suffisamment bref et adéquat pour que les propriétés observées ne soient pas altérées en fonction de la durée d'observation.

Par un changement d'échelle spatiale on peut observer les composants d'un objet et la façon dont ils sont organisés en structures plus ou moins complexes, ou répétitives selon des textures différentes.

On peut dénombrer le nombre de vaches d'un pré, mesurer la longueur d'une tige, peser la proportion de sable d'un sol, mesurer la distance entre deux noeuds sur la tige, ou entre deux particules de sable dans le sol... Il s'agit à chaque fois de quantifications de composants, et éventuellement, de description de leur organisation réciproque, ces composants étant des parties d'un ensemble plus vaste et étant éventuellement susceptibles d'être à leur tour analysés en composants plus fins. Ces observations se rapportent à des objets descriptibles qualitativement et quantitativement, voire de façon floue, à un moment donné. L'observation est considérée comme instantanée (ou sinon on adopte une autre échelle spatio-temporelle).

Toutes ces observations se rapportent à des formes d'individus, de sols, de sédiments, de dispositions spatiales des éléments. Si le "temps" est évoqué c'est en tant qu'élément descriptif par une date à un instant : il n'y a pas de durée mais seulement situation temporelle.

La morphologie écologique s'appliquera donc à tout état portant sur un objet écologique observé instantanément (la modification de l'objet durant l'observation est négligée) à une date situant temporellement l'observation.

Les cinèses sont, à l'opposé des états, les observations sur les mouvements d'objets, quantité ou/et qualité changeant de situations spatio-temporelles. Ces phénomènes nécessitent des observations en fonction de la durée du temps et s'expriment en fonction de celle-ci dans l'espace. L'espace peut ne pas être traduit sous forme de coordonnées cartésiennes mais par référence à un objet (exemple : l'absorption d'oxygène par un individu vivant implique la mesure vers cet être vivant pendant un temps donné d'un mouvement gazeux). Il s'agit de mouvements exprimant une quantité en fonction du temps, c'est-à-dire un **débit**. L'une des propriétés de ce débit peut être une vitesse par unité de surface orientée perpendiculairement au vecteur vitesse (= un flux). Un **flux** ne porte pas sur des quantités mais sur des vitesses.

Ces cinèses jouent un rôle essentiel en écologie, qu'il s'agisse de mouvements d'éléments chimiques indispensables aux êtres vivants, des déplacements des êtres vivants et des diverses régulations chimiques ou biochimiques de ces cinèses. En fait, toute transformation fonction du temps (évolution) des systèmes écologiques est nécessairement liée à des cinèses. Ici encore d'importants changements d'échelle dans l'espace et le temps existent.

3.4.3. Relations entre états et cinèses

Les cinèses, mouvements des objets, modifient les états (les formes). Les formes, ou structures, sont l'organisation spatiale des états. Les relations entre états et cinèses sont celles de systèmes. Par exemple : la texture sableuse (état ou forme) d'un sol permettra le débit des fluides (cinèse) et réciproquement l'intensité de l'anabolisme (cinèse) permettra l'élaboration d'une nouvelle tige (forme). Le dénombrement d'un troupeau dans un pré est modifié par l'arrivée ou le départ (cinèses) de quelques individus de ce troupeau dans le pré ; l'assimilation chlorophyllienne (CO_2 + H_2O) accroît la masse végétale ; le lessivage d'un sol modifie sa composition, sa texture et sa structure, ... Toutes ces modifications s'observent en deux temps t_1 et t_2 entre lesquels des cinèses (parfois contradictoires) se produisent.

3.5. ACQUISITION D'INFORMATION

3.5.1. L'élaboration des connaissances écologiques

L'élaboration est prise ici, vis-à-vis d'une science, au plein sens du terme : il s'agit d'élaborer les connaissances écologiques, c'est-à-dire d'en accroître l'étendue, d'en promouvoir la qualité et d'en assurer la disponibilité, notamment à des fins socio-économiques.

Il s'agit, au plan quantitatif, du travail nécessaire pour obtenir de nombreux résultats au coût minimal et, au plan qualitatif, d'améliorer en permanence notre compréhension de notre milieu.

Pour cela, il faut constamment développer ce qui est littéralement essentiel aux faits, en organisant faits et interprétations pour un progrès permanent vers des connaissances plus rigoureuses et efficaces. Cette continuité opérationnelle impliquant à tout instant un plein accès à l'état juste antérieur des connaissances et une critique réelle et objective des interprétations sur l'objet : c'est-à-dire tout ou partie de l'écosphère.

L'étendue des connaissances est avant tout un phénomène quantitatif : c'est accumuler les données écologiques vis-à-vis d'un milieu particulier, en fonction d'une action humaine, autour d'une espèce quelconque ou par rapport à un processus défini..., mais c'est aussi un problème d'organisation des données pour les rendre accessibles et donc exclure leur perte pour la communauté humaine.

La qualité de l'écologie est évidemment dépendante du savoir acquis, donc de la quantité des données disponibles susceptibles d'interprétations scientifiques réitérées (aucune interprétation ne pouvant être considérée comme achevée), mais elle résulte aussi de la pertinence des interprétations qui en sont issues.

Cette pertinence est elle-même fonction à la fois de la qualité des données, de leur étendue (quantité) et de l'importance de l'analyse-synthèse pratiquée permettant une élimination systématique des interprétations spéculatives au profit de conclusions réfutables. C'est enfin permettre une utilisation rationnelle des connaissances par l'homme agissant dans l'écosphère.

Ces connaissances sont d'abord basées sur des faits constatés et enregistrés sous forme de données initiales ; celles-ci, après corrections (= les Données Initiales Contrôlées), sont la

partie perçue des systèmes écologiques. C'est sur elles que s'appuient les interprétations. Les données initiales (après élimination d'erreurs et avant toute interprétation) constituent la partie intangible de la connaissance écologique.

Les données ainsi acquises se présentent comme une somme d'atomes factuels, une masse qui ne prend signification que par la constitution d'ensembles de données interprétées, puis par mise à disposition de ces connaissances.

La grande diversité des faits observés et la carence des moyens de synthèse ont longtemps été si contraignantes que l'utilisation de ces connaissances s'est faite, et se fait encore largement, selon un mode archaïque conduisant à un gaspillage considérable des acquis et à une grande carence de réfutabilité scientifique effective.

L'acquisition des données objectives (sur l'objet d'étude écologique) permet de poursuivre un ou plusieurs objectifs finalisés. C'est la motivation des travaux de recherche car :

1) transcendant tous les objectifs, les acquisitions de données visent à nous apporter une connaissance sur l'écosphère, c'est-à-dire sur notre environnement,

2) chaque objectif particulier tend à répondre à une question (plus ou moins explicite) et le premier acte n'est pas nécessairement (ou plutôt ne devrait pas être) l'acquisition de données nouvelles mais d'abord l'utilisation des connaissances déjà acquises... ce qui implique une acquisition, une gestion, une exploitation logique et rigoureuse de données obtenues antérieurement pour d'autres finalités.

Les démarches d'acquisition, de gestion et d'interprétation des dōnnées procèdent mutuellement l'une de l'autre pour donner des réponses temporairement locales (limitées dans l'espace-temps) mais pouvant aussi s'intégrer dans notre connaissance globale de l'écosphère.

3.5.2. La donnée : l'élément d'information sur le champ écologique

L'approfondissement des connaissances implique de toute manière une accumulation de données relatives à l'objet d'étude de l'écologie : l'écosphère.

En pratique l'écologue, nécessairement spécialisé, exerce des choix en réduisant cet objectif global à ce qu'il peut effectivement appréhender. Il effectue d'emblée une série

d'options réductionnistes ou simplificatrices avec l'espoir que leurs conséquences (= élimination des autres informations pouvant être acquises) auront des effets négatifs minimaux sur les interprétations qu'il désire.

Mais la donnée est rarement étudiée individuellement, elle l'est comme élément d'ensembles de données, aujourd'hui possibles à gérer par des Bases de Données Relationnelles permettant d'optimiser l'usage des liens écologiques. De sorte que l'acquisition de données se fait pour compléter ces ensembles ; les données peuvent alors être appréciées par rapport à ce qu'elles apportent à ces ensembles après avoir été saisies sur un support informatique.

En fait, ce jugement est relatif aux modalités d'interprétation lesquelles peuvent, pour une même donnée, varier considérablement en fonction de la constitution des ensembles de données et des objectifs poursuivis.

Les combinaisons de données peuvent être très diverses, une même donnée pouvant être réutilisée pour de nombreuses études de finalités écologiques différentes grâce à son référentiel la liant au champ écologique (par exemple une donnée relative à un animal observé pour un recensement quantitatif dans une forêt ... peut contribuer à la cartographie européenne de l'espèce observée).

Les liens naturels entre données écologiques permettent de constituer pour ces dernières un référentiel unique permettant d'explorer tous types de données objectivement écologiques.

En définitive, et au-delà des objectifs spécifiques qui ont leurs exigences immédiates, la donnée écologique contribuant directement à notre connaissance de notre propre milieu, peut être recombinée indéfiniment et de façon optimale, moyennant le respect de quelques principes élémentaires.

L'acquisition de données est la base de tout travail scientifique et de tout progrès technique. Ces données étant ensuite interprétées. L'interprétation des données conduit à des conclusions, des décisions techniques ou de nouvelles connaissances et à de nouvelles hypothèses scientifiques. Cette interprétation exige enfin d'être refutée par un retour éventuel aux premières données ou/et par de nouvelles acquisitions de données.

L'acquisition d'une donnée est l'ensemble des opérations concrètes que l'on fait antérieurement et jusqu'à l'obtention d'une donnée initiale enregistrée correctement (y

compris son contrôle). Cette définition porte sur le concret et exclut donc les démarches hypothétiques conduisant à cette acquisition d'informations.

3.5.3. Donnée et liens

Une donnée est un atome (au sens littéral : insécable) d'information écologique. Elle peut être qualitative : rouge, *Lumbricus terrestris*, France ; ou quantitative : 22 ; 684,5...; ou enfin floue : vieux, grand, lourd.

Isolément une donnée n'a aucun sens : "rouge", "g" ou 2,5 n'ont aucun sens alors que *Lumbricus terrestris* rouge commence à signifier quelque chose grâce au lien entre *Lumbricus terrestris* et rouge, et encore plus entre *Lumbricus terrestris* rouge (de) 2,5 g.

Une donnée pour être signifiante est donc reliée par des liens à d'autres données ; la description de ces liens logiques constitue une propriété naturelle complémentaire aux données.

Une donnée fait partie d'un ensemble : 2,5 g font partie de l'ensemble des nombres (2,5) et g des unités de mesures pondérales lui-même sous-ensemble des unités du Système Métrique International. Les données des ensembles peuvent en effet appartenir par leurs liens à plusieurs ensembles, et chaque ensemble être éventuellement subdivisé en sous-ensembles (jusqu'à la donnée) et être réciproquement lui-même un des sous-ensembles d'ensembles plus globaux. Les liens sont alors des propriétés d'ensembles de données.

Une donnée "atome d'information" fait donc partie d'ensemble(s) où chaque donnée-atome est lié à d'autres données-atomes. Une certaine configuration "moléculaire" multidimensionnelle des liaisons de ces "atomes" en résulte.

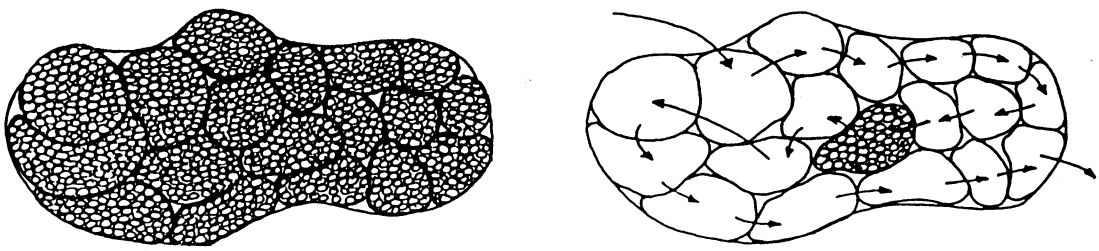


Figure 3-3 : *Représentation "ésotérique" d'un ensemble, sous-ensembles et liens entre les données.*

Exemple : *Lumbricus terrestris* - rouge - 2,5 g - pourra constituer une "molécule" (vue simple) qui peut devenir une "macromolécule" extrêmement complexe. Par exemple, *terrestris* est lié au genre *Lumbricus* (ensemble des espèces de ce genre), ..., l'individu de 2,5 g en cause est lié à une portion d'espace liée à un dispositif expérimental, un type de sol, un climat, un dossier de recherche, ce dernier à un contrat, un chercheur responsable...

Une information peut porter sur un objet particulier, précis, par exemple : un individu de *Lumbricus castaneus* (Sav. 1826) pèse 360 mg. Ceci comporte une information objective, le poids de l'animal, mais la phrase décrit également un dénombrement (=1), une espèce : *Lumbricus castaneus*, une relation générique (*castaneus* est dans le genre *Lumbricus*), le nom d'auteur abrégé (Savigny), l'indication que cet auteur n'a pas attribué cette espèce initialement dans le genre *Lumbricus* (la parenthèse) et l'année de première publication de la description par Savigny : 1826. Il s'agit d'autant de données qui peuvent être contrôlées et exploitées selon des logiques dépendantes ou indépendantes.

En pratique, on acquiert des données pour au moins un <u>objectif</u> de sorte que, par exemple, l'on étudiera la variation pondérale de l'individu décrit ci-dessus en fonction du temps. Toutes les informations autres restent constantes et l'on enregistre donc celles-ci une fois et une seule tandis que le poids et le temps, variables, seront enregistrés plusieurs fois. Ils sont liés à l'ensemble poids et liés à la série des dates.

On constate par cette description que <u>deux concepts</u> vont caractériser l'information écologique : les données sont <u>enregistrées</u> et elles sont <u>organisées</u> par des liens en ensembles. L'information a deux supports : les données et les liens (ou relations).

Ainsi dans notre exemple du *Lumbricus castaneus*, l'individu sera à chaque fois extrait de son milieu, nettoyé, porté sur une balance... et le poids ainsi que la date de collecte seront enregistrés en relation avec son identité. On <u>contrôlera</u> ensuite que la série de valeurs pondérales obtenue n'a pas de données aberrantes (= dans l'échelle des poids possibles pour le taxon : détection d'erreur d'enregistrement et éventuellement d'identification) et l'on enregistrera définitivement les données ainsi contrôlées comme "la meilleure perception objective".

L'acquisition d'une donnée peut faire appel à des techniques très simples (observation directe) ou très sophistiquées (analyses chimiques complexes pour séparer un constituant précis avant sa quantification par des systèmes automatiques d'étalonnage en fonction d'une concentration de référence). Dans ce deuxième cas, bien que l'objectif ne soit que d'acquérir une donnée, on enregistre en fait de très nombreuses "prédonnées" relatives à la

techniques d'acquisition car chaque modalité technique modifie le résultat final, lequel n'apparaîtra qu'après un calcul sur d'autres prédonnées préalablement enregistrées (à moins que le dispositif soit informatisé et effectue directement le calcul de la donnée recherchée). On conçoit dès lors que l'acquisition de données puisse faire appel à des techniques délicates et complexes.

Le mot **donnée** très général désigne donc un atome d'information décrivant pour partie un objet concret. Cette information a en quelque sorte deux raisons d'être :

1) elle est un acquis assuré initialement pour un objectif scientifique ou technique et de ce point de vue, elle est le résultat d'options ayant une motivation initiale précise ;

2) elle est par ailleurs une information en général et de ce point de vue, doit être gérée dans une Base de Données pour permettre de futures utilisations et un éventuel travail synthétique.

Ce dernier aspect est essentiel en écologie (science globale) où la synthèse à partir de données analytiques est fondamentale. Cela implique que la donnée enregistrée puisse être individuellement indépendante de l'objectif initial d'acquisition non dans son choix mais dans son référentiel, ce qui n'exclut pas une liaison explicite à cet objectif.

Il faut maintenant établir une distinction très nette entre la Donnée Initiale Contrôlée (DIC) issue du champ écologique et la Donnée Secondaire Elaborée (DSE) issue d'interprétations, même minimes. La première doit être le reflet le plus exact et rigoureux possible de la portion du champ qu'elle est sensée représenter et être directement référée à celui-ci. La seconde résulte d'un choix interprétatif impliquant une proposition (par exemple : paramètre, total d'effectifs, moyenne,...).

Les DICs résultent d'options de sorte qu'un ensemble de DICs (par exemple, 3 valeurs de la biomasse végétale d'une prairie) peuvent être considérées selon des appréciations externes comme insuffisantes pour représenter un tout (une estimation de la biomasse de cette prairie). C'est l'ensemble qui est insuffisant, c'est la moyenne, une DSE calculée sur 3 DICs, qui est suspecte, mais chaque DIC prise isolément est une valeur représentant une biomasse d'un prélevat de ce champ et reste valable pour toute ré-interprétation différente (notamment dans ce cas par ajout de données pour établir une DSE "moyenne" plus fiable).

Les DICs sont les données objectives issues du champ écologique ; les DSEs sont des interprétations des DICs. L'une et l'autre donnent une perception de la réalité écologique. Les DICs dépendent seulement de l'acquisition, tandis que les DSEs dépendent aussi des

hypothèses de l'interprétation (= propositions). On conservera donc les DICs comme informations fondamentales sur les systèmes.

A côté de ces études écologiques, le chercheur peut être conduit à simuler dans des dispositifs artificiels (que nous nommerons microcosmes) des modalités expérimentales reflétant les hypothèses qu'il porte sur certains mécanismes. Ces microcosmes fournissent des Données Expérimentales sur les Potentialités (DEPs) des processus proprement écologiques. Ces DEPs, acquises hors champ écologique, n'ont pas de référentiel avec celui-ci et ne dépendent que du dispositif artificiel choisi et des hypothèses de ce choix : DSEs et DEPs dépendent intrinsèquement de propositions et non les DICs.

Il existe aussi des prédonnées, n'ayant qu'une valeur transitoire au niveau des dispositifs techniques d'acquisition. Ces prédonnées (telle la tare d'un récipient) sont sans signification par rapport à l'objet d'étude (le prélevat) mais permettent d'intervenir dans le calcul d'une DIC ou DEP. DIC, DSE, DEP sont des abstractions, la DIC est directement concrétisable par son lien au réel, DSE et DEP sont réfutables.

3.5.4. Notion de prélevat

Le prélevat est un objet concret quelconque, donc une portion de l'espace-temps du champ écologique. Par relevé, prélèvement ou/et interception, on obtient directement, ou après analyse (au laboratoire), une ou plusieurs données caractérisant la composition de cet objet qui est vis-à-vis de cette(ces) caractéristique(s) insécable ; pour un même prélevat chaque donnée est de nature (caractéristique) différente. Réciproquement, le prélevat est l'objet concret ayant fourni pour chaque caractéristique une DIC et une seule [BOUCHé 1990].

Le prélevat est défini d'abord par l'information écologique (la ou les DICs) qu'il fournit et est limité dans l'espace-temps par ses cotes (ses dimensions spatio-temporelles). C'est l'objet du champ écologique auquel la donnée est liée. Il est dans l'écosphère la limite spatio-temporelle inférieure de l'analyse vis-à-vis de l'espace-temps. C'est le "grain" le plus fin de l'information disponible de cette portion spatio-temporelle. Le choix des cotes est donc une option limitant les futures interprétations. Niveau composition, le prélevat peut fournir plusieurs données mais chacune appartenant à une caractéristique différente. Par exemple, une même prise de sol peut donner un pH, un dénombrement des acariens oribatides *Oppia minus*, un taux de carbone, une proportion de ^{15}N dans l'azote du prélevat, son propre volume, ...

Les données sont toujours liées à un prélevat ayant potentiellement des cotes spatio-temporelles. Dans l'absolu, la forme (espace) et les dimensions associées à cette forme font partie des caractéristiques (cotes spatiales) liés aux autres données du prélevat. Le temps de début de collecte et celui de fin permettent de situer temporellement le prélevat par la date (par exemple dans les jours du calendrier) et sa durée (par exemple de 6h 30 à 7h 10 = 40 mn). Le prélevat porte donc sur les trois référendaires du champ écologique : espace, temps et composition.

Le prélevat est l'objet concret sur lequel on effectue une acquisition d'informations. Il n'est pas nécessairement le prélèvement physique de l'objet concret (par exemple dans les relevés de la végétation, dans les mesures directes de cinèses d'un anémomètre, ...). C'est même parfois l'inverse: un apport dans le système écologique (par exemple la pluie ou encore l'acte correspondant au semis).

Le prélevat est donc liée à une acquisition d'informations traduites finalement sous forme de DICs. Ses caractéristiques spatio-temporelles, c'est-à-dire sa situation et ses cotes dans le champ écologique, le caractérise.

Le prélevat étant la portion d'espace-temps permettant d'obtenir une donnée d'une caractéristique, ou plusieurs données de caractéristiques à chaque fois différentes, a une importance littéralement essentielle puisqu'il est la source de DICs.

Derrière une apparente grande variété de "méthodes de prélèvement" il n'y a en fait qu'une opération : on acquiert l'information dans le champ écologique.

Pour comprendre la notion de prélevat, on pourrait imaginer la démarche suivante qui est l'inverse de la démarche expérimentale : On a une donnée, liée à une caractéristique. D'où provient-elle ? Quel est son référentiel ? Conclusion : on a déterminé son prélevat.

Ce concept joue un rôle clé en écologie opérationnelle. La donnée d'une caractéristique sera ensuite liée aux autres propriétés de son prélevat. Une information (par exemple pH = 5,7) est indissociablement liée aux caractéristiques de son prélevat (pH de quoi ?). La donnée, atome d'information, n'a de sens que grâce aux liens permis par le prélevat.

Fondamentalement un prélevat a les caractéristiques spatio-temporelles d'une portion du champ écologique dans lequel il est délimité par ses cotes définissant son volume (par exemple : x, y, z) et sa durée (d) auxquelles on peut lier les éléments étudiés de sa composition (nature des données issues de ce prélevat). Dans la pratique, ces cotes ne sont

pas toujours précisées car négligées par les options qui limiteront les possibilités d'interprétation.

Le prélevat peut être formé de sous-prélevats. Dans un certain nombre de cas, les observateurs prennent plusieurs sous-prélevats et les mélangent avant d'en exprimer les caractéristiques mêlées. C'est notamment le cas lorsque les moyens d'analyses font défaut et obligent à médiatiser le milieu pour réduire le caractère ponctuel d'une valeur. Ce choix, très peu souhaitable, est un compromis introduisant un flou important sur la signification spatio-temporelle des données issues de tels prélevats. Présentement, la non-réflexion sur le prélevat, en tant que source de données, conduit très souvent à une perte d'information pure et simple.

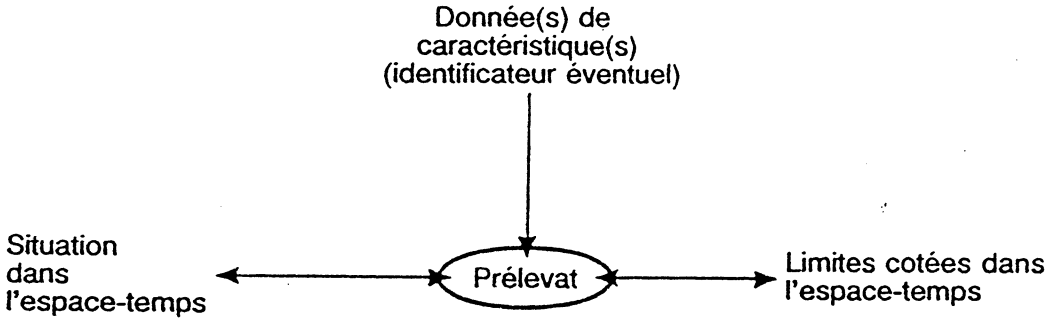


Figure 3-4 : *Caractéristiques, situation et limites d'un prélevat.*

La définition des caractéristiques du prélevat (= situation X, Y, Z, D et cotes x, y, z, d) est donc cruciale pour toute l'écologie car elle est une des limites essentielles à tout transfert, synthèse ou globalité de l'information.

3.5.5. Sélection des données et options

Les choix effectués lors de l'acquisition résultent d'hypothèses dénommées options. Ces options portent sur la pertinence des caractéristique retenues pour représenter au mieux la portion du champ écologique sélectionnée.

Si l'on considère le champ écologique total de l'écosphère, aucun écologue, aucune équipe de recherche ne peut le prendre en compte. Obligatoirement une problématique (de préférence explicite) conduira à sélectionner un domaine d'étude puis à focaliser sur certaines variables.

D'un point de vue plus opérationnel, on précisera même des objectifs de recherche et l'on constatera que l'on peut :

- déjà disposer des données,
- acquérir de nouvelles données nécessaires en tenant compte des moyens et techniques d'acquisition disponibles,
- ne pas disposer de données antérieures, ni acquérir de données nouvelles, car l'on manque de techniques adéquates et que ces techniques doivent préalablement être développées,
- manquer de moyens humains ou financiers (ou simplement manquer d'idée et/ou d'imagination).

Il y a donc successivement six "réductions" par :

1) option sur la problématique,
2) option sur l'objectif (concret),
3) option sur la localisation d'étude une portion d'espace-temps de l'écosphère,
4) option sur le prélevat : la modalité de sélection dans cet espace-temps global des portions d'espace-temps à laquelle chaque donnée est liée,
5) option sur la distribution spatio-temporelle des prélevats dans le site,
6) options analytiques : les techniques d'analyse chimique, physique ou biologique permettant de tirer de la composition globale des prélevats, les DICs de caractéristiques sélectionnées.

Enfin les modalités de saisie de données seront arrêtées ; celles-ci n'ont pas d'influence sur la qualité de la globalyse (fraction connue, observée du champ écologique) mais sur son volume (= efficacité d'acquisition).

Les 6 options doivent conduire à des données individuellement indépendantes des interprétations. Les options se font pour un objectif mais les DICs une fois acquises doivent être indépendantes intrinsèquement de cet objectif.

Il est en effet souvent très difficile de dissocier intellectuellement les interprétations des analyses car les recherches sont faites pour un objectif visant à une interprétation. Si cet objectif est légitime et doit être explicite, le mode d'acquisition une fois arrêté doit fournir des DICs (données analytiques) non pré-interprétées : si l'on y prend garde, c'est toujours le cas.

Avec les moyens modernes, les pré-interprétations doivent être éliminées très souvent. En écologie, avec l'informatique classique, les données ont été après analyse, codées avec perte d'information . Par exemple les pH des sols sont enregistrés par lots de valeurs, regroupées en 1/2 unités pH, alors que la meilleure information est celle de la technique utilisée (par exemple celle du pH-mètre : 5,6 ; 5,7 ; 5,8...). La définition du lot résulte d'une pré-interprétation, d'un jugement de valeur, qui peut toujours être remis en cause. Les contraintes d'espace sur fichier justifiaient ces pré-interprétations ; elles ne doivent plus aujourd'hui être pratiquées qu'au niveau de chaque interprétation et en aucun cas lors de l'acquisition puis de la gestion des DICs.

Dans certains cas, cette pré-interprétation est liée à l'option analytique. Par exemple, avec l'usage d'un technique donnant une caractéristique que l'on considère comme représentative d'une autre.

Cette transposition interprétative devra être évitée au stade acquisition et il y a lieu de noter la DIC obtenue et non celle souhaitée. Ceci doit être respecté même si à l'interprétation on considère la différence entre DIC obtenue et donnée souhaitée comme "négligeable" ou si l'on pratique ultérieurement cette conversion. Cette opinion (= "négligeable") n'a de sens que dans le cadre d'une interprétation spécifique ; la même DIC dans un autre cas pourra être considérée avec sa vraie signification.

Indiquer par exemple que le semis d'une céréale a eu lieu 153 jours avant la récolte est une information qui ne peut être utilisée que pour une interprétation de croissance. Noter l'acte semis le 10/2/88 et l'acte récolte le 10/6/88 permet d'autres interprétations (par exemple sur les dates de semis pratiquées en moyenne dans une région). L'acte semis lui-même est en fait :

a) un apport biologique : Exemple maïs 80000/grain/ha,

b) un passage de tracteur-semoir dont on peut noter, selon divers objectifs, la pression moyenne/m^2 sur le sol ou/et la durée du travail ou/et la consommation/ha de fioul.

Il est fréquent que l'on ne dispose plus que de données pré-interprétées qui, faute d'un référentiel cohérent, sont inexploitables. Alors que le travail d'acquisition avait été fait, ces résultats initiaux utilisables à tous points de vue, sont perdus ; ce qui ne se serait pas produit si l'enregistrement des DICs avait été correctement conduit.

La pré-interprétation est une pratique très courante, qui doit être débusquée, car elle constitue, sans contrepartie, une perte archaïque de nombreuses informations.

3.5.6. Liens des données au champ écologique

Notons que la donnée ne vaut extrinsèquement que par ses liens. Ce sont eux qui permettront de l'associer dans des ensembles de données. En outre, en écologie la donnée devra obligatoirement être écologiquement référée sinon elle ne contribuera pas à une interprétation.

Certaines données usuelles peuvent ne pas être écologiques, telles un insecte de collection sans étiquette : il peut faire l'objet d'une étude zoologique (morphologie comparée) mais non écologique ; c'est encore le cas d'un dispositif expérimental arbitraire (microcosme sans référence démontrée au champ écologique "naturel", par exemple les expérimentations sous serre chauffée artificiellement).

Il ne s'agit pas uniquement d'accumuler des données analytiques mais de le faire pour permettre d'interpréter synthétiquement la globalyse (somme de données objectives) puis confronter cette synthèse aux données globales que nous avons du champ écologique. Une donnée doit donc être initialement reliée à d'autres données à la fois pour atteindre un objectif particulier et à la fois pour éviter sa perte dans un quelconque ensemble de données non reliables au champ écologique.

La donnée est d'abord repérée dans l'espace-temps par sa situation qui dans l'écosphère pourrait se résumer au niveau général par la longitude, la latitude, l'altitude et la date ; en pratique des modalités précises et systématiquement méthodiques existent pour situer une donnée.

La donnée est concrètement issue d'un prélevat qui peut avoir fourni plusieurs données ; le prélevat est un lien entre les données décrivant ce prélevat.

Le prélevat peut être prélevé *in situ*, puis la prise analysée (cas de la collecte d'une prise de sol, d'eau, de bois, ...). Il peut être l'objet d'un relevé par analyse directe *in situ* ou transposition sur supports artificiels analysés ultérieurement (cas de la photo d'un groupe de canards en vol puis du dénombrement sur celle-ci ; cas d'un enregistrement multispectral par satellite artificiel). Dans tous les cas, le prélevat fournit une donnée d'une caractéristique ou des données sur chacune des caractéristiques différentes : il est l'origine de la donnée.

Les données issues d'un même prélevat sont spatio-temporellement liées : il s'agit d'un lien fondamental des données, extrêmement important pour l'étude future de l'information.

2) Le prélèvement

Il s'agit d'un acte séparant 1 ou n prélevats de leur contexte écologique pour les étudier souvent en mettant en oeuvre des moyens analytiques inapplicables directement *in situ*, et supposant usuellement la destruction de la prise, résultant du prélèvement, pour isoler la donnée de la caractéristique recherchée. Le prélèvement prend des formes très variées en fonction du but recherché. Sans précision, il a trait à des descripteurs d'état. Le produit du prélèvement est une prise.

3) L'interception

C'est un mode d'obtention des DICs-cinèses. Une interception-relevé permet d'obtenir directement une donnée ou un analogique cinétique (cas de l'anémomètre, du rhéomètre, du comptage d'oiseaux passant un col, d'une lecture radar de passages de papillons, etc.). Il y a traduction de débit (= quantité *vs* temps, ex : nombre d'oiseaux/heure) parfois uniquement exprimé en tant que vitesse (de l'air, de l'eau) en considérant la section du passage comme "quelconque" : il s'agit alors de flux (= vitesse *vs* surface). L'interception-prélèvement est l'interposition, dans une cinèse, d'un capteur accumulant l'élément étudié en le détournant de sa destinée "naturelle". Ceci prend les formes les plus diverses (filets de capture d'oiseaux, récepteurs de litière au cours de la chute de celle-ci depuis la frondaison vers le sol, piège-trappe d'animaux, prélèvement automatique d'aliquotes d'eau dans un réseau hydrique, pluviomètre, ...). L'interception-prélèvement permet ultérieurement des analyses poussées de la prise. Il ne se rapporte pas à un compartiment mais à une cinèse.

Si dans les interceptions-relevés, les données acquises sont souvent continues, les interceptions-prélèvements conduisent usuellement à des données discontinues dans le temps. Le choix de la durée du prélevat est important car il limitera la finesse de l'analyse temporelle.

4) La prise

C'est un objet concret qui n'est plus dans le champ écologique mais qui fournira, après analyse, des données sur ce champ. Ce sont, par exemple, les analogiques de relevé, telle une photo, les prises de prélèvements, tel un flacon d'eau contaminée, les analogiques d'interception-relevé, telle la bande d'un thermohygromètre et les produits de l'interception-prélèvement, tel le pot-piège. Avec les données non saisies, ils constituent (avant ou après analyses et saisies) des collections, c'est-à-dire des objets matériels issus de collectes et conservés pour de futures analyses.

5) Les actes (ou actions anthropiques)

Les actions anthropiques ou actes constituent les prélevats reliant les systèmes écologiques aux motivations socio-économiques qui ne relèvent pas que de la régulation spontanée de systèmes écologiques.

Le rôle des données qui en sont issues est essentiel car il permet, en respectant totalement le caractère scientifique et opérationnel de l'écologie d'articuler cette discipline aux préoccupations humaines.

Les actes ne diffèrent pas concrètement des modalités classiques de collecte de données : relevés, interception ou prélèvement, mais ils présentent intellectuellement des difficultés qui tiennent :

a) au fait que nous confondons la finalité de nos actions techniques avec l'acte agissant sur le système écologique considéré. Par exemple, on n'a pas "traité contre le parasite X" mais apporté *y* grammes/ha du pesticide *z*,

b) au fait que l'acte est souvent instantané et semble extérieur au système. Par exemple, dans le cas ci-dessus, c'est l'apport du pesticide *z* le jour *j* qui est l'acte... quels que soient les délais des réponses de cet acte dans le système écologique le recevant.

Les finalités apparaissent classiquement dans la terminologie et influent les choix analytiques en créant une confusion dans la globalyse (fraction connue, observée du champ écologique) ; on parle de semis alors qu'il s'agit d'un apport de graines lié à d'autres actes, mécaniques notamment (couteaux du semoir, charge des machines, etc.). L'acte technique doit être perçu par rapport au prélevat c'est-à-dire à la portion du champ écologique étudié. Très souvent en agriculture il s'agit de l'ensemble d'une placette choisie dans une parcelle cadastrale recevant un acte.

Il ne faut pas confondre l'acte et ses conséquences : l'apport de 400 g/ha de benomyl sur une céréale est quasi instantané, tandis que son action sur le système est étalé dans le temps (action fongique, en tant que fongicide, puis sur tout le système en tant que géodrilicide ayant des conséquences pendant des années).

Du point de vue de l'acquisition de données objectives, on notera simplement la nature de l'acte (= composition, par exemple benomyl), sa situation spatio-temporelle (date, durée, situation, par exemple : 22/06/84, 0 = instantané, placette B du lieu 1200) et sa quantification (par exemple : 400 g/ha). Il est à noter que l'acte peut avoir une durée contribuant à mesurer son intensité, par exemple : vache allaitante, 2/04/85 - 18/04/85 (ou

temporelle au cours de la durée de collecte est négligeable : nous ne souhaitons connaître que l'état à "l'instant" de collecte. Exemples :

1) le taux de sable d'un sol ne se modifie pas au cours du prélèvement et de l'analyse d'une terre ;

2) une photographie permet un dénombrement "instantané" des canards et des foulques sur un étang en Camargue ; etc. ;

3) constater qu'une parcelle est occupée par une prairie n'entraîne pas une modification au cours du constat.

En fait, cette option négligeant la durée est très fréquente dans les collectes sur les variables d'état : la durée du prélevat est usuellement négligeable par rapport à la modification des états pris en compte durant cette durée, et les données issues d'un prélevat sont alors considérées comme instantanées. Inversement, les données d'interception sont le résultats de phénomènes dépendant de la durée (d) : ici la connaissance des caractéristiques début-fin est essentielle. En conclusion toute donnée est associée à une date (D) (variables d'état et cinèse) et en outre toute cinèse est fonction de la durée (d) ; en revanche, beaucoup d'états sont seulement considérés à l'instant où l'état est constaté.

La date est la situation temporelle du prélevat et la durée, négligée ou non, sa cote temporelle.

3.5.9. Options spatiales : prélevats euclidiens, cartésiens, informes et ponctuels

Les caractéristiques spatiales des prélevats sont associées à toutes données écologiques.

Les prélevats euclidiens sont usuellement cotés en volumes cubiques, parallélipipédiques, cylindriques,.... Ils permettent au moment des interprétations , des inférences (interpolations et extrapolations) à un espace euclidien de dimensions supérieures au prélevat (par exemple, on interpole les dosages effectués sur 100 prélevats, mesurant chacun 1 litre, au volume total d'un étang). Ils sont un élément des relevés destinés à échantillonner, par exemple, une parcelle ou une région.

Très souvent la cote "hauteur" des prélevats, c'est-à-dire la composante verticale du champ écologique, est négligée car celle-ci est implicite. Les données sont alors exprimées en unité/surface (exemple : production de blé : 0,7 t/ha ; nombre d'animaux : 300 individus/m^2). Cette pratique est largement justifiée en raison du caractère essentiellement

superficiel de l'écosphère dont la dimension verticale est souvent négligeable par rapport à sa surface. De tels prélevats dont les données sont liées exclusivement à des cotes durée et horizontales sont des prélevats cartésiens.

Dans certains cas, les prélevats sont linéaires : l'information n'est plus acquise que vis-à-vis d'un transect (séries d'observations contiguës), les deux autres dimensions euclidiennes étant négligées (par exemple hauteur d'un arbre, pente, longueur en radiotrigonométrie d'un déplacement animal pendant la durée d, etc.).

Les prélevats informes négligent tout à fait l'expression des cotes spatiales, ils sont notamment utilisés dans les études typologiques où la coexistence spatiale de caractéristiques différentes est seule prise en compte. Ils tolèrent parfois une imprécision importante ; par exemple, il est possible de connaître l'existence de certaines espèces, difficiles à localiser par nos propres moyens, en étudiant le régime alimentaire d'un prédateur facile à étudier avec des moyens spécifiques (cas d'oiseaux prédateurs sachant trouver certaines proies dans un paysage où l'on ne peut pas nous-mêmes localiser ces proies). Un individu peut être également un prélevat lorsqu'il est source de DICs.

Les prélevats cotés ont, par définition, une forme spatio-temporelle, généralement simple (ligne, carré, rond), qui, une fois retenue dans un plan d'échantillonnage, ne varie souvent plus en raison du fait que la composition de la (des) caractéristique(s) observée(s) dans les prélevats n'est pas indépendante des cotes de ce prélevat et des pratiques techniques qui l'accompagnent.

A l'inverse, un prélevat ponctuel peut être informe, c'est-à-dire collecté sans forme fixée *a priori* dans les options d'échantillonnage. Il est alors possible de respecter les "formes" que la nature a donnée aux caractéristiques observées c'est-à-dire d'adapter le prélevat au terrain au cas par cas. Mais ce type de prélevat ne permet pas une extrapolation numérique dans l'espace-temps.

L'organisation d'une série de prélevats ponctuels techniquement de nature différente, mais caractérisant une même situation spatiale, accroît considérablement la somme des données analytiques caractérisant un point. Le lien spatio-temporel des prélevats joue un rôle essentiel et la situation doit être impérativement conservée.

On dénomme échantillon un ensemble de données susceptibles de représenter un tout. Le prélevats sont alors choisis en fonction d'un plan d'échantillonnage visant à tirer de ceux-ci,

par inférence des données, une description par DSEs d'un objet plus vaste (un tout) que les DICs elles-mêmes.

La signification (la qualité) de l'échantillon par rapport au tout que l'on cherche à connaître est une préoccupation ancienne traitée classiquement en terme de probabilité. Un ensemble de n données d'une caractéristique (= échantillon) issu d'un ensemble de n prélevats est jugé représentatif d'un tout par le calcul du gain d'information résultant de l'ajout de nouvelles données de la même caractéristique issue du même tout. Si l'on cherche une différence entre deux tout(s) (par exemple, dans les essais classiques témoin/traité) on teste les différences observées par rapport aux différences qui seraient dues à un pur hasard (aléatoire).

Les divers plans d'échantillonnage sont donc des compromis entre les probabilités d'imprécisions observées et les limites en moyens d'obtention de données. La gestion de l'information permet souvent, en fusionnant des ensembles de données au moment de l'interprétation, d'améliorer la qualité des ensembles de données (échantillons) pour certains objectifs de recherche.

3.5.10. Associations de prélevats (prélevats associés et l'associateur)

Dans la saisie de l'information certains objets (= portion quelconque réelle du champ écologique) sont étudiés de façon répétitive : ils fournissent donc une série de prélevats. Certaines caractéristiques relatives à ces objets sont toutefois invariantes et peuvent donc être décrites par un prélevat tandis que d'autres, variables dans l'espace-temps, sont décrites par une série de prélevats.

Le prélevat résumant les caractéristiques considérées comme invariantes est le prélevat associateur ; les prélevats résumant les caractéristiques variables sont dits prélevats associés.

Plusieurs séries différentes de prélevats associés peuvent porter sur un même associateur. La relation entre associateur et associé est assurée à l'aide d'un identificateur commun à l'associateur et aux divers associés. Afin d'éviter toute redondance hormis le lien à l'associateur seules les caractéristiques variables sont décrites ; elles forment entre elles des associations.

Ces associations sont de deux types :

1) L'association diachronique :

L'associateur est un prélevat décrivant un objet fixé spatialement et comporte les caractéristiques temporellement constantes. Les caractéristiques variant dans le temps, et observées à partir de ce même objet, sont décrites par une ou plusieurs séries d'associés (= séries de prélevats associés). Par exemple un plot peut être initialement défini par sa situation X, Y, Z et être décrit par un prélevat indiquant en outre à la date D des caractéristiques considérées comme invariantes (taux d'argile du sol, pente, orientation, etc.) et l'identificateur. Ceci forme le prélevat associateur auquel des séries de prélevats associés peuvent être rapportées grâce à l'identificateur commun ; par exemple à une série de date D_1, D_2, ..., D_n, la phénologie de végétaux de ce plot, état sanitaire, leur hauteur, etc. Une autre série diachronique peut décrire sur le même plot aux dates D_a, D_b, ... D_n les insectes homoptères présents, etc.

2) L'association stéréochronique :

Dans ce cas l'objet est spatialement instable, tout déplacement demandant un temps, cette instabilité est spatio-temporelle.

L'associateur comme précédemment a pour rôle de conserver l'information intangible. Par exemple, pour un animal : espèce, race, données sur lignées parentales, sexe, etc., et l'identificateur. La série de prélevats associés portera notamment sur les situations successives (incluant les dates et situations géodésiques) et d'éventuelles observations liées. Par exemple : notations sur un animal suivi spatio-temporellement par radiogoniométrie, nombre d'individus d'un troupeau ; poids vifs successifs d'un oiseau bagué, etc.

En conclusion, placette, plot (méso-, micro-, pico-plots) et objets instables seront souvent les associateurs auxquels sont liés des séries de prélevats associés, diachroniques pour les premiers et stéréochroniques pour les objets instables.

L'organisation d'associations de prélevats à travers ses liens facilite la constitution d'ensembles de données.

3.6. LES ENSEMBLES DE DONNEES

3.6.1. Objectifs, ensembles et caractéristiques

L'acquisition de données initiales est une analyse de l'objet d'étude écologique (milieu d'une espèce, caractéristiques d'un individus, analyse spatiale, etc.). Le chercheur isole de l'écosphère des groupes d'informations d'une même "nature" (en fait, de données propres à une caractéristique) : valeurs de pH, étapes de la croissance d'une plante, présence d'une espèce, apports d'engrais, etc.

En écologie analytique, cette démarche s'arrête toujours à un niveau qui pourrait être subdivisé en éléments encore plus analytiques. Ainsi certaines études vont dénombrer les vers de terre, d'autres vont compter chaque espèce de vers de terre, d'autres vont enregistrer chaque stade (adultes, juvéniles, cocons, etc.) de chaque espèce de lombriciens, d'autres vont faire trois lots de poids pour chaque individus de chaque stade de chaque espèce de lombriciens, d'autres vont faire 5, 7... n lots, ou analyser la structure morphologique ou la composition chimique des individus, etc.

En pratique, les moyens disponibles et <u>les options sur objectifs</u>, dictent le niveau analytique dans la globalyse et en définitive on doit préciser la liste des caractéristiques observées, par exemple :

- température maximale journalière
- température minimale journalière
- pH du sol
- présence/absence de *Dactylis glomerata*
- masse de l'animal au garrot
- carbone total du sol
- azote total du sol
- vitesse du courant de l'eau
- orientation du vent...

En général, une caractéristique sera l'objet de plusieurs observations instantanées comparables dans l'espace (données synchroniques) ou/et dans le temps (données diachroniques), parfois un seul état considéré comme constant est enregistré (présence d'une espèce d'arbres comme caractère du milieu où se situe une étude plus détaillée) enfin souvent, et c'est normal, les données de composition sont stéréochroniques. Les caractéristiques donnent à chaque observation une donnée et ces données peuvent être

organisées en ensembles de données. Ces ensembles entrent eux-mêmes dans des ensembles de niveau hiérarchique supérieur, etc.

Chaque donnée écologique doit être située par des liens à d'autres données elles-mêmes partie du champ écologique et résulte d'un choix d'acquisition dans ce champ : elle est partie de la globalyse (somme de données objectives disponibles) effectuée sur ce champ et est composante systémique de l'écosphère utilisable pour sa valeur intrinsèque et pour la possibilité d'être incluse en ensembles plus vastes pour interprétation.

3.6.2. Options spatio-temporelles sur les ensembles

- Un ensemble est synchronique lorsque les données de celui-ci sont considérées comme ayant été prises simultanément (les différences de dates étant jugées négligeables) (ex : ensemble des profils de sol utilisés pour constituer une carte pédologique).

- Un ensemble est diachronique lorsque les données de celui-ci sont organisées pour refléter des différences chronologiques (ex : les variations horaires du dégagement de CO_2 d'un sol ; le nombre de pucerons/m^2 observé à chaque mois de l'année ; les quantités de précipitations exprimées en mm/jour pendant un mois, une saison, une année).

- Un ensemble est stéréochronique lorsque les données de celui-ci sont organisées pour tenir compte de la situation totale des données, à la fois dans l'espace et le temps, (exemple : étude d'une migration saisonnière).

3.6.3. Qualité des données et des ensembles : les erreurs

La donnée peut être de qualité très inégale par rapport à divers points de vue : elle a intrinsèquement une qualité qui peut être définie par rapport à trois critères : adéquation, précision et médiation. Elle peut extrinsèquement être appréciée par rapport à l'ensemble auquel elle appartient. Elle peut enfin être considérée dans les démarches d'interprétation.

1) Qualités intrinsèques des données

L'adéquation d'une donnée est la qualité de son lien effectif à la caractéristique d'attribution. Une donnée est adéquate lorsque sa mesure, son observation correspondent bien au(x) lien(s) que l'on entend en faire. Ex : la couleur d'une feuille de hêtre doit bien se rapporter à l'ensemble des couleurs, à feuille et à hêtre (et non par erreur à un poids, une

branche et à un orme). Ex : l'incorporation du carbone dans un sol doit correspondre au carbone entrant dans le compartiment donné à l'instant dt, et non pas être le bilan positif de C entre t_1 et t_2 (car entre t_1 et t_2 il peut y avoir eu aussi "sortie"), etc. En d'autres termes, une donnée est inadéquate lorsqu'elle ne se rapporte pas à la caractéristique à laquelle on l'attribue.

La précision porte sur la classique "erreur de mesure" de physique, c'est-à-dire porte sur des données quantitatives (Ex. : 2,5 m + ou - 1 cm), en raison de l'imprécision des moyens techniques de mesure, ou des données qualitatives (Ex : *Lumbricus rubellus* subsp ?) en raison de la carence des moyens d'analyse. La précision n'est pas toujours explicitement décrite mais implicitement relevée par exemple sur la dernière décimale choisie (Ex. : 2,53 cm implique une erreur < 1 cm) ou de l'expression qualitative (Ex. *Lumbricus rubellus* subsp ?). Elle peut être explicitement indiquée et liée à la donnée, notamment lorsque cette donnée est enregistrée de façon apparemment plus précise que la précision effective (Ex : la situation, latitude longitude, d'un lieu arbitrairement attribué au centre d'un village car on ne connaît que le nom de la commune).

La médiation c'est le résultat de la lecture globale du prélevat permettant l'homogénéisation de l'hétérogénéité spatio-temporelle. La donnée obtenue correspond à une "moyenne" ; cette "moyenne" ou médiation spatio-temporelle de la caractéristique analysée conduise à une donnée sur celle-ci quelle que soit la complexité interne au prélevat qu'elle "homogénéise" par sa mesure globale. Ex : pH égal à 5,7 à moins 10 cm de profondeur pour un prélevat de $1m^2$; ou 28 ppm de Cd, dans un prélevat de 100 g de sol correspondant à la concentration dans un microsite d'un cristal pesant 25 mg... dans l'ensemble des 100 g de prélevat. Les cotes du prélevat associé à la donnée donnent l'ampleur de cette médiation qui "efface ces détails".

L'aberration enfin est une donnée proprement anormale, obtenue par un dérèglement évident dans l'acquisition (à commencer par un *lapsus*).

2) Qualités extrinsèques des données (dans leur ensemble)

La qualité des données peut encore être jugée d'après le résultat d'une interprétation, après constitution d'un ensemble de données. Une donnée est alors jugée comme cas particulier de cet ensemble. On considère alors sa position relative dans la dispersion des valeurs de l'ensemble (écart à la moyenne de l'ensemble), mesure de sa distance à la moyenne contribuant à la relative hétérogénéité de cet ensemble. Elle peut également être considérée en cas d'interpolations sous l'angle de sa portée, c'est-à-dire du volume spatio-temporel qui après analyse de l'ensemble lui est attribué par inférence.

Une donnée peut enfin ne pas avoir été acquise, mais être arbitrairement introduite dans la globalyse par supposition ce qui doit être toujours explicité par un indicateur complémentaire en raison du risque ainsi pris.

Si chaque donnée est partie d'ensembles et est d'une qualité appréciable, elle est nécessairement un atome d'information du champ écologique dans lequel elle doit être située pour devenir totalement explicite, ce qui est rarement fait sans confusion en raison du caractère à la fois hétérogène et en continuum de ce champ. Le lien aux trois référendaires sera donc une qualité essentielle.

3.7. LA SAISIE DES DONNEES INITIALES

3.7.1. La saisie

Comme nous l'avons vu, les prélevats sont susceptibles de fournir des données sur un très grand nombre de caractéristiques du champ écologique. Par nature ces caractéristiques sont repérables dans le champ écologique.

Nous avons vu également que ces données sont liées entre elles spatio-temporellement par le prélevat lui-même et que, parmi les caractéristiques d'un prélevat, les cotes temporelles et spatiales peuvent être essentielles.

Pour des raisons techniques, les données sont analysées jusqu'à un certain niveau qui peut être implicite c'est-à-dire qui est contenu dans une proposition sans être exprimé précisément (une espèce, une variété, une classe de granulométrie) ou explicite c'est-à-dire qui ne prête à aucune contestation (une longueur en décimètre).

La saisie des données devra toujours se faire selon la précision la plus fine acquise, même si l'objectif initial de la recherche ne le requiert pas. A partir du moment où un niveau précis de connaissances est atteint, il faut éviter de le perdre car les données n'ont pas une destinée unique en vue de l'interprétation. C'est après saisie et après gestion des données que, si nécessaire, l'on écartera, sans élimination, l'information "inutilement" précise pour une interprétation spécifique.

En effet, les données ne sont pas uniquement utilisables dans le seul cadre initial ayant motivé leur acquisition. Leur gestion en Bases de Données et leur usage par Intelligence

Artificielle en écologie peuvent requérir une précision imprévisible : les données sont multi-interprétatives.

La réciproque n'est pas vraie. S'imposer la contrainte d'une grande précision sans justification explicite serait une option maladroite freinant l'acquisition des données sans certitude de la moindre utilité. A l'inverse, l'adoption de codes restreignant l'information initialement acquise est aujourd'hui injustifiée et nuisible.

La règle est donc de bien associer la saisie de données au niveau d'analyse le plus précis, initialement justifié et rien de plus. Il faut conserver précieusement les DICs comme l'information la plus objective, ce qui est aujourd'hui économiquement et scientifiquement justifié, car jamais cette donnée ne se reproduira spatio-temporellement dans le système écologique qui "évolue" et se transforme.

Les données initiales doivent subir avant leur enregistrement définitif un ou plusieurs contrôles. Ceux-ci peuvent révéler des carences techniques et conceptuelles et doivent être inscrits "en routine" avant l'enregistrement définitif de Données Initiales Contrôlées.

La saisie manuelle ou automatique des données dépend du contexte matériel et informatique réalisable au niveau de la collecte, dont l'analyse des prises. Les techniques à cet égard évoluent beaucoup et dépendent du contexte technique, fort différent par exemple allant de la télédétection où la saisie automatique des données est considérable et quasi simultanée aux relevés sur le terrain qui imposent un travail fastidieux de dépouillement de microarthropodes à partir d'une prise de sol.

3.7.2. Le cas des données continues

Les données analogiques (enregistrement en continu dans le temps ou l'espace, tel un enregistrement d'anémomètre, de débitmètre, de thermomètre, de contours sur photographie, ...) sont non compatibles avec la structure d'enregistrement actuelle de l'information en ordinateur et obligent à une perte d'information par lecture discontinue selon un pas arbitraire (en général choisi comme optimal ou suffisant vis-à-vis de l'objectif d'interprétation). Dans ce cas les DICs représentent une information appauvrie par rapport à l'information enregistrée. Celle-ci peut être conservée hors-saisie informatisée, en collection.

Notons que les développements des mémoires informatiques permettent à un coût perpétuellement décroissant des enregistrements "quasi analogiques" de données discontinues et leur transmission numérique en réseaux intégrés.

3.7.3. Collections et enregistrements (données non saisies)

Le processus d'acquisition de données peut conduire à conserver des objets matériels porteurs d'informations potentielles inexploitées... c'est-à-dire ceux dont on sait que l'on pourra ultérieurement extraire des données qui n'ont pas été initialement observées.

Il peut s'agir d'échantillons (plantes, animaux, ...) conservés dans un muséum et dont la sauvegarde permettra de revenir à leur étude pour de nouvelles acquisitions de données (telle la révision d'un type). Il peut s'agir d'enregistrements analogiques à une ou n dimensions (photographie, film, bande multicanal de télédétection , ...) dont on a dû, pour une première saisie, choisir des cotes de prélevats, telles la durée ou le pas de temps. Ces cotes peuvent ultérieurement ne pas s'avérer optimales.

Les prises analogiques brutes continues doivent être transformées en données numériques discontinues. Le prélevat n'est défini ici, quant à sa cote durée, qu'après enregistrement avant la saisie informatique, et l'enregistrement initial peut être conservé pour une redéfinition de la durée du prélevat.

3.8. LES LIENS ENTRE DONNEES

3.8.1. L'organisation des données

L'écologie, par définition, aborde un nombre énorme de problèmes fondamentaux ou/et appliqués. Ses pratiques sont dispersées selon les trois référendaires écologiques :

- dans l'espace : les points d'application planétaire étant infiniment divers,

- dans la composition : le niveau d'analyse extrêmement variable condamnait à un défaut de rigueur ou d'explication s'il était "trop général" ou à une carence de généralisation s'il était "trop réduit". La quantité de données nécessaires augmente rapidement dans les synthèses complexes,

- dans le temps : allant de l'étude de l'évolution (des sédiments, des organismes, des écosystèmes) sur des millions d'années, à l'impact de l'industrialisation sur la gestion de l'espace rural (quelques décennies), à la croissance saisonnière d'un végétal ou aux échanges (instantanés) dans le système sol-racine.

Ainsi, la totalité de la connaissance scientifique écologique est descriptible et accessible.

Ceci a quatre conséquences immédiates :

- Les données et démarches, rendues accessibles permettent le contrôle, la vérification donc la scientificité des domaines concernés.
- L'extraordinaire perte d'informations, encore actuellement en cours, source de méconnaissances et d'incompréhension n'a plus aucune raison d'être.
- L'organisation des données constitue en soi une élaboration. Elle permet l'affranchissement du nucleus intellectuel initial d'acquisition, la ré-interprétation et la recombinaison *ad libitum* des connaissances initiales (DICs) ou élaborées (DSEs)
- Les bornes des spécialités devraient concrètement sauter. On passe d'une "écologie" à sommation multidisciplinaire à une véritable écologie globale "monodisciplinaire" des systèmes complexes favorisant les interactions entre approches holistique et réductionniste.

3.8.2. Situations des prélevats dans le champ écologique

Comme nous l'avons dit, toute donnée provient d'un prélevat. Pour la globalyse (information objective disponible) tout prélevat est limité par ses <u>cotes</u> (x, y, z, d) et sa composition. Réciproquement, toute donnée issue du prélevat se rapporte à la portion spatio-temporelle x, y, z, d dont elle est par définition l'analyse spatio-temporelle la plus fine.

Cette portion s'inscrit dans le champ écologique par sa <u>situation</u> (X, Y, Z, D). Pour cela, il suffirait logiquement de quatre repères géodésiques permettant de situer ce prélevat (et les données issues du prélevat) dans l'écosphère. Ce pourrait être longitude, latitude, altitude et "date" (L, l, A, D) où la date n'est pas prise qu'au sens calendaire, mais peut être précisée avec d'autres unités de temps même plus précises (année, quantième, heure, minutes, secondes). Le niveau de composition étant indiqué par la nature de la donnée (= caractéristique).

En pratique, malheureusement, la situation du prélevat dans le champ écologique est tout à fait imprécise. Les moyens de situer avec précision un prélevat sont généralement inutilement lourds. Il est préférable de situer, avec la plus grande précision possible les prélevats par leurs coordonnées géodésiques, et lorsque celles-ci sont quasi identiques de repérer ces prélevats entre eux par des coordonnées locales ou repères (X, Y, Z, D).

Pour le système de repérage généralisé, adopté dans la Base de Données Relationnelle ECORDRE, les placettes (surfaces ayant en commun des données), par exemple un mode de gestion (une sapinière, ou un itinéraire technique agricole, ou un traitement d'engrais dans une expérimentation au terrain) sont rattachés à des lieux repérés par les coordonnées géographiques (longitude et latitude). Elles comportent l'altitude et la date (normalement année et quantième de l'année) et les repères (X_1, Y_1, Z_1, D_1) situant les plots dans ces placettes, où sont effectués des relevés (prises). Si nécessaire, par exemple en micromorphologie spatiale un système de pointage X_2, Y_2, Z_2, permet de repérer chaque donnée d'un mésoplot, voire d'un microplot, dans le plot.

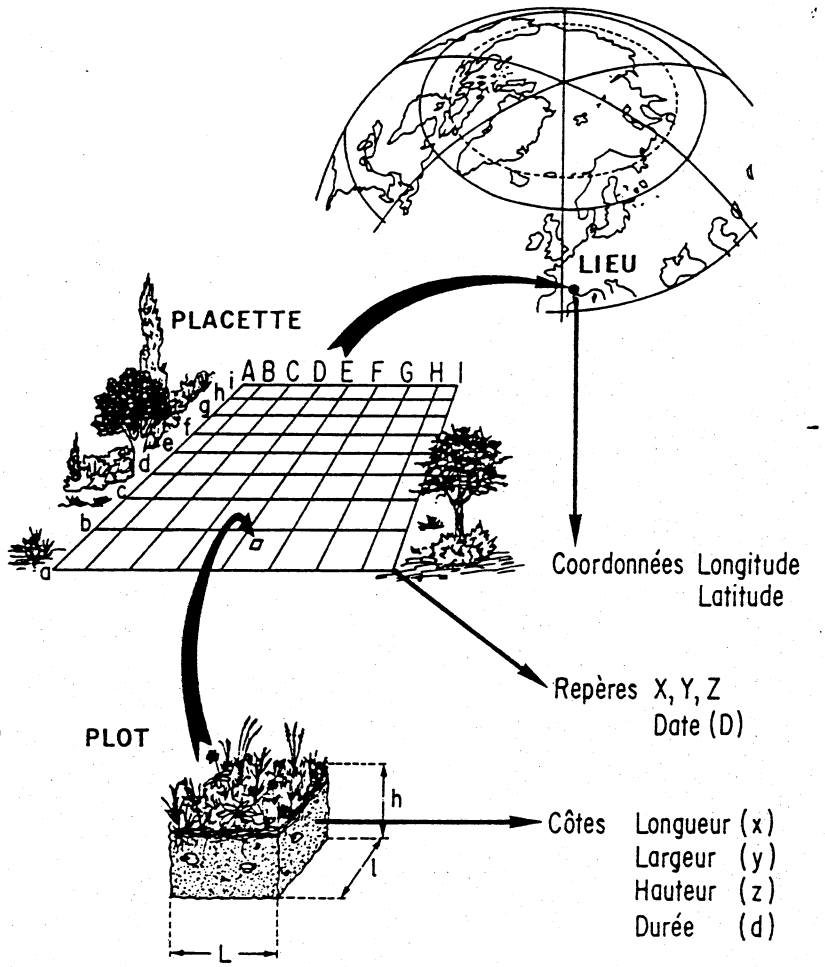


Figure 3-6 : *Intégration et repères spatio-temporels des données agro-écologiques dans l'écosphère.*

Si le lieu est la localisation géodésique et la placette la portion d'espace où l'on collecte des données à partir de plots, il arrive que l'on soit conduit à subdiviser l'espace de la placette après la définition des plots. Ces unités spatiales peuvent en partant du plot être décrites vers des unités plus englobantes en termes de macroplot, mégaplot, ... ou, en partant de la placette, en mésoplacette, ...

Ainsi, toute donnée de composition quelconque, peut être décrite en fonction de sa situation, de sa médiation spatiale (cote du prélevat), de sa modalité potentielle d'extrapolation (cartésienne ou typologique) et d'intrapolation ponctuelle mais également elle peut être repérée dans l'écosphère ou tout sous-système spatio-temporel de celle-ci.

3.8.3. Champ écologique et informations

Comme nous venons de le voir toute donnée est située *ipso facto* dans le champ écologique et est définie par sa composition et ses coordonnées spatio-temporelles.

Retenons simplement qu'une donnée acquise, quelle que soit la finalité initiale pour laquelle le travail d'acquisition a été effectué, doit circuler et être ré-interprétable en fonction d'autres données situées dans le champ écologique, voire du domaine socio-économique. En aucun cas, la DIC (obtenue directement du champ écologique concret) ne doit être perdue car elle est notre perception objective depuis notre environnement immédiat jusqu'à l'écosphère.

Cette remarque n'exclut nullement l'énoncé de la motivation pour justifier et organiser l'acquisition initiale. Si toute donnée doit être ré-interprétable, elle doit d'abord être interprétée en fonction de l'objectif explicité par la motivation de la recherche qui en a assuré l'acquisition.

Notons que la distinction entre recherche fondamentale et appliquée n'a de sens que socialement en fonction de cette première motivation d'acquisition mais que cette distinction perd tout sens dans le cadre de ré-interprétations.

3.8.4. Le référentiel des données

Quel que soit l'objectif de recherche ayant motivé la collecte d'une DIC, celle-ci résulte d'un processus d'options nous apportant un atome d'information sur l'écosphère. Cet atome

pourra être regroupé dans tout ensemble de données et contribuer ultérieurement à toute synthèse interprétative.

La saisie des données se fait pour un multi-usage des données grâce à la gestion de celles-ci par des Systèmes de Gestion de Bases des Données (SGBD). Cette modalité contribue à lever l'un des grands handicaps de l'écologie. Quelles que soient les modalités de collecte et d'analyse fournissant des données transitoires brutes sur supports papier ou magnétiques, en enregistrement automatique ou manuel, il y a intérêt à développer des procédures d'énoncé relationnel obligeant, au cours de la saisie sur support informatique, et avant le transfert dans une Base des Données, à l'énoncé de relations qui systématiquement lient la donnée au prélevat, à sa situation, ses cotes, son descriptif technique et sa motivation.

Le référentiel des données a pour objet de repérer la donnée acquise dans l'écosphère par ses liens à cette écosphère et ses liens aux données ayant même position dans l'espace-temps écologique. Celui-ci s'appuie sur cinq concepts génériques liés à la donnée :

1) Le prélevat : portion d'espace-temps indissocié ayant fourni la donnée. Le prélevat est limité par des cotes spatio-temporelles, x, y, z, d (d pour durée). Les cotes peuvent être négligées, non enregistrées, ou délocalisées, plus ou moins complètement en exprimant directement par exemple le volume d'un prélevat ou sa surface. Le principe de conservation des DICs devra toutefois être respecté. La durée peut également être "délocalisée" en exprimant les caractéristiques par rapport à une durée standard (g/s ou km/h, travail/mn, ...).

2) La caractéristique ou nature de la donnée que, par analyse de la composition du prélevat, l'on caractérise dans ce prélevat par la donnée. Les modalités de l'analyse portant usuellement sur tout ou partie du prélevat, sont décrites dans ses modalités par un descriptif technique (DT).

3) La situation des prélevats dans l'écosphère est enregistrée totalement par quatre données (X, Y, Z, D) : les trois euclidiennes et la date (au sens de l'instant dans l'écoulement du temps). Mais en pratique, ceci ne peut pas se faire à toute échelle sans difficulté. La date est toujours repérable avec les précisions souhaitées. La situation euclidienne ne peut pratiquement, pour toute donnée individuellement, être enregistrée directement il y a alors hiérarchie cumulative des situations et des prélevats.

La plus élevée est le lieu indiquant la situation géodésique : longitude, latitude, qui permet de situer la placette d'étude : altitude. Celle immédiatement inférieure est la

situation des plots de données dans la placette, grâce à des repères euclidiens (X_1, Y_1, Z_1, etc.). Celle encore inférieure peut être un mésoplot (coupe microscopique ayant ses repères X_2, Y_2, Z_2).

4) La motivation de l'acquisition des données permet de référer la responsabilité de l'acquisition de la donnée, la propriété initiale de celles-ci, les raisons de leur acquisition (problématique, objectif, réalisation, essai, ...), voire le dossier administratif (contrat, thèse, ...).

5) Si nécessaire, l'identificateur assurant le lien entre prélevats associés et associateur.

La donnée est ainsi dans un "contenant" analytique (la caractéristique), un contenant physique (le prélevat) inclus dans un contenant général (l'écosphère) et sa justification sociale initiale (la motivation).

3.8.5. L'acquisition de données : principes d'optimisation

On peut tenter de résumer des régles générales d'acquisition des données écologiques pour une stratégie d'acquisition optimale des Données Initiales Contrôlées (DICs).

1) Par un principe de sélection des options : que l'on pourrait aussi appeler loi du moindre effort. C'est le choix de sites, de placettes, de prélevats, de moyens d'analyse aussi performants que possible pour un objectif particulier motivant l'acquisition.

2) Par un principe de conservation des DICs et de leurs liens : c'est à chaque étape du travail le souci de ne pas "perdre" une information acquise, même si celle-ci "ne semble plus utile". Ce jugement utilitaire sur les DICs est nécessairement entaché d'une interprétation particulière résultant de l'objectif de recherche parcellisé et souvent d'hypothèses-propositions qui, par essence, sont subjectives et insuffisantes. Puisque le travail d'acquisition est fait et les moyens modernes de gestion aisés, il faut conserver. La conservation est aussi le seul moyen d'une critique objective (*stricto sensu*) du travail.

3) Par un principe d'objectivité. Pouvoir représenter l'objet grâce au référentiel des données aussi complètement que possible (malgré le principe de sélection et grâce au principe de conservation) avec des caractéristiques adéquates au champ écologique et des données adéquates aux caractéristiques.

3.9. CONCLUSION

Les Bases de Données ouvrent de nouvelles voies à l'écologie : le traitement et l'accès aux données peuvent maintenant être maîtrises par l'utilisateur qui peut interroger toutes les données antérieurement acquises. De plus, les données étant disponibles, il peut définir leur présentation, les organiser pour des applications avec des logiciels statistiques, mathématiques, graphiques, etc. Enfin, et c'est essentiel, la mise en commun de la partie objective, les DICs, est effectivement possible et aisée.

On peut constater que pour l'instant seuls de "gros" logiciels implantés sur des "grosses machines" présentent toutes les propriétés souhaitables et nécessaires aux bases de données pour leur plein usage : contrôle d'intégrité, confidentialité, interconnexion télématique à des réseaux, mais aussi à des micro-ordinateurs où une part importante du travail peut être faite. Par interfaçage avec une base de faits, incluse elle-même dans une base de connaissances, ces données peuvent enfin contribuer à un dialogue formalisé en Systèmes-Experts. Les Systèmes de Gestion de Bases de Données peuvent jouer alors un rôle stratégique entre la gestion des données et la gestion des connaissances.

Tant les logiciels que les matériels sont dans ce domaine en évolution rapide et l'intérêt de telles bases ne peut pas être jugé à partir des contraintes actuelles en matériels, connexions, logiciels... ces éléments techniques, rapidement obsolètes, se remplaçant par des matériels et logiciels plus performants. Il serait notamment à ce niveau incompréhensible d'opposer "grosse informatique" à "petite", le "gros" actuel étant destiné à devenir petit et de toute façon l'un étant complémentaire de l'autre.

De la même façon, les Bases de Données, tout comme le dénombrement d'oiseaux ou la pesée d'une récolte, ne sont pas une fin en soi, mais des moyens d'élaboration dans un contexte intégré en rapide évolution, notamment en écologie par le développement de l'aide à la modélisation actuellement en cours.

Il ne s'agit pas au niveau des Bases de Données d'un problème de traitement biométrique de données, qui dépend des objectifs spécifiques d'interprétation, mais de conceptualisation de la logique d'acquisition des données par les écologues, les agronomes et les environiciens et de mise en forme des ensembles de données et de formalisation des relations par les informaticiens.

Gérant des données, un Système de Gestion de Bases de Données n'a aucun rôle vis-à-vis de la qualité intrinsèque de chaque donnée, ni des interprétations que l'on fera sur ces

données. On peut y gérer de mauvaises mesures, ou également après avoir sélectionné des séries de données correctes, tirer des conclusions erronées sur ces ensembles. Ceci n'est pas du ressort d'un SGBD, mais bien de l'intelligence naturelle.

Par contre, l'utilisation d'un SGBD rend possible le partage des connaissances entre spécialistes en langage quasi-naturel, ce qui pose aux utilisateurs un problèmes de partager les mêmes concepts avec la même sémantique.

Si le choix actuel des logiciels et matériels peut effectivement limiter ou faciliter le travail du spécialiste, dès à présent il est possible par celui-ci de "dé-sectorialiser" ses pratiques, et, tout en restant un bon spécialiste d'un domaine restreint appliqué ou fondamental, de contribuer à la croissance de nos connaissances objectives sur l'écosphère.

En tout cas, la stratégie générale de recherche s'en trouve profondément modifiée. Devant une nouvelle problématique écologique, environnementale ou agronomique, les objectifs de recherche débuteront non plus par l'acquisition de données, mais par l'exploitation des données pré-existantes. La gestion des données, permet ainsi d'optimiser l'éventuel échantillonnage qu'il faudra conduire. Une base de données de ce type, ne marque qu'une étape vers un stade plus avancé, où l'Intelligence Artificielle prendra en compte la connaissance des experts qui deviendra électroniquement accessible devant tout nouvel objectif de recherche ou demande sociale dans des bases de connaissances.

En résumé, l'organisation et la gestion des données écologiques est possible actuellement grâce aux Systèmes de Gestion de Bases de Données, et à l'adoption d'un **Schéma Conceptuel Global** qui doit être cohérent avec l'écosphère pour permettre une organisation rationnelle des ces données, assurée par le processus de normalisation des relations, dont les règles doivent absolument être respectées.

La conception et l'évolution des telles bases doivent se faire à la fois vers une simplification permettant une bonne vulgarisation scientifique et vers une sophistication accrue des systèmes répondant aux désirs des spécialisations pour lesquelles elles doivent être conçues. Leurs avenir passent donc en premier lieu par la structuration et la formalisation (dynamique) de nos connaissances dans un travail d'équipe pour parvenir à des systèmes aptes à se développer au service de tous.

CHAPITRE QUATRIEME

ECORDRE : UN OUTIL TRANSDISCIPLINAIRE AU SERVICE DE L'ENVIRONNEMENT

4.1. INTRODUCTION

Les problèmes d'environnement sont à la fois multiples, "interdisciplinaires" et interdépendants.

MULTIPLES, en ce sens que notre responsabilité est engagée pour tout acte humain (choix d'un nouvel emballage, ... qu'il faudra traiter comme déchet ; usage d'un pesticide ; travaux d'hydraulique ; sécurité d'une fabrication ; etc.). Outre les processus spontanés se produisant dans l'écosphère, toutes les conséquences, la puissance et la fréquence croissantes des actes humains devraient pouvoir être appréciées, si possible *a priori*, pour éviter des désordres environnementaux fâcheux et les inadéquations des agrotechniques.

"INTERDISCIPLINAIRES", en ce sens que l'aptitude intellectuelle de tout chercheur étant très limitée, il est obligé de restreindre son expertise à un domaine spécifique étroit pour assurer une compétence. A l'inverse, les systèmes écologiques, où tôt ou tard les actes humains exercent leurs conséquences, "ignorent" les limitations des chercheurs et des disciplines : ils sont simultanément bio-physico-chimiques. Toute tentative pour les réduire à un petit nombre de disciplines (approches paucidisciplinaires) est un choix *a priori* qui n'a présentement pas de fondement scientifique car pour en avoir un il faudrait d'abord connaître ces systèmes (au sens exact du terme), ce qui n'est pas le cas. Interdisciplinaire est donc insuffisant ; on utilisera "transdisciplinaire" pour souligner l'aptitude d'intégrer un nombre illimité de disciplines.

Il faut ici distinguer les systèmes contrôlés des systèmes spontanés. Les systèmes contrôlés sont caractérisés par une organisation d'une unité de production dont on peut identifier toutes les entrées de matières premières et reconnaître toutes les sorties (produits recherchés, effluents gazeux, liquides et solides). Dans ces systèmes un nombre limité de spécialistes interviennent et maîtrisent, en principe, le système de production dont tous les éléments peuvent être décrits (exemples : aciérie, porcherie industrielle, centrale nucléaire, brasserie, cimenterie). Les systèmes spontanés sont, à l'inverse, ceux des espaces généraux de la planète où les intrants des systèmes contrôlés sont prélevés et où toutes leurs sorties aboutissent ; ils peuvent être modelés par l'homme (ex : agriculture) mais dépendent largement de l'énergie naturelle (surtout solaire) et d'un état antérieur toujours très partiellement connus. Parmi les systèmes spontanés ceux qui sont biophysicochimiques sont par définition des systèmes écologiques.

INTERDEPENDANTS, en ce sens que les systèmes spontanés sont des systèmes totalement ouverts sur les systèmes adjacents et que les systèmes contrôlés ne fonctionnent

qu'avec entrée et sortie vers ces systèmes spontanés. Si l'on peut à l'intérieur d'un système contrôlé exercer un art relativement maîtrisé d'un nombre identifiable de disciplines techniques et scientifiques, on ne peut le faire dès que le contexte environnemental est pris en compte. En conséquence, aucune discipline ne peut prétendre poser un problème environnemental, d'où la multitude de constats *a posteriori* divers (croissance du CO_2, du CH_4, des NO_x, des déforestations, de l'enfrichement, des déserts, ...) dont on sait qu'ils ne sont pas indépendants ni socialement (mondialisation de l'économie et interaction des motivations des actes humains) ni au niveau de l'écosphère (mouvements spontanés des fluides et plus limité des solides, transports volontaires).

En résumé,

1) Aucun problème d'environnement ne peut être abordé isolément. Si certains thèmes, bien perçus par le public, peuvent constituer des moteurs politiques stimulant les études environnementales, en aucun cas ces études ne sauraient être organisées de façon séquentielle.

2) Si toute étude technique ou scientifique doit pouvoir être appréciée pour sa qualité dans le cadre de sa (ou ses) discipline(s) elle n'a de sens environnemental que si elle s'inscrit effectivement (= opérationnellement et non par le discours) dans une synthèse assurant son usage multiple, interdisciplinaire et multithématique car ses résultats dépendent de l'interaction des systèmes contrôlés ou spontanés et contribuent à leur compréhension en général.

3) L'écologie, qui est définie originellement comme la science globale des systèmes biophysicochimiques, ne saurait être confondue avec les techniques de l'environnement qui sont des choix humains de modes d'action impliquant des aspects juridiques, des approches focalisées dans les systèmes contrôlés (le fonctionnement d'une centrale nucléaire n'implique pas de biologie !), des jugements de valeur, etc. Tout comme la médecine ne saurait être confondue avec la biologie, l'environique ne saurait dont l'être avec l'écologie. Toutefois, il y a homologie ou analogie de ces deux pratiques sur certains plans :

a) Toute étude environnementale s'inscrit nécessairement dans les systèmes biophysicochimiques, dont le plus grand d'entre eux, l'écosphère ; à ce stade, il y a identité d'objet d'étude.

b) L'écologie étant une science globale et les problèmes environnementaux étant interconnectés, il y a dans les deux cas nécessité d'outils opérationnels de synthèse (et non comme actuellement de discours ou commentaires généraux). L'absence de ces outils de synthèse a détourné, tant les techniciens que les scientifiques, des échanges transdisciplinaires réellement opérationnels, seule valorisation objective de leurs disciplines ; il faut passer de l'affirmation de l'"importance" d'une discipline à la

démonstration de cette importance en considérant sa contribution à la compréhension des systèmes où s'inscrivent les problèmes environnementaux.

c) Enfin toute approche environnementale s'inscrit dans un contexte économique et social. Le fait de mieux comprendre les systèmes écologiques et les impacts environnementaux des techniques humaines permet de passer des filières de monoproduction à la multiproduction. Tirer parti des espaces pour diverses productions qui se succèdent ou se chevauchent en exploitant par exemple différents compartiments des systèmes écologiques, tels la récolte végétale, les produits de la chasse, l'exploitation arborée, fruitière, etc. est possible à condition de ne pas détruire la complexité de ces richesses par une technique uniformisante volontaire ou non. Inversement, concevoir des démarches scientifico-techniques d'optimisation des études environnementales ne peut se faire hors du contexte socio-économique tel qu'il se traduit par des actes humains dans les systèmes spontanés ou contrôlés. Il y aurait risque à rechercher seulement *a posteriori*.

Les progrès récents de l'informatique en général sont très rapides et nous placent dans la situation de chercheurs qui n'ont pas su, ou pas pu encore, mettre en oeuvre ces outils dans toute leur plénitude et dynamique de création. En permettant à l'homme un accès exhaustif et orienté à la connaissance humaine, ces outils réduiront considérablement les effets pervers non intentionnels de nos actes ou règlements.

En conclusion, on dira que le projet ECORDRE vise, sur la base d'acquis concrets et d'une discipline d'intégration effective à montrer la faisabilité de l'acquisition, de la gestion, de l'interprétation explicite des faits à des fins environnementales. Il s'agit de promouvoir une conceptualisation transdisciplinaire, un schéma conceptuel global de gestion des données et une explicitation des modalités d'interprétation qui ne trouvent pour l'instant place nulle part dans les structures "verticales" de la recherche.

Par ailleurs, le système d'information ECORDRE est par essence transdisciplinaire. L'actuel développement concret de la base de données ECORDRE est déjà transinstitutionnel. Les progrès de l'informatique (bases de données réparties, architecture client/serveur, réseau national de télécommunications pour la recherche à gros débit, atelier de génie logiciel, systèmes d'information géographique, programmation orientée objets, hypertextes ...), de l'écologie opérationnelle (conceptualisation transdisciplinaire permettant la fédération des différentes disciplines étudiant ou agissant sur les écosystèmes réels) et les prises en compte environnementales (réseau national de bassin, "valeur" économique du patrimoine environnemental, ...) créent la possibilité de l'indispensable approche intégrée transdisciplinaire.

4.2. APPROCHE DES SYSTEMES COMPLEXES

4.2.1. L'approche systémique

La compréhension des phénomènes du monde dans lequel nous vivons est très importante car elle peut nous en permettre une certaine maîtrise. Cette notion de "monde" est relative car *chacun a son propre monde*. Il s'agit donc de savoir si nous pouvons avoir une bonne estimation des ces mondes qui interagissent les uns avec les autres au travers de frontières diffuses, pour en connaître le dynamisme, les activités, les comportements et de voir si ce que nous percevons est proche ou tout au moins conforme à la réalité. Autrement dit, il s'agit de trouver un instrument, une ou plusieurs méthodes permettant de les imaginer au mieux.

Ces mondes peuvent s'appeler systèmes (*ex. : physicochimiques, vivants, économiques*), genres de boîtes noires ou réservoirs à l'intérieur desquels il semble se passer interactivement quelque chose en temps réel et dans lesquels circulent divers flux : énergie, matière, hommes, informations, ... Ils se transforment et évoluent, bien ou mal.

La notion même de système évolue et deux étapes importantes de cette évolution sont intéressantes à distinguer :

1) Le rationalisme cartésien fondé sur "*le discours de la méthode pour bien conduire sa raison*" de Descartes en 1637, qu'il exprime en quatre préceptes :

- **Le précepte d'évidence :** "... *ne recevoir jamais aucune chose pour vraie que je ne la connaisse évidemment pour telle ...*"
- **Le précepte réductionniste :** "... *de diviser chacune des difficultés que j'examinerais en autant de parcelles qu'il se pourrait et qu'il serait requis pour les mieux résoudre ...*"
- **Le précepte d'explication de cause à effet :** "... *conduire par ordre mes pensées en commençant par les objets les plus simples et les plus aisés à connaître, pour monter peu à peu comme par degrés jusqu'à la connaissance des plus composés ...*"
- **Le précepte d'exhaustivité :** "... *faire partout des dénombrements si entiers et de revue si générale que je fusse assuré de ne rien omettre ...*"

Ce rationalisme est impressionnant de par sa rigueur, mais il s'oppose à la notion de variation des choses dans le temps : "*tout recommencer lorsque nous serions supposés avoir tout bien fait ...*" Cette approche rationnelle est analytique car elle va rechercher les

éléments les uns après les autres et isole leur action alors que, pendant ce temps : **le monde bouge !**

Une autre approche, plus globale était nécessaire pour élargir cette vision du monde. Vers 1930, dans son ouvrage General System Theory, Von Bertanlanffy propose un nouvel axiome, dit systématique, que nous pouvons énoncer de la façon suivante : un objet est doté d'une structure ; il est plongé dans un environnement ; il fait quelque chose dans une certaine finalité ; enfin il évolue dans le temps.

A partir de ce modèle J.L. Le Moigne, en 1977, présente une nouvelle méthode dans la Théorie du système général. A l'instar de Descartes, il propose quatre préceptes pour conduire une étude : précepte de pertinence, précepte d'approche globale, précepte téléologique et précepte d'agrégativité, qu'il oppose aux préceptes d'évidence, réductionniste, de causalité et d'exhaustivité du Discours de la méthode. Cette méthode entraîne des bouleversements profonds de nos raisonnements par rapport à notre culture. C'est la systémique.

- **Le précepte de pertinence :** Ne plus accepter "d'évidences" mais retenir ce qui, dans un cadre donné, à un moment donné et dans un but donné, paraîtra pertinent. Une part de subjectivité est ainsi introduite dans l'étude et il faudra convenir que d'autres observateurs peuvent avoir d'autres intentions. Il ne faudra pas les rejeter *a priori*, mais plutôt chercher à en comprendre le sens et l'intérêt.

- **Le précepte d'approche globale :** car la somme des éléments qui composent le système n'est pas le système lui-même. Ce précepte est tout à fait en accord avec le raisonnement de Pascal qui "... *tient pour impossible de connaître les parties sans connaître le tout*" ... pas plus que "... *de connaître le tout sans connaître particulièrement les parties*". Ce tout fait supposer l'existence d'un plus que les différentes parties n'ont pas elles-mêmes mais qui émerge d'elles et qui est en rapport avec ce qu'on peut appeler : **complexité.**

Adopter une démarche globale en considérant l'objet dans son ensemble, à travers les échanges qu'il a avec son environnement et non pas en lui même. Il est plus important de comprendre d'abord le comportement de l'objet plutôt que sa structure. Il en découle qu'au lieu de chercher à analyser un objet pas à pas, nous essaierons de **concevoir un modèle** de cet objet.

- **Le précepte de comportement du système :** encore appelé précepte téléologique qui tient compte de l'interaction entre les éléments du système qui vont en modifier le comportement. Ceci s'oppose à la notion de cause à effet cartésienne qui ne tient pas

compte d'une autre action réciproque et que nous pourrions assimiler à une réaction. Plus généralement, on peut appeler ces actions : **interactions**.

- **Le précepte d'agrégativité :** pour en simplifier la représentation. Alors que la méthode rationaliste cherchait à simplifier les systèmes en supprimant les choses inconnues, aléatoires ou apparemment aléatoires, la systémique conserve cette complexité propre à certains systèmes dont on dira qu'ils sont **complexes** en gardant rassemblés leurs divers composants, ce qui correspond bien à une simplification sans réduction.

La complexité est une propriété des systèmes possédant un grand nombre d'éléments variés ayant chacun des fonctions propres et spécialisées, des niveaux internes de hiérarchie étant associés par de nombreuses relations. Les systèmes sont dits ouverts lorsqu'ils ont des interactions nombreuses avec leur environnement, ce qui accroît leur degré de complexité. Ainsi, un système sera plus complexe que les sous-systèmes qui le composent et la simplification de la complexité d'un tel système s'arrêtera lorsque sera atteint le niveau du sous-système identifié hiérarchiquement le plus simple. A ce moment, son intelligibilité se traduira par une certaine compréhension. Ce qui fait que la complexité se simplifie jusqu'à une limite qui est l'unité indécomposable, irréductible. Aussi, la complexité caractérise la richesse en information d'un système.

On pourra dire, que la complexité est en quelque sorte l'âme d'un système et, lorsqu'il a perdu sa complexité il n'est plus un système, il est un simple objet.

4.2.2. La modélisation des systèmes

La modélisation n'est qu'une représentation d'un système réel, quelle qu'en soit sa forme : physique, mathématique, graphique, verbale ou mentale. Cette représentation intelligible est indispensable pour assurer la compréhension de systèmes naturels complexes. Par ailleurs, dans une représentation mathématique, l'introduction de données numériques permet à la fois d'assurer la résolution de problèmes (*ex. : compréhension du comportement physique ou de la recherche de paramètres les plus pertinents*) et la simulation de situations possibles à envisager (*ex. : influence d'un dysfonctionnement, d'une panne sur un ensemble*).

En cela, il est donc nécessaire de disposer de modèles. Un modèle possède la structure et la fonction que l'observateur lui aura définies et des paramètres ajustables, de façon à s'adapter aux contraintes liées aux conditions de son fonctionnement. La modification des valeurs de ces paramètres d'entrée et internes au système modélisé,

correspond à des stimulations appliquées à la simulation du phénomène étudié. Lorsque le modèle est proche de la réalité, les valeurs de sortie obtenues correspondent à celles des observations que nous obtiendrions pour le système réel soumis aux mêmes conditions d'étude. Ainsi, lorsqu'une telle similitude existe, nous pourrons utiliser un tel modèle pour faire des prévisions.

4.2.3. Les étapes de la modélisation

La méthodologie d'élaboration d'un modèle à créer passe par plusieurs étapes qui font appel à la théorie et à l'observation. Ce sont **l'identification** des éléments du modèle, leur classification en fonction de leurs similitudes ou de leurs propriétés puis, la mise en relation de ces éléments par l'intermédiaire d'un graphe ou d'une matrice des liens, **la caractérisation** qui correspond à la phase théorique du choix de la structure mathématique, **la détermination** des paramètres qui minimisent les différences entre les observations et les valeurs obtenues et, enfin, **la limitation** pour en déterminer ses contraintes, ses limites.

Une fois le modèle élaboré, il sera procédé à une tentative de son remplacement par la **généralisation**, par extension, d'un modèle existant. Interviendront ensuite la **validation** et, si possible, **l'induction** qui est le remplacement de plusieurs modèles existants par un seul modèle universel, acceptable.

4.2.4. Les méthodes pour l'approche pratique

1) Des outils généraux

Si la systémique peut être considérée comme une recette qui consiste à laisser agir le système pour le comprendre, par définition, elle ne peut également prétendre opérer sans une méthode stricte prédéfinie. Les outils qui matérialisent cette approche amènent à des réflexions de groupe, facilitent la compréhension, favorisent la motivation et permettent l'accréditation du projet par les différents participants nécessaires à son succès. En cela, la systémique peut être considérée elle-même comme un outil de communication.

Concernant les relations entre les sous-systèmes identifiés, il est intéressant de pouvoir définir dynamiquement les différents liens qui existent entre eux. Les systèmes (progiciels) de gestion de base de données relationnelles et, surtout, les langages dits de 4ème génération (L4G), sont pour cela des outils remarquables. Il s'agit là d'une aide à la conception qui permet de rajouter mais également de supprimer des rubriques ou des liens.

Lorsque les rubriques, statiques, sont remplacées par des procédures dynamiques ou des traitements, cela correspond à de véritables tâches. Nous obtenons là une maquette qui correspond déjà à une organisation.

Concernant les moyens en ressources matérielles et humaines, les durées de temps et les coûts affectés aux différentes tâches et aux différentes phases d'un projet, les progiciels de gestion de projets seront des outils remarquables. Ils permettront la planification, le suivi et la présentation de projets relativement complexes. Tout projet pourra donc être approché de façon à en connaître la faisabilité avant même la prise de décision de son lancement. De plus, la planification assurera la détermination du chemin critique et les marges ou tampons de temps pour chaque tâche. En outre, elle permettra l'établissement de tableaux complets avec la répartition des ressources et des tâches au cours du temps. Comme à chaque ressource est affecté un coût et à chaque tâche, des ressources et du temps, nous pourrons déduire aisément le coût global et les effets sur l'organisation elle-même, en fonction de ses propres hypothèses, hautes, moyennes et basses.

Les résultats concrets qui s'en suivent vont permettre la mise en évidence de certaines incohérences ou anomalies et la modification à temps d'un projet au cours de son élaboration lorsque des événements non prévus interviennent, qu'ils soient internes ou liés à l'environnement et seraient en mesure d'affecter l'avancement du projet. Pourvus de tels outils, nous pourrons examiner si un rattrapage est possible sur l'ensemble du projet, pour le mener à terme. L'utilisation de tels outils est une aide certaine autant pour l'élaboration et le développement d'un système d'information en termes de qualité et de sécurité que pour son examen *a posteriori*.

2) Des outils orientés

Le formalisme entité-association permet la description des aspects conceptuels statiques concernant les données lors de la conception d'un système d'information. Conforme à la norme ISO, il est utilisé par des méthodes telles que MERISE, AXIAL et SADT avec de variantes propres à chaque méthode.

Le modèle relationnel introduit par E. F. Codd en 1970 est le mieux formalisé des modèles de données. De plus en plus de système de gestion de bases de données l'utilisent. Il est fondé sur la notion mathématique de relation.

4.3. LE SYSTEME D'INFORMATION ECORDRE

4.3.1. De la base de données ECORDRE ...

Plusieurs Unités et Laboratoires de différents organismes de recherche (INRA, CNRS, CIRAD, ENSA, USTL) ont décidé d'unir leurs efforts depuis quelques années, pour concevoir et réaliser une véritable Base de Données Agro-écologique et Environnementale (appelée ECORDRE, acronyme de ECOlogie et ORDRE), en apportant leur savoir faire au niveau des disciplines concernées. Cette base de données a donc une vocation fédératrice et d'intégration transdisciplinaire pour accroître les synergies institutionnelles au niveau régional, national et international.

L'évolution de cette base se fait à la fois vers une simplification permettant une bonne vulgarisation scientifique et vers une sophistication accrue des systèmes répondant aux désirs des spécialisations pour lesquelles elle doit être conçue. Son avenir passe donc en premier lieu par la structuration et la formalisation ouverte et dynamique de nos connaissances, dans un travail d'équipe, pour parvenir à un système apte à se développer au service de tous.

La base de données ECORDRE permet, après la saisie des données, d'assurer la **gestion** ordonnée de celles-ci. Il s'agit de leur prise en charge exhaustive, de leur conservation et de leur restitution, cette dernière opération doit être optimisée pour chaque demande spécifique quelles que soient la nature de la demande et l'origine des données.

Cette base de données, pour son importance et originalité a reçu en 1990 le **label de Agropolis** (pôle international agronomique de la région Languedoc-Roussillon).

Diverses études sont réalisées, par exemple : sur le rôle agronomique et écologique des lombriciens (vers de terre), le dénombrement de canards et foulques en Camargue, la biologie des populations d'oiseaux, la culture du riz, l'étude de la végétation, l'étude de la forêt et le rôle de l'arbre, la physique du sol et le rôle du machinisme dans le travail du sol, l'étude économique de la production viti-vinicole, l'étude des helminthes parasites de poissons, ...

De nouvelles équipes ont manifesté leur intérêt à nous rejoindre et en même temps plusieurs équipes déjà participantes ont vu l'intérêt d'élargir leurs actions vers d'autres institutions au niveau régional, national et international dans le cadre de cette application informatique.

4.3.2. ... Au système d'information ECORDRE

Le projet informatique ECORDRE était conçu tout au début (1984) comme un projet orienté principalement vers la gestion interdisciplinaire et ouverte des données écologiques sur le gros ordinateur du CNUSC avec un système de gestion de base de données moyen et centralisé le SGBD SQL/DS sous VM/CMS.

Aujourd'hui, d'une part, l'intérêt d'une intégration transdisciplinaire de la connaissance écologique et environnementale, l'importance en volume et en diversité de données à gérer en général sur la base de données ECORDRE, les nouvelles demandes de développement de la part des utilisateurs permettant la représentation et la formalisation de la connaissance (système d'information géographique, systèmes d'explicitation, système d'aide à la décision, ...), la nécessité d'atteindre un public plus large (construction de logiciels d'interrogation et de saisie de données, utilisation du Minitel, ...) et d'autre part, l'apparition de nouvelles méthodes techniques et outils de conception : base de données réparties, architecture client/serveur, réseau national de télécommunications pour la recherche à gros débit, atelier de génie logiciel, ... nous amènent à redéfinir le cadre général du projet informatique d'ECORDRE, et nous allons être obligés de prendre en considération des contraintes de coûts, de délais, de qualité et de sécurité au niveau des applications.

4.3.3. Le nouveau cadre du projet ECORDRE

L'élargissement du projet ECORDRE comprend :

a) l'utilisation du SGBD DB2 au CNUSC permettant une grande vitesse d'accès et la gestion répartie de gros volumes de données initiales et des informations descriptives, avec connexion en réseau avec d'autres grands sites possédant le même système (bases de données réparties homogènes) et avec des possibilités d'interrogation interactive par Minitel ;

b) l'utilisation d'un Système d'Information Géographique (SIG) permettant la gestion de l'information spatiale définissant les éléments graphiques (points, lignes, polygones, relief) et interconnecté avec le SGBD DB2 pour la construction de cartes dynamiques grâce à la récupération des données spatio-temporelles et descriptives de base, permettant l'aide à l'interprétation de l'information géographique ;

c) l'évolution vers l'intégration des systèmes, avec une extension progressive vers une architecture client/serveur (possibilité de bases de données réparties hétérogènes, par exemple SYBASE, ORACLE, DB2), dans la mesure où le réseau national de télécommunications pour la recherche est mis en place et les outils nécessaires mis à disposition ;

d) l'élaboration de logiciels polyglottes pour satisfaire à la réalité européenne, permettant l'explicitation des interprétations des spécialistes (hypothèses, suppositions, modèles, ...) servant d'outils critiques, de systèmes d'explicitation et de systèmes d'aide à la décision en utilisant principalement les possibilités d'ergonomie de la micro-informatique et des stations de travail avec des outils de développement principalement sur l'environnement Windows et XWindows ;

e) l'utilisation des outils du génie logiciel, permettant l'optimisation du système d'information ECORDRE qui obligent à accroître les synergies et aussi à minimiser les coûts des opérations, grâce à la gestion de projets associés à la production des applications informatiques.

L'Atelier de Génie Logiciel TELON, disponible actuellement au CNUSC, sera utilisé pour répondre aux besoins de développement d'applications dans cet environnement. Cet AGL devra être complété à notre avis en amont au niveau analyse et conception avec l'AGL MEGA de Gamma International (si possible sur des stations de travail) en prenant les couches Merise et Axial, et en aval avec l'outil LCM permettant la gestion de sources et de load-modules et l'outil PANAPT permettant l'automatisation de la mise en production des applications.

Cette dernière démarche est fondée sur les principes suivants :

- permettre une couverture aussi complète que possible du cycle de vie des projets,
- alléger tout de suite la charge de réalisation et le temps de développement,
- valoriser l'utilisation des SGBD répartis, la technologie SQL, et l'architecture client/serveur,
- améliorer la qualité des applications informatiques et répondre aux besoins de maintenance,
- répondre aux exigences de standardisation et pérenniser les investissements à forte valeur ajoutée.

4.3.4. Les moyens pour la vulgarisation scientifique

Un projet de ce type implique de gros moyens informatiques et réseaux, problème que nous avons résolu en partie grâce à l'effort financier réalisé principalement par les différents partenaires et aux facilités offertes par le CNUSC à Montpellier. La base de données ECORDRE est interrogeable actuellement à distance, via : Transfix (liaison numérique spécialisée), Transpac (réseau commuté) et bientôt par le réseaux RENATER. Elle peut être utilisé interactivement par Minitel (terminaux standards supportant l'affichage en 80 colonnes).

En outre, la synthèse informatisée des connaissances environnementales au service de la société est de plus en plus prioritaire, d'où la nécessité d'intégrer le système informatique réparti et fédérateur d'ECORDRE dans le cadre Agropolis à Montpellier et de développer le partenariat avec l'Institut Français de l'Environnement (IFEN) au niveau national, pour être capables de prendre en compte simultanément des informations appartenant au différents domaines au niveau régional, national et international et pour mener à bien cette action de vulgarisation scientifique et de travail coopératif.

4.4. QUELQUES EXEMPLES DES SCHEMAS D'APPLICATIONS

4.4.1. Sous-schéma d'application tableaux communs : quelques relations de la base de données communes aux différents sous-schémas d'application

a) Enoncé de l'étude

Un des buts principaux de la constitution de la Base de Données ECORDRE est de permettre de coordonner les efforts entre les différentes disciplines concourantes dans le champ écologique, en accumulant des données non redondantes et en faisant profiter un plus grand nombre d'utilisateurs, dans de meilleures conditions grâce à la mise en commun, l'organisation, la gestion et la circulation de l'information.

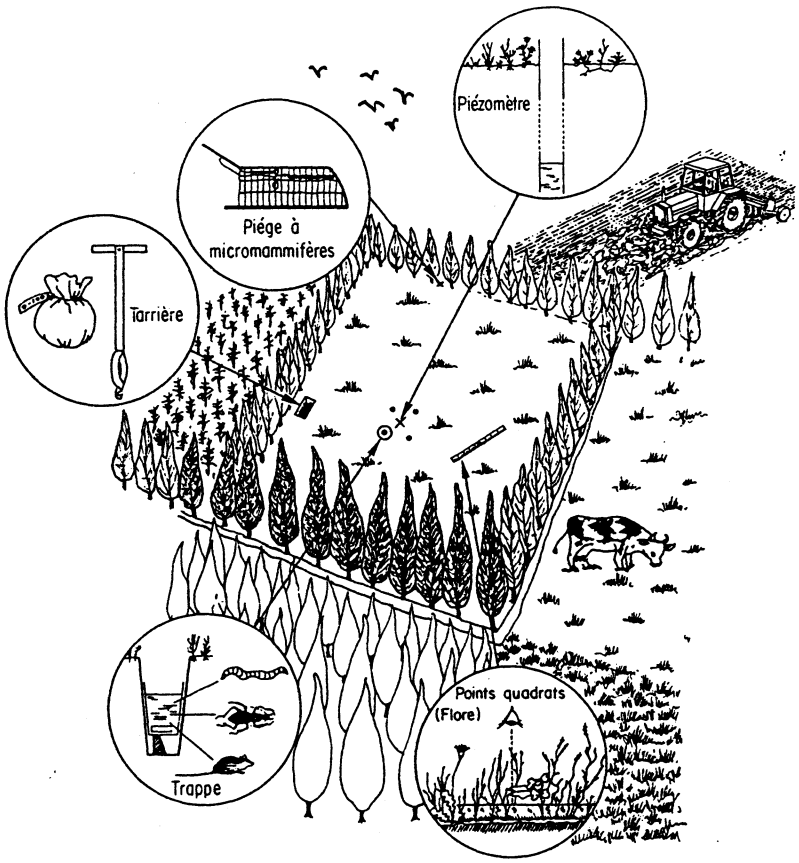


Figure 4-1 : *L'interdisciplinarité, lors de l'acquisition des données dans une portion de l'espace-temps, permet l'échange et la gestion conjointe d'information objective entre les différents spécialistes.*

Le schéma conceptuel global, qui est applicable à toute donnée venant du champ écologique ayant une origine dans l'espace-temps et à différents niveaux de composition et

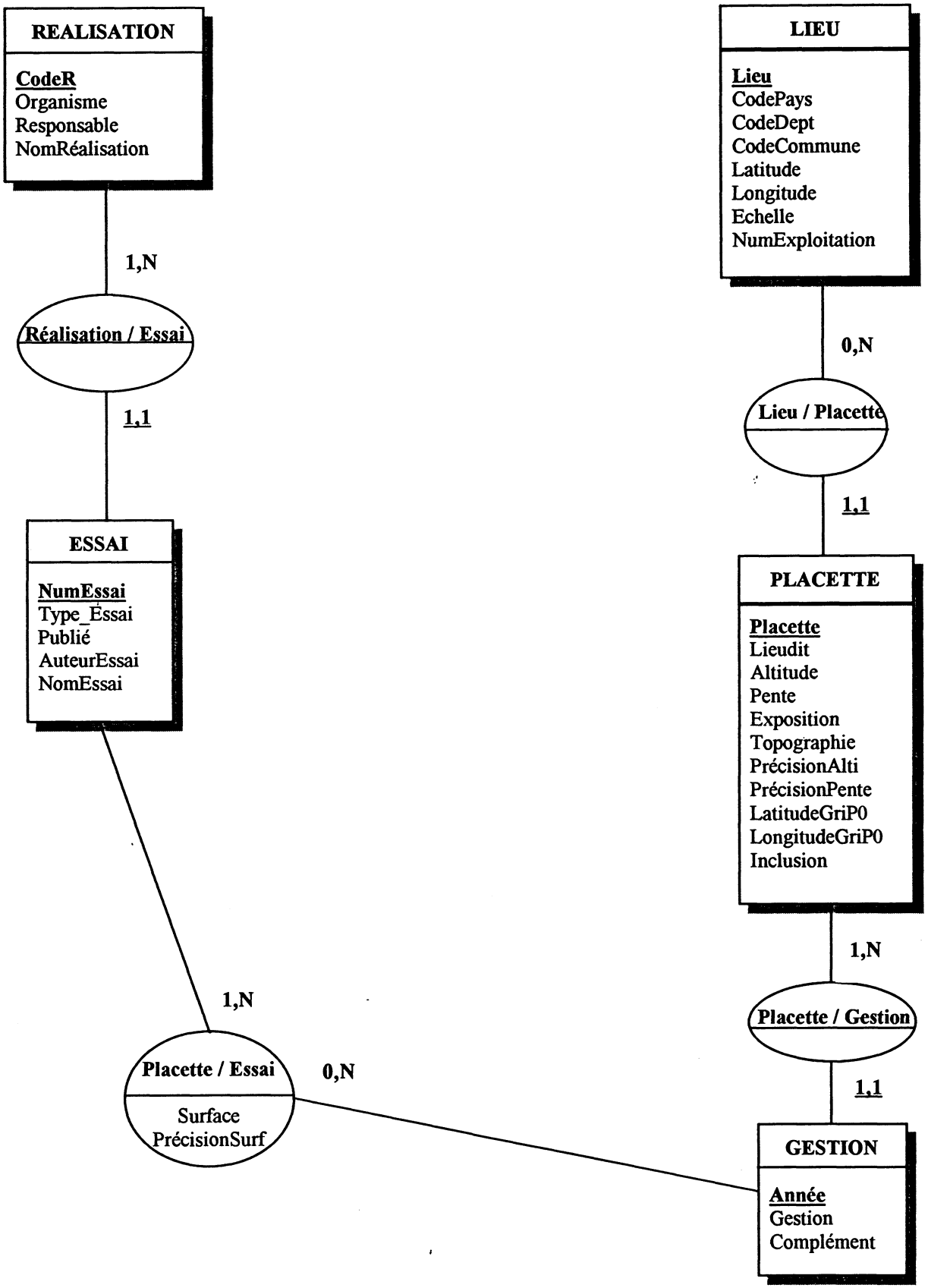
de motivation, doit permettre de définir les objectifs spécifiques de chaque discipline ainsi que ceux communs au travail transdisciplinaire. Cela se fait en mettant en oeuvre un langage univoque, assurant de cette manière la circulation des données objectives et une ré-utilisation de celles-ci indépendamment de toute considération "disciplinaire" ou mono-objectif. Ainsi toutes les données acquises, par les agronomes, les écologues, les environiciens,... sur une portion d'espace-temps quelconque, deviennent regroupables, échangeables et administrées conjointement.

Le fait de rendre possible le partage des connaissances entre spécialistes pose la nécessité de partager en même temps non seulement les mêmes concepts de base, avec la même sémantique, mais également, en termes pratiques au niveau du développement conceptuel de la Base de Données, de partager, de plus en plus, des tableaux communs.

Retenons simplement qu'une donnée acquise, quelle que soit la finalité initiale pour laquelle le travail d'acquisition a été effectué, doit circuler et être réinterprétable en fonction d'autres données situées dans le champ écologique. Cette remarque n'exclut nullement l'énoncé de la motivation pour justifier et organiser l'acquisition initiale. Si toute donnée peut être ré-interprétable, elle doit d'abord être interprétée en fonction de l'objectif explicité par la motivation de la recherche qui en a assuré l'acquisition.

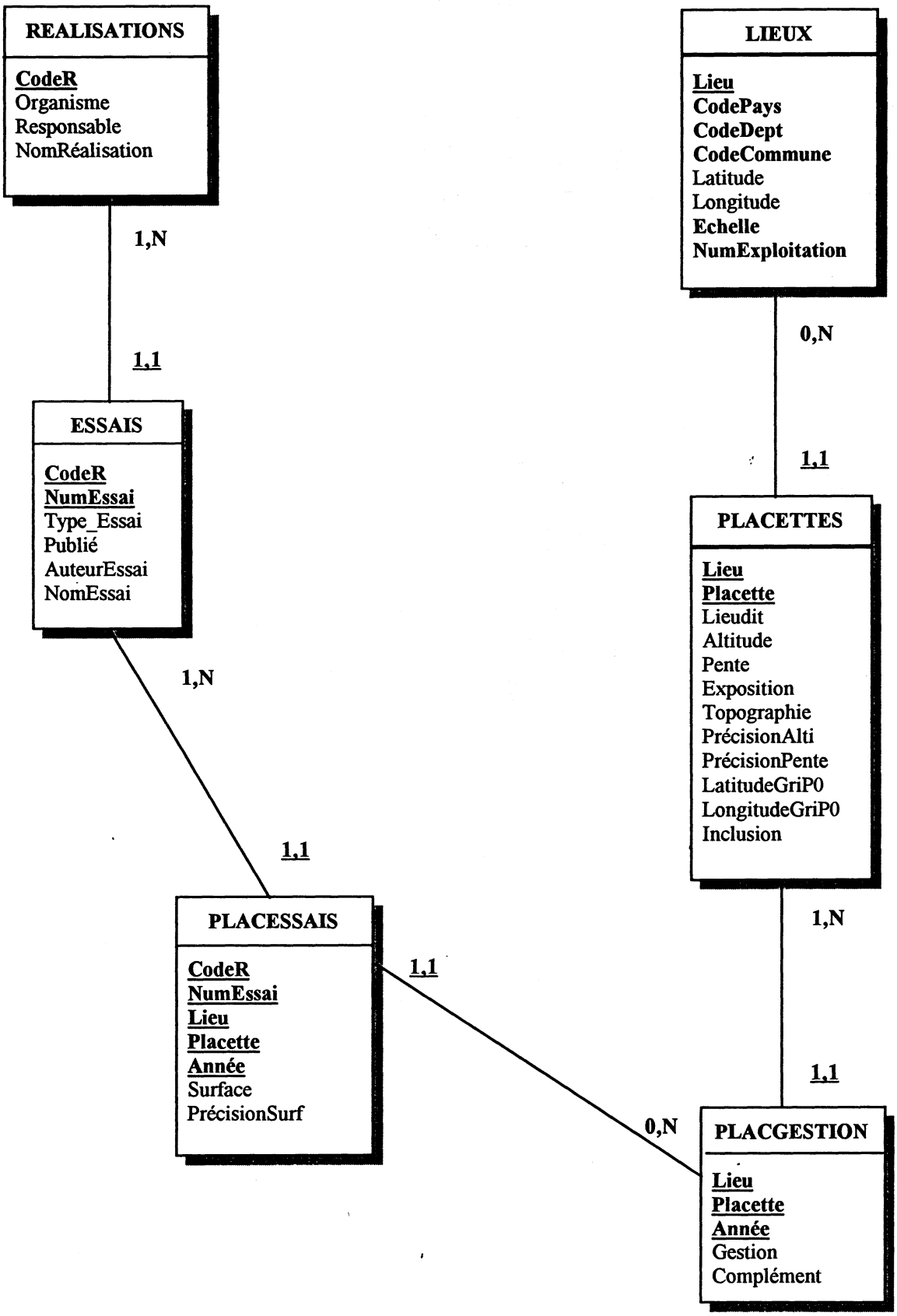
MODELE CONCEPTUEL DE DONNEES

Tableaux communs obligatoires de base



MODELE RELATIONNEL DE DONNEES

Sous-schéma tableaux communs obligatoires de base



b) Schémas relationnels en troisième forme normale

REALISATIONS (CodeR, Organisme, Responsable, NomRéalisation)

ESSAIS (CodeR, NumEssai, Type_Essai, Publié, AuteurEssai, NomEssai)

LIEUX (Lieu, CodePays, CodeDept, CodeCommune, Latitude(grades), Longitude(grades), Echelle, NumExploitation)

PLACETTES (Lieu, Placette, LieuDit, Altitude(m), Pente(%), Exposition, Topographie, PrécisionAlti, PrécisionPente, LatitudeGriP0, LongitudeGriP0, Inclusion)

PLACGESTION (Lieu, Placette, Année, Gestion, Complément)

PLACESSAIS (CodeR, NumEssai, Lieu, Placette, Année, Surface(m2), PrécisionSurf)

c) Clés des relations du sous-schéma tableaux de base

REALISATIONS	CodeR
ESSAIS	CodeR, NumEssai
LIEUX	Lieu
PLACETTES	Lieu, Placette
PLACGESTION	Lieu, Placette, Année
PLACESSAIS	CodeR, NumEssai, Lieu, Placette, Année

4.4.2. Sous-schéma d'application canards: étude des dénombrements de canards et de foulques en Camargue

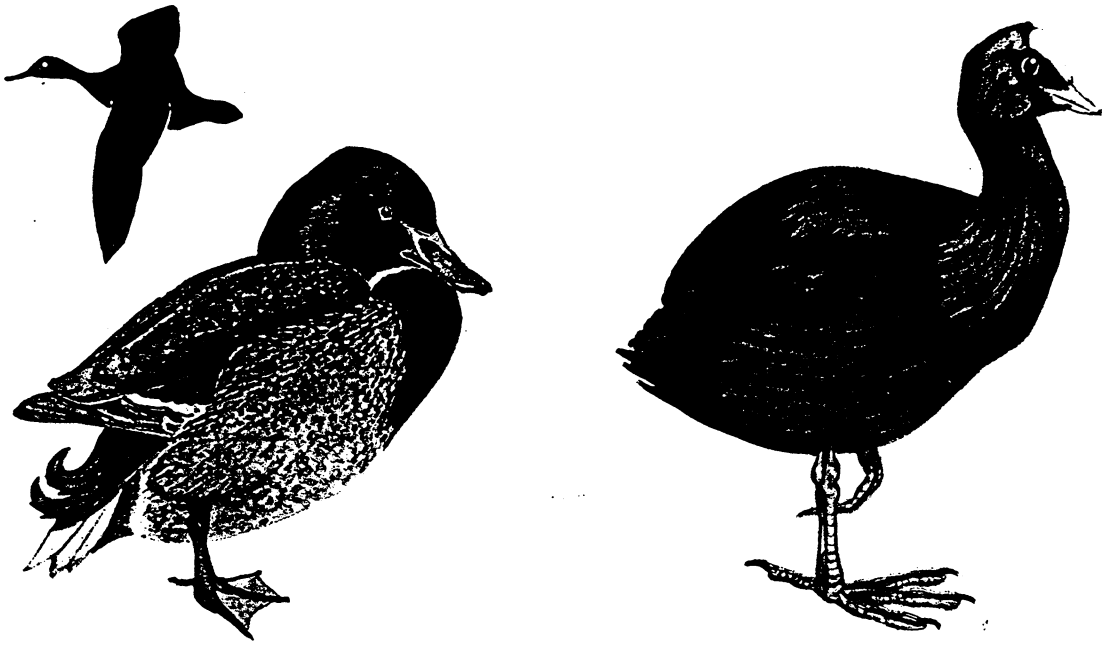


Figure 4-2 : *Canard colvert et foulque macroule*

a) Enoncé de l'étude

Cette modélisation de l'organisation des données ornithologiques concernant l'étude des dénombrements de canards et de foulques en Camargue s'appuie sur les travaux de recherche menés dans le cadre de zones humides, par une équipe d'ornithologues du Centre d'Ecologie Fonctionnelle et Evolutive (CNRS) de Montpellier. Les données collectées dans 85 placettes d'étude (59 lieux géographiques) par des dénombrements mensuels aériens depuis 1964 sont gérées actuellement dans la Base de Données "ECORDRE".

L'objectif est d'essayer de déterminer avec une précision satisfaisante la taille réelle des populations et de mesurer avec sécurité leurs fluctuations d'abondance.

L'analyse statistique permet au niveau de l'interprétation des données, d'estimer les correctifs à apporter, c'est-à-dire de tester à la fois la justesse et la fidélité des dénombrements de type exhaustifs réalisés en avion sur de concentrations de canards (Anatidés) et de foulques (*Fulica adra*) pendant leur hivernage en Camargue (où les effectifs maximaux oscillent entre 100000 et 180000 canards et entre 20000 et 40000

foulques), et la variabilité des résultats (marge d'erreur) liée à la fidélité de la mesure, aussi bien pour des groupes isolés que pour l'ensemble des effectifs.

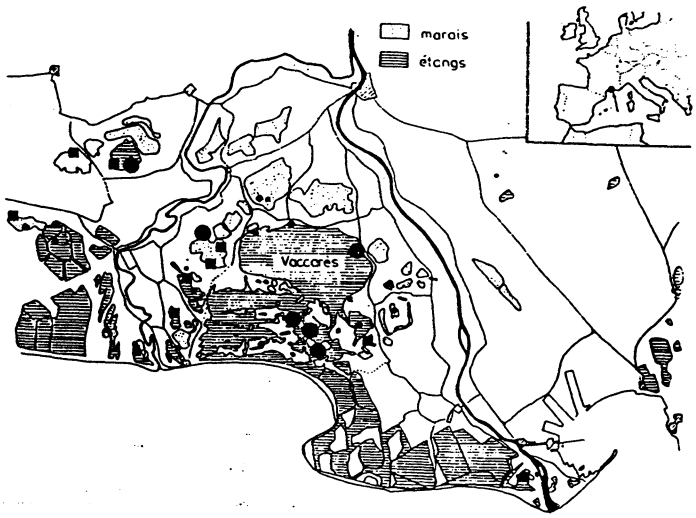


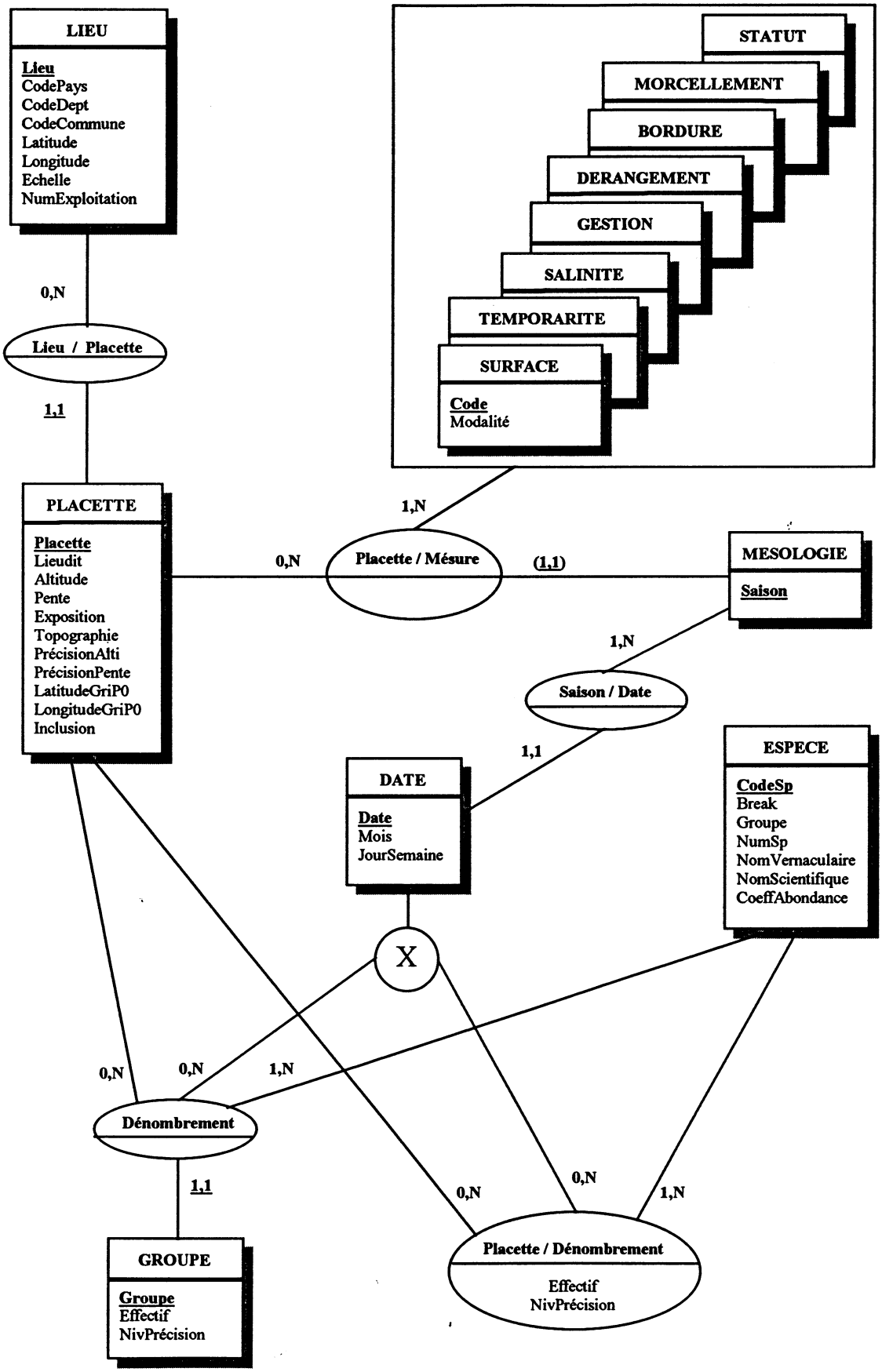
Figure 4-3 : *Localisation géographique des sites de recherche en Camargue (Midi de la France)*

Ces dénombrements sont réalisés une fois par mois pendant la saison hivernale (septembre à mars) aux dates fixées par le Bureau International de Recherches sur la Sauvagine depuis 1964 par le même observateur.

La modélisation proposée est destinée au développement d'un outil de gestion informatique des données permettant l'étude de l'évaluation quantitative des populations de canards et de foulques en Camargue. L'établissement des relations nécessaires permet d'aider à comprendre l'évolution des effectifs de ce type de population animale dans le temps et la capacité d'accueil de ce milieu.

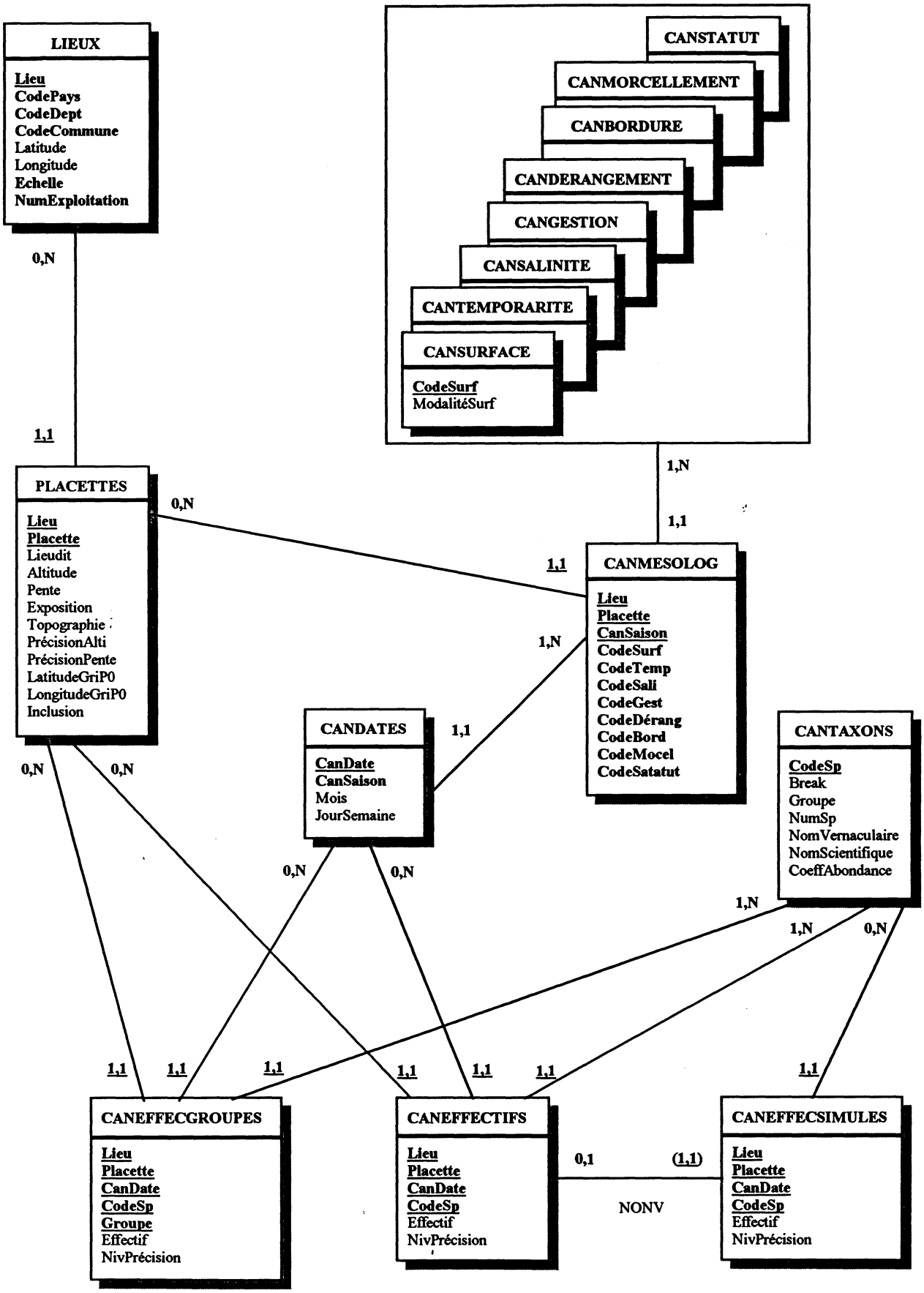
MODELE CONCEPTUEL DE DONNEES

Sous-schéma d'application Canards



MODELE RELATIONNEL DE DONNEES

Sous-schéma d'application Canards



b) Schémas relationnels en troisième forme normale

CANTAXONS (CodeSp, Break, Groupe, NumSp, NomVernaculaire, NomScientifique, CoeffAbondance)

CANDATES (CanDate, CanSaison, Mois, JourSemaine)

CANEFFECTIFS (Lieu, Placette, CanDate, CodeSp, Effectif, NivPrécision)

CANEFFECGROUPES (Lieu, Placette, CanDate, CodeSp, Groupe, Effectif, NivPrécision)

CANMESOLOG (Lieu, Placette, CanSaison, CodeSurf, CodeTemp, CodeSali, CodeGest, CodeDérang, CodeBord, CodeMorcel, CodeStatut)

CANSURFACE (CodeSurf, ModalitéSurf)

CANTEMPORARITE (CodeTemp, ModalitéTemp)

CANSALINITE (CodeSali, ModalitéSali)

CANGESTION (CodeGest, ModalitéGest)

CANDERANGEMENT (CodeDérang, ModalitéDérang)

CANBORDURE (CodeBord, ModalitéBord)

CANMORCELLEMENT (CodeMorcel, ModalitéMorcel)

CANSTATUT (CodeStatut, ModalitéStatut)

c) Clés des relations du sous-schéma canards

CANTAXONS :	CodeSp
CANDATES :	Candate
CANEFFECTIFS :	Lieu, Placette, CanDate, CodeSp
CANEFFECGROUPES :	Lieu, Placette, CanDate, CodeSp, Groupe
CANMESOLOG :	Lieu, Placette, Saison
CANSURFACE :	CodeSurf
CANTEMPORARITE :	CodeTemp
CANSALINITE :	CodeSali
CANGESTION :	CodeGest
CANDERANGEMENT :	CodeDerang
CANBORDURE :	CodeBord
CANMORCELLEMENT :	CodeMorcel
CANSTATUT :	CodeStatut

4.4.3. Sous-schéma lombriciens : étude du rôle agronomique et écologique des lombriciens

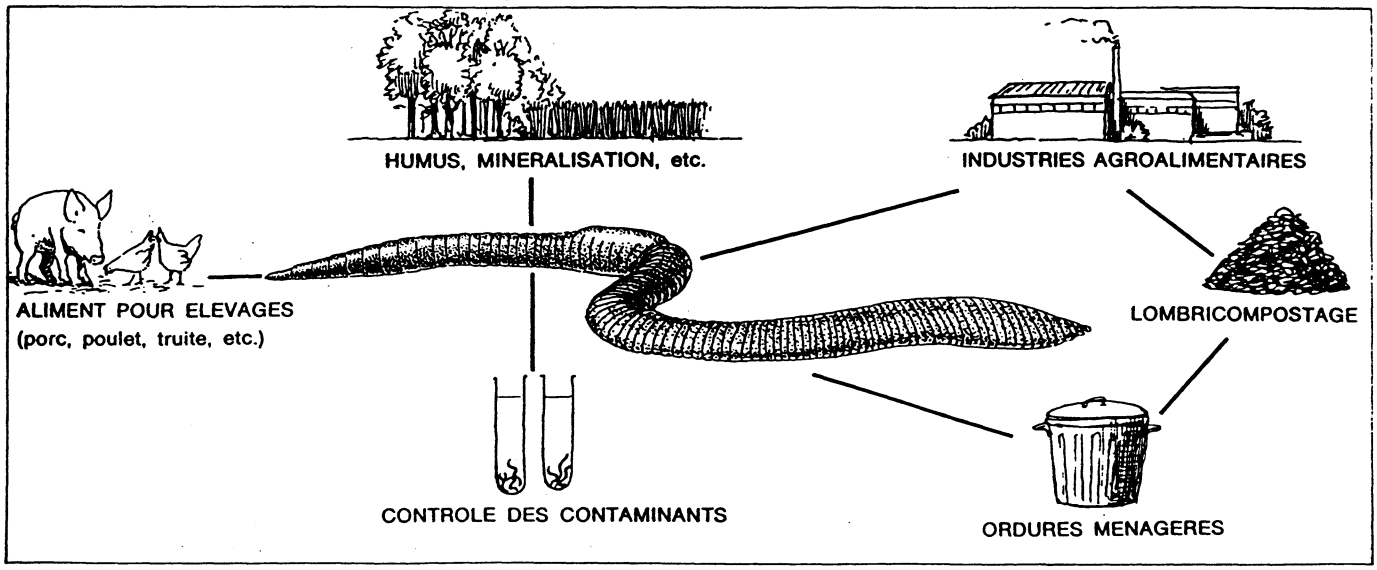


Figure 4-4 : *Rôle des lombriciens*

a) Enoncé de l'étude

Cette modélisation de l'organisation des données relatives aux lombriciens s'appuie sur les travaux menés principalement en France par le laboratoire de zooécologie du sol (INRA). Cette étude est focalisée sur la recherche du rôle agronomique, donc écologique des lombriciens.

Les lombriciens, d'un point de vue taxonomique appartiennent aux *Annelida*, classe des *Oligochaeta* ; à l'intérieur de cette classe, ils rentrent dans l'ordre des *Opisthopora*.

L'établissement, le développement, le jugement des diverses agropratiques ignorent normalement l'un des principaux auxiliaires de nos cultures : les lombriciens. Toutefois, cette situation change lentement à la fois par un début de prise de conscience de leur rôle et par les difficultés croissantes de l'application aveugle des pratiques agricoles qui conduisent à des impasses préoccupantes.

De nombreux problèmes relatifs à l'érosion hydrique et éolienne des sols, à la perte de structure de ceux-ci (battance, semelles de labour), aux pollutions azotées, à l'état phytosanitaire, résultent de la destruction ou/et mauvais usage de ces animaux. Ceux-ci

représentent en effet entre 60 et 80% de la masse animale active totale dans le sol et ont donc un impact réel considérable usuellement négligé en raison de la sectorialisation des connaissances.

Les lombriciens ont des rôles très différents, ce rôle dépend beaucoup plus de l'aptitude d'une population locale que des traits de chaque espèce (il peut y avoir, sous le même nom, des populations ayant des rôles très différents). Pour cette raison, il est souhaitable (au niveau de l'interprétation des données) de regrouper ces animaux en ensembles ayant le même rôle (= catégories écologiques), car c'est le rôle de ces animaux qui nous intéresse au plan technique et économique.

En pratique, le poids de chaque individu est le meilleur niveau d'analyse car il permet ensuite par regroupement successif de connaître le poids des différentes espèces et catégories.

Chaque point élémentaire d'étude des lombriciens (plot) est repéré spatialement à l'intérieur d'un lieu-placette donné par des repères orthogonaux dites "coordonnées locales" (X, Y, Z) à un moment donné (date) ; quand il y a lieu sont notées pour chaque point élémentaire ses cotes ou dimensions (x, y, z) et la durée de prélèvement.

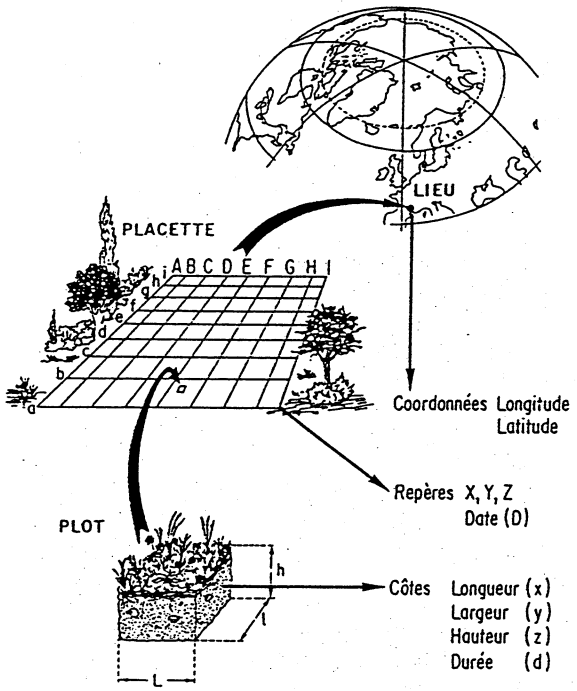


Figure 4-5 : *Situation spatiale et dimensions d'un plot*

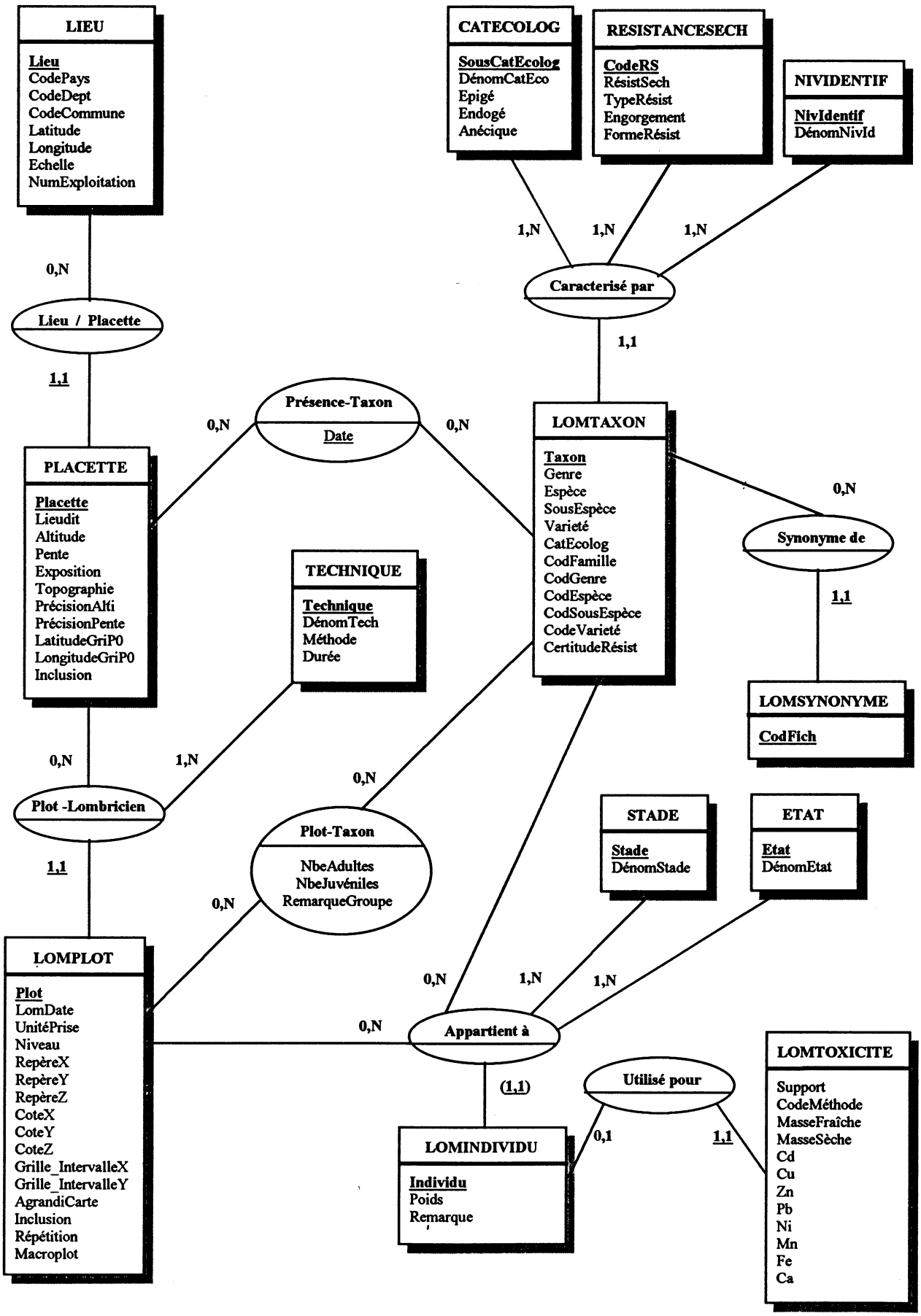
Les divers rôles des lombriciens sont usuellement rapportés à la biomasse de catégories écologiques fonctionnelles. Ils peuvent aussi être quantifiés principalement in situ en terme de travail physique de ces animaux (agrégats organo-minéraux, galeries,...) et ses conséquences (anti-lessivage, percolation de l'eau,...), en terme du cycle des éléments biogènes ; celui-ci couvre jusqu'à 80% des assimilats des végétaux en azote (observations en prairie permanente non perturbée), en terme de transfert d'éléments traces, y compris indésirables (organohalogénés, métaux lourds,...) et en terme de production de "viande" assurant la diététique (lysine, méthionine, iode, etc.) de nombreux omnivores et carnivores vertébrés.

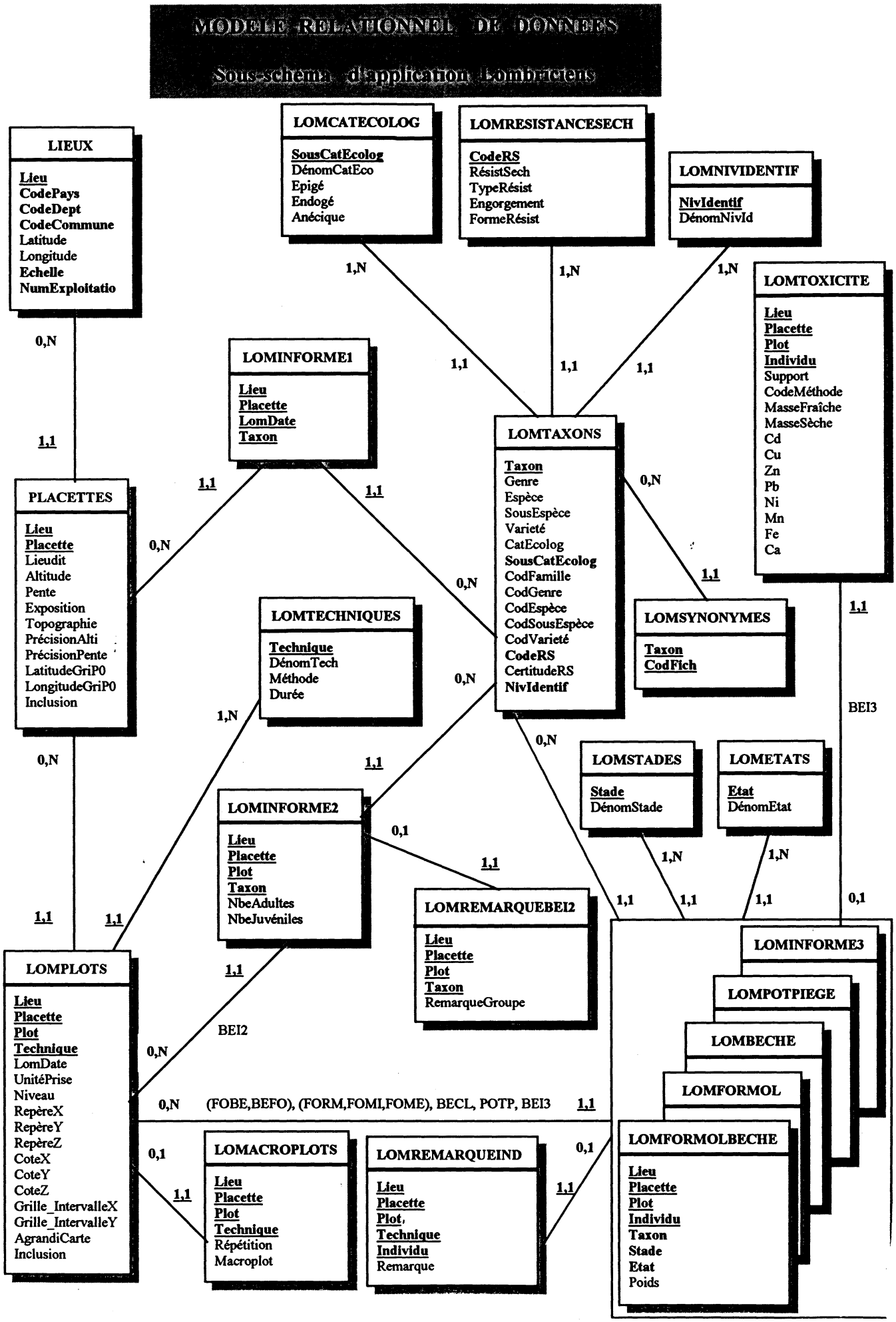
En raison de l'omniprésence des lombriciens dans les divers milieux (forêt, prairie, cultures), leur étude permet une approche située entre différentes disciplines comme la zoologie, la physique du sol, l'agronomie, la biologie du sol, l'étude du gibier et de la faune sauvage, la foresterie, le pastoralisme, etc.

La modélisation proposée dans cette partie est destinée au développement des relations constituant l'évaluation de la biomasse lombricienne dans les différents milieux naturels. L'approche avec les autres disciplines sera décrite dans les exemples correspondants. Notons toutefois que l'objectif est de désectorialiser les approches des diverses disciplines en leur fournissant l'information ayant trait à ces animaux et à leurs rôles en **relation** avec les agropratiques développées à partir des objectifs de chacune d'entre elles.

MODELE CONCEPTUEL DE DONNEES

Sous-schéma d'application Lombricien




MODELE RELATIONNEL DE DONNEES
Sous-schéma d'application Lombricens
LIEUX
Lieu
CodePays
CodeDept
CodeCommune
Latitude
Longitude
Echelle
NumExploitatio
LOMCATECOLOG
SousCatEcolog
DénomCatEco
Epigé
Endogé
Anécique
LOMRESISTANCESECH
CodeRS
RésistSech
TypeRésist
Engorgement
FormeRésist
LOMNIVIDENTIF
NivIdentif
DénomNivId
LOMTOXICITE
Lieu
Placette
Plot
Individu
Support
CodeMéthode
MasseFraîche
MasseSèche
Cd
Cu
Zn
Pb
Ni
Mn
Fe
Ca
LOMINFORME1
Lieu
Placette
LomDate
Taxon
LOMTAXONS
Taxon
Genre
Espèce
SousEspèce
Varieté
CatEcolog
SousCatEcolog
CodFamille
CodGenre
CodEspèce
CodSousEspèce
CodVarieté
CodeRS
CertitudeRS
NivIdentif
PLACETTES
Lieu
Placette
Lieudit
Altitude
Pente
Exposition
Topographie
PrécisionAlti
PrécisionPente
LatitudeGriP0
LongitudeGriP0
Inclusion
LOMTECHNIQUES
Technique
DénomTech
Méthode
Durée
LOMSYNONYMES
Taxon
CodFich
LOMSTADES
Stade
DénomStade
LOMETATS
Etat
DénomEtat
LOMINFORME2
Lieu
Placette
Plot
Taxon
NbeAdultes
NbeJuvéniles
LOMREMARQUEBEI2
Lieu
Placette
Plot
Taxon
RemarqueGroupe
LOMPLOTS
Lieu
Placette
Plot
Technique
LomDate
UnitéPrise
Niveau
RepèreX
RepèreY
RepèreZ
CoteX
CoteY
CoteZ
Grille_IntervalleX
Grille_IntervalleY
AgrandiCarte
Inclusion
LOMINFORME3
LOMPOTPIEGE
LOMBECHE
LOMFORMOL
LOMFORMOLBECHE
Lieu
Placette
Plot
Individu
Taxon
Stade
Etat
Poids
LOMACROPLOTS
Lieu
Placette
Plot
Technique
Répétition
Macroplot
LOMREMARQUEIND
Lieu
Placette
Plot
Technique
Individu
Remarque
BEI2
BEI3
(FOBE,BEFO), (FORM,FOMI,FOME), BECL, POTP, BEI3

b) Schémas relationnels en troisième forme normale

LOMCATECOLOG (SousCatEcolog, DénomCatEco, Epigé(%), Endogé(%), Anécique(%))

LOMRESISTANCESECH (CodeRS, RésistSech, TypeRésist, Engorgement, FormeRésist)

LOMNIVIDENTIF (NivIdentif, DénomNivId)

LOMTAXONS (Taxon, Genre, Espèce, SousEspèce, Varieté, CatEcolog, SousCatEcolog, CodFamille, CodGenre, CodEspèce, CodSousEspèce, CodVarieté, CodeRS, CertitudeRS, NivIdentif)

LOMTECHNIQUES (Technique, DénomTech, Méthode, Durée(mn))

LOMPLOTS (Lieu, Placette, Plot, Technique, LomDate, UnitéPrise(cm^2), Niveau(-cm), RepèreX, RepèreY, RepèreX, RepèreY, RepèreZ, CoteX(cm), CoteY(cm), CoteZ(cm), GrilleIntervalleX(cm), GrilleIntervalleY(cm), AgrandiCarte, Inclusion)

LOMMACROPLOTS (Lieu, Placette, Plot, Technique, Répétition, MacroPlot)

LOMSTADES (Stade, DénomStade)

LOMETATS (Etat, DénomEtat)

LOMINFORME1 (Lieu, Placette, LomDate, Taxon)

LOMINFORME2 (Lieu, Placette, Plot, Taxon, NbeAdultes, NbeJuveniles)

LOMINFORME3 (Lieu, Placette, Plot, Individu, Taxon, Stade, Etat, Poids(mg))

LOMFORMOL (Lieu, Placette, Plot, Individu, Taxon, Stade, Etat, Poids(mg))

LOMBECHE (Lieu, Placette, Plot, Individu, Taxon, Stade, Etat, Poids(mg))

LOMFORMOLBECHE (Lieu, Placette, Plot, Individu, Taxon, Stade, Etat, Poids(mg))

LOMPOTPIEGE (Lieu, Placette, Plot, Individu, Taxon, Stade, Etat, Poids(mg))

LOMTOXICITE (Lieu, Placette, Plot, Individu, Support, CodeMéthode, MasseFraîche(mg), MasseSèche(mg), Cd(ppm), Cu(ppm), Zn(ppm), Pb(ppm), Ni(ppm), Mn(ppm), Fe(ppm), Ca(ppm))

LOMSYNONYMES (Taxon, CodFich)

LOMREMARQUEBEI2 (Lieu, Placette, Plot, Taxon, RemarqueGroupe)

LOMREMARQUEIND (Lieu, Placette, Plot, Technique, Individu, Remarque)

c) Clés des relations du sous-schéma lombriciens

LOMCATECOLOG :	SousCatEcolog
LOMRESISTANCESECH :	CodeRS
LOMNIVIDENTIF :	NivIdentif
LOMTAXONS :	Taxon
LOMTECHNIQUES :	Technique
LOMPLOTS :	Lieu, Placette, Plot, Technique
LOMMACROPLOTS :	Lieu, Placette, Plot, Technique
LOMSTADES :	Stade
LOMETATS :	Etat
LOMINFORME1 :	Lieu, Placette, LomDate, Taxon
LOMINFORME2 :	Lieu, Placette, Plot, Taxon
LOMINFORME3 :	Lieu, Placette, Plot, Individu
LOMFORMOL :	Lieu, Placette, Plot, Individu
LOMBECHE :	Lieu, Placette, Plot, Individu
LOMFORMOLBECHE :	Lieu, Placette, Plot, Individu
LOMPOTPIEGE :	Lieu, Placette, Plot, Individu
LOMTOXICITE :	Lieu, Placette, Plot, Individu
LOMSYNONYMES :	Taxon, CodFich
LOMREMARQUEBEI2 :	Lieu, Placette, Plot, Taxon
LOMREMARQUEIND :	Lieu, Placette, Plot, Technique, Individu

4.4.4. Sous-schéma d'application machinisme agricole : apport du machinisme à la modélisation globale de l'information sur l'exploitation agricole

a) Enoncé de l'étude

Le CIRAD (Centre de Coopération Internationale en Recherche Agronomique) est un organisme public implanté à Montpellier qui a pour vocation principale l'aide au développement agricole des pays défavorisés du sud (Afrique et Amérique latine). Pour mener à bien cette tâche il est divisé en départements ayant chacun une spécialité et des compétences particulières concernant l'agriculture de ces pays. C'est le cas du SAR (Systèmes Agro-alimentaires et Ruraux).

Pour la partie concernant le machinisme, le SAR joue un rôle de conseil pour équiper, au mieux, les exploitations agricoles en tenant compte des contraintes locales. L'accomplissement de cette mission à permis d'accumuler au fil du temps (le ex-CEEMAT a été créé en 1962) une expertise et un savoir-faire étayés par une très importante documentation commerciale sur le machinisme agricole.

Dans le cadre d'une coopération interinstitutionnelle et interdisciplinaire entre le CIRAD, l'INRA et l'ENSA de Montpellier nous avons pu élaborer un schéma conceptuel global du traitement de l'information sur le machinisme agricole comme étape nécessaire pour avancer dans la modélisation globale de l'information sur l'exploitation agricole. Simultanément nous avons imaginé un schéma d'ordonnancement des opérations qui nous a permis d'avancer dans les réalisations informatiques et de définir les limites du travail coopérative entre les différents partenaires.

A l'origine, un besoin exprimé par les acteurs du développement : équiper une exploitation en fonction des systèmes techniques de production agricole et des contraintes locales. Et un problème réel : développement de la demande et diminution du nombre d'experts capable d'y répondre.

La définition de ces objectifs nous a conduit à l'utilisation d'outils informatiques performants pour d'avancer dans la réalisation des applications nous permettant de déboucher sur un système d'aide à la décision.

Ce système doit être "alimenté" par trois composantes :

- les connaissances des experts,
- la connaissance des matériels et des fournisseurs existant sur le marché international,
- des données expérimentales (technico-économiques) actualisées provenant d'expérimentations sur le terrain et d'enquêtes.

Dans ce document nous décrirons la mise en place de cette dernière composante.

Grâce aux travaux de conception de systèmes d'informations réalisés par l'INRA, en collaboration avec le CIRAD-SAR depuis 1988 dans ce domaine, et en s'appuyant sur :

- le fonds documentaire de l'ex-CEEMAT,
- les rapport des expérimentations sur site,
- et les enquêtes des chercheurs sur le terrain ;

une base de données performante et informatisée, interrogeable à distance, multi-utilisateurs et autofinancée en phase d'exploitation a été entamée.

Enfin, les informations (données expérimentales) sur le machinisme agricole gérées actuellement dans la base de données ECORDRE sont également à envisager dans leur complémentarité avec les données sur les agro-équipements qui font déjà l'objet d'un schéma d'application dans la base de données AGRIMAF.

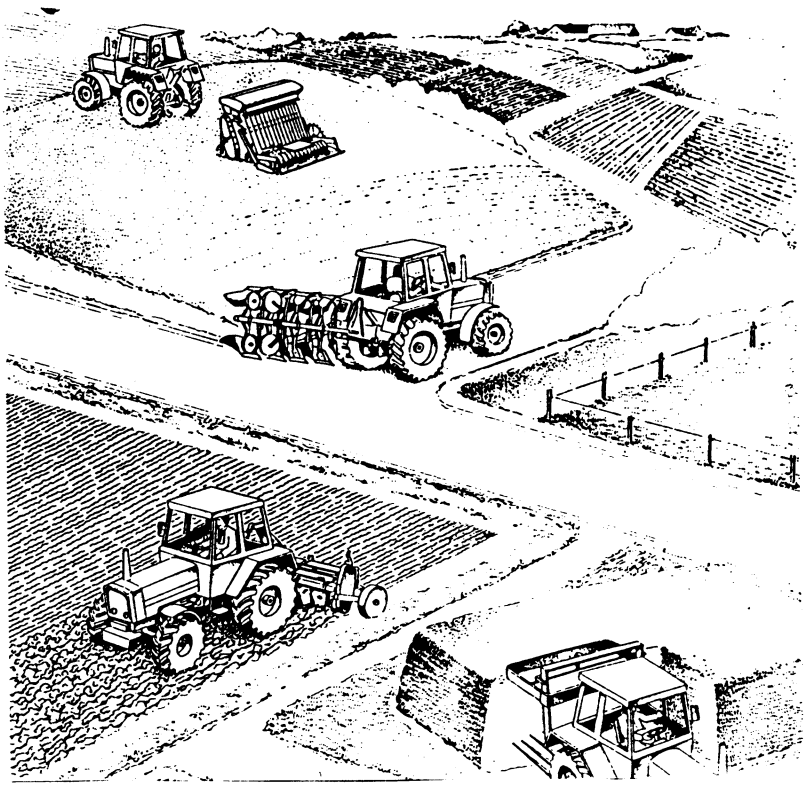
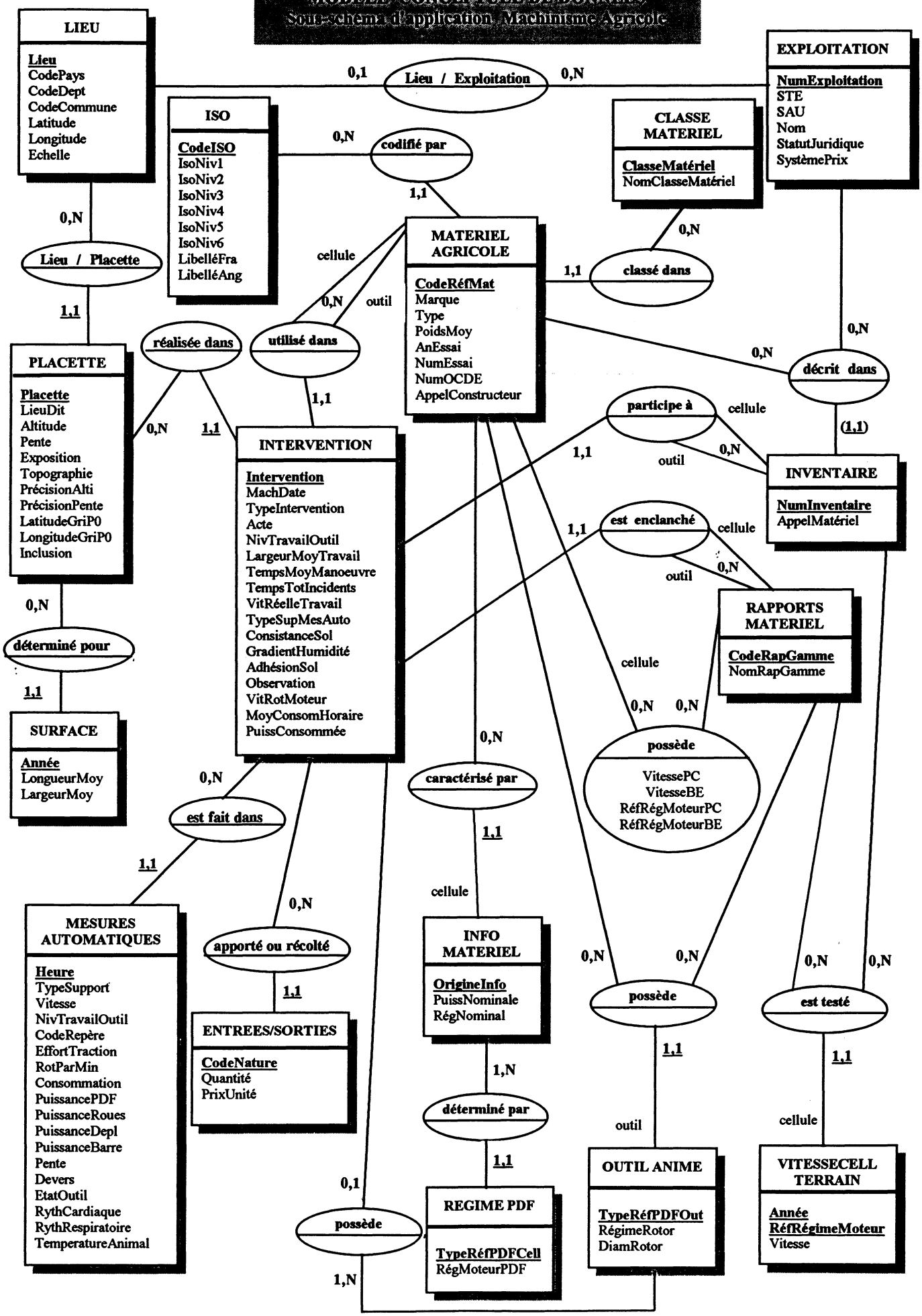
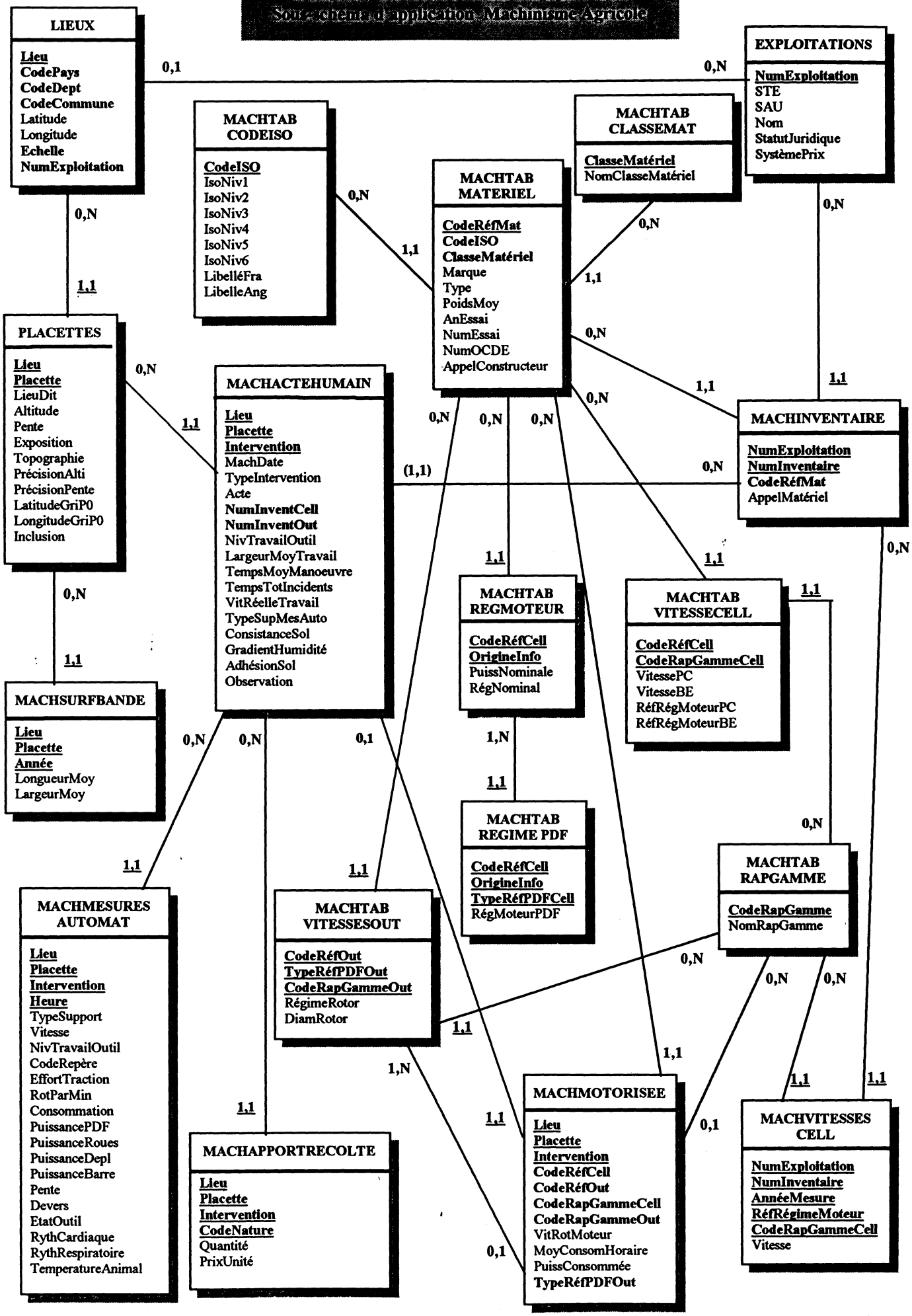


Figure 4-6 : *Le machinisme agricole*

MODELE CONCEPTUEL DE DONNEES
Sous-schéma d'application Machinisme Agricole



MODELE RELATIONNEL DE DONNEES
Sous-système d'application Machinisme Agricole



b) Schémas relationnels en troisième forme normale

MACHINVENTAIRE (NumExploitation, NumInventaire, CodeRéfMat, AppelMatériel)

MACHSURFBANDE (Lieu, Placette, Année, LongueurMoy(m), LargeurMoy(m))

MACHACTEHUMAIN (Lieu, Placette, Intervention, MachDate, TypeIntervention, Acte, NumInventCell, NumInventOut, LargeurMoyTravail(cm), TempsMoyManoeuvre(s), TempsTotIncidents(mn), VitRéelleTravail(km/h), ConsistanceSol, GradientHumidité, AdhésionSol, Observation)

MACHMOTORISEE (Lieu, Placette, Intervention, CodeRéfCell, CodeRéfOut, CodeRapGammeCell, CodeRapGammeOut, VitRotMoteur(tr/mn), MoyConsomHoraire(l/h), PuissConsommée(Kw), TypeRéfPDFOut)

MACHVITESSESCELL (NumExploitation, NumInventaire, AnnéeMesure, RéfRégimeMoteur, CodeRapGammeCell, Vitesse(k/h)

MACHMESURESAUTOMAT (Lieu, Placette, Intervention, Heure, TypeSupport, Vitesse, NivTravailOutil, CodeRepère, EffortTraction, RotParMin, Consommation, PuissancePDF, PuissanceRoues, PuissanceDepl, PuissanceBarre, Pente, Devers, EtatOutil, RythCardiaque, RythRespiratoire, TemperatureAnimal)

MACHAPPORTRECOLTE (Lieu, Placette, Intervention, CodeNature, Quantité, PrixUnité)

MACHTABCODEISO (CodeIso, IsoNiv1, IsoNiv2, IsoNiv3, IsoNiv4, IsoNiv5, IsoNiv6, LibelléFra, LibelléAng)

MACHTABCLASSEMAT (ClasseMatériel, NomClasseMatériel)

MACHTABMATERIEL (CodeRéfMat, CodeIso, ClasseMatériel, Marque, Type, PoidsMoy, AnEssai, NumEssai, NumOCDE, AppelConstructeur)

MACHTABREGMOTEUR (CodeRéfCell, OrigineInfo, PuissNominale, RegNominal)

MACHTABREGIMEPDF (CodeRéfCell, OrigineInfo, TypeRéfPDFCell, RegMoteurPDF)

MACHTABRAPGAMME (CodeRapGamme, NomRapGamme)

MACHTABVITESSECELL (CodeRéfCell, CodeRapGammeCell, VitessePC, VitesseBE, RéfRégMoteurPC, RéfRégMoteurBE)

MACHTABVITESSEOUT (CodeRéfOut, TypeRéfPDFOut, CodeRapGammeOut, RégimeRotor, DiamRotor)

c) Clés des relations du sous-schéma machinisme

MACHINVENTAIRE :	NumExploitation, NumInventaire
MACHSURFBANDE :	Lieu, Placette, Année
MACHACTEHUMAIN :	Lieu, Placette, Intervention
MACHMOTORISEE :	Lieu, Placette, Intervention
MACHVITESSESCELL :	NumExploitation, NumInventaire, AnnéeMesure, RéfRégimeMoteur, CodeRapGammeCell
MACHMESURESAUTOMAT :	Lieu, Placette, Intervention, Heure
MACHAPPORTRECOLTE :	Lieu, Placette, Intervention, CodeNature

MACHTABCODEISO :	CodeIso
MACHTABCLASSEMAT :	ClasseMatériel
MACHTABMATERIEL :	CodeRéfMat
MACHTABREGMOTEUR :	CodeRéfCell, OrigineInfo
MACHTABREGIMEPDF :	CodeRéfCell, OrigineInfo, TypeRéfPDFCell
MACHTABRAPGAMME :	CodeRapGamme
MACHTABVITESSECELL :	CodeRéfCell, CodeRapGammeCell
MACHTABVITESSEOUT :	CodeRéfOut, TypeRéfPDFOut, CodeRapGammeOut

4.4.5. Sous-schéma d'application viticulture : étude des mécanismes de l'économie viti-vinicole et du marché du vin au niveau national et international

a) Enoncé de l'étude

Cette modélisation prend en compte une partie de l'organisation des données concernant l'étude de l'économie viti-vinicole au niveau national et international. Elle s'appuie sur les travaux de recherche menés par l'équipe viticole de la Station d'Economie et Sociologie Rurales de l'INRA de Montpellier.

Figure 4-7 : *L'économie viti-vinicole*

L'objectif global des recherches est de comprendre les mécanismes de l'économie viti-vinicole dans sa complexité et diversité, en prenant en compte principalement la coexistence de prospérité et de crise, la difficulté d'ajustement offre/demande et les politiques aux effets contradictoires.

Dans le cadre de cet objectif, des recherches sont poursuivies pour comprendre les mutations en cours, pour préciser les contraintes qui s'imposent et pour identifier les évolutions possibles.

Pour la réalisation des ces travaux une importante documentation statistique doit être mise en oeuvre. Des mises à jour régulières sont nécessaires, de même qu'un enrichissement permanent pour de nouvelles séries de données.

Les données fournies par différentes sources depuis 1980 sont gérées actuellement dans la Base de Données "ECORDRE".

Ces données sont toutes de type statistique, c'est-à-dire qu'elles correspondent, dans le cadre du réel d'ECORDRE, à des données secondaires élaborées (DSE). Donc l'ensemble des tables de cette application contient actuellement, soit des données secondaires élaborées, soit des données descriptives.

La modélisation proposée est destinée au développement d'un outil de gestion informatique des données permettant l'étude de l'évaluation des fluctuations du marché du vin.

b) Identification des objets de l'application (le dictionnaire de données)

Nous disposons des informations de base suivantes :

I.- Production de vin par pays
- nom des pays producteurs de vin,
- années de production (depuis 1970),
- superficie totale annuelle en hectares par pays,
- production totale annuelle en hectolitres par pays,
- organisme source de l'information (sigle).

II.- Production de vin en France
- code INSEE des départements de la France,
- années de production (depuis 1980),
- nombre annuel de déclarants par département,
- nom des différents types ou catégories de vin,
- couleur par type ou catégorie de vins,
- superficie totale annuelle en hectares par département et type de vin,

- volume total annuel en hectolitres par département et par couleur
  pour chaque catégorie,
- organisme source de l'information (sigle).

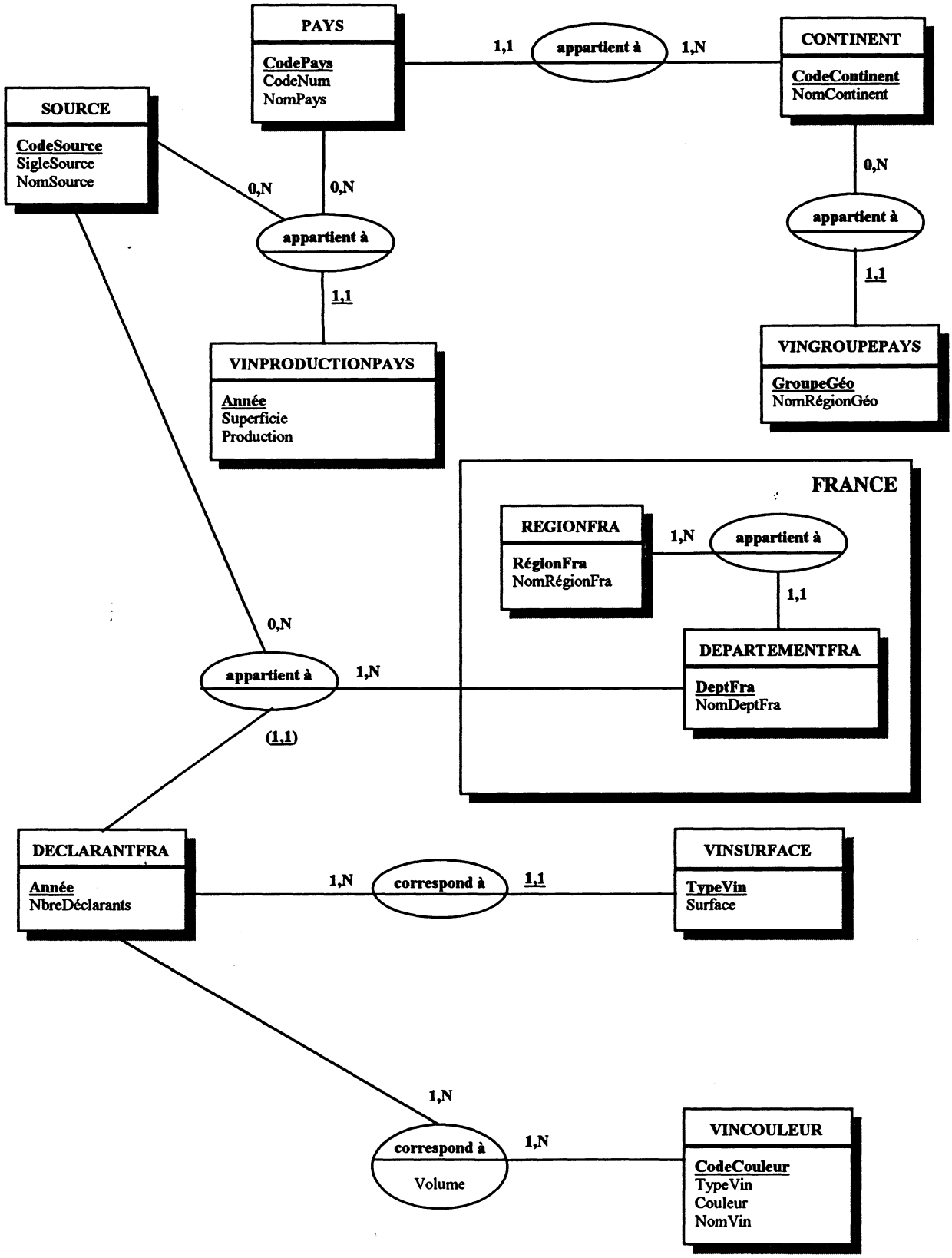
Pour des raisons de stockage de données, de rapidité d'accès à l'information, de liens entre les différentes tables physiques et pour des raisons d'édition et présentation des différents types de données, on enrichira le dictionnaire de données d'une part, avec des identificateurs ou informations clés (numériques ou alphabétiques) et d'autre part, avec des descripteurs simples décrivant en clair les codes ou attributs clés. Tous ces attributs bien entendu reliés aux objets (relations) qu'ils décrivent le plus directement.

Par le dictionnaire de données décrit précédemment, nous connaissons :

- Les **noms des pays** producteurs de vin (NomPays) en clair. Nous les échangeons par les codes ISO des pays (Pays), comme identificateur pour chaque pays, nous utilisons le code ISO à trois caractères alphabétiques (exemple : France = FRA) ; le nom en clair de chaque pays, fait partie de la table de référence PAYS à l'intérieur du sous-schéma d'application *administration géographique*. En même temps chaque pays doit être lié pour des raisons d'édition à un continent et à une région géographique (à l'intérieur d'un continent). Nous ajouterons un code groupe (Groupe) et un code sous groupe (SousGroupe) comme identificateur pour les continents et les régions géographiques et leurs dénominations en clair comme descripteurs simples (NomContinent et NomRégionGéo).
- Les **sigles des organismes source** d'informations (SigleSource). Nous ajouterons un code numérique comme identificateur de la source (Source) et le nom en clair de l'organisme comme attribut simple (NomSource).
- Les **noms des vins** en France (NomVin). Nous ajouterons un code identificateur à quatre caractères alphabétiques (TypeVin) pour chaque type ou catégorie de vin (exemple : vin de table = TVDT).
- Pour chaque type de vin en France nous connaissons **les couleurs** (Couleur). Nous ajouterons un code identificateur à quatre caractères alphabétiques (CodeCouleur) pour identifier de manière unique chaque couleur liée à un type de vin (exemple : Vin de Consommation Courante (vin de table) Blanc = VCCB).
- Nous ajouterons l'attribut VolumeCat correspondant à la somme de la production par catégorie pour optimiser les accès aux données.

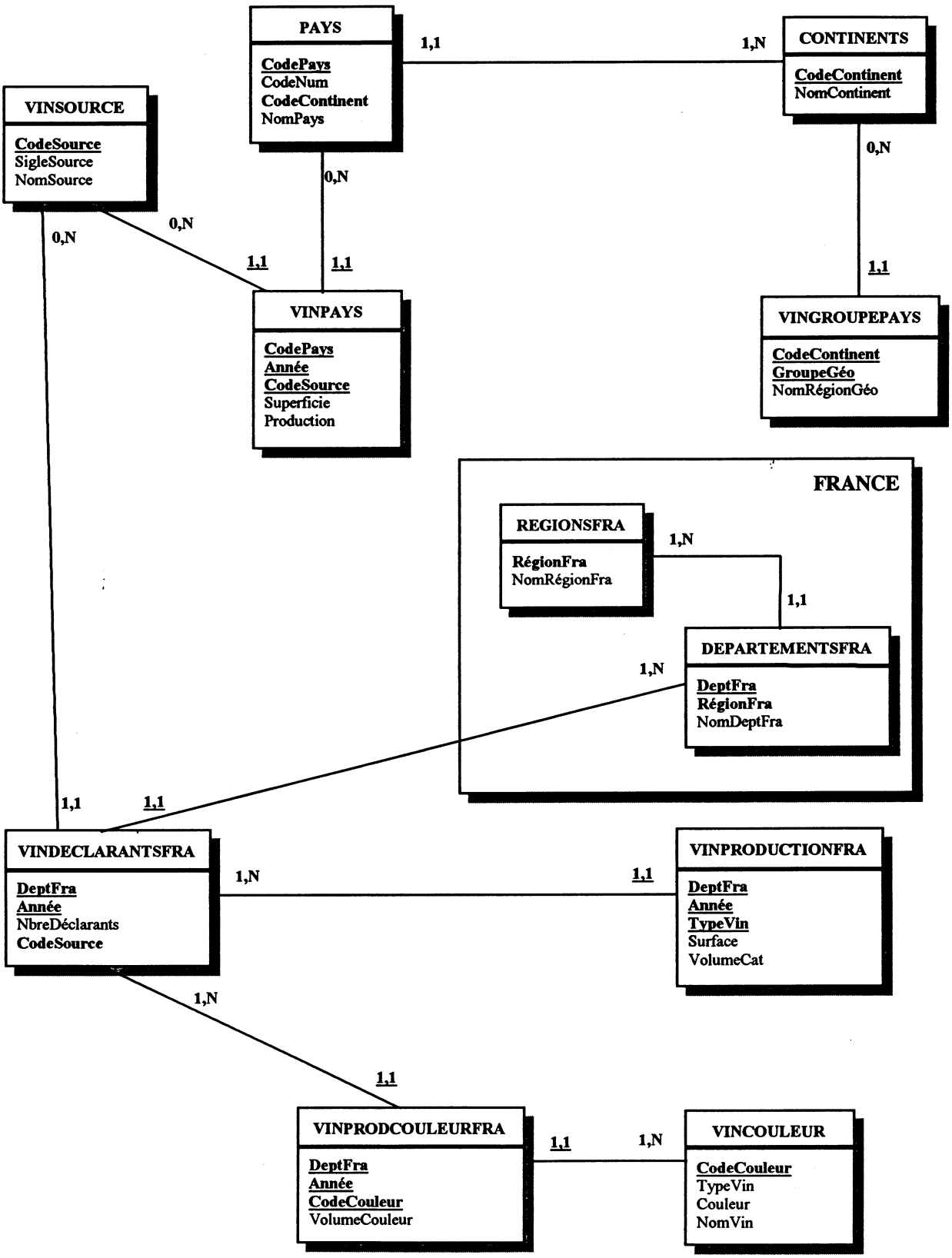
MODELE CONCEPTUEL DE DONNEES

Sous-schéma d'application économie viti-vinicole



MODELE RELATIONNEL DE DONNEES

Sous-schéma d'application Economie viti-vinicole



c) schémas relationnels en troisième forme normale

Le but de la normalisation est de renforcer l'intégrité des données en supprimant les redondances, ce qui facilite la mise à jour des données. On obtient ainsi une meilleure gestion des propriétés rattachées à un identifiant et une redondance limitée aux clés. La normalisation permet aussi d'éviter que certains attributs ne prennent trop souvent la valeur nulle (absence d'information), ce qui prend une place inutile dans les tables.

Une table en 1NF satisfait la condition que chaque attribut dans la table est atomique, c'est-à-dire en valeur unique, et admet au moins une clé. Une table en 2NF est une table en 1NF qui satisfait en plus la condition que tous les attributs ne formant pas la clé dépendent globalement de la clé. Une table en 3NF est en 2NF, et chaque attribut non clé dépend directement de la clé primaire.

VINPAYS (CodePays, Année, CodeSource, Superficie(ha), Production(hl))

VINDECLARANTSFRA (DeptFra, Année, NbreDéclarants, CodeSource)

VINPRODUCTIONFRA (DeptFra, Année, TypeVin, Surface(ha), VolumeCat(hl))

VINPRODCOULEURFRA (DeptFra, Année, CodeCouleur, VolumeCouleur(hl))

VINCOULEUR (CodeCouleur, TypeVin, Couleur, NomVin)

VINSOURCES (CodeSource, SigleSource, NomSource)

VINGROUPEPAYS (CodeContinent, GroupeGéo, NomRégionGéo)

d) Clés des relations du sous-schéma viticulture

La clé primaire d'une relation peut être constituée d'un attribut ou de plusieurs attributs identifiant de façon unique les valeurs des attributs non clé de la relation.

VINPAYS :	CodePays, Année, CodeSource
VINDECLARANTSFRA :	DeptFra, Année
VINPRODUCTIONFRA :	DeptFra, Année, TypeVin
VINPRODCOULEURFRA :	DeptFra, Année, CodeCouleur
VINCOULEUR :	CodeCouleur
VINSOURCES :	CodeSource
VINGROUPEPAYS :	CodeContinent, GroupeGéo

CONCLUSION GENERALE

Pourquoi ce rapport sur l'organisation du système informatique de l'IFEN ? Pour trois raisons principales :

D'abord avec deux objectifs personnels :

- faire le point de l'expérience et des connaissances acquises dans ce domaine au cours de ces dernières années d'activité professionnelle et de formation. Cette expérience a été diverse ; par exemple : des constats personnels faisant suite à des résultats positifs ou négatifs (ces derniers ayant le plus souvent été plus bénéfiques que les précédents), d'autres sont le fruit d'observations et expériences diverses avec mes supérieurs, mes collègues et mes collaborateurs dans les domaines de la recherche scientifique et technique aux niveaux de l'écologie, de l'environnement, de la modélisation, de l'évaluation des systèmes d'information, de l'analyse, de la programmation et du génie logiciel.
- préparer les bases d'une éventuelle évolution dans ma prise de responsabilités au niveau professionnel. Or à ce point de ma carrière comme ingénieur informaticien à l'INRA je suis très sollicité. Malgré des horaires assez chargés (mais non exagérés), je me suis imposé un temps de réflexion et de prise de recul en mettant en chantier ce travail.

Ensuite, pour des raisons moins égoïstes, ce travail pouvant servir à comprendre la réalité et l'impact des modifications induites par les actes humains sur l'environnement dans lequel nous vivons. Apprécier l'impact des actes, les avantages/inconvénients de toutes variante technologique industrielle ou agricole devrait se fonder sur une connaissance des écosystèmes exploités ou influencés par l'homme. L'absence totale actuelle d'une telle connaissance globale (réellement intégrée) de ces systèmes écologiques laisse nos évaluations cantonnées à des approches réductionnistes se fondant sur les apports de quelques disciplines scientifiques et techniques.

Ensuite, pour des raisons permettant de faire resortir une valeur ajoutée du travail d'équipe dans le domaine de :

- une problématique transdisciplinaire qui n'est pas placée usuellement au centre des préoccupations des organismes de recherche classiques,
- une dynamique d'un organisme en création où il est possible, sans avoir à remettre en cause de pratiques sectorielles antérieures quelquefois sclérosantes, de mettre en oeuvre une approche moderne, notamment systémique dans l'organisation de l'entreprise,
- un développement de la réflexion où les problèmes d'éthique et de propriété de données sont centraux,

- une plateforme de mise en valeur au niveau national, puis communautaire, de nos travaux en écologie, en environnement et en suivi de projets informatiques dans ces domaines.

Ce travail essaie aussi de montrer les limites des organisations informatiques d'aujourd'hui par rapport au monde du possible et du rêvé. Il pourra ensuite être utile aux décideurs de l'Institut qui ont à dialoguer avec les managers informatiques et institutionnels au niveau national, régional et international et établir avec eux des synergies constructives pour le devenir de l'Institut.

Je dirai enfin, que la mise en oeuvre de la stratégie informatique de l'IFEN est une ambition collective qui dépasse largement ses frontières institutionnelles et qui requiert, pour sa maîtrise, des compétences multiples, un budget précis, une organisation adaptée aux changements permanents (aux niveaux conceptuel, technologique, institutionnel, politique), des outils de pilotage fins et des leaderships performants.

Les trois domaines d'action principaux de cette stratégie sont le développement applicatif, l'architecture technique et la production. Pour le premier, les efforts essentiels portent sur le choix des solutions d'acquisition de logiciel, l'architecture applicative, la mise en oeuvre des méthodes disponibles, le recours aux compétences externes, la maîtrise de la complexité et de la maintenance applicative. Le second domaine concerne la veille technologique, les choix d'architecture technique, la recherche de la performance des moyens en place, la maîtrise des changements matériels et logiciels de base et des dossiers de financement. La production enfin couvre les aspects d'automatisation de l'exploitation, de gestion de la périphérie et de l'environnement, le recours aux facilitées de management et la facturation des services. En plus des techniques managériales générales, les responsables de ces domaines doivent avoir recours à des méthodes et outils de management spécifiques.

L'efficacité de la stratégie informatique de l'IFEN passe également par l'insertion et l'ouverture dans son environnement. Celui-ci touche les utilisateurs des systèmes informatiques avec tous les problèmes d'organisation et de méthodes de travail, ainsi que ses partenaires avec tous les aspects image de marque de l'Institut. Il couvre aussi tous le moyens mis en oeuvre autour de l'informatique de production (infocentre et bureautique notamment) et avec lesquels l'informatique doit s'articuler efficacement.

BIBLIOGRAPHIE

ACKERMAN, T.P., et W.P. CROPPER Jr, 1988 - Scaling global climate projections to local biological assessments. Environment, **30**, 5, 31-34.

ADIBA, M.; GARDARIN G.; ROLLAND C.; DEMOLOMBE R.; SCHOLL M.; ROHMER J. 1986 - Nouvelles perspectives des bases de données. Travaux dans le contexte du programme de recherche coordonné bases de données de troisième génération (PRC-BD3), Ed. Eyrolles, Paris, 1-302.

A.E., 1972 - Alpha encyclopédie. Ed. Grange batelière, Paris, 1-6240.

ALLEGRE, C., 1983 - L'écume de la terre. Ed. Fayard, Paris, 1-366.

ANDRO, T., J. BACHACOU, P. DELHOTAL, J.C. FAYET, J. LE RENARD, J.-P. RELLIER, C. THIOLON et V. WAHL, 1988 - SEPV : un ensemble de Systèmes Experts pour l'aide au diagnostic des maladies des plantes cultivées en France. In proc. VIIth int. workshop "Expert Systems and their applications" (Avignon, France, 13-15/05/87), 173-186.

ANSI, 1975 - ANSI/X3/SPARC, Study group on data base management systems interim report 785-02-08. ACM SIGMOND Newsletter, Vol 7, N°2, 1975.

ANSI, 1986 - ANSI/X3/135, Database Language SQL. Also available as International Standards Organization Document ISO/TC97/SC21/WG3 N117.

ARMSTRONG, W.W., 1974 - Dependency structures of database relationships. In proceedings, IFIP, North-Holland, Amsterdam 1974, pp. 580-583.

ASTRAHAN, M.M. et al., 1976 - System R : Relational Approach to Database Management. ACM transactions on Database Systems 1, N°2, June 1976.

ATZENI, P. et E.P.F.CHAN, 1987 - Independent database schemes under funtional and inclusion dependencies. In proceedings of 13th VLDB Conference, Brighton 1987, pp 159-166.

AZAR, N., 1987 - Conception d'un système d'information utilisant les modèles Entité-Association ou Relationnel avec inclusions. Thèse de Doctorat, Université de Montpellier II, 1987.

BERNARD, J., et M. PARKER, 1980 - Macroscopie de l'entreprise, Systémique appliquée. Ed. d'Organisation, 1-214.

VON BERTALANFY L., 1973 - Théorie générale des systèmes. Dunod, Titre anglais : General system theory, foundation, development, applications. Braziller, New-York, 1968.

BERTELLIN, J., et F. TOUTAIN, 1979 - Biologie des sols. *In* Bonneau et Souchier (dir.), éd. Masson, Paris, 123-160.

BERTHET, C., 1982 - Les graphes et leur visualisation : réalisation d'un outil d'aide à la décision. Thèse de Docteur-ingénieur, Université Claude Bernard Lyon I, 1982.

BLONDEL, J., 1979 - Biogéographie et écologie. Ed. Masson, Paris, coll. écologie, **15**, 1-173.

BLONDEL, J., 1979 - Ecologie et gestion de l'espace naturel, l'apport du "modèle oiseaux". Coll. CNRS, Ecologie et Développement, 135-146.

BLONDEL, J. et J.P. CHOISY, 1983 - Biogéographie des peuplements d'oiseaux à différentes échelles de perception : de la théorie à la pratique. Acta Oecologica, Oecol. gener., 4, 89-110.

BLONDEL, J., 1987b - Biogéographie évolutive. Ed. Masson, Paris, coll. écol., 20, 1-221.

BOLIN, B., 1981 - Carbon cycle modelling. Ed. John Wiley, Chichester, 1-390.

BOUCHE, M.B., 1972 - Lombriciens de France. Ecologie et systématique. Ed. INRA, Ann. zool. - Ecol. anim., n° spéc., 72-2, 1-671.

BOUCHE, M.B., 1984b - Ecotoxicologie des lombriciens. II. Surveillance de la contamination des milieux. Oecol. applic., **5**, 4, 291-301.

BOUCHE, M.B., 1984 - Les modalités d'adaptation des lombriciens à la sècheresse. Bull. soc. bot. Fr., 131, actual. bot., 319-327.

BOUCHE, M.B., 1984 - Les vers de terre. La Recherche, 156, juin 1984, Vol 15, 796-804.

BOUCHE, M.B. et R. ALIAGA, 1986 - Contre une dégradation physique et chimique des sols et pour leur optimisation économique, l'échantillonnage des lombriciens : une urgente nécessité. La Défense des Végétaux, 242, 30-36.

BOUCHE, M.B., P. AUPETIT et M.-A. PHILIPPE, 1988 - Etude agroenvironnementale intégrée du Marais Poitevin. Doc. pédozool., 1, 3, 1-126.

BOUCHE, M.B., 1990 - Ecologie opérationnelle assistée par ordinateur. Ed. Masson, Paris, 1-572.

BOUGUERRA, M.-L., 1986 - Les pesticides et le tiers monde. La Recherche, **17**, 176, 544-553.

BRILLOIN, L., 1956 (1959) - Science and information theorie. Ed. Acad. Press New York, trad. franç. 1959 - La science et la théorie de l'information, Ed. Masson, Paris, 1-302.

CABRERA, Ch., 1987 - Normalisation d'une base de données prenant en compte les dépendances fonctionnelles et les dépendances d'inclusion. Thèse de Doctorat, Université de Montpellier II, 1987.

CALKINS, H., TOMLINSON R.F., 1977 - Geographic information systems, methods and equipments for land use planning. Int. commission on geographical data sensing and processing, Reston, 1977.

CALKINS, H., 1983 - A pragmatic approach to GIS design. Int. commission of geographical data sensing and processing, New York, 1983.

CALOW, P., 1987 - Toward a definition of functional ecology. Functional ecology, **1**, 57-61.

CANCELA DA FONSECA, J.-P., 1974 - Théorie des systèmes : son application à l'étude des écosystèmes. Ed. lab. zool. écol., Rennes, 1-116.

CASIMIR, D.G., 1982 - Ecosystème prairial. Acta oecol. Oecol. gen., **3**, 1, 1-215.

C.D.A.F., 1842 - Complément du Dictionnaire de l'Académie Française. Ed. F. Didot, Paris, 1-1281.

CHAMBERLIN, D.D. et al., 1981 - A History and Evaluation of System R. Communications of the ACM 24, N°10, Octobre 1981.

CHEMIN, J.-L., 1976 - Savoir préserver l'environnement. La Recherche, **7**, 63, 83-85.

CHEN, P.P., 1976 - The Entity-Relationship Model : Toward a Unified View of Data. ACM Transactions on Database Systems, Vol.1 N°1, January 1976, pp. 9-36.

CODD, E.F., 1970 - A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, Vol.13 N°6, June 1970, pp. 377-387. Reprinted in Communications of the ACM 26, N°1, January 1983.

CODD, E.F., 1972 - Further Normalization of the Data Base Relational Model. In Data Base Systems (R. Rustin Editor), Prentice-Hall, Englewood Cliffs, New Jersey, 1972, pp. 33-64.

COLLONGUES, A. et J. HUGUES, B. LAROCHE, 1987 - MERISE méthode de conception, Ed. Dunot 2ème édition, Paris, 1-240.

DAGET, P. et GODRON, M., 1982 - Analyse fréquentielle de l'écologie des espèces dans les communautés. Ed. Masson, Paris, 1-163.

DARWIN, C.R., 1859 - The origin of species by means of natural selection, or the preservation of favored races in the struggle for life (6ème éd., corrigée 1878). Ed. J. Murray, London, 1-458.

DATE, C., 1989 - Introduction au standard SQL. Ed. InterEditions, Paris, 1-239.

DAVID, B., 1988 - Le modèle Spatiarel : une extension du modèle relationnel pour gérer les données spatiales surfaciques. 4e Journées Bases de données Avancées, Bénodet, mai 1988.

DELATTRE, P., et H. THELLIER, 1979 - Elaboration et justification des modèles, applications en biologie. Ed. Maloine, Paris, 1-370.

DELOBEL, C. et M. ADIBA, 1982 - Bases de Données et Systèmes Relationnels. Dunod, Paris, 1-449.

DI CASTRI, F., 1984 - L'écologie. Les défis d'une science en temps de crise. Rapport au Ministre Ind. Rech., éd. La Documentation française, Paris, 1-116.

DI CASTRI, F., et M. HADLEY, 1986 - Enhancing the credibility of ecology : is interdisciplinary research for land use planning useful ? Geojournal, **13**, 4, 299-325.

DI CASTRI, F., et M. HADLEY, 1988 - Enhancing the credibility of ecology : interacting along and accross hierarchical scales. Geojournal, **17**, 1, 5-35.

DURAND, D., 1979 - La systémique. Que sais-je ? PUF, 1-125.

EDWARDS, C.A., (red.) 1973 - Environmental pollution by pesticides. Ed. Plenum press, London, 1-542.

EHRLICH, P.R., A.H. EHRLICH et J.P. HOLDEN, 1977 - ECOSCIENCE : Population, resources, environment. Ed. W. H. Freeman and Co, San Francisco, CA, USA, 1-1051.

ESCOUFIER, Y., 1987 - Méthodes statistiques basées sur un usage intensif de l'ordinateur. Rapp. techn. 8702, Unité Biométrie, ENSAM, INRA/USTL, Montpellier, 1-16.

FISCHER, J.-C., 1972 - La paléoécologie. La Recherche, **3**, 22, 335-343.

FORMAN, R.T.T., et M. GODRON, 1986 - Landscape ecology. Ed. Wiley, New York, 1-619.

FRONTIER, S., 1982 - Stratégies d'échantillonnage en écologie. Masson, Paris, 1-494.

GARDARIN, G. et P. VALDURIEZ, 1985 - Bases de Données Relationnelles : analyse et comparaison des systèmes. Ed. Eyrolles, Paris, 1-326.

GARDARIN, G., 1989 - Bases de Données : les systèmes et leurs langages -. Ed. Eyrolles, 6ème édition, Paris, 1-265.

GARDARIN, G. et P. VALDURIEZ, 1990 - SGBD Avancés : bases de données objets, déductives, réparties -. Ed. Eyrolles, Paris, 1-255.

GILLARD, A., 1969 - On the terminology of biosphere and ecosphere. Nature, **223**, 500-501.

GODRON, M., Ph. DAGET, L. EMBERGER, G. LONG, E. LE FLOC'H, J. POISSONET, Ch. SAUVAGE et J.-P. WACQUANT, 1968 - Code pour le relevé méthodique de la végétation et du milieu. Ed. CNRS, Paris, 1-292.

GODRON, M., 1984 - Ecologie de la végétation terrestre. Ed. Masson, Paris, 1-196.

GOUZE, J.L. et P. VIGNARD, 1984 - EDORA : un système intelligent d'aide à la modélisation. AMSE Review, 2, 3, 9-26.

GUTING, R.H., 1988 - Geo-relational algebre : a model and a query language for geometric database. EDBT Conf., Venise, mars 1988.

GRANVAL, Ph., et R. ALIAGA, 1988 - Analyse critique des connaissances sur les prédateurs de lombriciens. Gibier et faune sauvage, **5**, 71-94.

HAECKEL, E., 1866 - Generelle Morphologie der Organismen. 2 tomes, Ed. G. Reiner, Berlin, Tome 1 : 1-463. Tome 2 : 1-574.

HAECKEL, E., 1884 - Histoire de la création des êtres organisés d'après les lois naturelles (traduit de l'allemand : 1866). Ed. Reinwald, Paris, 1-606.

HAWKING, S., 1989 - Une brève histoire du temps. Ed. Flammarion, 1-236.

IGL Technology, 1991 - SADT, un langage pour communiquer. Ed. Eyrolles, 1-336.

JAULT, C., 1986 - Les Bases de Données Relationnelles -ou le libre accès aux informations-. Les Editions d'Organisation, Paris, 1-109.

KENT, W., 1981 - Consequences of Assuming a Universal Relation. ACM transactions on Database Systems, Vol.6 N°4, Decembre 1981, pp. 539-556.

KORTH, H.F. et A. SILBERSHATZ, 1986 - Database System Concepts. McGraw-hill, 1986.

KOUTCHOUK, M., 1989 - SQL et DB2 : le relationnel et sa pratique. Ed. Masson, Paris, 1-226.

LAMARCK, J.-B. de, 1809 - Philosophie zoologique ou exposition des considérations... Ed. Dentu, **1**, 1-428 ; **2**, 1-475.

LAMARCK, J.-B. de, 1815 - Histoire naturelle des animaux sans vertèbres, ... (introduction). Ed. Verdière, Paris, **1**, 1-462.

LAMARCK, J.-B. de, 1822 - Histoire naturelle des animaux sans vertèbres, ... Ed. Verdière, Paris, **7**, 1-711 (cité par G. Laurent, 1987).

LAMBERT, G., 1987 - Le gaz carbonique dans l'atmosphère. La Recherche, **18**, 189, 778-787.

LAROUSSE, 1964 - Grand Larousse encyclopédique. **10**, Ed. Librairie Larousse, Paris, 1-302.

LAURINI, R., 1988 - Modélisation et manipulation des objets et des connaissances spatiales en géomatique. Lab. Informatique Appliquée, INSA Lyon, 1988.

LEBRETON, J.D. et C. MILLIER, 1982 - Modèle dynamique déterministe en biologie. Ed. Masson et Cie, Paris, 1-208.

LEGAY, J.-M., 1986 - Quelques réflexions à propos de l'écologie : défense de l'indisciplinarité. Acta oecol., Oecol. gen., **7**, 4, 391-398.

LEMAIRE, F.C., 1972 - Côuts et rentabilités de l'informatique. Dunod Economie.

LE MOIGNE J-L., 1984 - La théorie du système général, théorie de modélisation. PUF

LE MOIGNE J-L., 1990 - La modélisation des systèmes complexes. Dunod, AFCET Systèmes, 1-178.

LEONARD, M., 1988 - Structures des bases de données. Ed.Dunod, Paris, 1-285.

LONG, G., 1974 - Diagnostic phytoécologique et aménagement du territoire. I. Principes généraux et méthodes. Ed. Masson et Cie, Paris, 1-252.

MANOLA, F., et al., 1987 - Geographic information processing in the PROBE database system. Proc 8th Int. Symp. on Automation in the Cartography, Baltimore, 1987.

MARGALEF, R., 1957 - La teoria de la informacion en ecologia. Mem. roy. acad. Barcelona, **32**, 13, 373-449.

MARGALEF, R., 1968 - Perspectives in ecological theory. Univ. Chicago press. public., 1-111.

MAREE, C. et G. LEDANT, 1989 - SQL initiation, programmation et maitrise. Ed. Armand Colin, Paris, 1-264.

MARTIN, J.L., 1988 - Variation géographique, adaptation et spécialisation : l'exemple de Parus caerulens (Aves). Thèse de doct., USTL, Montpellier, 29/11/88, 1-160.

MEINADIER, J.P., 1990 - Intégration des systèmes. Cycle de formation de Chef de projet en ingénierie informatique. INFI année universitaire 1990-1991, 1-85.

MINGASSON, M., 1991 - Systèmes d'information. Cycle de formation de Chef de projet en ingénierie informatique. INFI année universitaire 1990-1991, 1-96.

MIRANDA, S. et BUSTA J-M., 1987 - Introduction aux Bases de Données. Ed. Eyrolles, 2ème édition, Paris, 1-227.

MITCHELL, J.C., 1983a - Inference Rules for Functional and Inclusion Dependencies. In Proceedings, 2nd ACM SIGACT-SIGMOD Symp. on Principles of Database Systems, Atlanta, 1983, pp. 58-69.

MITCHELL, J.C., 1983b - The Implication Problem for Functional and Inclusion Dépendencies. Information and control, N° 56, pp. 154-173.

MORIN, E., 1985 - La méthode. Ed. Seuil, Paris. 1. La vie de la vie, 1-399. 2. La nature de la nature, 1-470.

ORENSTEIN, J., 1988 - Spatial query processing in an object-oriented database system. Proc. ACM SIGMOD Conf., 1988.

OZIMA, M., et B. MARTY, 1983 - L'origine de l'atmosphère. La Recherche, **14**, 141, 176-184.

PAVE, A., 1987 - Systèmes experts : application à la modélisation en biologie. Matapli, **11**, 5-19.

PIAGET, J., 1968 - Le structuralisme. Que sais-je ? PUF, 1-125.

PICHAT, E. et A. ROCHE, 1989 - Extension de la normalisation aux dépendances d'inclusion. In Actes du Congrès INFORSID 89, Nancy, 1989, pp. 57-77.

PICHAT, E. et R. BODIN, 1990 - Ingénierie des données : Systèmes d'information Modèles et bases de données. Ed. Masson, Paris, 1-428.

PRACONTAL, M. de, 1987 - L'imposture scientifique en dix leçons. Ed. La Découverte, Paris, 1-257.

PRIGOGINE, I., 1972 - La thermodynamique de la vie. La Recherche, **24**, 547-562.

RAMADE, F., 1984 - Eléments d'écologie, écologie fondamentale. Ed. McGraw Hill, Paris, 1-397.

REEVES, H., 1990 - Malicorne, réflexions d'un observateur de la nature. Ed. du Seuil, Paris, 1-215.

REGNER, A., 1974 - La crise du langage scientifique. Ed. Anthropos, Paris, 1-399.

ROCHE, A., 1990 - Définition et réalisation d'un logiciel d'aide à la conception de systèmes d'information. Thèse de Doctorat, Université de Montpellier II, 1-134.

ROUX, C., et M. ROUX, 1967 - A propos de quelques méthodes de classification en phytosociologie. Rev. stat. appl., **14**, 2, 59-72.

ROSNAY de, J., 1975 - Le macroscope, vers une vision global. Ed. du Seuil, Paris, 1-346.

SAUVAGE, Ch., 1965 - Taxinomie ou taxonomie. Bull. soc. bot. Fr., **112**, 3/4, 180-182.

SCHALLER, J., 1987 - The geographic information system ARC/INFO. Euro Carto VI, Brno, 1987.

SCHERRER, B., 1984 - Biostatistique. Morin (ed), Montréal, 1-850.

SCHOLL, M., A. VOISARD, 1989 - Modeling of thematic maps : an application to geographic databases. 5e Journées Bases de Données Avancées, Genève, septembre 1989.

SMITH, T., 1987 - Requirements and principles for the implementation and construction of large-scale geographic information system. Geographical Information System, vol. 1, N° 1, 1987.

SOMMERVILLE, I., 1988 - Le génie logiciel et ses applications. Ed. InterEditions, Paris, 1-330.

SOTO, P. et M.B. BOUCHE, 1989 sous presse - ECORDRE : Une Base de Données Relationnelles pour toutes données écologiques, y compris les actes humains. *In* C.R. Réunion de travail internationale, Etudes pluridisciplinaires et systèmes d'information environnementale, Grenoble (France), 30-31 mai 1989.

SOTO, P., 1990 - ECORDRE : une base de données relationnelle pour toutes données agro-ecologiques. Lettre d'information trimestrielle CNUSC, ISSN 0757-4371, N°35, 19-26.

SOTO, P.et A. ALIAGA, 1990 sous presse - ECORDRE : a relational database for all the agro-ecological data. ISEE 4, Avignon, 11-15 juin 1990.

STONEBRAKER, M.et L. ROWE, 1986 - The design of POSTGRES. Proc. ACM SIGMOD Conference, Washington D.C., juin 1986.

STONEBRAKER, M.et G. KEMNITZ, 1991 - The POSTGRES next-generation database management system. Communication of the ACM, vol. 34, N° 10, october 1991.

TANSLEY, A.G., 1935 - The use and abuse of vegetational concepts and terms. Ecology, **16**, 3, 284-307.

TARDIEU, H. et A. ROCHFELD, R. COLLETI, 1983 - La Méthode MERISE : Principes et Outils, Editions d'Organisation, Paris.

TARDIEU, H. et D. NANCI, D. PASCOT, 1984 - Conception d'un système d'information. Editions d'Organisation, Paris.

ULLMAN, J.D., 1988 - Principles of Database and Knowledge-Base Systems. Volume I, Computer Science Press, Rockville, Maryland.

VARET, J., 1990 - Institut Français de l'Environnement : connaître pour choisir. Etude préliminaire. Paris, 1-132.

WOLF, A., 1989 - The DASDBS GEO-kernel, Concepts, Experiences, and the Second Step. LCNS, N° 409, Design and Implementation of large Spatial Databases, 1989.

YANG, C.-C., 1986 - Relational Databases. Prentice-Hall, Englewood Cliffs, New Jersey.

ANNEXES

SECRÉTARIAT D'ÉTAT
AUPRÈS DU PREMIER MINISTRE
CHARGÉ DE L'ENVIRONNEMENT
ET DE LA PRÉVENTION
DES RISQUES TECHNOLOGIQUES
ET NATURELS MAJEURS

Cartes p 70

INSTITUT FRANÇAIS DE L'ENVIRONNEMENT

Connaître pour choisir

Etude préliminaire

Avril 1990

Jacques VARET

Ce rapport a été établi avec l'appui
du Bureau de Recherches Géologiques et Minières

TABLEAU 3

INSTITUTIONS TECHNIQUES EN CHARGE DE LA SURVEILLANCE DE L'ENVIRONNEMENT DANS UN SCHEMA ORGANISATIONNEL OPTIMISE

NIVEAU	INSTITUTION EN CHARGE	PARTENAIRES THEMATIQUES (listes non exhaustives)
INTERNATIONAL ET GLOBAL	PNUE : PMRC, IGBP, ... ONU : IDNDR,...	
EUROPEEN	Agence Européenne de l'Environnement	Centres thématiques européens (15 répartis dans différents pays)
NATIONAL	Institut Français de l'Environnement	Opérateurs thématiques Français : (à labéliser) AQA, CEMAGREF, ANRED, IFREMER, BRGM, MUSEUM, etc Ministères (santé, industrie, agriculture, etc)
"FLUVIAL"	Commissions internationales de bassins "Agence de Bassins"	
REGIONAL	Institut (ou Observatoire) Régional de l'Environnement	Associations thématiques DRAE, DRIR, DDAS, DDA, DDE
LOCAL	Associations locales ou municipales	Associations de gestion des réseaux de mesure de la pollution atmosphérique (24) Associations écologiques (correspondants ZNIEFF)

.../

TABLEAU 6 : OBSERVATION ET INFORMATION :

INSTITUTS NATIONAUX ET INSTITUTS REGIONAUX DE L'ENVIRONNEMENT DANS QUELQUES PAYS MEMBRES DE LA COMMUNAUTE

PAYS	POPULATION (M Hab.)	INSTITUT NATIONAL DE L'ENVIRONNEMENT	NOMBRE DE PERSONNES	BUDGET P :propre I :Intervention	INSTITUTS REGIONAUX	CANDIDATURES EUROPEENNES
RFA	62	U.B.A. (Berlin)	524 p	P : 224 MF(1) I : 272 MF	dans chaque länder	Berlin (AEE)
		B.N.L. (Bonn)	100 p	P : 31 MF I : 150 MF		
GB	56	NERC : I.T.E. NEPA en projet	240 p			
ITALIE	56	SIMA en création (Loi de 1988)		(P+I) 322 MF	dans chaque province	
FRANCE	55	En projet (ce rapport)			à créer	Strasbourg (AEE)
ESPAGNE	38				dans chaque province	Madrid (AEE)
PAYS BAS	14	IRVM (1500 p : Santé + environnement)	300 p(2)	P : 105 MF(3) I : 150 MF		IRVM (AEE) pollution atmosphérique
DANEMARK	5	NAEP	360 p	P : 100 MF I : 72 MF		NAEP)) (AEE)
		NERI	250 p	P : 150 MF		NERI)

(1) Hors BMFT (recherche environnement : 725 MF).
où → (2) Sans l'épidémiologie, l'écotoxicologie et l'administration.
(3) Hors frais de personnels.

TABLEAU 13

INDICATEURS DE L'ENVIRONNEMENT (D'APRES L'OCDE, 1989)

1) Conditions et tendances

- de pression sur l'environnement

  ex. :
  - industrie
  - agriculture
  - exploitation de ressouces
  - émission de polluants

- de qualité de l'environnement

  ex. :
  - concentration de polluants dans l'air
  - pollution des eaux
    des sols
  - bruit des transports et des équipements ménagers
  - perturbation des biotopes
  - impact sur la santé

- de réponse de la part des :
  - gouvernements
  - entreprises
  - institutions
  - individus

  ex. :
  - dépenses et investissements
  - opinions

2) Politiques

- indicateurs sectoriels :

  - transports : stock de machines, état des infrastructures, efficacité, etc.

3) Economiques

- comptabilité économique :
  - stocks) des
  - flux) ressources
  - valeur) naturelles

S.R.E.T.I.E.

Tableau 15

B U D G E T 9 0
PROGRAMMES
RECHERCHE ENVIRONNEMENT

1.	ECLAT (CLIMAT ET ATMOSPHERE)	6,6 MF
2.	DECHETS TOXIQUES	3,3 MF
3.	TECHNOLOGIES PROPRES - ECO-PRODUITS TRAITEMENT - ASSAINISSEMENT	7,4 MF
4.	GENIE GENETIQUE ET ENVIRONNEMENT	1 MF
5.	GRANDS FLEUVES	1,8 MF
6.	CONNAISSANCES ET DYSFONCTIONNEMENTS DES ESPACES MARINS LITTORAUX	4,2 MF
7.	MUTATIONS DE L'ESPACE RURAL ET ENVIRONNEMENT	3 MF
8.	ACIDIFICATION DE L'ATMOSPHERE ET SES EFFETS SUR LES ECOSYSTEMES (DEFORPA ET EUROTRAC)	3,8 MF
9.	BRUIT ET VIBRATIONS	4,7 MF
10.	ENVIRONNEMENT-SANTE	6,6 MF
11.	RECHERCHE ECOLOGIQUE A LONG TERME	5,4 MF
12.	QUALITE ET CONSERVATION DES SOLS	2,8 MF
13.	INTERACTIONS SOCIETE-ENVIRONNEMENT APPROCHES SOCIO-ECONOMIQUES ET INSTITUTIONNELLES ET PROSPECTIVE	6,4 MF
14.	SOUTIEN A LA COOPERATION SCIENTIFIQUE BILATERALE ET MULTILATERALE	3 MF

TOTAL : 60 MF

TABLEAU 17

DGRT/TOEE/LAL

Janvier 1990
(document de travail)

BUDGET CIVIL DE RECHERCHE
CONSACRE A L'ENVIRONNEMENT
(estimation)

		AP+DO (MF 1990)	EFFECTIFS (équivalent temps plein)
EPST:			
	INRA	250	800
	CEMAGREF	50	150
	CNRS et Inst.	380	1000
	INSERM	75	200
	ORSTOM	95	200
EPIC:			
	CEA	50	100
	AFME	36	20
	IFREMER	80	140
	CIRAD	25	50
FRT:		50	
MINISTERES:			
	AGRICULTURE	12	90
	DOM TOM	15	
	EDUCATION/UNIV		
	ENVIRONNEMENT	51	
	EQUIPEMENT		
	INDUSTRIE		
	TRANSPORT et MER		
		1200 MF	3000

TABLEAU 21 : STRUCTURE DE GESTION DE LA PROCEDURE EUROPEENNE DE LABELISATION DES ECOPRODUITS : PROPOSITION DE LA DG XI (1989)

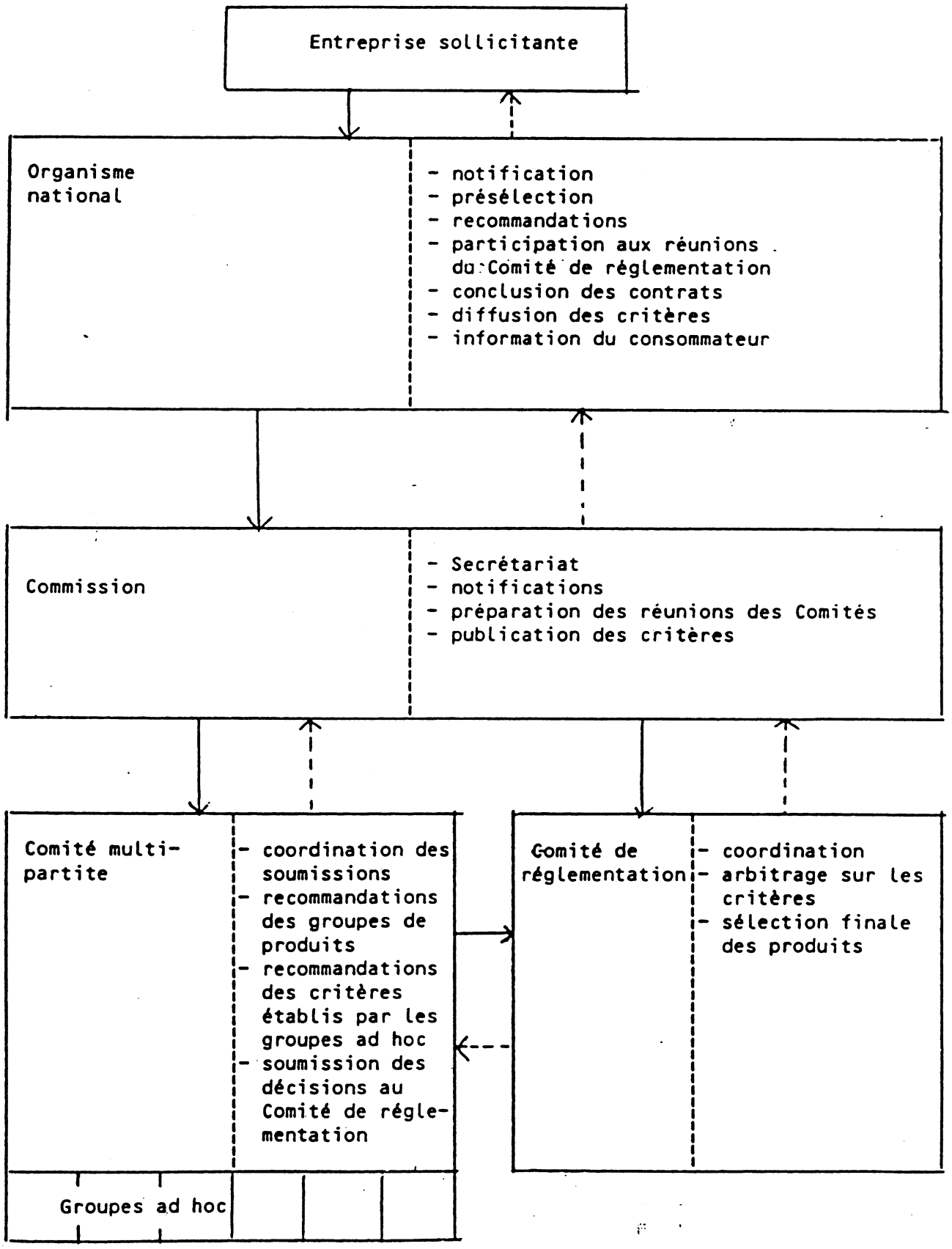


Tableau N° 22

IFE : Première tentative d'évaluation financière. (Fonction de base)

Moyens propres	Nombre d'agents
- Milieux et cycles	20 - 24
- Informatique et réseaux physiques	12 - 20
- Etudes économiques et sociales et santé	8 - 12
- Maîtrise écologique du territoire (O.S.P.N.)	20 - 24
- Banque centrale de données, statistiques, synthèse	35 - 40
- Produits : éco-produits, éco-procédés, éco-industries	10 - 12
- Affaires internationales	20
- Actions régionales et locales (10 en région, 5 au siège)	15
- Diffusion et communication	20 - 10
- Adm. Fin. Juridique - Personnel - Commercial	30 - 40
- IHEE	7 - 10
- Direction	5 - 6
TOTAL (personnels)	182 - 238
soit une valeur médiane de	210
soit pour (0,5 MF/agent)	105 MF
Stagiaires (de thèse, la 3ème année de G.E.) : 40	8 MF
Stagiaires et invités étrangers : 40	8 MF
Total général	121 MF

dont 15 % environ consacrés à la recherche (soit environ 20 MF).

Tableau 22 (suite)

Pour remplir ses missions, l'institut devra recourir largement à des partenaires extérieurs. Les moyens d'incitation suivants seront nécessaires :

Moyens d'incitation :

- Réseaux et systèmes d'informations géographiques	100 MF(42)
- Information, documentation, communication	50 MF
- Instituts régionaux (23 x 2 MF)	46 MF
- IHEE	10 MF
- Recherche - Développement	50 MF
	256 MF

Soit au total, un budget de 377 MF, en provenance :

- du budget du SEE (ressources nouvelles)	237 MF
- du budget du SEE (43) (redéploiements)	70 MF
- du budget du MRT (R et D interne et externe)	70 MF

(42) Chiffre évalué sur la base d'une prise en charge par l'IFE des réseaux actuellement financés par le SEE, et de ceux qui font défaut ; n'inclu pas les réseaux financés par les autres ministères (industrie, équipement, agriculture, santé) dont le coût n'a pu être évalué.

(43) Système d'information : 40 MF ; recherche et études économiques : 24 MF ; personnel : 6 MF (12 personnes)

Tableau N° 23

IFE : Première tentative d'évaluation financière. (Fonction de base + option recherche)

Moyens propres	Nombre d'agents
- Milieux, cycles et modélisation	40 - 50
- Informatique, réseaux physiques et normalisation	25 - 30
- Etudes économiques et sociales et santé	15 - 20
- Maîtrise écologique du territoire (O.S.P.N.)	40 - 50
- Banque centrale de données, statistiques, synthèse	35 - 40
- Produits : éco-produits, éco-procédés, éco-industries	15
- Affaires internationales	20
- Centres permanents d'observation des espaces et des milieux (CPOM)	35 - 45
- Actions régionales et locales (10 en région, 5 au siège)	15
- Diffusion et communication	10
- Adm. Fin. Juridique - Personnel - Commercial	40 - 50
- IHEE	7 - 10
- Direction	5 - 6
TOTAL (personnels)	300 - 360
soit une valeur médiane de	330
soit pour (0,5 MF/agent)	165 MF
Stagiaires (de thèse, la 3ème année de G.E.) : 40	12 MF
Stagiaires et invités étrangers : 40	10 MF
Total général	187 MF

dont 40 % environ consacrés à la recherche.

TABLEAU N° 24

Proposition budgétaire pour l'Institut Français de l'Environnement dans les 4 hypothèses étudiées = répartition des ressources : personnel, moyens propres et moyens d'incitation

	Moyens propres			Moyens pour l'incitation (conventions, sous-traitance, subventions, etc)						TOTAL MF
	% S.I.	% Recher.	% Techno.	Observation réseaux SIG	Information communication	Instituts régionaux	IHEE	Recherche développement	Technologie Ind./Aménage.	SEE MRT MIAT, etc
	TOTAL en MF									TOTAL
Fonction de base (1)	70 %	15 %	15 %							307 70 0
	121			100	50	46	10	50		377
Option 1 (+ Recherche) (2)	55 %	40 %	15 %							307 232 0
	187			100	50	92	10	100		539
Option 2 (+ Technologie) (3)	60 %	15 %	25 %							307 70 165
	136			100	50	46	10	50	150	542
Option 1+2 (4)	202			100	50	92	10	100	150	307 232 165
										704

Tableau N° 25

Proposition budgétaire pour l'Institut Français de l'Environnement
dans les 4 hypothèses étudées : origine des ressources

			Ministère de l'Environnement	Ministère de la Recherche	Ministère de l'Industrie	TOTAL
1	Fonction de base - "Noyau dur"	Ressources nouvelles	237	70	0	307
		Redéploiement	70		0	70
2	Option 1 + Fonction de base (Recherche)	Ressources nouvelles	237	182	0	419
		Redéploiement	70	50	0	120
3	Option 2 + Fonction de base (Technologie Industrie)	Ressources nouvelles	237	70	115	422
		Redéploiement	70		50	120
4	Fonction de base + Options 1 + 2	Ressources nouvelles	237	182	115	484
		Redéploiement	70	50	50	120

TABLEAU N° 26
Positionnement de l'IFE parmi les établissements publics français

STATUT	ORGANISMES	PERSONNEL	BUDGET TOTAL *	dont BCRD *	Autres Minis.	TAXE PF	Ress. P
EPST	CNRS	26 090	10,40	10,30	-	-	0,10
EPIC	CEA	13 100	9,50	3,00	3,70	-	2,80
EPIC	CNES	2 317	7,00	5,57	-	-	1,40
EPST	INRA	8 256	2,90	2,45	-	-	0,45
EPST	INSERM	4 485	1,82	1,82	-	-	-
EPST	IFP	1 866	1,33	10	-	0,90	0,39
EPIC	CIRAD	1 900	1,00	0,55	-	-	0,45
EPIC	IFREMER	1 191	0,93	0,78	-	-	0,14
EPIC	BRGM	1 690	0,87	0,10	0,2	-	0,57
EPST	ORSTOM	1 566	0,75	0,75	-	-	-
EPST	CEMAGREF	586	0,24	0,14	0,06	-	0,45
EPST	INRIA	547	0,3	0,3	-	-	-
EPST	INRETS	378	0,15	0,15	-	-	-
EPIC	AFME	380	0,52	0,15	0,37	-	-
EPIC	IFE minimum	210	0,38	0,07	0,31	-	-
	maximum	360	0,70	0,23	0,47	-	-
EPST	INED	158	0,19	0,19	-	-	-
EPIC	ANRED	100	0,12	0	0,06	0,05	-
EPIC	AQUA	20	0,03	0	0,02	-	-

Tableau N° 27

Positionnement de l'IFE par rapport à des instituts équivalents
de quelques pays européens

Pays	Institut	Nombre de personnes	Institus Nat. d'environnement : budget propre	Budget d'incitation géré par ces inst.	Autres budgets publics d'incitation Environnement (R & D)
RFA	UBA (Berlin)	524	224	272	
	BNL (Bonn)	100	31	150	
	Total		255 MF	422 MF	725 MF (BMFT)
ITALIE		n.d.	322 MF (propre + incitation)	n.d.	n.d.
PAYS-BAS	IRVM (part Env)	500	105 MF (hors personnel)	150 MF	250 MF (STIPT)
	NOVEM			75 MF	
	Total			225 MF	
DANEMARK	NAEP	360	100		
	NERI		250		n.d.
	Total		350 MF	72 MF	
FRANCE	IFE 1	210	121	256	
	IFE 2	330	187	352	50 MF (SRETIE)
	IFE 3	240	136	406	
	IFE 4	360	202	502	

Rapports et notes consultés

- Prospection de l'environnement et du cadre de vie. LESOURNE, 1981. Rapport "Environnement" du IXème Plan.

- Politique de l'Environnement en Asie du Sud et de l'Est. Problèmes économiques et sociaux, n° 453, 1982, p. 1-40.

- Les statistiques de l'Environnement. Rapport Ader, 1982, 46 p.

- L'Ecologie : les défis d'une science en temps de crise. Ministère de l'industrie et de la recherche. DI CASTRI, 1983, 116 p.

- Organisation de la recherche et de la formation pour la maitrise écologique du territoire. Claude HENRY, 1984, 82 p.

- RNO : Réseau National d'Observation de la qualité du milieu marin, dix années de surveillance. IFREMER, 1984, 8 p. + annexes.

- Répertoire des formations de recherche "environnement". CNRS-PIREN, 1984, 123 p.

- Politiques comparées de Recherche-innovation dans le domaine de l'Environnement. R.D. International, 1985, 126 p.

- Les politiques publiques de l'Environnement - Chine, France, Vénézuéla, Méditerranée, OCDE, ONU, Revue Française de l'Administration Publique, n° 35, 1985, p. 393-455.

- Environnement : recherche et formation dans l'enseignement supérieur. Yves PIETRASANTA, 1986, 19 p, 9 annexes.

- Les comptes du Patrimoine naturel. Les collections de l'INSEE, C 137-138, 1986, 552 p.

- Environnement et Politique en RFA. Problèmes économiques et sociaux, n° 533, 1986, p. 1-36.

- L'environnement et les ressources au XXIème siècle. Futurible N° 115, J. THEYS, 1987.

- Qualité des eaux douces et superficielles. Procédure commune d'échange d'informations (1982-1986). CCE, rapport EUR 12043 FR, 1987, 138 p.

- Le programme des Nations Unies pour l'Environnement. 1987, 36 p.

- Le rôle des technologies de mesure, de détection et de protection, dans la gestion des ressources naturelles renouvelables . N. LELEU, FAST n° 186, 1987, 120 p.

- IRENE : an integrated modeling and information system for natural resource management. L.C. BRAAT, A.J. GILBERT et J. KREYSA, FAST n° 177, 1987, 47 p.

- Politique Européenne de l'Environnement. Communautés Européennes, comité economique et social, Bruxelles, 1987, 52 p.

-VL'Etat de l'Environnement dans la communauté européenne. CCE, DG XI, EUR 10633, 1987, 379 p.

- Espace et environnement. Groupe de réflexion à long terme. CNIS. HENRY et THEYS, 1988, 52 p.

- Le financement de la Recherche-Développement Environnement en France. Cabinet P. LOISILLIER, 1988, 65 p. et annexes.

- Le programme FAST II (1984-1987). Résultats et recommandations. Volume 6 : l'usage des ressources agricoles et forestières en Europe : les modèles de développements dominants en question. CCE - DGXII, 1988, 187 p. + 2 annexes.

- Appréciation du potentiel technique de l'IRCHA. Orientations possibles. J.C. BALACEANU, 1988, 21 p.

- Surveillance de l'environnement. OCDE : monographie sur l'environnement, n° 16, 39 p., 1988.

- Guide de la valorisation économique des ressources locales. Fédération des parcs naturels de France, Ed. Syros, 1988.

- IFREMER. Rapport annuel 1988, 24 p. + annexes (dont Environnement et Recherche océanique, 16 p.).

- Les structures et l'organisation de la Recherche au Ministère de l'environnement. Evaluation et propositions, Ministère de l'Environnement, SRETIE, J. THEYS, 25 p. + 3 annexes.

- Environnement Canada. Rapport annuel 1987-1988. 20 p. + annexes.

- Rapport d'étape du groupe de travail sur l'environnement et l'économie, oct. 1988, 12 p.

- Etat de l'environnement. Edition 1988. Secrétariat d'Etat auprès du Premier ministre chargé de l'Environnement, 1989, 348 p.

- Données économiques de l'environnement. Edition 1989. Secrétariat d'Etat auprès du Premier Ministre chargé de l'Environnement, 1989, 149 p.

- Cartographie de la pollution industrielle. DEPPR, 1989, 63 p.

- Rôle "Service public" de l'IRCHA dans le cadre d'un "Institut de l'environnement". R. CABRIDENC, 1989, 22 p.

- Mission recherche-développement dans le domaine de la prévention des risques technologiques et naturels majeurs. Compte rendu de mission. Conclusions et recommandations. J.C. BALACEANU, 1989, 39 p.

- L'évolution du climat et de l'environnement global. La contribution française en 1989. M.R.T., 1989, 30 p.

- Reporting on the state of the environment. A review of the experience in the OECD countries. C.P. AVEROUS, 1989, 11 p.

- Il était une fois la Pollution, ou l'enfer du Nord. Apperçu global des pollutions du Nord - Pas-de-Calais. Spécial Nord Nature, 1989, 60 p.

- Risques majeurs : proposition pour la mise en oeuvre d'une politique d'information et de prévention à l'échelon du département. Rapport du groupe de travail présidé par le Préfet Jean MINGASSON, 1989, 99 p.

- Rapport sur la modernisation de l'administration territoriale de l'environnement. Ministère de l'intérieur. Inspection générale de l'Administration. J.F. LORIT et M. LE FUF 1989, 54 p. + annexes.

- L'Environnement : L'approche économique. J. Ph. BARDE, L'observateur de l'OCDE, n° 158 (1089), p. 12-15.

- L'utilisation des investissements économiques pour la protection de l'Environnement dans les pays de l'OCDE, Note OCDE, 1989

- United States EPA initiative for better environmental measures T.R. TITUS, EPA, OCDE, environment committee, déc. 89, 10 p.

- Summary conclusions of the OCDE Woskshop on environment indicators OCDE, Environment committee, déc. 89, 15 p.

- Les banques de données en environnement. Secrétariat d'Etat à l'Environnement, SRETIE, 1989, 229 p.

- Etude comparative sur les systèmes de traitements d'images (STI) associés à des systèmes d'informations géographiques (SIG). S. THIRION. Rapport BRGM 09DT007DED, 1989, 73 p.

- Répartition et chronologie de la migration prénuptiale et de la reproduction en France des oiseaux d'eau gibier. Museum National d'Histoire Naturelle, Office National de la Chasse, 1989, 86 p.

- L'évolution de l'Environnement global. Rapport au Comité des programmes scientifiques du CNES, 1989, 88 p.

- La science, la technologie, l'innovation : une politique globale, dite Rapport Farge Commissariat Général du Plan, La Documentation Française, 1989, 194 p,

- Programme Agotech : Contribution à la connaissance et à la gestion de la biosphère domestiquée pour l'agriculture de demain (1990-1992), J.C. REMY, INRA, 1989, 19 p.

- ORSTOM. Rapport d'activité 1987-1988, 1989, 64 p.

- Inventaire national des formations concernant l'environnement. SEE-DQV, 1989, n° 9 dossier de présentation, 28 p. ; et 1989, n° 4 liste des formations par diplôme.

- To choose or to loose. National environmental policy plan, LA HAGUE, 1989, 258 p.

- PIREN. Orientations et programmes pour la période 1990-1992, CNRS, 1989, 29 p.

- Groupe interministèriel sur la réorganisation des systèmes de collecte et de mise à disposition des données concernant les précipitations atmosphériques et les débits et hauteur d'eau en rivière. Ministere de l'Environnement, 45 p. + 7 Annexes 1989.

- Accord cadre pour la mise en place et le fonctionnement de la banque nationale des données hydrométriques, dite banque "Hydro", 11 p. + Annexes - Envrionment data system of Finland, Ministry of Environment, Helsinski, 1989, 14 p.

- Pour un institut national de recherche appliquée à la gestion du patrimoine naturel. J.C. LEFEUVRE, Aménagement et nature n° 66 1989, p. 25-26.

- Un plan à long terme pour l'environnement : Le Plan Vert. Premier document d'étape, note du Ministère de l'environnement - J. THEYS. 4 décembre 1989, 27 p. + Annexes.

- ANRED. Plan d'entreprise. 7 p. + annexes, 09/02/90.

- Etats-Unis. Le budget Fédéral de R et D 1990 voté par le Congrès M. AUBRY, Ambassade de France aux Etats Unis, MA/rg 837, janv. 90, 12 p.

- Efficacité économique et impératifs éthiques : environnement en copropriété C. HENRY Lab. d'Econométrie, n° 324, 1990, 31 p.

- Dossier scientifique et technique, Mission scientifique française aux Etats Unis, 1990, 19 p. + annexes.

- Energie et environnement. Colloque du 30 au 31 octobre 1989, synthèse. Ed. AFME, 1990, 40 p.

- Plan national pour l'Environnement : objectif. Ministère de l'environnement - J. THEYS. 5 février 1990, 34 p.

- Objectifs et moyens de la recherche et des systèmes d'information dans le plan national de l'environnement. SEE-SRETIE, 29 mars 1990, 17 p.