



HAL
open science

Modélisation du système d'informations d'un programme de recherches par la méthode MERISE : l'exemple d'une enquête écopathologique sur les vaches laitières

Laurent Perochon, Françoise Lescourret

► To cite this version:

Laurent Perochon, Françoise Lescourret. Modélisation du système d'informations d'un programme de recherches par la méthode MERISE : l'exemple d'une enquête écopathologique sur les vaches laitières. *Veterinary Research*, 1994, 25, pp.115-119. hal-02701922

HAL Id: hal-02701922

<https://hal.inrae.fr/hal-02701922v1>

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les simulations utilisées dans cette étude ont démontré l'impact des erreurs de mesure. Ces erreurs peuvent expliquer, au moins en partie, que l'application prospective des recommandations issues des résultats d'études dans des programmes ne soient pas toujours suivie des résultats escomptés. On peut aussi conclure qu'il n'est pas facile de prendre en considération les erreurs de mesure lors du traitement de données. Il faut pourtant s'assurer que les laboratoires et les enquêteurs évaluent la précision de leurs mesures et que les épidémiologistes la prennent en considération lors du traitement de données pour produire des informations médicales.

Vet Res (1994) 25, 115-119
© Elsevier/INRA

RÉFÉRENCES

- Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S (1991) *Measurement errors in surveys*. John Wiley and sons, New York
- Freund JE, Walpole RE (1980) *Mathematical statistics*. Prentice-Hall, Englewood Cliffs, New Jersey, 3rd ed
- Fuller WA (1987) *Measurement error models*. John Wiley and sons, New York
- Mallows CL (1987) *Design, data and analysis by some friends of Cuthbert Daniel*. John Wiley and sons, New York
- Martin SW, Meek AH, Willeberg P (1987) *Veterinary epidemiology: principles and methods*. Iowa State University Press, Ames, Iowa

Modélisation du système d'informations d'un programme de recherches par la méthode MERISE : l'exemple d'une enquête écopathologique sur les vaches laitières

L Pérochon *, F Lescourret

INRA de Theix, laboratoire d'écopathologie, 63122 Saint-Genès-Champanelle, France

Résumé — Nous présentons la méthode MERISE de conception d'un système d'informations, en détaillant plus précisément la modélisation des données. Nous utilisons comme exemple un programme de recherches exploratoires sur la pathologie du péri-partum en élevage bovin laitier intensif. Le système d'informations est découpé en sous-systèmes thématiques appelés «domaines», que l'on représente par un schéma conceptuel global. Pour chaque domaine, à partir de règles de modélisation précises, on détermine ensuite le modèle conceptuel détaillé. Ce découpage facilite la validation des modèles par les différents partenaires du projet et permet de réutiliser les domaines dans d'autres programmes de recherche.

écopathologie / modélisation conceptuelle / vache

* Correspondance et tirés à part.

Summary — Modelling the information system of a research program using the MERISE method: the example of an ecopathological survey on dairy cows. *We present the MERISE method for the design of an information system, detailing data modelling. An exploratory research program on peri-partum pathology in intensive dairy herds is used as an example. The information system is divided into thematic subsystems called 'fields', which are represented by a global conceptual scheme. For each field, a detailed conceptual model is then determined through precise modelling guidelines. This division facilitates the validation of models by the partners of the project, and makes it possible to re-use the fields in other research programs.*

ecopathology / conceptual modelling / cow

INTRODUCTION

Les programmes de recherches actuels en épidémiologie, du fait de leur aspect multifactoriel, nécessitent un volume et une variété d'informations importants. Ils réunissent de nombreux partenaires spécialisés dans divers domaines. Pour résoudre les problèmes de communication et de définitions inhérents, il est nécessaire d'adopter une démarche rigoureuse. Scherrer (1984) propose la démarche suivante : définir la problématique ; établir un bilan des connaissances par une revue de la littérature ; concevoir un prémodèle conceptuel d'explication ; fixer les objectifs particuliers permettant de vérifier les hypothèses du prémodèle ; choisir une approche (expérimentale, descriptive ou quasi expérimentale) ; définir le système d'informations nécessaire aux objectifs.

Nous nous intéressons ici à la définition du système d'informations. Celle-ci nécessite l'utilisation d'une méthode d'analyse de système. Le choix peut être fait dans 3 grandes catégories de méthodes (Bouzeghoub, 1993). Les méthodes d'analyse et de décomposition hiérarchique (SADT, JACKSON...), apparues dans les années 1970, décomposent hiérarchiquement un problème en sous-problèmes pour mieux appréhender la complexité. Les méthodes systémiques (MERISE, AXIAL...), apparues dans les années 1980, décrivent les fonctions et la structure du système d'informations. Elles modélisent concurremment les données et les traitements. Les méthodes objets

(HOOD, OOSE, OMT...) ont fait leur apparition dans les années 1990. Elles lient la structure des objets composant le système d'informations à leur dynamique. Nous avons choisi la méthode MERISE (Tardieu *et al*, 1983) par intérêt pour son modèle des données et par sa large diffusion en France, garante de sa pérennité, des possibilités d'aide qu'a son utilisateur, et de son adaptabilité à des problèmes divers.

Nous présentons les concepts de la méthode MERISE, puis nous les appliquons à un programme de recherches exploratoire, sur la pathologie du péri-partum en élevage bovin laitier intensif : l'enquête écopathologique Bretagne (Faye *et al*, 1989). Les objectifs de ce programme sont de définir les facteurs de risque des maladies du péri-vêlage à 2 échelles de perception : l'animal et l'élevage.

MÉTHODES

MERISE comprend 3 grandes étapes résumant la vie du système : i) la phase de conception du système qui est composée d'un schéma directeur, d'une étude préalable et d'une étude détaillée ; ii) la réalisation ; et iii) la maintenance. Seule la conception du modèle des données est décrite ici.

Le schéma directeur fait le lien entre les objectifs fixés et le besoin en données. Pour cela, l'univers du discours est décomposé en domaines. Chacun d'eux représente un thème de la problématique. Les domaines doivent avoir entre eux le moins de relations possibles. A ce niveau de précision, les domaines sont décrits de manière conceptuelle. Cette étape se termine

obligatoirement par la validation par les partenaires.

L'étude préalable doit déterminer si le projet est réalisable. Pour chaque domaine, on extrait un sous-ensemble représentatif, sélectionné en fonction des volumes de données en jeu, de la fréquence d'accès aux données et de la priorité des objectifs. Le niveau conceptuel est affiné. A ce stade, le problème de la représentation des données est pris en compte : c'est le niveau logique. Généralement on arrive à un modèle hiérarchique, réseau ou relationnel. Cette description permet d'évaluer les ressources humaines, matérielles et financières nécessaires. L'étude préalable est surtout utile pour des systèmes d'informations très importants : elle permet de modifier, voire d'arrêter un projet avant d'avoir trop investi. Son niveau de finesse permet également, si besoin est, le découpage du projet d'étude d'un domaine en sous-projets plus facilement accessibles. La validation des partenaires termine cette phase.

L'étude détaillée élargit le sous-ensemble représentatif au reste des données du projet ou du sous-projet. On retrouve les niveaux conceptuel et logique. Les moyens nécessaires sont clairement déterminés et l'ordonnement dans le temps est défini. Cette étape est également validée.

Le formalisme graphique utilisé au niveau conceptuel que nous aborderons ici repose sur le modèle individuel, équivalent européen du modèle entité-relation (Chen, 1976) et inspiré du modèle Ansi-Sparc d'après Rochfeld et Morejon (1988). Il a subi de nombreuses extensions allant jusqu'aux concepts de généralisation et d'héritage propres aux méthodes objets. Seule sa forme originelle sera décrite ici. Le système d'informations est décomposé en individus, en relations définissant les liens entre individus et en propriétés associées à un individu ou à une relation. Les propriétés identiques sont réunies en types. Puis des types d'individus sont définis, qui réunissent les individus ayant les mêmes types de propriétés. Enfin, on définit des types de relations en réunissant les relations mettant en jeu les mêmes individus, possédant les mêmes propriétés et ayant le même sens (fig 1).

RÉSULTATS

Le schéma directeur a commencé par plusieurs réunions, qui ont permis d'établir un

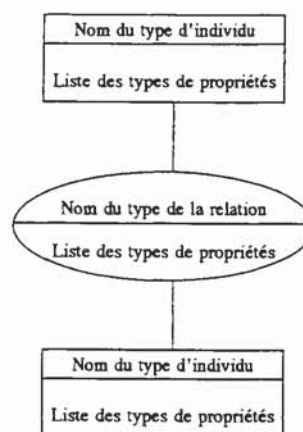


Fig 1. Formalisme graphique de l'individu et de la relation.

schéma conceptuel grossier représentant les domaines (fig 2). Nous avons remarqué que le facteur temporel ne ressortait pas. En revanche, les 2 échelles de perception étaient très nettes. Ce schéma a été validé par les partenaires avant de passer à l'étape suivante. Il faut bien remarquer que pour un même système d'informations, plusieurs découpages étaient possibles. L'important était d'arriver à un consensus.

L'étude détaillée a ensuite directement été appliquée. En effet, notre programme n'était pas suffisamment important pour nécessiter une étude préalable. Comme exemple, nous avons pris le domaine «alimentation» (fig 3). Nous avons retrouvé bien sûr les 2 échelles de perception, mais cette fois le temps est apparu. Il était représenté de 2 manières différentes : le temps physiologique avec l'individu «Numéro de lactation», et le temps chronologique avec les individus «Date», «Intervalle de temps» et «Année». Pour la période autour du vêlage (axe principal du programme de recherches), l'information était connue à l'échelle de l'animal. En dehors de cette période, l'information concernait l'échelle de l'exploitation. Elle était bornée par des dates (individu «Intervalle de temps» et

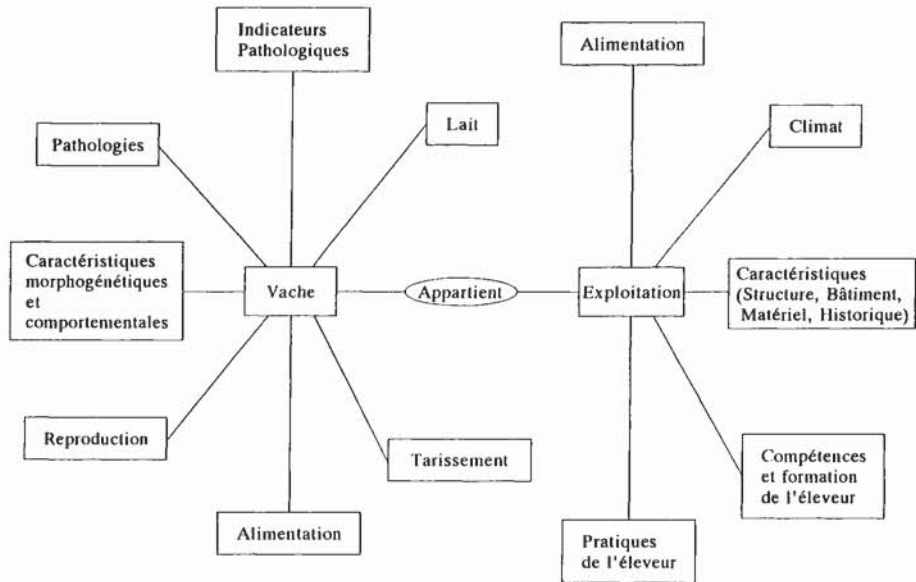


Fig 2. Modèle conceptuel global.

«Date») ou annuelle (individu «Année»). Prenons comme exemple la distribution d'ensilage. Elle se lit : pour une exploita-

tion, une catégorie d'animal et un intervalle de temps donnés, il y a eu distribution d'un ensilage, dont on a fait un prélèvement à

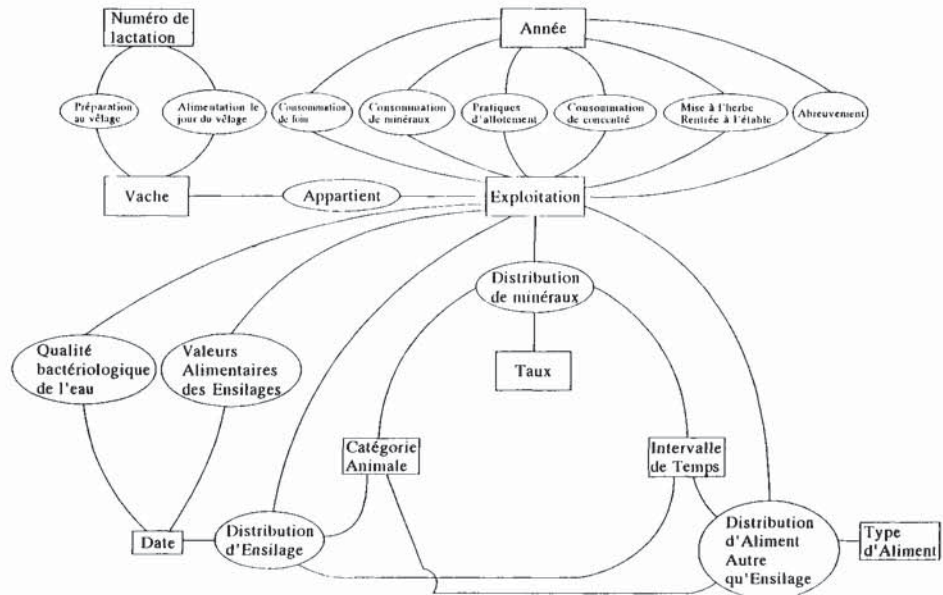


Fig 3. Modèle conceptuel du domaine «alimentation».

une certaine date. Cette relation ne donnait pas la valeur alimentaire de l'ensilage. Cette valeur existait pourtant, pour une exploitation et une date de prélèvement données comme ensemble de propriétés de la relation «Valeurs alimentaires des ensilages». Pourquoi une telle séparation ? Parce que les partenaires ont considéré que cette valeur dépendait largement de l'exploitation et du temps chronologique. *A contrario*, la distribution d'aliment autre que les ensilages, par ailleurs similaire à la relation «Distribution d'ensilage», comportait cette valeur alimentaire (individu type d'aliment), parce qu'on a fait l'hypothèse qu'elle était indépendante de l'exploitation et du temps chronologique. Ces exemples nous ont montré combien était longue la mise au point de ces modèles, tant les aspects à prendre en compte étaient nombreux, et combien les problèmes de recueil de données et de coût rentraient en jeu dans ces choix. Cette étape a également été validée par les partenaires.

CONCLUSION

La démarche MERISE est intéressante à plusieurs titres. Tout d'abord le découpage en domaines simplifie le système d'informations à appréhender. La stabilité des domaines offre une bonne base pour déterminer les niveaux ultérieurs et permet qu'on réutilise leurs modèles dans des études ayant des problématiques similaires (Lescourret *et al*, 1992, 1993). D'autre part, la validation des étapes par tous les partenaires est un gage de réussite. Il est difficilement imaginable de préciser un domaine sans validation. On court le risque de voir ce domaine modifié, avec des répercussions importantes à un niveau plus fin et donc une perte de temps considérable. Cependant, autant cette méthode est intéressante pour modéliser les données, autant il nous semble que son modèle des traitements

n'est pas adapté à la recherche exploratoire. Les méthodes objets semblent mieux adaptées à la modélisation du comportement dynamique d'un système.

Une fois les modèles stabilisés, la suite du travail concerne plus l'informaticien. Il devra effectuer des modifications de représentation des modèles en garantissant l'intégrité de l'information désirée. Ce souci de mémorisation de l'information s'appuie principalement sur les règles de normalisation de Codd (1972).

RÉFÉRENCES

- Bouzeghoub M (1993) Introduction aux méthodes de conception orientées objets. In: *Actes de l'école d'été 1993 : Autour de l'objet*. Informatique
- Chen PP (1976) The entity-relationship model. Toward a unified view of data. *ACM T Database Syst* 1, 9-36
- Codd EF (1972) Further normalization of the data base relational model. In: *Courant Computer Science Symposia, Vol 6 : Data base systems* (Rustin R, ed). Prentice-Hall, Englewood Cliffs, 33-64
- Faye B, Barnouin J, Lescourret F (1989) Objectifs principaux et stratégie de l'enquête écopathologique bretagne sur la vache laitière. *Epidémiol Santé Anim* 15, 23-31
- Lescourret F, Pérochon L, Coulon JB, Faye B, Landais E (1992) Modelling an information system using the MERISE method for agricultural research: the example of a database for a study on performances in dairy cows. *Agric Syst* 38, 149-173
- Lescourret F, Genest M, Barnouin J, Chassagne M, Faye B (1993) Data modeling for database design in production and health monitoring systems for dairy herds. *Dairy Sci* 76, 1053-1062
- Rochfeld A, Morejon J (1988) Extensions au modèle conceptuel des données. *Modèle et bases des données* 8, 37-56
- Scherrer B (1984) *Biostatistique*. Morin, Chicoutimi, Québec
- Tardieu H, Rochfeld A, Coletti R (1983) *La méthode MERISE Tome I, Principes et Outils*. Les Éditions d'Organisation, Paris