

## ENREGISTREMENT ET STOCKAGE DES DONNÉES

J.C. FAYET\*, Annie BOUVIER\*\* et Nicole CHONION\*\*\*

\* Laboratoire d'Eco-Pathologie, I.N.R.A. Theix - 63110 Beaumont

\*\* Laboratoire de Biométrie, C.N.R.Z. de Jouy-en-Josas - 78350 Jouy-en-Josas

\*\*\* F.N.G.D.S.B.

---

---

### RESUME

L'enquête écopathologique porte sur 125 exploitations situées dans 7 départements. Les données relevées sont très variées : effectifs, pathologie, analyses sanguines, événements de la reproduction, fichier socio-économique, etc. Il est évident qu'une enquête de ce type doit avoir pour but de dégager une hiérarchie de la pathologie mais aussi de fournir des pistes de travail en mettant en évidence des liaisons entre la pathologie et les différents paramètres de l'environnement. Il ne faut pas croire que ce type d'observations puisse déboucher (sauf cas exceptionnel) sur des "recettes" directement utilisables par l'agriculteur à brève échéance. Mais il faut que la masse de données recueillies soit suffisante pour que leur analyse aboutisse à des résultats pas trop éloignés du cas général et non pas confinés au cas particulier étudié.

### I NECESSITE DE L'INFORMATIQUE

A partir du moment où il est question d'une masse de données, il devient nécessaire de recourir à l'informatique pour pouvoir y accéder à tout moment dans des délais acceptables.

#### Caractéristiques des données à traiter

Les données enregistrées dans l'enquête le sont avec des fréquences variables.

Tous les 4 ans : analyse de sol

Tous les ans : fichier annuel socio-économique

Tous les semestres : analyses sanguines biochimiques et sérologiques.

Toutes les quinzaines : sont enregistrés tous les événements touchant aux effectifs, à la pathologie, à la reproduction, à l'alimentation, à la production de lait.

Toutes les décades : relevés météorologiques

Le genre de questions qu'on allait être amené à poser supposait qu'on puisse accéder à n'importe quel type de donnée conditionnellement à n'importe quel autre type. Il est évident que de telles exigences ne pouvaient être satisfaites à partir de fichiers informatiques classiques et indépendants les uns des autres.

Devant cette nécessité de prévoir un mode d'accès très diversifié et performant, donc le plus indépendant de leur organisation physique, les données ont été organisées en base\*.

\* Une base de données est un ensemble d'informations exhaustives et non redondantes connues par un logiciel qui en assure la gestion (S.G.B.D.)

## II. LA BASE DE DONNEES VETO

Le Système de Gestion de Base de Données implanté sur l'IRIS 80 est connu sous le nom SOCRATE 2-CII. Il a la particularité d'admettre plusieurs structures différentes de données : ponctuelle, linéaire, arborescente, en réseau.

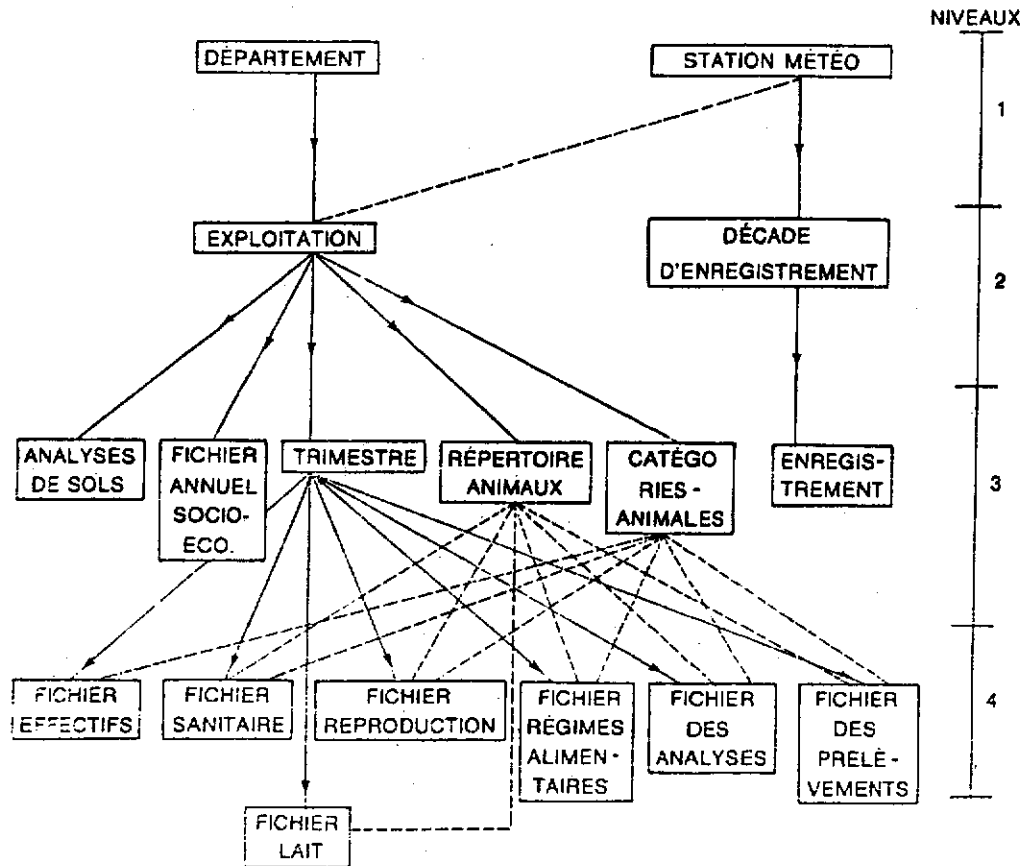
La base VETO a été la première à être mise en oeuvre par nos collègues de la Biométrie de Jouy. Sa structure peut être représentée sous forme d'un diagramme (Figure 1). La lecture de cette figure permet de se rendre compte des différents liens unissant entre eux les différentes catégories d'«objets» enregistrés.

- Ainsi une exploitation n'est accessible qu'à l'intérieur d'un département.

Elle est rattachée aussi à une station météo.

- Pour consulter un fichier-analyses d'une exploitation donnée on peut le faire soit en passant par trimestre soit par le répertoire des animaux soit par la catégorie animale.

Cette structure de la base VETO a été programmée et enregistrée par le S.G.B.D. avant l'introduction des données et elle sert de guide "navigationnel" aussi bien pour le chargement que l'extraction des informations (Figure 2).



— La structure logique de la base vétérinaire, telle qu'elle est enregistrée par SOCRATE.

Legende liaison « en sens unique » ———  
liaison « aller-retour » - - - - -

FIGURE 1

00072 Entité 12 Département  
00073 Début  
00074 NOBPT NON CODEE (12 2) (07 29 41 46 52 53 63 73 87)  
00075 Entité 50 EXPLOITATION  
00076 Début  
00077 NEXP Binaire 1 de 1 à 50  
00097 STATION-METEO refere une station-météo par CH-METEO  
00112 Entité 16 Fichier-sol  
00121 Entité 5 Fichier-animal  
00401 Entité 19 Trimestre  
00402 Début  
00403 NOTRIM CODEE (19 5) (77 NDJ..... 82 MJJ)  
00409 Entité 30 Fichier-effectif  
00422 Entité 100 Fichier-sanitaire  
00423 Début  
00425 Valeur binaire 2 de -9 à 999  
00426 QZE binaire 2 de 7721 à 8214  
00427 Code réfère un code-sani  
00428 Race réfère un code-race  
00429 Nocie réfère une catégorie d'une exploitation par cat-san  
00434 Fin/ de fichier-sanitaire /  
00437 Entité 100 Fichier-reproduction  
00453 Entité 100 Fichier-alimentation  
00470 Entité 150 Fichier-complémentaire  
00484 Entité 150 Fichier-prélèvement  
00498 Entité 150 Fichier-laitier  
00511 Fin / de Trimestre /  
00555 Entité 8 catégorie  
00556 Début  
00561 CAT-SAN chaîne 8 Tout relie-cat-san  
00562 de une exploitation par nocie trié par QZE  
00579 Fin / de catégorie /  
00581 Fin / de exploitation /  
00582 Fin / de département /

FIGURE 2 : EXTRAIT DE LA DESCRIPTION DE LA STRUCTURE DE LA BASE VETO

### III. MODE ACTUEL D'ENREGISTREMENT DES DONNEES

L'enregistrement des données dans la base est soumis à un cheminement qui peut être résumé par un organigramme représenté sur la Figure 3 que nous allons analyser brièvement.

Le stockage des données se fait selon un processus qu'on peut décomposer en 3 étapes.

ETAPE\_1 Au cours de ce premier temps les données passent de l'état brut à un état informatisé. Cela signifie en particulier que l'information qui était détaillée au niveau de l'animal va être regroupée. En effet dans la conception actuelle de l'enregistrement des données c'est l'exploitation qui est privilégiée en tant que véritable individu, et non pas l'animal. Ainsi au cours de cette étape, au point 2, on va dénombrer, à l'intérieur de l'exploitation, les animaux ayant présenté les mêmes symptômes dans une quinzaine donnée et aussi compter le nombre d'individus malades.

Ex : Description de la pathologie sur un relevé trimestriel.

N° de l'exploitation 40 N° du département 29 Trimestre Fév. Mar.  
Avri. 81

Nom de la vache	Date	Maladie	N° de code*	Catégorie
Jib...	16/4	Non délivrance	206	V Laitière (1)
Jip...	23/4	Kyste ovarien	414	V L
		Métrite	410	
Lar...	26/4	Hyperthermie	140	V L
Mir...	18/4	Métrite	410	V L
Nig...	28/4	Fièvre	140	V L
Abo...	18/4	Métrite	410	V L
Ois...	21/4	Kyste ovarien	414	V L
		Cycle irrégulier	203	
Par...	30/4	Kyste ovarien	414	Génisse (5)
Pam...	30/4	Kyste ovarien	414	Génisse
		Cycle irrégulier	203	
Pas...	24/4	Kyste ovarien	414	V L
		Cycle irrégulier	203	

\*N° de code donné au moment de l'encodage

Après codage ce relevé trimestriel va devenir :

N° de l'explo.	Catég.	Qze	N° de code	Nbre	Catég.	Qze	N° de code	Nbre
2940	1	8108	206	1	1	8108	140	2
2940	1	8108	410	3	1	8108	203	2
2940	1	8108	414	3	1	8108	999**	8
2940	5	8108	414	2	5	8108	203	1
2940	5	8108	999	2				

\*\*Le code 999 regroupe tous les animaux malades ou morts au cours de la qze.

ETAPE\_2 Au cours de cette étape les différents fichiers se retrouvent sur un disque de l'IRIS 80, comme des fichiers ordinaires. Ils ont donc les mêmes propriétés, en particulier celle de pouvoir être corrigés manuellement d'une façon simple et rapide grâce à l'éditeur de textes.

ETAPE\_3 Les fichiers vont faire partie de la base et donc en acquérir les propriétés, à savoir que toute modification y compris leur création où leur destruction, ne peut avoir lieu que par programmation dans le langage SOCRATE. Les corrections à ce niveau nécessitent donc dans le cas le plus simple la mise en oeuvre d'un programme, dans les cas plus complexes l'écriture et l'exécution d'un nouveau programme ; les deux possibilités sont toujours plus longues et délicates qu'une correction sous éditeur d'où la nécessité de pratiquer la majorité des corrections à l'étape précédente.

#### IV EVOLUTION PREVISIBLE DE L'ENREGISTREMENT DES DONNEES

Le stockage des données sous forme de base est tributaire du matériel et du logiciel informatiques disponibles. A l'heure actuelle, l'ordinateur de Jouy est un IRIS 80 et le SGBD est SOCRATE. Si, comme il est probable, l'IRIS 80 est remplacé en 1983 pour un autre matériel n'acceptant pas nécessairement SOCRATE, alors il faudra réécrire toute la description de la base en fonction du nouveau logiciel et la réimplanter sur le nouvel ordinateur. Ce travail sera probablement long mais impératif. On voit ici une des difficultés de l'organisation en base de données par rapport à un système de fichiers classiques.

Mais compte-tenu des incertitudes pesant encore sur le choix du matériel, il serait vain de développer ce point.

En revanche, il faut être attentif aux améliorations qu'on peut apporter au schéma d'ensemble.

Un des reproches majeurs qu'on peut faire est, à mon avis, non pas tellement la condensation de l'information au niveau de la quinzaine mais plutôt l'absence de saisie informatisée au niveau de l'individu animal. De la sorte il est impossible, à partir des renseignements contenus dans la base, de reconstituer la carrière d'une vache sur le plan des événements de la reproduction ou de la pathologie. Il serait pourtant important de pouvoir détenir ces informations et de les communiquer à l'éleveur. Celui-ci

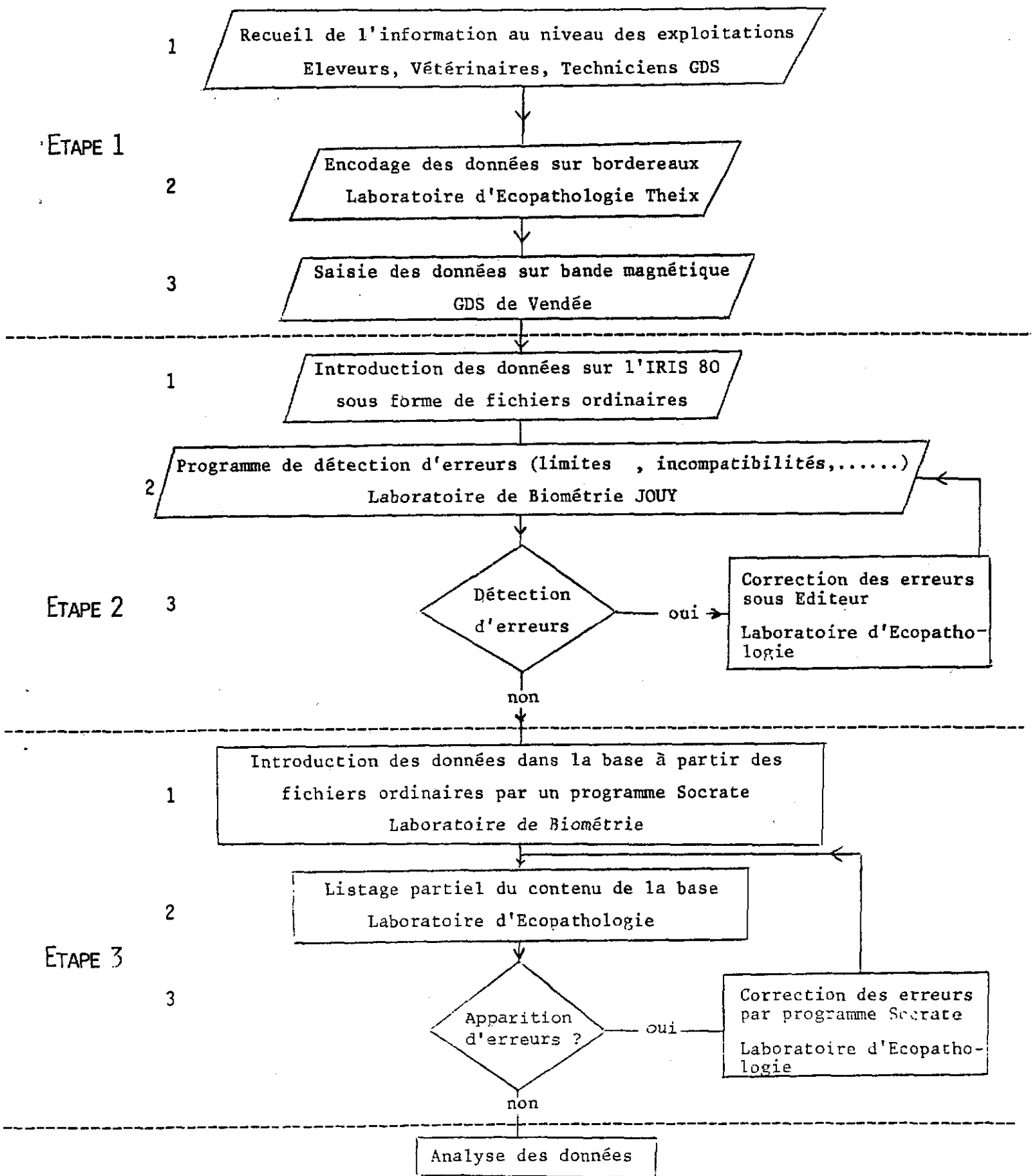


FIGURE 3 . ORGANIGRAMME DE L'ENREGISTREMENT DES DONNEES

s'apercevrait par exemple que telle vache a eu besoin de 8 inséminations pour 2 vélages ; il pourrait éventuellement en tirer des conclusions.

Or depuis Février 1982, il existe à Theix un ordinateur MINI 6 susceptible de porter remède à cet état de chose. Ainsi on peut envisager que les informations de base soient entrées à l'état brut (donc au niveau individuel) sur le MINI 6. Un programme passant sur ces fichiers individuels générerait un fichier susceptible d'entrer dans la base sur le gros ordinateur. Mais le fichier de base resterait sur le MINI 6 et à partir de là un programme de tri permettrait de reconstituer rapidement les carrières individuelles de toutes les vaches. Ainsi l'étape 1 serait complètement bouleversée. Les points 2 et 3 auraient lieu à Theix. Les fichiers de base étant l'image informatisée des données brutes, les vérifications en seraient facilitées.

#### V CONCLUSIONS

Un des problèmes majeurs dans l'enregistrement et le stockage des données est la conservation sans dénaturation de l'information. Pour atteindre cet objectif, on a probablement intérêt à avoir le moins de manipulations possible au niveau de l'encodage et de la saisie. Pour cela, la constitution de fichiers de base très proches des données du terrain faciliterait les vérifications. L'exécution de programmes sur ces fichiers fournirait des données en accord avec le choix d'agrégation retenu pour alimenter la base. L'information individuelle subsisterait sur un support informatique et deviendrait plus facilement accessible.