

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Analyse d'un système de gestion de fichiers de données statistiques pour la mise au point de modèles économétriques
analyse réalisée dans le cadre du bureau du plan

Pletinckx, M.

Award date:
1980

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Facultés Universitaires
Notre-Dame de la Paix
NAMUR
Institut d'Informatique

Année académique 1979-1980

ANALYSE D'UN SYSTEME DE GESTION DE
FICHIERS DE DONNEES STATISTIQUES
POUR LA MISE AU POINT DE MODELES
ECONOMETRIQUES

ANALYSE REALISEE DANS LE CADRE DU BUREAU DU PLAN.

M. PLETINCKX

MEMOIRE PRESENTE EN VUE
DE L'OBTENTION DU GRADE
DE LICENCIE ET MAITRE
EN INFORMATIQUE.

LBS 3590672

77167

REMERCIEMENTS

Je tiens à exprimer ma plus profonde gratitude à Messieurs HAINAUT, BOGAERT et BODART sans qui ce travail n'aurait pu être réalisé.

Leur patience, leurs conseils avisés et leur aide constante tant sur le plan scientifique qu'humain m'ont permis d'aborder le problème, de le formaliser et de rédiger ce mémoire.

Je tiens à remercier aussi tout le personnel du Bureau du Plan qui a participé d'une manière quelconque à la réalisation de ce travail.

Enfin, que toutes les personnes qui m'ont aidée durant cette année trouvent ici l'expression de ma plus vive reconnaissance.

SOMMAIRE

	<u>Page</u>
INTRODUCTION	
Ière PARTIE : ETUDE DE L'EXISTANT	
CHAPITRE I - Présentation du Bureau du Plan	1
1.1. Le Bureau du Plan dans la procédure de planification en Belgique	1
1.2. Les moyens mis en oeuvre au Bureau du Plan pour établir le projet de Plan	3
CHAPITRE II - Le système TSAR	6
2.1. Les fichiers TSAR	7
2.2. Les applications	13
2.3. Critiques du système TSAR et extension	23
CHAPITRE III - Les objectifs et la méthode de travail	25
IIème PARTIE : ANALYSE FONCTIONNELLE	
CHAPITRE I - Description des informations	30
1.1. Un modèle de la vie économique et sociale	30
1.2. Etude détaillée des relations du modèle	39
CHAPITRE II - Architecture des informations	45
2.1. Les modèles descriptifs de la vie économique et sociale	45
2.2. Les modèles économétriques	46
2.3. Synthèse	56
CHAPITRE III - Structuration des traitements	59
3.1. Les phases de traitement	59
3.2. Les fonctions	60
CHAPITRE IV - Schéma fonctionnel de la mise au point d'un modèle économétrique	81

	Page
IIIème PARTIE : ANALYSE D'IMPLEMENTATION	84
CHAPITRE I - Le système informatique du Bureau du Plan	85
1.1. Le langage APL	85
1.2. Les softwares existants	86
1.3. Objectif de l'analyse d'implémentation	87
CHAPITRE II - Etude de l'identifiant des variables et classes de variables	90
2.1. Etude théorique de l'identifiant des variables	90
2.2. La cofication de la clé de la relation "variable 2"	94
2.3. Un nom de classe de variables	97
2.4. Traduction d'un nom de classe de variables	99
CHAPITRE III - Les équations génératrices d'équations économétriques	101
CHAPITRE IV - Structure physique d'un fichier projet	105
CHAPITRE V - Evaluation du travail, conclusion et possibilités d'extension	109

Bibliographie

Annexe

INTRODUCTION

Le cadre de référence de ce travail est constitué par le système mis en place au Bureau du Plan pour la création et la mise au point de modèles économétriques.

L'objectif du mémoire est de repenser les fichiers de données statistiques nécessaires à la construction des modèles économétriques et ce tant du point de vue des informations qu'ils doivent contenir que du point de vue des procédures de manipulation des fichiers et des informations de ces fichiers.

Ce travail se subdivise en trois grandes parties.

La première présente le système informatique mis en place au Bureau du Plan pour la création et la mise au point d'un modèle économétrique et en particulier le système TSAR de gestion des fichiers de données statistiques. On y présente aussi les critiques des utilisateurs concernant le système TSAR.

La deuxième partie de ce mémoire constitue l'analyse fonctionnelle du système informatique. Les étapes de cette analyse sont :

- la modélisation des informations d'un fichier de données statistiques ;
- la présentation des grandes phases de l'élaboration d'un modèle économétrique et une étude plus détaillée de la phase de construction qui a permis de définir les types de fichiers de données statistiques du Bureau du Plan ;
- la structuration des phases de traitement nécessaires au système de gestion de fichiers de données statistiques et à la mise au point de modèles économétriques.

Enfin, la troisième partie concerne l'implémentation de la phase de construction d'un modèle économétrique.

Ière Partie - Etude de l'existant

Chapitre I - Présentation du Bureau du Plan

- 1.1. Le Bureau du Plan dans la procédure de planification en Belgique
- 1.2. Les moyens mis en oeuvre au Bureau du Plan pour établir le projet de Plan

Chapitre II - Etude du système TSAR

- 2.1. Les fichiers TSAR
 - 2.1.1. Exemples de fichiers TSAR du Bureau du Plan
 - 2.1.2. Caractéristiques des variables TSAR
 - 2.1.3. Structure physique d'un fichier TSAR
 - 2.1.3.1. Le fichier "index"
 - 2.1.3.2. Le fichier "série"
 - 2.1.3.3. Le fichier "commentaire"
 - 2.1.3.4. Schéma de la structure physique d'un fichier TSAR
- 2.2. Les applications
 - 2.2.1. Les procédures Batch
 - 2.2.1.1. La création d'un fichier TSAR
 - 2.2.1.2. La MAJ d'un fichier TSAR
 - 2.2.1.3. L'édition d'un fichier TSAR
 - 2.2.2. Les procédures interactives
 - 2.2.2.1. La mise au point d'un modèle économétrique
 - 2.2.2.2. Fonctions d'accès et de manipulation standard
 - 2.2.3. Schéma de l'existant

2.3. Critiques du système TSAR et extensions souhaitées

2.3.1. Les critiques

2.3.2. Les améliorations

2.3.3. Les extensions

Chapitre III - Les objectifs et la méthode de travail

CHAPITRE I - PRESENTATION DU BUREAU DU PLAN

1.1. Le Bureau du Plan dans la procédure de planification en Belgique

Le Bureau du Plan a été créé par la Loi du 15 juillet 1970, c'est un organisme d'intérêt public chargé d'élaborer le Plan, il est placé sous l'autorité du Premier Ministre et du ministère des Affaires économiques.

Le processus théorique d'élaboration d'un Plan se décompose selon la Loi en 2 phases : (1)

- "- la phase initiale, ou phase des options, permet aux décideurs de fixer les orientations fondamentales du futur Plan, à partir d'un inventaire et d'une confrontation des choix nationaux dits "techniquement possibles" et des options régionales ;
- "- les options retenues doivent ensuite être transposées lors de la phase de confection du Plan, en un ensemble cohérent et plus détaillé d'objectifs et de moyens. L'originalité - voulue par le législateur - de la procédure de planification en deux stades et les exigences de concertation qu'il a voulu imposer sont illustrées par le schéma ci-après :"

(1) Extrait du mémoire "Etude d'interface entre un Software d'estimation et de simulation écrit en APL "IESEM" et un Software de simulation écrit en PL1 "SIMUL" de Yvan POMES.

PROCEDURE D'ELABORATION DU PLAN

PREMIERE ETAPE : LES CHOIX DU PLAN

BUREAU DU PLAN

LA DIRECTION GENERALE LA DIRECTION SECTORIELLE

- LES 3 SECTIONS REGIONALES (en collaboration avec les C.E.R. et sur base de l'inventaire dressé par les S.D.R.), préparent • LES 3 C.E.R. adoptent

propose LES CHOIX DE NIVEAU GLOBAL NIVEAU SECTORIEL NIVEAU REGIONAL

confrontation
AVANT-PROJET D'OPTIONS

GOUVERNEMENT

Recueille les AVIS du Conseil Central de l'Economic, du Conseil National du Travail et des Conseils économiques régionaux

les CONCLUSIONS du Comité National de l'Expansion Economique

Délibère, compte tenu de ces avis et conclusions et
Adopte un PROJET D'OPTIONS

Le transmet au Parlement avec l'avant-projet du Bureau du Plan et les avis et conclusions prévus ci-dessus

PARLEMENT

Discute, amende et adopte les OPTIONS du PLAN

DEUXIEME ETAPE : LE PLAN LUI-MEME

BUREAU DU PLAN

LA DIRECTION GENERALE LA DIRECTION SECTORIELLE

- LES 3 SECTIONS REGIONALES (en collaboration avec les C.E.R.), préparent • LES 3 C.E.R. adoptent

prépare LE PROJET DE PLAN AU NIVEAU GLOBAL

prépare LES PROJETS DE PROGRAMMES SECTORIELS

LES PROJETS DE PLANS REGIONAUX

confrontation et intégration
AVANT-PROJET DE PLAN ARBITRÉ

GOUVERNEMENT

Recueille les AVIS - des Conseils régionaux - des Conseils professionnels

les CONCLUSIONS du Comité National de l'Expansion Economique

Délibère, compte tenu de ces avis et conclusions et
Adopte le PROJET DE PLAN

Le transmet au Parlement avec les avis et conclusions

PARLEMENT

Discute, amende et adopte le PLAN

1.2. Les moyens mis en oeuvre au Bureau du Plan pour établir le projet de Plan

Un projet de Plan s'établit sur base d'un modèle économétrique.

Un modèle économétrique se construit sur 2 types d'informations :

- les séries de données statistiques qui sont les variables du modèle
- les équations économétriques

Une série de données statistiques représente la valorisation d'une composante de la vie économique et sociale sur une période déterminée.

Ex.: la consommation privée annuelle totale de produits alimentaires exprimée en milliards de francs

1970	: 22.283
1971	: 23.471
1972	: 24.908
1973	: 26.809
1974	: 31.356
1975	: 35.392
1976	: 38.161
1977	: 43.918

Les équations économétriques décrivent les interactions entre les différentes composantes de la vie économique et sociale et ce, dans le but de pouvoir prévoir leur évolution.

Ex.: modèle économétrique de la consommation privée totale :

$$\left\{ \begin{array}{l} C = \alpha_0 + \alpha_1 (Y-T) \\ Y = C + I \end{array} \right.$$

où
les variables du modèle

{	C = consommation privée totale
	Y = revenu
	T = taxes
	I = investissements

α_0 et α_1 = les coefficients du modèle à estimer.

Dans ce modèle,

C et Y sont les variables endogènes ou expliquées, c'est-à-dire, des variables dont la valeur est définie par le modèle

T et I sont les variables exogènes, c'est-à-dire, des variables dont la valeur est définie en dehors du modèle.

"L'élaboration d'un modèle économétrique peut comprendre 1 ou plusieurs des étapes suivantes : (1)

STEP0 : conception du modèle

- quelles sont les variables
- comment les relier

STEP1 : - construction et recherche des variables
- construction des équations économétriques

STEP2 : - estimation des coefficients des équations économétriques

STEP3 : - structuration en modèle simulable
- simulations statiques et dynamiques sur le passé
- prévisions des exogènes + coefficients
- établissement d'une projection de référence
- étude de projections alternatives
- étude de scénarios".

(1) Extrait du mémoire cité.

Dans ce travail, on s'intéressera tout particulièrement au STEP1 de construction des variables et des équations d'un modèle économétrique.

On ne s'intéresse au STEP2 d'estimation et au STEP3 de simulation que dans la mesure où ils s'alimentent au fichier des équations économétriques et au fichier de données statistiques construits pour les besoins du modèle.

1. Construction des données statistiques

Les informations dont on dispose pour la construction et la recherche des variables d'un modèle économétrique sont les fichiers de données statistiques de base construits en général en dehors du Bureau du Plan.

La construction des variables d'un modèle économétrique nécessite la construction d'un fichier de données statistiques propre aux besoins du modèle économétrique.

Les fichiers de données statistiques du Bureau du Plan sont les fichiers TSAR. Le système TSAR comprend :

- les procédures de gestion des fichiers de données statistiques
- les fichiers de données statistiques.

L'objectif du mémoire est de redéfinir, de reconcevoir le système TSAR, tant du point de vue des informations des fichiers de données statistiques que du point de vue des procédures de gestion de fichiers.

2. Construction des équations économétriques

Les équations économétriques sont construites en interactif ; elles seront écrites en LEC Langage Econométrique Condensé qui a été créé au Bureau du Plan.

La syntaxe du LEC répond à 2 objectifs principaux :

- être aussi proche que possible d'une notation économétrique courante, être, en outre, d'une syntaxe arithmétique normale et indépendante d'un langage de programmation ;
- être facilement manipulable par le concepteur d'équations économétriques ; le jeu d'opérateurs économétriques propre à cette nouvelle syntaxe doit pouvoir être étendu, par l'utilisateur lui-même, au fur et à mesure de ses besoins.

Une équation économétrique sera écrite en fonction d'opérateurs LEC, de numériques et de noms de variables.

Les équations économétriques d'un modèle constitueront le fichier d'équations du modèle.

CHAPITRE II - ETUDE DU SYSTEME TSAR

Les informations des fichiers de données du Bureau du Plan sont des séries statistiques construites :

- à l'extérieur du Bureau du Plan par des bureaux d'études statistiques tels que INS, OCDE, CEE,.....
- au Bureau du Plan pour les besoins d'un modèle économétrique.

Le système d'informations des données statistiques du Bureau du Plan s'appelle le système TSAR et comprend :

- les fichiers TSAR
- des procédures de manipulation de fichiers TSAR
- des fonctions de manipulation des variables d'un fichier TSAR.

2.1. Les fichiers TSAR

Un fichier TSAR est constitué de l'ensemble des données statistiques relatives à un aspect cohérent de la vie économique et sociale.

Il existe actuellement une trentaine de fichiers TSAR au Bureau du Plan qui peuvent être de 2 types :

- les fichiers construits en général à l'extérieur du Bureau du Plan et que l'on appelle les fichiers de base ;
- les fichiers construits au Bureau du Plan pour répondre aux besoins en données statistiques d'un modèle économétrique ; données qui seront construites à partir des données en provenance des fichiers de base.

2.1.1. Exemples de fichiers TSAR du Bureau du Plan

fichiers
de
base

- BPEMPL0I qui contient les données statistiques concernant l'emploi sectoriel et régional
- NACCOUNT qui contient les données statistiques de la comptabilité nationale
- REGINV qui contient les données statistiques concernant les investissements industriels par province et par région
- COSECAIB qui contient les données statistiques relatives aux comptes sectoriels et annuels pour l'industrie belge
- COMCOR qui contient des données statistiques représentant des agrégats macro-économiques pour la C.E.E.

fichiers
de
modèles
écono-
métriques

{ - MARIFILE qui est le fichier de données statis-
tiques du modèle économétrique
MARIBEL
- SERENA qui est le fichier de données statis-
tiques du modèle économétrique SERENA

2.1.2. Caractéristiques des variables TSAR

Un fichier TSAR est constitué de l'ensemble des données statistiques relatives à un aspect cohérent de la vie économique et sociale.

Une série de données statistiques relatives à la valorisation d'une composante de la vie économique et sociale sur une période déterminée sera une "variable TSAR".

Ex.: - la consommation privée totale des ménages étudiée sur la période 1953-1970

- l'indice des prix à la consommation de pain et produits de boulangerie de 1953 à 1970.

Une variable TSAR est caractérisée par :

- un nom de variable qui est unique dans le fichier. C'est un mot de 6 caractères au maximum commençant par une lettre
- une série chronologique d'observations statistiques relatives à une composante de la vie économique et sociale sur une période déterminée

La série chronologique d'une variable annuelle sur la période 70-78 aura 9 éléments, chaque élément représentera l'observation statistique de la variable pour une année.

Si l'on ne dispose pas d'observations pour une période déterminée, on doit prévoir un "trou" dans la série chronologique. "Un trou" dans une série chronologique sera représenté par un élément vide dans la série.

- un en-tête de série "header" qui est constitué de PPMMMDD où

PP = un indicateur de périodicité :

= 00	série cross section
01	" annuelle
02	" semestrielle
04	" trimestrielle
12	" mensuelle
52	" hebdomadaire

MMMM = millésime de la lère observation

DD = période dans l'année de la lère observation

- des informations de commentaires comprenant :

un texte en format libre dans lequel on donne, en général, les références exactes des publications d'où proviennent les observations et

une série de codes reprenant :

- * l'auteur, c'est-à-dire, la personne responsable au Bureau du Plan de l'introduction de la variable dans le fichier
- * la date de dernière MAJ (JJMMAA)
- * le code d'origine bibliographique de la variable (B.P., I.N.S., O.C.D.E., C.E.E.,)
- * un code de complétion de la variable indiquant si les observations se succèdent sans interruption ou s'il en manque
- * la date de création de la variable (MMAA)
- * le champ géographique d'application indiquant le territoire auquel se rapporte la variable

- * le type de la variable : en valeur absolue ou en indice
- * le millésime de référence nécessaire pour indiquer l'année de base de l'indice ou des prix constants
- * la nature de la variable indiquant si la variable est à prix courants, à prix constants, en prix, en salaires, en taux d'intérêt, en population, emploi
- * l'identification des unités dans lesquelles la variable est exprimée
- * l'ordre de grandeur par rapport aux unités : en unité, en milliers d'unités ...

ex. d'une série de codes pour une variable

"TDB" : responsable de l'introduction de la variable dans le fichier

031078 : date dernière MAJ : JJMAAA

02 : source des observations : I.N.S.

1 : la série est complète

1074 : date création sous la forme MMAA

01 : champ géographique d'application : la Belgique

0 : type de la variable, elle est exprimée en valeur absolue

6 blancs : qui indiquent l'absence d'année de base de l'indice des prix

1 : nature de la variable : en prix courants

03 : unité dans laquelle sont exprimées les observations : le FB

09 : ordre de grandeur par rapport aux unités : en milliards.

2.1.3. La structure physique d'un fichier TSAR

Un fichier TSAR est implémenté en 4 fichiers physiques :

- le fichier "index" reprenant essentiellement le nom des variables du fichier TSAR ;
- le fichier "série" contenant les observations des variables du fichier ;
- le fichier "commentaire" contenant toutes les autres caractéristiques des variables ;
- un fichier "historique" dont on ne sait rien.

2.1.3.1 Le fichier "index"

Le fichier "index" d'un fichier TSAR est utilisé en APL pour traduire le nom de variable en accès aux caractéristiques de la variable dans le fichier "série" et dans le fichier "commentaire".

Il faut savoir, qu'en APL, on peut accéder aux fichiers :

- en séquentiel

ou

- en accès direct sur le numéro de record qui est la position relative du record par rapport au début du fichier.

Le fichier "index" est un fichier séquentiel contenant pour chaque variable du fichier TSAR :

- le nom de la variable
- le numéro du 1er record concernant la variable dans le fichier "série"
- le nombre de records concernant la variable dans le fichier "série"

L'accès au fichier "série" sur le nom de variable nécessitera un balayage séquentiel du fichier "index" suivi de la lecture des records concernant cette variable dans le fichier "série".

Le fichier "commentaire" d'un fichier TSAR contient 1 record par variable, la position relative du nom de variable dans le fichier "index" sera donc le numéro de record du commentaire de la variable dans le fichier "commentaire".

Pour accélérer l'accès aux caractéristiques intrinsèques des variables (contenues dans les fichiers "série" et "commentaire"), en début de session APL, on copie le fichier "index" en mémoire sous la forme d'un tableau ; chaque ligne du tableau correspond aux informations d'accès à 1 variable (qui sont le nom de variable, le numéro du 1er record concernant la variable dans le fichier "série" et le nombre de records concernant cette variable dans le fichier "série").

2.1.3.2 Le fichier "série"

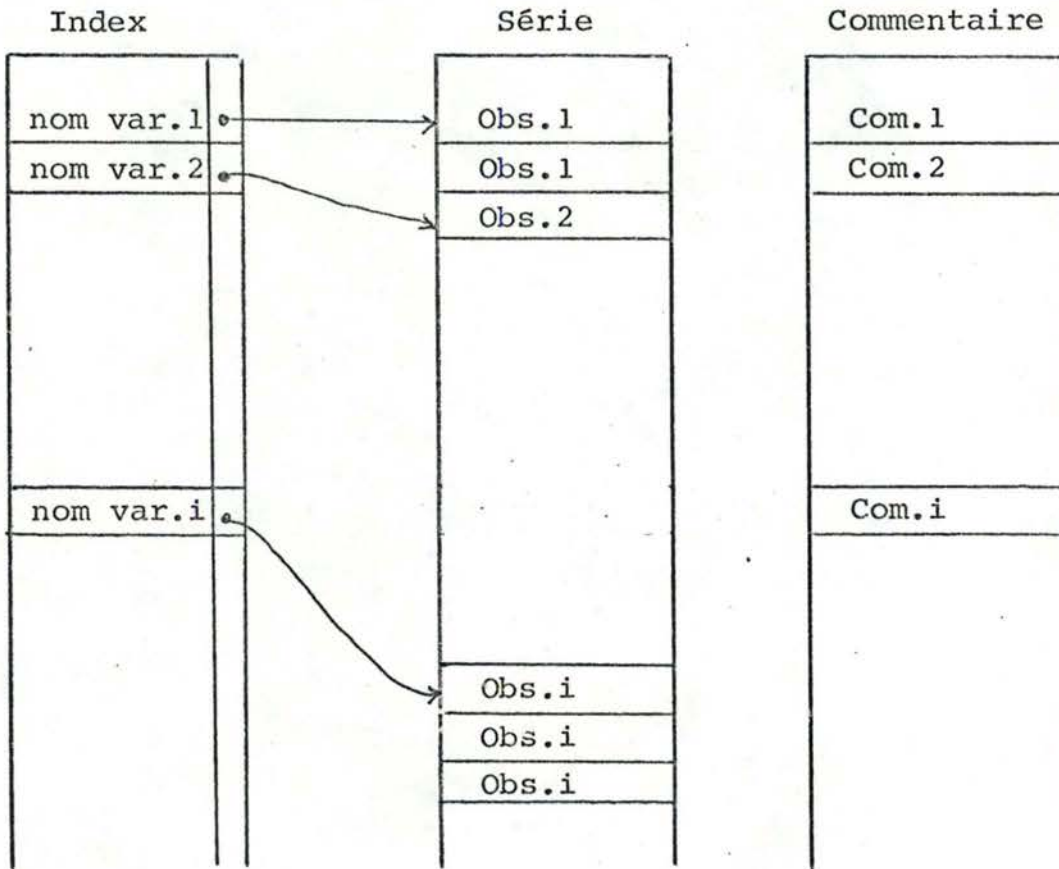
Le fichier "série" d'un fichier TSAR contient les observations des variables du fichier TSAR. La longueur d'un record est de 400 bytes ce qui permet de justifier que pour une variable TSAR, on peut avoir plusieurs records dans le fichier "série".

Le fichier "série" est accessible sur le nom de variable ou sur le numéro de série. Le numéro de série est la position de la variable dans le fichier TSAR.

2.1.3.3 Le fichier "commentaire"

Le fichier "commentaire" d'un fichier TSAR est un fichier dont les records ont 640 bytes, chaque record contient le "commentaire" d'une variable du fichier TSAR. Ce fichier est accessible sur le nom de variable et sur le numéro de la variable dans le fichier.

2.1.3.4 Schéma de la structure physique d'un fichier TSAR :



2.2. Les applications

Les fichiers TSAR peuvent être utilisés en batch ou principalement en interactif.

Les procédures batch consistent en :

- la création d'un fichier TSAR
- la MAJ d'un fichier TSAR
- l'édition d'un fichier TSAR

Les procédures interactives sont :

La mise au point d'un modèle économétrique

Fonctions d'accès et de manipulation standard

2.2.1 Les procédures batch

2.2.1.1 La création d'un fichier TSAR

La création d'un fichier TSAR doit toujours se faire à partir d'un fichier sous format carte. Chaque variable créée doit être construite comme suit :

- une carte en-tête contenant le nom de la variable, un indicateur de lecture de la carte format et des cartes commentaires, un header contenant un indicateur de périodicité, l'année de départ de la série et la période de départ dans l'année ;
- une carte format si l'indicateur est positionné (le format implicite d'une observation est F10.0) ;
- les cartes séries qui contiennent les observations de la variable dans le format précisé par la carte format (ou par le format implicite) ;
- les cartes commentaires dont la 1ère et la dernière sont obligatoires. La 1ère carte contient le nombre de cartes commentaires $2 \leq \text{nombre} \leq 9$. Les cartes optionnelles (max 7) décrivent la série en texte libre. La dernière carte, la carte ** est une carte d'informations codées des caractéristiques de la variable par ex.: le responsable de la variable, la date de la dernière MAJ de création, code de l'origine bibliographique de la variable, le champ géographique de la variable, son type (en valeur absolue ou en indice), l'année de référence
.....

La procédure de création d'un fichier TSAR crée 4 fichiers physiques distincts :

- le fichier "index" séquentiel, 1 record = 13027 pos.
- le fichier "série" en accès direct, 1 record = 400 pos.
- le fichier "commentaire" en accès direct, 1 record = 640 pos.
- le fichier historique

Le programme de création est un programme batch écrit en PL1.

2.2.1.2 La MAJ d'un fichier TSAR

- a. d'un fichier de base : cette MAJ est le plus souvent annuelle et consiste en :
- ajouter 1 (ou plusieurs) observations à chaque variable avec correction éventuelle des observations des années précédentes ;
 - éventuellement, s'il y a un changement de base pour les séries indices de prix, il faut recalculer les séries indices et les séries en quantité, modifier le millésime de référence ;
 - modifier le commentaire en texte libre, la date de dernière MAJ.

- b. d'un fichier construit au Bureau du Plan : les fichiers construits au Bureau du Plan doivent être mis à jour quand on change un des fichiers de base intervenant dans leur création.

L'autre type de MAJ pour ces fichiers est directement lié à l'évolution du modèle économétrique pour lequel ils sont créés. Ces MAJ peuvent devoir s'effectuer à n'importe quel stade de l'élaboration du modèle économétrique.

On peut être amené à introduire de nouvelles variables dans le fichier, à modifier la définition d'une variable, à supprimer une variable. Les MAJ des variables existantes dans un fichier TSAR doivent être faites sur le fichier format cartes et on recrée alors le fichier à partir du nouveau fichier format cartes. Si la MAJ consiste uniquement en l'ajoute de nouvelles variables, on peut utiliser le programme de fusion des deux fichiers.

2.2.1.3 L'édition d'un fichier TSAR en batch par cartes commandes on peut

- imprimer la liste des séries
- imprimer la documentation des séries
- créer un fichier standard à partir d'un fichier enregistré et ce, pour tout le fichier TSAR ou pour une liste de variables identifiées par leur nom.

2.2.2 Les procédures interactives

2.2.2.1 Mise au point de modèles économétriques

Un modèle économétrique est un ensemble d'équations reliant des variables économiques dans le but d'expliquer l'économie. Comment s'élabore un modèle ?

Par exemple, on veut créer un modèle de la consommation totale en Belgique, quelles sont les différentes étapes à suivre ?

1. Rechercher les variables qui permettent de définir la consommation :

la consommation totale est fonction de :

- la consommation privée
- la consommation publique
- l'investissement

Toutes ces variables sont définies dans le fichier NACCOUNT, on doit donc lire le fichier NACCOUNT pour y chercher les valeurs des variables nécessaires ;

2. Construction des données statistiques du modèle :

Les besoins ayant été définis à l'étape précédente, on va construire les données utilisables par le modèle; pour l'exemple on procédera comme suit :

C = variable Naccount de la consommation privée

G = variable Naccount de la consommation publique

I = variable Naccount de l'investissement

$Y = C + G + I$ Y = revenu observé

On doit établir aussi les relations qui existent entre les variables pour l'exemple choisi , on définira les équations suivantes :

(1) $C = \underline{C}1 + \underline{C}2 \times Y$

(2) $Y = C + G + I$

(1) : on définit la consommation comme étant une fonction du revenu

$\underline{C}1$ et $\underline{C}2$ sont des coefficients que l'on doit estimer.

(2) : on identifie le revenu à la somme des consommations et de l'investissement.

Ces équations sont écrites en LEC. Ces équations sont alors traduites par la fonction TRANSLEC qui a pour but de créer essentiellement :

- la forme APL de l'équation (input estimation)

- la liste des variables utilisées.

Ces informations sont écrites dans un fichier d'équations.

3. Après la création du fichier d'équations, on doit estimer les équations, c'est-à-dire qu'il faut trouver les valeurs des coefficients \underline{C}_1 , \underline{C}_2 . C'est le rôle de la fonction ESTEQ. Cette fonction crée dans le fichier d'équations des informations statistiques dont notamment la liste des valeurs trouvées pour les coefficients.
4. Quand on a trouvé des valeurs statistiquement acceptables pour les coefficients des équations, on peut passer à l'étape suivante qui est la simulation du modèle créé. Les programmes de simulation nécessitent en entrée :
 - un fichier d'équations qui doit être trié dans l'ordre de résolution du modèle. Ces équations seront traduites en sous-routines FORTRAN ;
 - un fichier de variables trié par ordre alphabétique avec les variables endogènes en premier lieu, les exogènes ensuite ;
 - des paramètres.

On peut faire soit une simulation statique, soit une simulation dynamique du modèle. Pour une simulation statique, on prend la valeur observée pour une variable retardée.

(Ex. $WP = \underline{C}_1 + \underline{C}_2 \times \underline{L} (YD-CP) + \underline{C}_3 \times (Li - .9 \times Li(1))$, $Li(1)$ est une variable retardée, c'est-à-dire que la variable WP dépend de la valeur de la variable Li au temps i et au temps (i-1). Donc, dans une simulation statique, on prendra les valeurs observées pour la variable Li au temps i et au temps (i-1). Pour une simulation dynamique, on prend la valeur simulée de la variable avec retard (dans l'exemple, on prendra pour $Li(1)$ la valeur trouvée à l'itération précédente)).

Après la résolution du modèle, les résultats trouvés sont stockés dans un fichier résultat qui pourra être consulté et analysé soit en batch soit au terminal. Il peut servir notamment à l'édition de tableaux variés et comme entrée pour une nouvelle simulation.

Remarques :

1. Une simulation du passé permet de tester la validité du modèle créé. Une simulation du futur sera toujours une simulation dynamique.
2. Le fichier d'entrée au programme de simulation est construit en interactif et est obtenu à partir du fichier de données du modèle (uniquement le fichier série). Il faut sélectionner dans le fichier de données du modèle les variables concernées par la simulation, les extrapoler (et éventuellement les "rétropoler") jusqu'à l'année voulue, insérer des séries externes (par exemple des estimations de l'évolution du marché mondial). Le fichier de données du modèle est un fichier d'observations alors que le fichier de simulation contient aussi des prévisions.

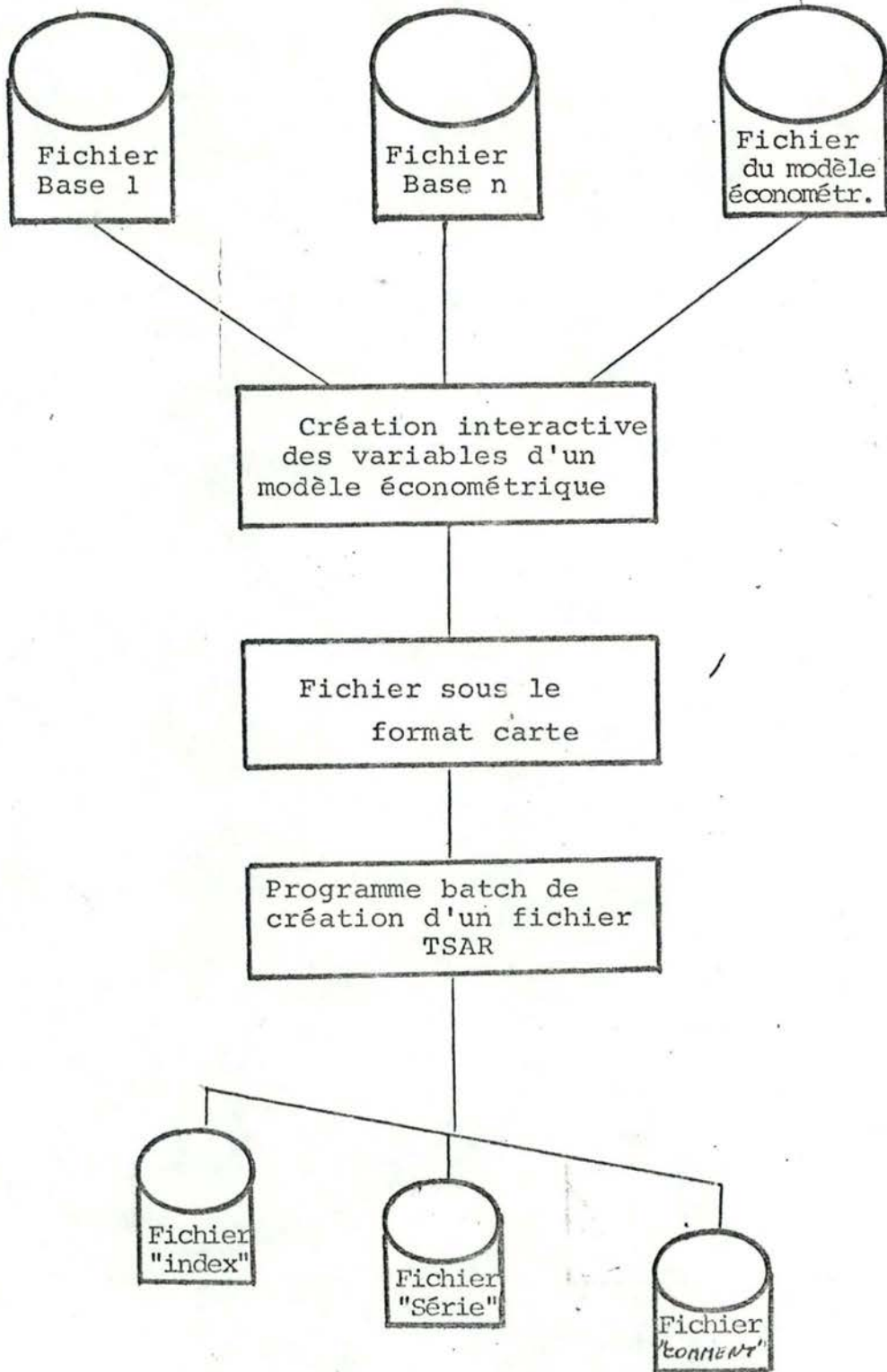
2.2.2.2 Fonctions d'accès et de manipulation standard

- La fonction d'ouverture d'un fichier TSAR (GO) copie le fichier "index" du fichier TSAR en mémoire.
- La fonction de lecture des observations d'une variable ou d'une liste de variables (SEARCH). Les observations d'une variable lue sont rangées dans la variable APL (nom variable TSAR) où (nom variable TSAR) est l'argument de la fonction SEARCH ou FIND.
- La fonction de lecture du commentaire d'une variable ou d'une liste de variables (DEFINE).
- La fonction de lecture du commentaire et des observations d'une variable ou d'une liste de variables (FIND).
- Les commentaires d'une variable lue sont rangés dans la variable APL (H⟨nom de variable TSAR⟩ où ⟨nom variable TSAR⟩ est l'argument de la fonction FIND ou DEFINE.
- Une fonction qui permet de retrouver le numéro d'ordre d'une variable par rapport au début du fichier en donnant son nom (CLE).

- Des fonctions qui permettent de retrouver les numéros d'ordre des variables par rapport au début du fichier en donnant une partie du nom de variable (GEN et GENGEN).
- On peut afficher à l'écran les observations des variables ou les commentaires des variables présentes en mémoire (LISTASERIES, LISTADEF, LISTAALL).
- On peut éditer les observations des variables, les commentaires sous forme de tableaux soit au terminal ou au remote ou au CTI (PRINTASER, PRINTADEF).
- On peut créer un fichier temporaire des variables présentes en mémoire (WORK△FILE).
- On peut lister le contenu du fichier temporaire et ce, soit au terminal, ou remote ou au CTI (LIST△WORK).
- Des fonctions qui permettent d'effacer de la mémoire les observations et/ou les commentaires des variables (CLEARASER, CLEAR△DEF, CLEAR).
- Une fonction de destruction du fichier temporaire (DELETE).
- Des fonctions qui permettent d'ajouter de nouvelles variables aux variables présentes en mémoire (ADD△SER, ADD△DEF).
- Une fonction qui permet de mettre le header d'une variable comme première observation de la variable (ADD△HEAD).
- Une fonction qui crée une matrice des observations des variables (SER△MAT).
- Une fonction qui donne la longueur maximale d'un ensemble de séries (MAXLEN).
- Une fonction de fermeture des fichiers ouverts (CLOSE).
- Une fonction qui donne le header d'une variable (HEADER).
- Une fonction qui imprime la liste triée alphabétiquement des noms des variables du fichier avec leur numéro d'ordre dans le fichier (LISTANAMES).
- Une fonction qui donne la liste des noms de toutes les variables du fichier dans l'ordre où elles se trouvent dans ce fichier (RECODE).

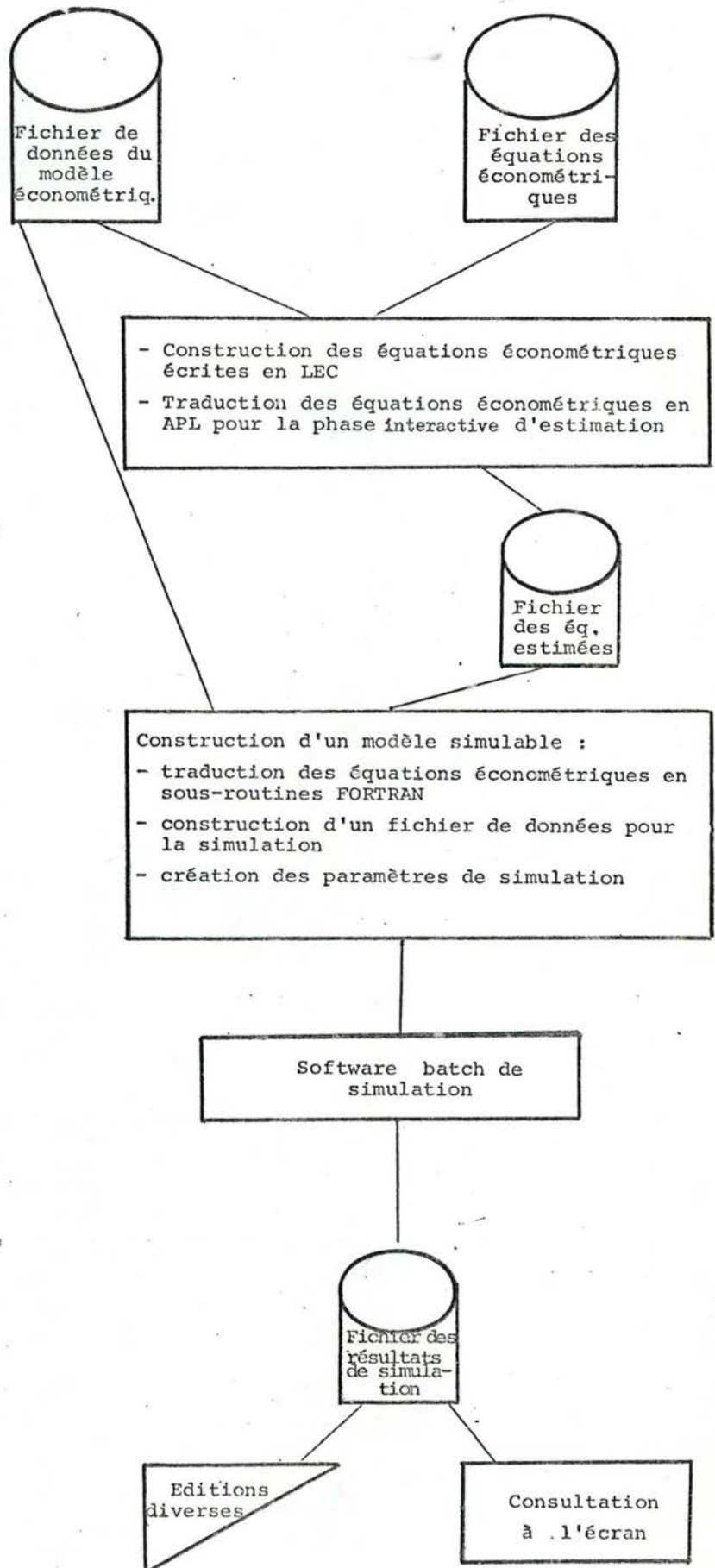
2.2.3 Schéma de l'existant

Création et MAJ d'un fichier TSAR



La création et la MAJ d'un fichier TSAR se fait quasiment toujours à partir d'un fichier de cartes perforées; même la création ou la MAJ d'un fichier de données d'un modèle économétrique passe par une phase de perforation de cartes.

Mise au point d'un modèle économétrique



2.3 Critiques du système TSAR et extensions souhaitées

Le système TSAR a l'avantage d'exister et permet de répondre assez bien aux besoins des utilisateurs du Bureau du Plan. Les fichiers TSAR sont, bien entendu, utilisés au Bureau du Plan mais aussi à l'extérieur du Bureau du Plan.

Les critiques du système TSAR ont été dégagées uniquement par consultation des utilisateurs du Bureau du Plan.

2.3.1 Les critiques

Ces critiques sont de plusieurs ordres :

- l'accès aux variables d'un fichier TSAR se fait uniquement sur base de l'identifiant de variable : le nom de variable. Ne pourrait-on utiliser d'autres caractéristiques d'accès aux variables et lesquelles ?
- les données statistiques d'un modèle économétrique se construisent en interactif à partir de variables des fichiers de base mais :
 - on ne peut accéder à plusieurs fichiers TSAR en concurrence : la fonction d'ouverture d'un fichier TSAR copie en mémoire tout le fichier "index" de ce fichier TSAR et la place mémoire disponible pour un utilisateur est limitée ;
 - si le nombre de variables d'un fichier TSAR est supérieur à ± 3000 , il devient très difficile de le manipuler en interactif : la place occupée par le fichier "index" en mémoire étant trop importante ;
 - la MAJ d'un fichier TSAR doit nécessairement se faire par la procédure de création du fichier, il faut modifier le fichier TSAR en format cartes puis exécuter le programme de création d'un fichier TSAR. Si la MAJ consiste uniquement à rajouter des variables en fin de fichier, on peut utiliser le programme batch de fusion de fichiers.

2.3.2 Les améliorations souhaitées

Les améliorations que l'on devrait introduire dans le nouveau système de gestion de fichiers se situent au niveau de l'édition d'un fichier de données statistiques et de la MAJ de ces même fichiers.

2.3.2.1 Il faudrait prévoir des éditions paramétrables des fichiers de données permettant :

- de sélectionner les variables à éditer
- de sélectionner les caractéristiques des variables à éditer par exemple l'édition d'une table des matières reprenant :
 - le nom de série
 - le nombre d'observations
 - la période de début des observations
- de créer des caractéristiques pour l'édition telles que :
 - calcul de la moyenne des observations d'une variable
 - calcul de l'écart-type
- de présenter un fichier trié
- d'imprimer un fichier soit à l'écran du terminal, soit sur papier.

2.3.2.2 Il faudrait étudier le problème de la MAJ des fichiers de données statistiques :

- des fichiers de base
- des fichiers de données statistiques d'un modèle économétrique.

2.3.3 Les extensions

Les extensions que l'on peut apporter au système TSAR ont été perçues en cours d'analyse : ce sont les équations génératrices d'équations de définition de variables et d'équations économétriques qui seront expliquées plus loin.

CHAPITRE III - LES OBJECTIFS ET LA METHODE DE TRAVAIL

L'objectif initial du travail a été de reconcevoir le système TSAR, il est à la base de l'analyse fonctionnelle.

Cette analyse a essayé de modéliser :

- ce qu'est un fichier de données statistiques pour le Bureau du Plan
- ce qu'est un modèle économétrique.

De cette modélisation, des applications existantes et des critiques des utilisateurs, on a pu découvrir les phases de traitements nécessaires à la mise au point d'un modèle économétrique et celles qui ne sont pas directement liées à la mise au point d'un modèle économétrique.

L'analyse d'implémentation a été orientée dans deux directions qui toutes les deux concernent uniquement les phases de construction d'un modèle économétrique (construction des données statistiques et des équations économétriques).

IIe Partie

ANALYSE FONCTIONNELLE

PLAN

CHAPITRE I - Description des informations

1.1. Un modèle de la vie économique et sociale

1.1.a. exemple

1.1.b. définition de concepts

1.1.b.1. objets économiques

1.1.b.2. liens économiques

1.1.b.3. modèle

1.2. Etude détaillée des relations du modèle

1.2.a. au niveau conceptuel

1.2.b. au niveau de la réalisation du modèle

1.2.c. au niveau de la collecte des données

CHAPITRE II - Architecture des informations

2.1. Les modèles descriptifs de la vie économique et sociale

2.2. Les modèles économétriques

2.2.1. Les grandes phases d'élaboration d'un modèle économétrique

2.2.2. Etude plus détaillée de la phase de construction d'un modèle économétrique

2.2.2.A. La construction du fichier de données statistiques d'un modèle économétrique

2.2.2.A.1. La construction des variables d'un modèle économétrique

- 2.2.2.A.2. Le fichier projet utilisateur
et le fichier projet global
- 2.2.2.B. La construction du fichier des
équations du modèle économétrique
- 2.2.2.B.1. Les étapes de construction des
équations économétriques
- 2.2.2.B.2. Le fichier des équations d'un
modèle économétrique

2.3. Synthèse

- 2.3.1. Les types de fichiers de données statis-
tiques du Bureau du Plan et leurs prin-
cipales caractéristiques
- 2.3.1.1. Le fichier projet utilisateur
- 2.3.1.2. Le fichier projet global
- 2.3.1.3. Un fichier de base
- 2.3.2. Un modèle économétrique

CHAPITRE III - Structuration des traitements

- 3.1. Les phases de traitement
- 3.1.1. Les phases de traitement concernant
la mise au point d'un modèle écono-
métrique
- 3.1.2. Les autres phases de traitement
- 3.2. Les fonctions
- 3.2.1. Description des fonctions par phase
de traitement
- 3.2.1.1. Les phases de traitement concer-
nant la mise au point d'un modèle
économétrique
- a. L'introduction de variables ou de
classes de variables dans le fichier
projet utilisateur.

- b. La construction du fichier des équations économétriques
 - b.1. La génération d'équations économétriques
 - b.2. L'insertion des équations générées dans le fichier des équations économétriques
 - c. L'estimation du fichier des équations économétriques
 - d. La structuration des équations d'un modèle économétrique en un modèle simulable
 - e. La création des données de simulation d'un modèle économétrique
 - f. La simulation du modèle
 - g. Les tests des résultats de simulation
 - h. L'intégration du fichier projet utilisateur dans le fichier projet global
- 3.2.1.2. Les autres phases de traitement:
- a. La création d'un fichier de base
 - b. Les contrôles de cohérence d'un fichier de données statistiques
 - c. L'édition d'un fichier de données statistiques
 - d. La MAJ d'un fichier de base
 - e. L'actualisation d'un fichier projet
- 3.2.2. Récapitulation des fonctions
- 3.2.2.a. Sélection de variables
 - 3.2.2.b. Recherche des nomenclatures référencées dans un identifiant de classe de variables

- 3.2.2.c. Constitution d'une liste d'identifiants de variables apparaissant dans une équation
- 3.2.2.d. Génération d'équations
- 3.2.2.e. Calcul des observations d'une variable
- 3.2.2.f. Affectation de valeur aux domaines de définition de variables
- 3.2.2.g. Remplacement d'une variable dans un fichier de données statistiques
- 3.2.2.h. Insertion d'une variable dans un fichier de données statistiques
- 3.2.2.i. Suppression de variable
- 3.2.2.j. Création des entrées pour le programme d'estimation
- 3.2.2.k. Recherche d'équations économétriques dans le fichier d'équation
- 3.2.2.l. Test de cohérence d'une nomenclature
- 3.2.2.m. Test de cohérence d'une variable
- 3.2.2.n. Test de cohérence d'un fichier de données statistiques
- 3.2.2.o. Calcul de caractéristiques statistiques de variables

CHAPITRE IV - Schéma fonctionnel de la mise au point d'un modèle économétrique

CHAPITRE I - DESCRIPTION DES INFORMATIONS

1.1 Un modèle de la vie économique et sociale

1.1.a Exemple

La comptabilité nationale est un modèle de la vie économique et sociale de Belgique qui a pour objectif le calcul du produit national brut (PNB). Le modèle utilisé permet de calculer le PNB selon 3 optiques différentes mais dont les résultats devraient être égaux.

Le PNB selon l'optique des revenus est =

- Σ des revenus des ménages
- + Σ des revenus des entreprises (= bénéfice net distribué aux ménages ou investi)
- + Σ des revenus de l'état (= impôts directs)

Revenu national
+ Amortissements
+ Impôts indirects
- Subventions

PNB₁

Le PNB selon l'optique des dépenses est =

- Σ des dépenses de consommation privée
- + Σ des dépenses de consommation publique
- + Σ des investissements
- + Σ des exportations
- Σ des importations

PNB₂

Le PNB selon l'optique des valeurs ajoutées est =

- Σ V.A. par secteur
- + Σ des impôts indirects
- subventions

PNB₃

$$\text{PNB}_1 = \text{PNB}_2 = \text{PNB}_3$$

Chaque facteur intervenant dans le calcul du PNB doit être mesuré, observé dans la vie économique et sociale réelle.

Il faut donc pour chaque rubrique se demander :

- que recouvre le concept énoncé
- quel est le niveau de détail voulu
- où et comment va-t-on recueillir les informations de base

Ex. : - les dépenses de consommation privée :

- ce sont les dépenses des ménages en vue d'acquérir des biens ou des services
 - on ne considère que les dépenses des ménages pris globalement
 - on se définit une liste, une nomenclature des biens et services consommés par les ménages
-
- les investissements publics par secteurs d'activité :
 - ce sont les dépenses de l'état au bénéfice des entreprises regroupées par secteurs et par activités
 - l'état est pris globalement
 - on se définit une nomenclature des entreprises regroupées par secteurs et par activités.

Pour la collecte des données, on devra percevoir les entreprises par la nomenclature définie sur les secteurs + activités et sur les biens produits. Le niveau de perception le plus fin de l'entreprise pour le modèle considéré dans l'exemple sera donc le bien produit.

Ex. de nomenclature des entreprises

industrie alimentaire

niveaux de perception du réel	}	- pain et boulangerie
		- pâtisserie et biscuits
		- biscotte, pâtes alimentaires, riz
		- poisson
		- lait
		- fromage
		- oeufs

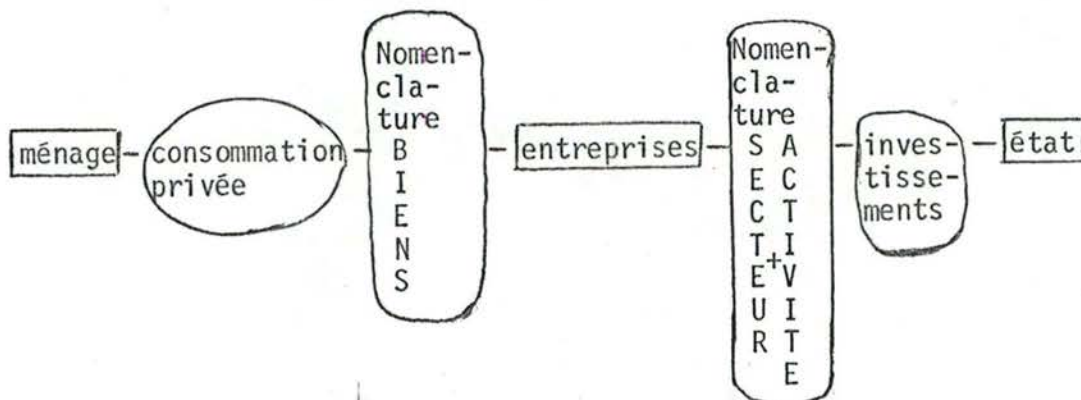
industrie extractive

⋮

Pour la construction des comptes de la consommation privée, on utilisera les observations réelles avec éventuellement des agrégations au niveau des biens (ex.: on peut avoir des comptes : "lait", "fromage" et "oeufs" et aussi un compte regroupant les 3 rubriques "lait + fromage + oeufs").

Pour la construction des comptes des investissements publics, on va agréger les informations réelles sur l'activité (ex.: le compte investissement dans l'industrie manufacturière alimentation, boisson et tabac).

Si l'on veut représenter graphiquement ces 2 comptes en fonction des agents économiques mis en présence par les transactions économiques étudiées, on aura le schéma suivant :



On remarque donc que l'agent économique "entreprises" est vu différemment suivant l'opération économique qu'il réalise "consommation privée" ou "investissements". L'analyse que l'on vient de faire pour deux rubriques de la comptabilité nationale, il faudra la faire pour chaque rubrique intervenant dans les différentes méthodes de calcul du produit national brut. Schématiquement, l'analyse de la comptabilité nationale devra comporter les étapes suivantes :

- identification des rubriques retenues : (consommation privée, consommation publique, investissements, revenu des ménages ...)
- pour chaque rubrique prise individuellement :
 - identifier la manière de collecter les données dans le monde réel, c'est-à-dire où et comment va-t-on recueillir les informations de base (enquête auprès des entreprises, étude de la comptabilité des entreprises...)
 - identifier le niveau de détail souhaité pour la grande rubrique, c'est-à-dire établissement de la nomenclature des agents économiques à la base de la rubrique considérée ;

- au niveau du modèle de la comptabilité nationale, reprendre les différentes nomenclatures. Pour chaque agent économique (entreprise, ménage, état ...) retenir la nomenclature la plus détaillée ou construire une nomenclature appropriée pour la collecte des données en fonction de laquelle on pourra construire toutes les nomenclatures utilisées dans le modèle pour cet agent ;
- la collecte des données proprement dite ;
- la construction, le calcul des différents comptes, c'est-à-dire, le regroupement des observations réelles au moyen des nomenclatures utilisées pour le compte ;
- la constitution du fichier comptabilité nationale qui est constitué de l'ensemble des comptes construits.

1.1.b Définition de concepts

La vie économique et sociale est perçue selon un modèle qui la décrit comme un ensemble d'"opérations", ou "liens" entre des "objets économiques".

Dans l'exemple de la comptabilité nationale on aurait :

des opérations telles que :

- consommation privée
- consommation publique
- revenu des ménages
- revenu des entreprises
- revenu de l'état
- investissement public

des objets économiques tels que :

- ménages
- entreprises
- état
- reste du monde

1.1.b.1. Les objets économiques

- Les objets économiques d'un modèle sont des entités qui peuvent être perçues à 3 niveaux différents :

1. le niveau théorique et général de conception du modèle qui manipule des TYPES D'OBJETS ECONOMIQUES ou types d'entités. Un type d'objets économiques du modèle correspond à la notion d'agents économiques. Le modèle de la comptabilité nationale reconnaît les agents économiques suivants: l'état, les entreprises, les ménages et le reste du monde.
2. le niveau de réalisation du modèle détermine les nomenclatures utilisées pour chaque type d'objets économiques ; c'est la structure propre au modèle des agents économiques. Un agent économique perçu à l'aide d'une nomenclature définit une CLASSE d'OBJETS ECONOMIQUES ou classe d'entités. Un type d'objets économiques peut être à l'origine de plusieurs classes d'objets économiques au sein d'un même modèle.

Ex.: dans le modèle de la comptabilité nationale, le type d'objets économiques "entreprises" peut être perçu par une nomenclature des secteurs d'activité et par une nomenclature des biens.

3. le niveau de la collecte des données est le niveau de perception élémentaire des objets économiques retenus par le modèle. Une classe d'objets économiques est une agrégation d'objets économiques suivant une nomenclature donnée.

Ex.: dans le modèle de la comptabilité nationale, les entreprises doivent être perçues au niveau du bien qu'elles produisent pour la phase de collecte des données ; ce qui permettra par agrégation sur le critère

"bien" et sur le critère "secteur d'activité" de construire les classes d'objets économiques "entreprises nomenclature biens" et "entreprises nomenclature secteurs d'activité"

1.1.b.2. Les liens économiques

- Les liens ou opérations entre les objets économiques sont les relations du modèle. Les liens économiques du modèle étant définis par rapport aux objets économiques, on les percevra à 3 niveaux parallèles aux niveaux de perception des objets économiques :

 1. Le niveau de conception du modèle manipule les TYPES DE LIENS ou types de relations qui sont définis sur les types d'objets économiques qu'ils relient.
 2. Le niveau de réalisation du modèle manipule des CLASSES DE LIENS ou classes de relations définies sur des classes d'objets économiques.
 3. Le niveau de la collecte des données manipule des LIENS ou relations définies sur des objets économiques.

Un type de relations peut être entièrement caractérisé par un nom de lien économique et par les types d'objets économiques présents.

Ex.: de types de relations

- consommation privée (ménages, entreprises)
- revenu des ménages (ménages, entreprises)
- investissements publics (état, entreprises)

Une classe de relations sera essentiellement caractérisée par un nom de lien économique, par les classes d'objets économiques (donc, référence à une nomenclature par classe) qu'elle met en présence et par une quantification du lien.

Ex.: de classes de relations

consommation privée (ménages, entreprises (nomenclature biens), valeur)

revenu des ménages (ménages (nomenclature type de revenus) entreprises (nomenclature activité), valeur)

investissements publics (état, entreprises (nomenclature secteur d'activité), valeur)

(Les caractéristiques des classes de relations d'un modèle seront étudiées en détail dans la suite de l'exposé).

Une relation sera essentiellement caractérisée par un nom de lien, par les objets économiques qu'elle met en présence et par une quantification du lien.

Une "relation" du modèle peut être :

- une "relation" unaire, c'est-à-dire, une "relation" où il n'y a qu'un seul objet économique concerné.

ex.: la "relation" "population" d'un modèle est la quantification d'une propriété de l'objet économique "ménages" ;

- une "relation" binaire qui représente un flux entre 2 objets économiques.

Les "relations" binaires sont les plus fréquentes dans les modèles de la vie économique et sociale.

Ex.: consommation privée (ménages, entreprises)
revenu des ménages (ménages, entreprises)
investissements publics (état, entreprises)

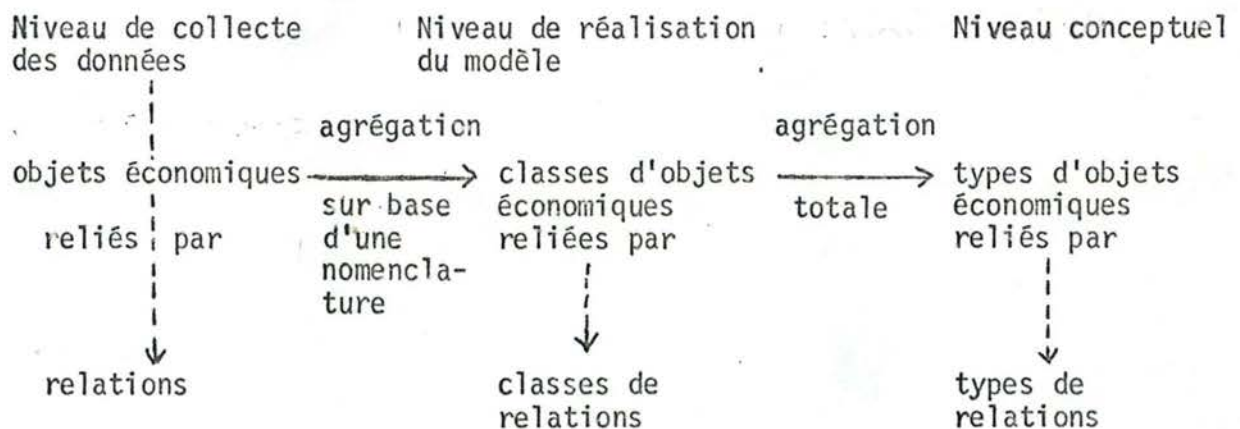
- une "relation" naire qui représente une relation entre n objets économiques.

Ex.: de "relation" ternaire :

importations (des biens, d'un pays, en provenance d'un autre pays)

Les propriétés des "relations" énoncées ci-dessus sont valables pour les types de relations, les classes de relations et les relations d'un modèle.

Les concepts de liens et d'objets économiques définis plus haut peuvent être schématisés comme suit :



1.1.b.3 Un modèle

- Un modèle de la vie économique et sociale ne retient en fait qu'un sous-ensemble des types de relations et des types d'objets économiques existant dans la vie économique et sociale réelle.

Les modèles de la vie économique et sociale se distinguent :

- par les types de relations et les types d'objets économiques retenus
- par les classes d'objets économiques et donc de relations qu'ils constituent

- L'ensemble des classes de relations créées par un modèle de la vie économique et sociale constitue ce que l'on appellera un fichier de données.

1.2. Etude détaillée des relations du modèle

Ce paragraphe va analyser les types de propriétés reconnus aux relations et ce, pour chaque niveau de perception défini au point 1.1.b.2. les liens économiques

1.2.a. Le niveau de conception du modèle

Cette étape a pour objectif de définir les types de relations à analyser entre des types d'objets économiques.

Un type de relation pourra être entièrement caractérisé par un nom de type de relation et par les types d'objets économiques concernés.

Un modèle de la vie économique et sociale pourra être représenté au niveau de la conception du modèle par un tableau de ce type :

type de relation	type d'objet 1	type d'objet 2	type d'objet n
consommation privée	ménages	entreprises		
revenu	ménages	entreprises		
investis. publics	état	entreprises		

1.2.b. Le niveau de réalisation du modèle

A ce niveau-ci, on doit :

- se constituer des nomenclatures pour les types d'objets économiques du modèle, c'est-à-dire, se constituer les classes d'objets économiques sur lesquelles sont définies les classes de liens du modèle;

- construire la valorisation des classes de liens pour une période déterminée.

Comme on l'a déjà défini plus haut, l'ensemble des classes de liens d'un modèle est décrit par un fichier de données. Pour la facilité de l'exposé et pour se conformer à la terminologie en vigueur au Bureau du Plan, on introduit ici la notion de "variable économique" ou encore "variable" :

- une variable économique est une occurrence dans une classe de relations
- une classe de variables est une classe de relations

Un type de variable est une relation au sens informatique du terme, c'est-à-dire, définie sur un ensemble de domaines ; chacun de ces domaines constitue un type de propriété reconnu au type de variable.

Les domaines de définition du type de variable sont :

1. Le nom de modèle
2. Le nom de classe de relations
3. De 1 à n noms de classe d'objets économiques

Le domaine des noms de classe d'objets économiques est constitué par les domaines suivants :

- nom de type d'objet économique
 - nom de nomenclature
 - niveau dans cette nomenclature
4. La valorisation de la variable, c'est-à-dire, un vecteur d'observations
 5. Les propriétés du vecteur des observations
 - la localisation temporelle, c'est-à-dire,
 1. l'année et la période dans l'année de la première observation du vecteur

2. l'année et la période dans l'année de la dernière observation
 3. la périodicité des observations (annuel, trimestriel, mensuel ...)
- un code de complétion du vecteur d'observations qui déterminera si le vecteur des observations est complet ou s'il manque des observations
 - la nature des observations, c'est-à-dire :
 - les observations sont exprimées en unités courantes (ex.: en francs courants, c'est-à-dire, dont le pouvoir d'achat n'est pas constant)
la variable est dite en valeur
 - les observations constituent un vecteur d'indice (par exemple un indice des prix)
la variable est dite en prix
 - les observations sont exprimées par rapport à une année de référence
la variable est dite en volume ou en qualité
(1 var. en volume = $\frac{\text{var. correspondante en valeur}}{\text{var. correspondante en prix}}$)
 - un millésime de référence pour les variables en quantité ou en prix
 - la localisation dimensionnelle des observations, c'est-à-dire :
 - la dimension des observations
(ex.: Monnaie, Réel)
 temps temps
 - les unités de la dimension
(ex.: FB)
 an
 - l'échelle des unités utilisées dans les observations (ex.: milliard, millier...)

- une formule de calcul des observations définie si la variable
 - est construite au Bureau du Plan en fonction de variables existantes
 - correspond à un niveau non élémentaire d'une nomenclature
 - la localisation géographique des observations.
6. Le libellé de la variable qui est un texte libre expliquant brièvement ce que représente la variable.
7. Les caractéristiques de MAJ de la variable, c'est-à-dire,
- le responsable de la variable au Bureau du Plan
 - la date de création de la variable
 - la date de dernière MAJ
 - l'origine bibliographique de la variable qui définit l'organisme qui a construit les observations de la variable
(INS, Bureau du Plan, OCDE, CEE)
 - un commentaire qui est un texte libre où l'on donnera, en général, les références exactes des publications où l'on a trouvé les observations de la variable
(ex.: de 53 à 65 : Bulletin statistique n° 21
de 66 à 75 : " " " " n° 40
de 75 à 70 : " " " " n° 42

Les domaines identifiants d'une variable ou d'une classe de variable .

L'identifiant d'une classe de variables dans un modèle sera constitué sur les domaines suivants :

- nom de classe de relations
- de 1 à n - nom de type d'objet économique
 - nom de nomenclature

- de 0 à $n-1$ niveaux dans les nomenclatures
- nature des observations

L'identifiant d'une variable décrite dans un modèle sera constitué sur les domaines suivants :

- nom de classe de relations
- de 1 à n - nom de type d'objet économique
 - nom de nomenclature
 - niveau dans la nomenclature
- nature des observations

Une même variable ou classe de variables pouvant être étudiée par plusieurs modèles différents, l'identifiant unique d'une variable ou classe de variables dans l'ensemble des modèles utilisés au Bureau du Plan sera constitué du nom du modèle et de l'identifiant de la variable ou classe de variables dans le modèle.

Synthèse des domaines de définition du type de variable :

- nom de modèle
- nom de classe de relations
- de 1 à n noms de classe d'objets économiques ; chacun étant composé des domaines :
 - nom de type d'objet économique
 - nom de nomenclature
 - niveau dans cette nomenclature
- vecteur des observations
- les propriétés du vecteur des observations
 - localisation temporelle, c'est-à-dire, année et période de la première observation
année et période de la dernière observation
périodicité
 - code de complétion

 - nature des observations
 - millésime de référence
 - localisation dimensionnelle, c'est-à-dire,
 - la dimension
 - les unités
 - les échelles des unités
 - localisation géographique
 - formule de calcul des observations
- libellé
- propriétés de MAJ :
 - date de création
 - date de dernière MAJ
 - responsable au Bureau du Plan
 - origine bibliographique
 - commentaire

1.2.c Le niveau de la collecte des données

La collecte des données d'un modèle de la vie économique et sociale peut se faire :

- soit par enquête directe dans la vie économique et sociale réelle ;
- soit par l'utilisation d'informations déjà perçues par un modèle de la vie économique et sociale (utilisation de variables économiques existantes pour en créer de nouvelles)

Au Bureau du Plan, on ne procède en général pas par enquêtes directes dans la vie économique, on utilise donc les variables économiques créées par d'autres organismes (tels que INS, OCDE, CEE...) Ce fait est important à souligner dans la mesure où il conditionne les classes de variables que l'on peut définir dans les modèles construits au Bureau du Plan.

CHAPITRE II - ARCHITECTURE DES INFORMATIONS

Les domaines de définition du type de variables économiques définis au point 1.2.b. permettent de définir toutes les classes de variables utilisées au Bureau du Plan. On peut donc dire que le Bureau du Plan ne manipule qu'un seul type de variables.

Dans ce chapitre, on va définir les 2 types de modèles de la vie économique et sociale utilisés au Bureau du Plan : les modèles descriptifs de la vie économique et sociale et les modèles économétriques. Une étude plus détaillée de la phase de construction d'un modèle économétrique permettra de découvrir les types de fichiers du Bureau du Plan qui sont :

- les fichiers de données statistiques ;
- les fichiers d'équations économétriques.

L'étude de la phase de construction des données statistiques d'un modèle économétrique permettra de découvrir les types de fichiers de données statistiques du Bureau du Plan qui sont :

- les fichiers de base
- le fichier projet global
- le fichier projet utilisateur

2.1 Les modèles descriptifs de la vie économique et sociale

Un modèle de la vie économique et sociale de type descriptif est un modèle dont le seul objectif est la création d'un ensemble de variables économiques représentant un aspect de la vie économique et sociale. Un modèle descriptif est souvent un modèle comptable. Le modèle utilisé pour créer le fichier de la comptabilité nationale est un modèle des-

criptif de la vie économique et sociale. Les grandes étapes de construction d'un modèle descriptif de la vie économique et sociale sont les suivantes :

- * conception du modèle, c'est-à-dire :
 - quels sont les types de liens économiques
 - quels sont les types d'objets économiques concernés
- * construction proprement dite du modèle :
 - construction des classes d'objets économiques (établissement de nomenclature)
 - construction des observations :
 - soit par enquêtes directes auprès des objets économiques réels
 - soit par utilisation de variables existantes
 - construction des variables
- * création du fichier proprement dit.

2.2 Les modèles économétriques

Un modèle économétrique est un modèle qui a pour but d'expliquer la vie économique et sociale par un ensemble d'interactions entre les variables économiques ; un modèle économétrique sera donc caractérisé par :

- l'ensemble des variables économiques qu'il étudie ;
- l'ensemble des équations qui expliquent les interactions entre les variables économiques. Cet ensemble d'équations doit permettre de prévoir l'évolution des variables économiques.

2.2.1 Les grandes phases d'élaboration d'un modèle économétrique

Ce sont :

- * la conception du modèle qui définit :
 - les types de relations économiques étudiées par le modèle et donc les types d'objets économiques et les types de liens

- les types d'interactions entre les types de relations économiques

Ex.: la conception d'un modèle économétrique des exportations définira :

- les exportations = F (des importations) où F est une certaine fonction
- les exportations sont des relations entre les types d'objets économiques "biens exportés", "pays exportateurs", "pays importateurs"
- les importations sont des relations entre les types d'objets économiques "biens importés", "pays importateurs", "pays exportateurs".

* construction des données et des équations du modèle, c'est-à-dire :

- création des variables du modèle
 - définition des nomenclatures
 - construction des équations de définition de variable
 - calcul des observations
 - création des variables
- création des équations économétriques qui traduisent les interactions entre les variables (ex.: d'équation économétrique de la consommation privée: $\text{Volume consommation privée} = a + b \text{ volume de revenu des ménages}$)

* estimation des équations économétriques du modèle: étant donné les observations de la consommation privée, trouver les coefficients a et b de l'équation économétrique qui permettent d'ajuster au mieux la consommation calculée à la consommation observée

- * structuration des équations économétriques en modèle simulable et création des données pour la simulation.
- * simulations statiques et dynamiques du modèle sur le passé qui permettent de tester la validité du modèle économétrique : les résultats de la simulation doivent être proches des observations réelles.

Pour expliquer la différence entre une simulation statique et une simulation dynamique, on peut prendre l'exemple d'un type d'équations économétriques de la consommation privée qui peut être de la forme

$$c_t = f (Y_{t-1}) \text{ où } c_t = \text{consommation privée au temps } t$$

f = une certaine fonction

$$Y_{t-1} = \text{revenu des ménages au temps } t-1$$

Pour calculer la consommation privée au temps t

- dans une simulation statique on prendra l'observation réelle du revenu au temps $t-1$
- alors que dans une simulation dynamique, on prendra le résultat du calcul du revenu au temps $t-1$

- * Préviation des exogènes et des coefficients. Une variable ou classe de variables est dite exogène si elle n'est pas expliquée par le modèle, c'est-à-dire, qu'il n'existe pas d'équation économétrique qui permet de prévoir son évolution.
- * Etablissement d'une projection de référence, c'est-à-dire, une simulation du modèle pour laquelle on suppose que le futur est entièrement déterminé par le passé.

* Etude de projections alternatives, c'est-à-dire, étude des répercussions d'une politique économique sur l'évolution des variables expliquées par le modèle.

Une politique économique détermine une évolution possible pour une ou plusieurs variables exogènes.

Ex.: quelles sont les répercussions sur notre économie belge d'une dévaluation du F.B. ?

* étude de scénarios

* projections prévisionnelles

2.2.2 Etude plus détaillée de la phase de construction d'un modèle économétrique

La complexité d'un modèle économétrique est fonction :

- du nombre de classes de variables qu'il étudie
- du nombre d'occurrences de variables dans les classes de variables étudiées
- de la connaissance que l'on a des interactions entre les variables.

Par exemple, le modèle économétrique du produit national brut sera beaucoup plus complexe que celui de la consommation privée du fait que le modèle du produit national brut devra contenir entre autre un modèle de la consommation privée.

Lorsque l'on doit créer un modèle complexe du type de celui du produit national brut, on découpera d'abord le modèle complexe en sous-modèles (du type de celui de la consommation privée). Ensuite, il faut relier les sous-modèles entre eux pour obtenir le modèle complet. Un sous-modèle peut :

- soit être considéré comme élémentaire et on peut alors passer à la phase de construction des données et des équations économétriques puis à l'estimation des coefficients des équations et enfin aux simulations du sous-modèle ;
- soit être considéré à son tour comme un modèle complexe décomposable.

La conception d'un modèle économétrique complexe se fait donc par une méthode top-down : de l'objectif initial, on déduit des objectifs intermédiaires qui, à leur tour, peuvent induire des objectifs soit intermédiaires, soit élémentaires.

Quant à la phase de réalisation du modèle, on utilise une méthode bottom-up, c'est-à-dire, que l'on construit d'abord les données et les équations des modèles élémentaires puis des modèles intermédiaires jusqu'au modèle complexe global.

La construction d'un modèle économétrique nécessite la construction :

- d'un fichier de données que l'on appellera dorénavant un fichier projet et qui contiendra toutes les données nécessaires au modèle économétrique. Les données seront les variables économiques que l'on a définies plus haut ;
- d'un fichier d'équations économétriques.

Tous les fichiers de variables économiques du Bureau du Plan qui ne sont pas définis en fonction des besoins en données d'un modèle économétrique seront appelés des fichiers de base.

2.2.2.A La construction du fichier de données statistiques d'un modèle économétrique

2.2.2.A.1 La construction des variables d'un modèle économétrique

Le fichier projet contient l'ensemble des données, c'est-à-dire, l'ensemble des variables économiques, nécessaires au modèle économétrique. Les variables économiques définies au sein d'un fichier projet sont du même type que celles expliquées précédemment et elles sont évidemment définies sur les mêmes domaines. Au départ, un fichier projet est vide, l'introduction de variables dans le fichier est conditionnée par les besoins en données d'un modèle économétrique élémentaire.

Exemple d'un modèle économétrique élémentaire de la consommation privée en volume et des besoins en données correspondants :

$$QC0_t = \underline{C}_1 \frac{YDH0_t}{PC0_t} + \underline{C}_2 QC0_{t-1}$$

- où
- QC0 = volume de la consommation privée
 - YDH0 = revenu disponible en valeur
 - PC0 = déflateur de la consommation privée
 - QC0_{t-1} = volume de la consommation privée au temps t-1
 - \underline{C}_1 et \underline{C}_2 sont les coefficients du modèle à estimer.

Les besoins en données de ce modèle sont les variables ou classes de variables QC0, YDH0, PC0. Il faut donc se poser les questions suivantes :

- où sont définies les variables ou classes de variables volume de la consommation privée, revenu disponible en valeur et déflateur de la consommation privée ;
- ces variables ou classes de variables peuvent être définies dans plusieurs fichiers différents et donc résulter d'analyses différentes de la vie économique et sociale ou peuvent aussi ne pas exister comme telles dans les fichiers de données existants .

Dans les 2 cas, il faut prendre une décision :

- on va chercher la variable ou classe de variables dans un et un seul fichier → choix du modèle à la base de la création de la variable ou de la classe de variables.
- on construit la variable ou classe de variables non encore existante en fonction de variables ou de classes de variables définies dans les fichiers de données existants
→ on crée une équation de définition, de construction de variable ou de classe de variables.

Une préoccupation constante lors de l'introduction de variables ou classes de variables dans un fichier projet est que le fichier doit être un ensemble cohérent de données. La cohérence des données doit être réalisée à 3 niveaux :

- au niveau de chaque variable, c'est-à-dire :
 1. que les propriétés de cette variable doivent appartenir aux différents domaines de définition des variables
 2. que la variable doit représenter une réalité économique
- au niveau de chaque classe de variables: une classe de variables utilise une et une seule nomenclature par classe d'objets économiques concernés
- au niveau du fichier, c'est-à-dire :
 1. que le fichier représente effectivement un aspect cohérent de la vie économique et sociale
 2. que les résultats de simulation du modèle économétrique pour lequel il est construit sont acceptables

2.2.2.A.2 Le fichier projet global et le fichier projet utilisateurs

L'élaboration d'un modèle économétrique complexe peut être faite par plusieurs utilisateurs simultanément, chacun d'eux étant responsable d'un modèle économétrique déduit du modèle global.

Exemple : le modèle complexe du PNB peut se décomposer en 2 sous-modèles relativement autonomes :

- le modèle du PNB optique revenu et le modèle du PNB optique dépense.

Chaque sous-modèle est lui-même complexe et pourrait à son tour être décomposé en sous-modèles.

Chaque sous-modèle a ses besoins en données, besoins, qui peuvent être propres au sous-modèle en question ou qui peuvent être partagés avec d'autres sous-modèles.

Pour permettre à chaque utilisateur de travailler en toute sécurité, c'est-à-dire, pour ne pas créer d'interférences involontaires entre 2 sous-modèles, on propose de créer un fichier projet par utilisateur qui devra répondre à tous les besoins en données d'un utilisateur à un moment donné pour l'élaboration d'un sous-modèle défini.

Les fichiers projet utilisateur répondent aux besoins en données de modèles économétriques en construction, c'est-à-dire, que ce sont des fichiers en MAJ constantes, MAJ pouvant être l'introduction, la suppression, la modification de variables ou classes de variables. Ce sont des fichiers de travail. Le fichier projet global quant à lui représente un niveau d'accord entre les différents utilisateurs, il contient l'ensemble des variables jugées à un moment donné comme définitives pour le modèle global. Lorsqu'un sous-modèle est jugé satisfaisant, c'est-à-dire, que les résultats des simulations sont jugés bons, l'utilisateur peut demander au chef de projet d'introduire son fichier projet utilisateur dans le fichier global. Le fichier projet global résulte donc de l'intégration de fichiers projet utilisateur dans le fichier projet global existant. Cette intégration entraînera sans doute des conflits au niveau de la définition de variables ; conflits qui devront être réglés par le chef de projet et qui peuvent être :

- au même identifiant de variable dans le fichier projet utilisateur et dans le fichier projet global correspondent en fait des variables différentes, ce qui peut se traduire par exemple, par des observations différentes ou des équations de définition différentes.

2.2.2.B La construction du fichier des équations du modèle économétrique

2.2.2.B.1 Les étapes de construction des équations économétriques

Les étapes de construction des équations économétriques peuvent suivre très exactement les étapes de construction des variables économiques.

* conception des équations économétriques

Dans cette étape on définit les types d'équations du modèle économétrique. Un type d'équations économétriques sera écrit en fonction de types de relations économiques sur base de la théorie économétrique.

Ex.: Exportations (bien, pays exportateur, pays importateur)
= fonction (Importations (bien, pays importateur, pays exportateur))

* réalisation des équations économétriques

A ce niveau-ci, on construit les classes d'équations économétriques qui seront écrites en fonction :

- d'identifiants de classes de variables
- d'opérateurs
- de coefficients

Ex.: $LQEXPORT$ (Biens, Belgique, pays importateur)
 = $C1 \times L$ $QIMPORT$ (Biens, Belgique, pays exportateur)
 + $C2 \times L$ (P EXPORT (Biens, Belgique, monde)
 $\frac{\cdot}{\cdot}$ P EXPORT (Biens, monde, monde))

où

- les opérateurs de la classe d'équations économétriques sont L , $=$, X , $+$, $\frac{\cdot}{\cdot}$

- les coefficients C1 et C2
- les identifiants de classes de variables :
 - QEXPORT (biens, Belgique, pays import) est l'identifiant de la classe de variable exportation en quantité de biens de la Belgique en provenance de pays importateurs; "Biens" et "pays importateurs" sont les noms des nomenclature des types d'objets économiques "biens" et "pays" utilisées pour la construction de la classe de variables Exportation en quantité du fichier projet.
Belgique est une occurrence dans la nomenclature "pays exportateur" définie pour la classe de variables "exportation en quantité du fichier projet"
 - QIMORT (biens, Belgique, pays exportateur)
 - P EXPORT (biens, Belgique, monde)
 - P EXPORT (biens, monde, monde)

* construction proprement dite des équations économétriques

Les équations économétriques d'un modèle seront écrites en fonction :

- d'identifiants de variables définies par le fichier projet
- d'opérateurs
- de coefficients

Une équation économétrique est une occurrence dans une classe d'équations économétriques

Ex.: la classe d'équations économétriques définie plus haut pourra contenir les équations économétriques suivantes :

$$\begin{aligned} 1) \quad & \underline{LQ} \text{ EXPORT (biens, Belgique, Pays-Bas)} = \\ & \underline{C1} \times \underline{L} \text{ Q IMPORT (biens, Belgique, Pays-Bas)} + \\ & \underline{C2} \times \underline{L} \text{ (PEXPOR (biens, Belgique, monde)} \\ & \underline{\quad} \text{PEXPOR (biens, Belgique, monde))} \end{aligned}$$

où biens, Belgique, Pays-Bas, monde sont des occurrences dans les nomenclatures

$$2) \underline{LQ} \text{ EXPORT (service, Belgique, Pays-Bas)} = \underline{C1} \times \underline{L} \text{ Q IMPORT (service, Belgique, Pays-Bas)} + \underline{C2} \times \underline{L} (\underline{P} \text{ EXPORT (service, Belgique, monde)} - \underline{P} \text{ EXPORT (service, monde, monde)})$$

2.2.2.B.2 Le fichier des équations d'un modèle économétrique

Le fichier des équations économétriques est l'ensemble des classes d'équations économétriques construites pour un modèle économétrique.

L'identifiant d'une classe d'équations d'économétriques sera composé :

- de l'identifiant de la classe de variables qu'elle explique (membre de gauche de l'équation)
- d'un code Equation ; ce qui permet de construire plusieurs classes d'équations pour expliquer une même classe de variables.

L'identifiant d'une équation économétrique sera composé :

- de l'identifiant de la variable expliquée
- du code équation.

2.3 Synthèse

2.3.1 Les types de fichiers de données statistiques du Bureau du Plan et leurs principales caractéristiques :

2.3.1.1 Un fichier projet utilisateur

- répond aux besoins en données d'un utilisateur particulier pour la mise au point d'un sous-modèle économétrique

- fichier de travail temporaire sous la responsabilité d'un utilisateur. C'est dans un fichier projet utilisateur que l'on crée les variables économiques propres à un modèle économétrique
- tous les types de MAJ sont permis sans restriction (insertion, modification, retrait de variables et classes de variables)
- le taux d'utilisation du fichier sera voisin de 1. On définit le taux d'utilisation d'un fichier comme étant le rapport entre le nombre d'articles utilisés au cours d'une session de travail et le nombre total d'articles du fichier.

2.3.1.2 Un fichier projet global

- répond aux besoins en données d'un modèle économétrique
- fichier de référence pour l'ensemble des utilisateurs d'un projet sous la responsabilité d'un chef de projet
- se constitue par l'intégration de fichiers projets utilisateurs
- les MAJ sont sous la responsabilité du chef de projet et sont de 2 types :
 - intégration d'un fichier projet utilisateur dans le fichier projet global, ce qui entraînera probablement des conflits au niveau de la définition des variables ; conflits qui seront réglés par le chef de projet et qui peuvent nécessiter la révision du fichier des équations économétriques ;
 - l'actualisation des observations des variables du fichier projet suite aux MAJ survenues dans les fichiers de base
- 2 types d'utilisation :
 - en consultation (lecture) accessible à tous sans restriction
 - en MAJ (écriture) accès réservé au chef de projet.

2.3.1.3 Un fichier de base

- c'est un fichier qui a une existence non directement liée à l'existence d'un modèle économétrique ;
- représente l'ensemble des variables créées par un modèle descriptif de la vie économique et sociale ;
- l'ensemble des fichiers de base du Bureau du Plan constitue la source commune d'informations de tous les utilisateurs du Bureau du Plan quel que soit le projet auquel ils sont affectés ;
- est généralement construit hors du Bureau du Plan ;
- MAJ théoriquement annuelles et faisant suite à la parution de nouvelles observations pour les variables et à une décision du responsable du fichier au Bureau du Plan ;
- fichier stable quant au nombre de variables et à leur définition ;
- fichier consultable :
 - par tous en lecture
 - pour un seul en écriture

2.3.2 Un modèle économétrique

Un modèle économétrique est un ensemble de classes d'équations économétriques construites sur un ensemble de classes de variables.

Une classe d'équations économétriques définit un ensemble d'équations économétriques.

Une classe de variables définit un ensemble de variables.

L'estimation d'un modèle économétrique, c'est l'estimation d'un ensemble de classes d'équations économétriques sur base d'un ensemble de classes de variables.

La simulation d'un modèle économétrique, c'est la simulation d'un ensemble de classes d'équations économétriques sur base d'un ensemble de classes de variables.

CHAPITRE III - Structuration des traitements

La structuration des traitements se fera au moyen de deux concepts :

- les phases de traitement qui sont constituées d'un ensemble de fonctions ayant un objectif déterminé ; on remarque qu'en général, une phase de traitement modifiera des fichiers (de données ou d'équations économétriques) du Bureau du Plan ;
- les fonctions qui sont des opérations d'une logique "simple" travaillant sur les propriétés d'une variable ou d'une classe de variables.

3.1 Les phases de traitement

Les phases de traitement peuvent être de 2 types :

- les phases de traitement concernant la création et la mise au point d'un modèle économétrique ; elles seront essentiellement interactives ;
- les autres phases de traitement

3.1.1 Les phases de traitement concernant la mise au point d'un modèle économétrique sont :

- l'introduction de variables dans le fichier projet utilisateur
- l'introduction d'équations économétriques dans le fichier des équations économétriques
- l'estimation du fichier des équations économétriques
- la structuration des équations économétriques en modèle simulable
- la création des données de simulation

- la simulation du modèle
- les tests des résultats de simulation
- la constitution du fichier projet par intégration d'un fichier projet utilisateur

3.1.2 Les autres phases de traitement

- la création d'un fichier de base
- les contrôles de cohérence d'un fichier
- l'édition d'un fichier de données
 - édition complète du fichier ou sélection de variables
 - édition de toutes les propriétés des variables ou sélection de propriétés
- la MAJ d'un fichier de base suite à la parution d'un nouveau bulletin de statistiques
- l'actualisation d'un fichier projet suite aux MAJ des fichiers de base
- la création d'un back-up d'un fichier de données
- le chargement d'un fichier à partir d'un back-up

3.2. Les fonctions

La découpe des traitements en fonctions se fera en 2 temps :

- une description plus détaillée des phases de traitement dégagées au point 3.1 ;
- une récapitulation des fonctions mises en évidence par cette description détaillée

3.2.1 Description des fonctions par phase de traitement

3.2.1.1 Les phases de traitement concernant la mise au point d'un modèle économétrique

- a. L'introduction de variables ou classes de variables dans le fichier projet utilisateur.

Les variables ou classes de variables d'un fichier projet utilisateur sont définies sur base de variables ou classes de variables existantes dans l'ensemble des fichiers de données.

Les variables ou classes de variables que l'on veut insérer dans le fichier projet utilisateur peuvent :

- * exister comme telles dans l'ensemble des fichiers de données existants: elles sont déjà définies dans un fichier de base ou dans le fichier projet global.

Pour cela, on doit créer une fonction :

- de sélection de variables sur base d'un identifiant de variable ou de classe de variables
- d'écriture de variables ou classes de variables dans le fichier projet utilisateur

- * ne pas exister comme telles dans les fichiers existants, il faudra donc créer une variable ou une classe de variables
Que signifie créer une variable au Bureau du Plan ?

C'est affecter une valeur particulière à chaque domaine de définition des variables. Les valeurs affectées à certains de ces domaines ne peuvent être créées que par l'utilisateur (ex.: le libellé d'une variable) ; d'autres par contre pourront être générées automatiquement.

Les fonctions d'aide à la génération de valeurs pour les domaines de définition des variables sont les suivantes :

1. Fonction de génération d'équations de définition de variables. Pour cela, l'utilisateur doit fournir :
 - * Une équation génératrice écrite en fonction d'identifiants de classes de variables
 - * Les noms des nomenclatures à partir desquelles on pourra générer les identifiants des variables.

Exemple : en entrée :

* équation génératrice

$$QC[mén:M, bien:F] = VC[mén:M, bien:F] \div PC[mén:M, bien:F]$$

* les nomenclatures

$$M = 1, 2$$

$$F = 1, 2, 3,$$

où Q, V, P, = nature des classes de variables

C = nom de relation = consommation privée

mén. = objet économique 1 = les ménages

bien = objet économique 2 = les biens

M = une nomenclature des ménages

F = une nomenclature des biens

en sortie on devra générer :

$$\begin{aligned}
 QC[mén: M1, bien:F1] &= VC[mén:M1, bien:F1] \div PC[mén:M1, bien:F1] \\
 QC[mén: M1, bien:F2] &= VC[mén:M1, bien:F2] \div PC[mén:M1, bien:F2] \\
 QC[mén: M1, bien:F3] &= VC[mén:M1, bien:F3] \div PC[mén:M1, bien:F3] \\
 QC[mén: M2, bien:F1] &= VC[mén:M2, bien:F1] \div PC[mén:M2, bien:F1] \\
 QC[mén: M2, bien:F2] &= VC[mén:M2, bien:F2] \div PC[mén:M2, bien:F2] \\
 QC[mén: M2, bien:F3] &= VC[mén:M2, bien:F3] \div PC[mén:M2, bien:F3]
 \end{aligned}$$

le traitement :

- effectuer le produit cartésien des nomenclatures spécifiées : construction des n-uples (1,1), (1,2), (1,3), (2,1) !...
- générer autant d'équations qu'il y a de n-uples dans le produit cartésien

$$\rightarrow \left\{ \begin{array}{l} QC[mén:M, bien:F] = VC[mén:M, bien:F] \div PC \\ \quad \quad \quad [mén:M, bien:F] \\ \vdots \end{array} \right.$$

où n = nombre de n-uples dans le produit cartésien

- pour chaque n-uple du produit cartésien donc dans chaque équation générée, compléter les noms de nomenclatures par l'occurrence correspondante dans le n-uple

→ sur l'exemple :

on entre : (1,1) ,

$QC[mén:M, bien:F] = VC[mén:M, bien:F] \div PC[mén:M, bien:F]$

en sortie :

$QC[mén:M1, bien:F1] = VC[mén:M1, bien:F1] \div$

$PC[mén:M1, bien:F1]$

Remarque : pour que $QC [mén:M1, bien:F1] =$

$VC[mén:M1, bien:F1] \div PC[mén:M1, bien:F1]$ soit une équation de définition de variable, il faut que $VC[mén:M1, bien:F1]$ et $PC[mén:M1, bien:F1]$ soient des identifiants de variables existantes dans l'ensemble des fichiers de données statistiques du Bureau du Plan.

2. Fonction de calcul des observations et de la location temporelle d'une variable étant donné son équation de définition

Une équation de définition de variable est écrite en fonction d'identifiants de variables. Pour cette fonction, il faut :

- extraire du littéral de l'équation de définition de la variable les identifiants des variables apparaissant dans le membre de droite ;
- lire les variables dont l'identifiant se trouve dans la liste construite ci-dessus ;

- homogénéiser les observations des variables lues, c'est-à-dire, qu'elles doivent toutes avoir
 - la même localisation temporelle
 - la même périodicité
 - la même année et période de début
 - la même année et période de fin
 - le même ordre de grandeur par rapport aux mêmes unités
- effectuer les opérations décrites dans l'équation

3. Le code de complétion prendra la valeur incomplète si un des opérandes est incomplet.

* On peut aussi vouloir modifier la définition d'une variable, c'est-à-dire, modifier la valeur d'un ou plusieurs domaines de définition de la variable.

Pour cela il faut :

- une fonction d'accès aux variables du fichier projet
- permettre la modification des valeurs attribuées aux domaines de définition des variables
- une fonction d'écriture dans le fichier projet utilisateur d'une variable existante (replace, rewrite)
- mais aussi une fonction de destruction de variable dans le fichier utilisateur.

b. La construction du fichier des équations économétriques

Les besoins en données de cette phase de traitement déclenchent la phase de construction de variables du fichier projet utilisateur.

Tout comme pour les équations de définition des variables, on peut écrire des équations génératrices d'équations économétriques. Une équation génératrice d'équations économétriques sera écrite en fonction d'identifiants de classes de variables et générera une classe d'équations

économétriques. Une équation génératrice peut être définie sur des classes de variables complètes ou sur des sous-classes de variables. Les classes de variables complètes sont celles définies dans le fichier projet, on travaille sur des nomenclatures complètes.

Les sous-classes de variables sont toujours définies dans le fichier projet mais on travaille sur des nomenclatures incomplètes (on spécifie les occurrences intéressantes dans au moins une des nomenclatures).

Ex.: on peut imaginer des équations génératrices différentes pour les sous-classes de variables : exportations de la Belgique vers les pays de la C.E.E. et exportations de la Belgique vers les autres pays ;

On peut aussi vouloir générer les équations économétriques d'exportations de Belgique concernant uniquement certains types de biens mais pas tous.

Exemples d'équations génératrices d'équations économétriques

Ex.:1. \underline{L} Q EXPORT (biens:B, Pays exp: Belgique, pays imp: Pays)

$= \underline{C1} \times \underline{L}$ Q IMPORT (biens:B, Pays imp:Belgique, pays exp: Pays)

$+ \underline{C2} \times \underline{L}$ (P EXPORT (biens:B, Pays exp:Belgique, Pays imp: monde) \div

P EXPORT (biens:B, Pays exp:monde, Pays imp.: monde))

où \underline{L} = logarithme

$\underline{C1}$ et $\underline{C2}$ sont les coefficients de l'équation à estimer

Q EXPORT (biens:B, Pays exp.:Belgique, Pays imp: Pays) est l'identifiant de la classe de variables "exportations en quantité de la Belgique"; les nomenclatures "B" et "Pays" sont implicites, ce sont donc celles définies dans le fichier projet utilisateur alors que pour l'objet économique "Pays exp" on spécifie une occurrence dans la nomenclature qui peut être la seule définie ou pas.

P EXPORT (biens:B, Pays exp:monde, Pays imp:monde) est l'identifiant de la classe de variables "prix des exportations de biens sur le marché mondial)

Ex.: 2. \underline{L} Q EXPORT (biens:B1, Pays exp:Belgique, pays imp: Pays)
= $\underline{C1}$ x \underline{L} Q IMPORT (biens:B1, Pays imp:Belgique, pays exp: Pays)
+ $\underline{C2}$ x \underline{L} (P EXPORT (biens:B1, Pays exp:Belgique, pays imp: monde) \div P EXPORT (biens:B1, pays exp: monde, pays imp: monde))

où B1 = 'revenu net des facteurs, biens + service', c'est-à-dire, une nomenclature définie par l'utilisateur

La construction du fichier des équations économétriques pourra se faire en 2 étapes :

- la génération des équations économétriques à partir d'une équation génératrice ;
- l'insertion des équations générées dans le fichier des équations économétriques.

b.1. La génération d'équations économétriques

entrée : - l'équation génératrice
- les nomenclatures utilisées pour les classes de variables référencées dans l'équation qui peuvent être :
- soit implicites, ce sont alors celles définies lors de la construction du fichier projet,
- soit explicites, elles sont alors données par l'utilisateur sous la forme d'une (de) liste(s) d'occurrences dans une (plusieurs) nomenclature(s) ;

sortie : - les équations générées écrites en fonction d'identifiants de variables.

traitement : - rechercher les nomenclatures implicites
- effectuer le produit cartésien des nomenclatures référencées dans l'équation génératrice
- pour chaque n-uple du produit cartésien, construire une équation qui sera l'équation génératrice elle-même mais où les opérandes seront complétés par les occurrences dans les nomenclatures (les éléments du n-uple)

b.2. L'insertion des équations générées dans le fichier des équations économétriques.

Une équation économétrique générée est écrite en fonction d'identifiants de variables.

Pour la même variable expliquée, (opérande de gauche d'une équation économétrique), on peut être amené à construire plusieurs équations économétriques; l'identifiant d'une équation économétrique sera donc composé :

- d'un code équation
- de l'identifiant de la variable expliquée.

L'insertion d'équations dans le fichier d'équations économétriques devra vérifier que chaque identifiant d'équation est unique.

c. L'estimation du fichier des équations économétriques.

Cette phase de traitement travaille sur le fichier des équations économétriques et sur le fichier projet utilisateur.

On peut vouloir estimer une classe d'équations économétriques, une équation économétrique, une liste de classes d'équations économétriques ou une liste d'équations économétriques.

en entrée : 1. liste d'identifiants de classes d'équations
économétriques et/ou d'équations économétriques
2. le fichier des équations économétriques
3. le fichier projet utilisateur

en sortie : les équations économétriques estimées

traitement : - sélectionner dans le fichier des équations économétriques
les équations identifiées par la liste
d'identifiants donnée
- pour chaque équation sélectionnée, l'estimer,
c'est-à-dire,
extraire les noms des variables apparaissant
dans une équation économétrique
lire les variables dans le fichier projet
utilisateur dont l'identifiant figure dans la
liste créée au point précédent
sélectionner pour les variables lues, les ob-
servations et la localisation temporelle
exécuter un programme d'estimation.

d. La structuration des équations d'un modèle économétrique
en un modèle simulable c'est la création de sous-routines
FORTRAN contenant la traduction des équations économétri-
ques du modèle.

e. La création des données de simulation d'un modèle économétri-
que : c'est la création d'un fichier d'observations pour
les variables apparaissant dans les équations économétri-
ques du modèle.

f. La simulation du modèle, c'est l'exécution d'un software de simulation dont les entrées sont :

- les équations économétriques à simuler ;
- les variables qui interviennent dans les équations économétriques ;
- des paramètres de simulation.

Le résultat d'une simulation est un fichier de variables simulées, c'est-à-dire, un fichier d'observations calculées.

g. Les tests des résultats de simulation, c'est la comparaison des observations réelles et des observations simulées des variables économiques et l'analyse des différences.

h. L'intégration d'un fichier projet utilisateur dans le fichier projet global

Cette phase de traitement consiste :

- à vérifier que cette intégration n'introduit pas de double définition de variable dans le fichier global : à un identifiant de variable correspond une et une seule variable. Si à un identifiant de variable correspond deux variables différentes, c'est au chef de projet de décider s'il faut garder la définition du fichier projet global ou celle du fichier utilisateur et s'il y a lieu de réestimer les équations économétriques et de resimuler le modèle
- à procéder à la MAJ du fichier projet global proprement dite, c'est-à-dire :
 - insérer les nouvelles variables
 - remplacer les anciennes définitions des variables
 - éventuellement supprimer des variables.

3.2.1.2 Les autres phases de traitement

a. La création d'un fichier de base.

C'est le passage d'un modèle de la vie économique et sociale vu à l'extérieur du Bureau du Plan au modèle de la vie économique et sociale tel qu'il a été défini dans les chapitres précédents.

- C'est donc la création de variables à partir d'informations publiées, en général, à l'extérieur du Bureau du Plan et éventuellement la création de variables définies au Bureau du Plan (par ex.: création de variables récapitulatives pour les niveaux non élémentaires des nomenclatures)
- On peut aussi devoir créer un fichier d'équations de contrôle des observations des variables d'un fichier de base. Ce fichier d'équations reprendra :
 - les équations de définition de variables exprimant un niveau non élémentaire d'une nomenclature ;
 - des équations de contrôle proprement dit, par ex.: pour le fichier de la comptabilité nationale, on créerait l'équation $\text{PNB optique revenu} = \text{PNB optique dépense}$.

b. Les contrôles de cohérence d'un fichier de données statistiques

b.1. Contrôle de l'unicité des identifiants des variables dans un fichier

b.2. Contrôle de l'unicité des variables dans un fichier ex.: des variables ayant des identifiants différents pourraient avoir la même équation de définition des observations et donc les mêmes observations dans un fichier projet global ou utilisateur

b.3. Contrôle des nomenclatures au sein d'une classe de variables économiques

- une classe de variables définit pour chaque objet économique en cause une et une seule nomenclature
- une nomenclature a une structure hiérarchisée, c'est-à-dire, découpée en niveaux suivant le même principe que la découpe en niveaux d'une donnée COBOL.

b.4. Contrôle des observations des variables :

- d'un fichier de base, c'est-à-dire, l'exécution du fichier des équations de contrôle du fichier de base correspondant
- d'un fichier projet global ou utilisateur. On peut aussi se construire un fichier d'équations de contrôle pour un fichier projet.

Un fichier projet (global ou utilisateur) sera cohérent si les résultats de simulation sont acceptables. L'exécution d'un fichier d'équations de contrôle devra, pour chaque équation :

- lire les variables dont l'identifiant figure dans l'équation
- effectuer les opérations arithmétiques de l'équation sur les observations des variables (c'est-à-dire, calcul de la variable à partir d'une équation)
- tester l'égalité des observations calculées de la variable et des observations de la variable dans le fichier de données (on peut accepter une certaine différence entre les observations des variables due à des écarts statistiques ou à des erreurs d'arrondis).

b.5. Contrôle de cohérence de chaque variable du fichier

Une variable est définie sur un ensemble de domaines, le contrôle de cohérence d'une variable devra vérifier que les valeurs attribuées aux domaines de définition de la variable sont acceptables.

Un domaine est l'ensemble des valeurs que peut prendre une caractéristique d'une variable ; un domaine peut être défini soit, par rapport à un fichier particulier, soit pour l'ensemble des fichiers.

Ex.: le domaine des identifiants de variables est défini par rapport à un fichier particulier;

le domaine des noms de modèles est défini par rapport à l'ensemble des fichiers.

Un domaine peut être défini avec plus ou moins de précision.

Ex.: le domaine des observations des variables est défini comme étant une liste de nombres ;

le domaine des origines bibliographiques des observations des variables est défini par une liste d'origines reconnues par le Bureau du Plan.

c. L'édition d'un fichier de données statistiques

Cette phase de traitement peut comporter 3 parties :

- la sélection des variables que l'on veut éditer
- la sélection des propriétés des variables que l'on veut éditer et éventuellement le calcul de propriétés statistiques des observations telles que la moyenne, l'écart type ...
- la mise en page des variables à éditer.

Elle peut en outre être soit interactive soit batch.

d. La MAJ d'un fichier de base.

Cette phase de traitement peut comporter :

- un changement de base pour les variables indice des prix et en quantité ;
- les autres MAJ des caractéristiques des variables du fichier qui peuvent être ajouter et modifier les observations ; ;
remplacer éventuellement d'autres caractéristiques.

Elle doit être suivie de la phase de contrôle de cohérence du fichier.

* Le changement de base pour les variables indice des prix et en quantité implique la modification des observations de ces variables et du millésime de référence

- La MAJ des observations des variables indice des prix peut être calculée en utilisant pour chaque observation d'une variable la formule suivante :

$$\text{observ}' = \frac{100}{A} \times \text{observ.}$$

où observ' = nouvelle observation

observ = observation dans l'ancienne base

A = indice des prix de la nouvelle année de référence exprimé dans l'ancienne base

- la MAJ des observations des variables en quantité peut être faite en recalculant la variable en quantité dont l'équation de définition est :

$$\text{var. en quantité} = \frac{\text{variable en valeur}}{\text{variable indice des prix}}$$

* les autres MAJ des caractéristiques

- MAJ des observations, c'est-à-dire, ajoute de nouvelles observations et modification d'observations existantes
 - MAJ de la date de dernière MAJ
 - MAJ du commentaire donnant les références des publications où sont parues les observations de la variable
 - MAJ de l'identifiant de la variable implique la redéfinition complète de la variable.
- L'ajoute ou le retrait de variables dans un fichier de base est très rare.

e. L'actualisation d'un fichier projet suite aux MAJ des fichiers de base nécessite :

- le calcul des observations des variables du fichier projet sur base des équations de définition des variables ;
- la MAJ de la date de dernière MAJ des variables.

3.2.2 Récapitulation des fonctions

3.2.2.a Sélection des variables:

entrée : une liste d'identifiants de variables ou de classes de variables
un fichier de données statistiques

sortie : la liste des variables sélectionnées

traitement : lecture du fichier de données statistiques sur base d'une clé identifiante ou non.

3.2.2.b Recherche des nomenclatures référencées dans un identifiant de classe de variables

entrée : un identifiant de classe de variables
un fichier de nomenclatures

sortie : la liste des nomenclatures référencées dans l'identifiant d'une classe de variables
une nomenclature sera présentée sous la forme de la liste des occurrences dans cette nomenclature

traitement : - reconnaître les noms de nomenclatures dans l'identifiant de la classe de variables
- lire les nomenclatures

3.2.2.c Constitution d'une liste des identifiants des variables apparaissant dans une équation :

entrée : une équation

sortie : liste d'identifiants de variables

traitement : extraire du littéral d'une équation, les identifiants des variables.

3.2.2.d Génération d'équations

- entrée : - équation génératrice écrite en fonction d'identifiants de classes de variables
- les nomenclatures référencées dans les identifiants des classes de variables de l'équation génératrice
- sortie : - l'ensemble des équations générées écrites en fonction d'identifiants de variables
- traitement : - générer le produit cartésien des nomenclatures données
- générer autant d'équations qu'il y a d'occurrences dans le produit cartésien
 - pour chaque équation générée, donc pour chaque n-uple du produit cartésien, compléter les identifiants des classes de variables par les occurrences dans les nomenclatures

3.2.2.e Calcul des observations d'une variable

- entrée : - l'équation de définition de la variable
- les observations et la localisation temporelle des variables utilisées dans le membre de droite de l'équation
- sortie : - les observations calculées de la variable
- traitement : - homogénéiser les observations des variables en entrée, c'est-à-dire, :
- même périodicité
 - même année et période de début
 - même année et période de fin
 - même ordre de grandeur par rapport aux mêmes unités
 - effectuer les opérations sur les observations homogénéisées

3.2.2.f Affectation de valeur aux domaines de définition de variable

entrée : - une variable incomplètement définie

sortie : - une variable complétée

traitement : - l'utilisateur donne des valeurs à certains domaines de définition des variables

3.2.2.g Remplacement d'une variable dans un fichier de données statistiques

entrée : * une variable modifiée

* un fichier de données statistiques

sortie : fichier modifié

traitement : c'est le REWRITE ou REPLACE d'une variable ;
c'est une fonction dont il faut limiter l'utilisation pour les fichiers de base et les fichiers projets globaux

3.2.2.h Insertion d'une variable dans un fichier de données statistiques

entrée : - la nouvelle variable

- le fichier de données statistiques

sortie : - fichier MAJ

traitement : écrire la variable donnée dans le fichier ;
c'est aussi une opération protégée pour les fichiers de base et les fichiers projet globaux

3.2.2.i Suppression d'une variable

entrée : un identifiant de variable
un fichier de données statistiques
traitement : supprimer la variable du fichier

3.2.2.j Création des entrées pour le programme d'estimation

entrée : - une ou plusieurs équations économétriques à estimer
- le fichier projet utilisateur
sortie : - les variables du fichier projet utilisées dans
les équations économétriques et présentées selon
le format voulu
traitement : - lecture du fichier projet sur base des
identifiants des variables présents dans
une équation économétrique
- pour les variables lues conserver les obser-
vations et la localisation temporelle

3.2.2.k Recherche d'équations économétriques dans le fichier des
équations

entrée : - une liste d'identifiants d'équations économétri-
ques ou de classes d'équations économétriques
- le fichier des équations économétriques
sortie : la liste des équations économétriques demandées
traitement : sélectionner dans le fichier des équations
économétriques la liste des équations demandées

3.2.2.1 Test de cohérence d'une nomenclature

entrée : * une classe de variables
* une nomenclature d'un objet économique intervenant dans la classe de relations

sortie : variable OK ou pas
nomenclature OK ou pas

traitement : - tester la découpe en niveaux de la nomenclature
- tester que chaque niveau non terminal d'une nomenclature est à la base d'une variable de contrôle, c'est-à-dire, pour laquelle les observations sont égales à 1 fonction des observations des niveaux immédiatement inférieurs

3.2.2.m Test de cohérence d'une variable

Pour cela, il faut créer des fonctions qui vont tester l'appartenance d'une valeur de caractéristique au domaine de définition de cette caractéristique.

Ex.: - les observations de la variable sont numériques
- l'origine bibliographique est >0 et <13

3.2.2.n Test de cohérence d'un fichier de données statistiques

entrée : - fichier d'équations de contrôle
- fichier de données à contrôler

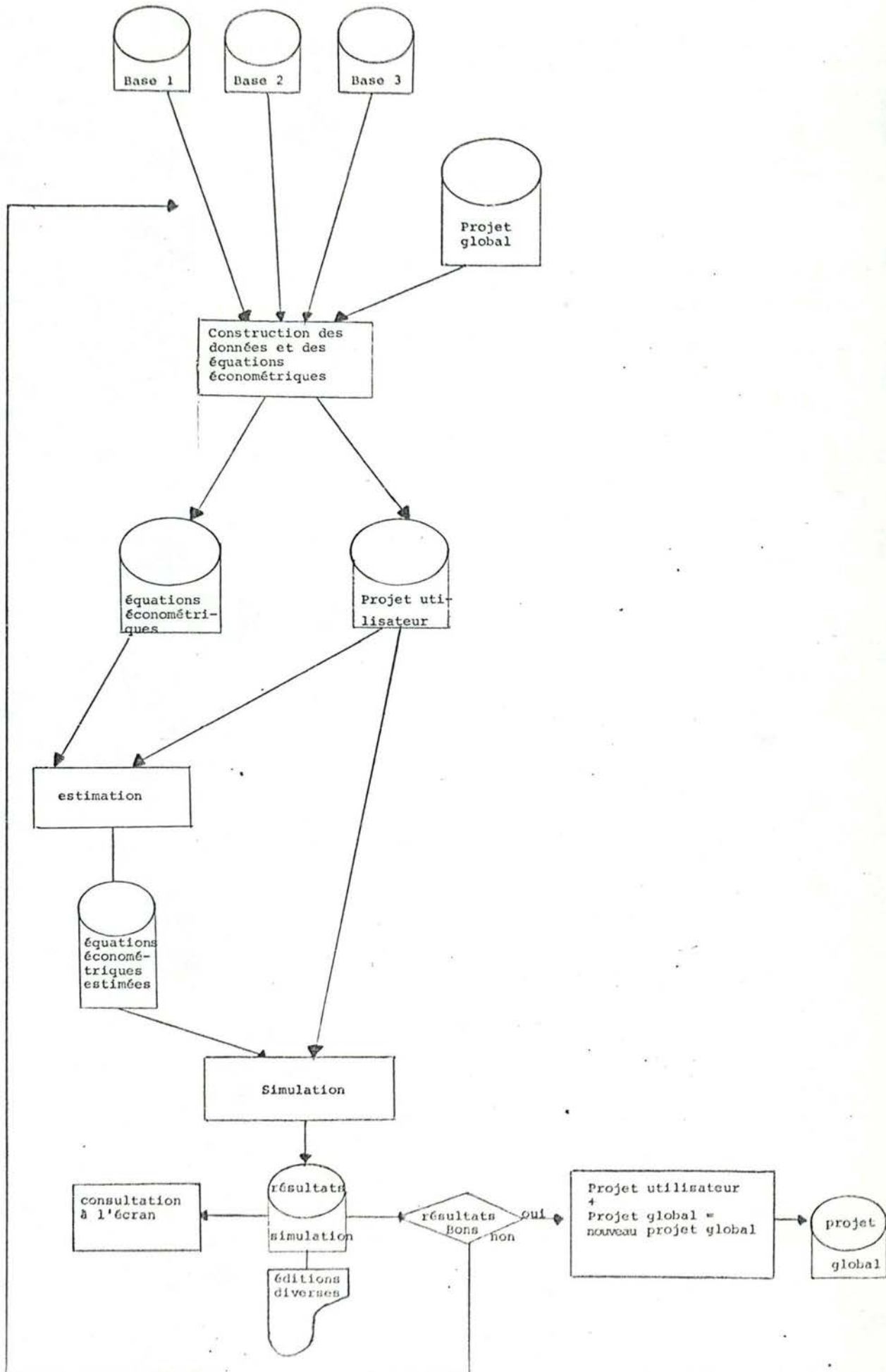
sortie : - les différences entre les observations du fichier de données et les observations calculées

traitement : - effectuer les équations de contrôle
- calculer la différence entre les observations calculées et les observations du fichier

3.2.2.0 Calcul de caractéristiques statistiques des variables

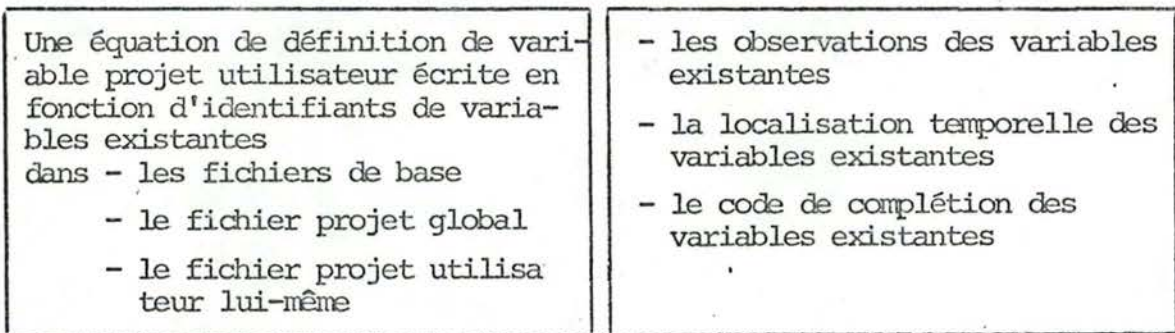
Création de fonctions telles que :

- calcul de la moyenne des observations d'une variable
- calcul de l'écart-type des observations d'une variable
- recherche de la plus petite, de la plus grande observation d'une variable.



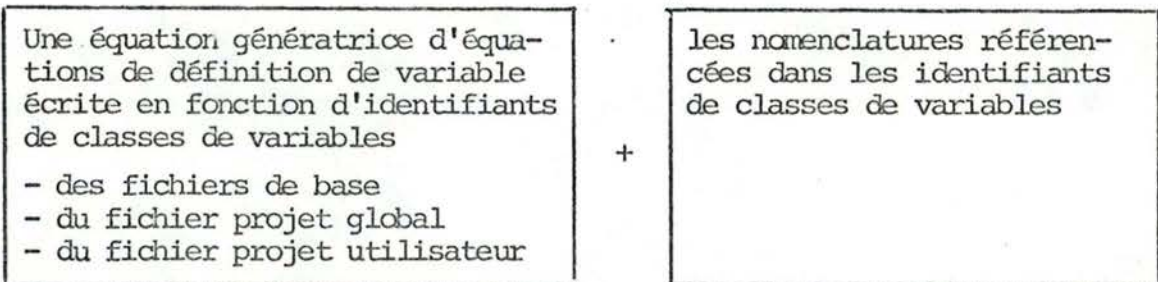
Les principales fonctions d'aide pour la phase de construction des données statistiques et des équations d'un modèle économétrique peuvent être schématisées comme suit :

1. Calcul des observations, de la localisation temporelle, du code de complétion d'une variable projet utilisateur.



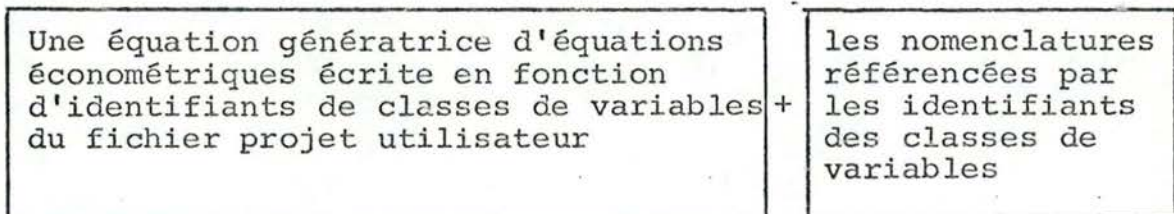
- Calcul - des observations d'une variable projet utilisateur
- de la localisation temporelle d'une variable projet utilisateur
 - du code de complétion d'une variable projet utilisateur

2. Génération d'équations de définition de variables



Génération d'une liste d'équations de définition de variables écrites en fonction d'identifiants de variables.

3. Génération d'équations économétriques



Génération d'une liste d'équations économétriques écrites en fonction d'identifiants de variables.

IIIe PARTIE : L'ANALYSE D'IMPLEMENTATION

PLAN

CHAPITRE I - LE SYSTEME INFORMATIQUE DU BUREAU DU PLAN

- 1.1. Le langage APL
- 1.2. Les softwares existants
- 1.3. Objectif de l'analyse d'implémentation

CHAPITRE II - ETUDE DE L'IDENTIFIANT DES VARIABLES ET CLASSES
DE VARIABLES

- 2.1. Etude théorique de l'identifiant des variables
- 2.2. La codification de la clé de la relation "variable₂"
- 2.3. Un "nom" de classe de variables
- 2.4. Traduction d'un "nom" de classe de variables

CHAPITRE III - LES EQUATIONS GENERATRICES D'EQUATIONS ECONO-
METRIQUES

CHAPITRE IV - STRUCTURE PHYSIQUE D'UN FICHIER PROJET

CHAPITRE V - EVALUATION DU TRAVAIL, CONCLUSION ET POSSIBILITES
D'EXTENSION

- 5.1. Evaluation du travail par rapport aux critiques et
souhais des utilisateurs et conclusion
- 5.2. Possibilités d'extension

CHAPITRE I - LE SYSTEME INFORMATIQUE DU BUREAU DU PLAN

La mise au point d'un modèle économétrique se fait essentiellement en interactif au Bureau du Plan, le langage interactif utilisé est l'APL.

Mais elle comporte aussi une phase de simulation qui, elle, est réalisée en FORTRAN.

Dans ce chapitre on présentera :

- le langage APL
- les logiciels existants :
 - le logiciel d'estimation
 - le logiciel de simulation
- l'objectif de l'analyse d'implémentation

1.1 Le langage APL

L'utilisation de l'APL est une contrainte imposée par le Bureau du Plan, on va dans un premier temps, donner quelques caractéristiques essentielles de l'APL concernant principalement l'utilisation des fichiers.

* L'APL est un langage interactif ce qui impose d'avoir un temps d'accès aux informations le plus court possible et de pouvoir accéder aux informations à partir d'identifiants aussi significatifs que possible pour l'utilisateur.

* L'APL reconnaît deux types de fichiers :

- les fichiers séquentiels
- les fichiers en accès direct sur le numéro de record qui est la position relative du record par rapport au début du fichier. Les fichiers en accès direct doivent être créés en séquentiel et avoir des records de longueur fixe non bloqués.

- * L'utilisateur APL communique avec le système de gestion de fichiers APL par l'intermédiaire de variables partagées. Une variable partagée doit être offerte par l'utilisateur au système et elle est affectée durant toute sa durée de vie à un et un seul fichier. Un fichier séquentiel nécessite une seule variable partagée alors qu'un fichier en accès direct en nécessite deux. Le nombre de variables partagées disponibles pour un utilisateur est limité et fixé lors de la génération du système ; ce qui détermine le nombre de fichiers concurremment accessibles pour l'utilisateur APL.
- * Un utilisateur APL travaille dans un espace mémoire appelé "Work Space" de taille fixée aussi à la génération du système. Cet espace mémoire doit pouvoir contenir toutes les fonctions et toutes les données nécessaires à un moment donné d'une session APL (+ 120K).

1.2 Les logiciels existants

1. Le logiciel d'estimation

L'estimation des équations économétriques est réalisée par un logiciel écrit en APL qui nécessite en entrée :

- les équations économétriques traduites en APL
- les observations et la localisation temporelle des variables référencées dans les équations économétriques.

2. Le logiciel de simulation

La simulation des équations économétriques est réalisée par un logiciel écrit en FORTRAN. La préparation des entrées du logiciel de simulation est faite en interactif, elle comprend :

- une phase de traduction des équations économétriques en sous-routines FORTRAN
- une phase de construction du fichier des données de simulation.

Les données de simulation seront constituées par :

- une sélection de variables dans le fichier projet dont on ne retient que les observations
- la prévision d'"observations" pour les variables exogènes (les variables exogènes sont celles qui ne sont pas expliquées par le modèle, on doit donc leur donner une série d'observations couvrant toute la période de simulation).

1.3 Objectif de l'analyse d'implémentation

Au terme de l'analyse fonctionnelle et vu le système informatique en place au Bureau du Plan, on peut conclure que le système TSAR répond assez bien aux besoins des utilisateurs. L'analyse d'implémentation du nouveau système de gestion de fichiers doit en fait résoudre 3 problèmes :

1. L'accès à un nombre quelconque de fichiers de base pour la phase de construction des variables du fichier projet utilisateur.
2. Trouver une implémentation efficace des identifiants des variables découverts par l'analyse fonctionnelle.
3. Générer des équations à partir d'une équation génératrice.

Dans un premier temps, on a étudié la phase de construction des variables du fichier projet utilisateur.

L'implémentation de cette phase devrait permettre de résoudre les problèmes de MAJ des fichiers qu'ils soient de base ou projet.

Le nombre de fichiers utilisés en concurrence durant la phase interactive de création d'un fichier projet utilisateur ne peut être quelconque et devra être minimisé.

Pour cela, on a commencé une étude concernant le regroupement des fichiers de base au sein d'une "Banque de Données" ; ce qui devrait permettre de n'avoir que la "Banque de Données" et le fichier projet global en entrée lors de la phase de construction du fichier projet utilisateur.

Cette "Banque de données" servirait aussi d'entrée pour la phase d'actualisation du fichier projet global.

En annexe, on trouvera un début d'analyse concernant la création et l'utilisation de la "Banque de Données" pour la phase de création du fichier projet utilisateur.

La deuxième étape de l'analyse d'implémentation s'est penchée sur le problème des équations génératrices d'équations économiques et ce, dans le cadre d'un fichier projet existant.

On a dû étudier les problèmes posés par l'implémentation des identifiants des variables en tenant compte des exigences des utilisateurs le concernant et des logiciels existants.

L'identifiant des variables découvert par l'analyse fonctionnelle est constitué de :

- la nature de la variable
- le nom de la relation économique
- de 1 à n {
 - nom d'objet économique
 - nom de nomenclature
 - occurrence dans la nomenclature

Les relations économiques actuellement utilisées au Bureau du Plan sont au maximum ternaires ; on aura donc

de 1 à 3 {

- nom d'objet économique
- nom de nomenclature
- occurrence dans la nomenclature

pour un identifiant de variable.

L'utilisateur veut pouvoir manipuler des "noms" de variable ; un "nom" de variable économique devra répondre aux exigences suivantes : (1)

- être significatif pour l'utilisateur
- être un identifiant de variable au sein d'un fichier de données statistiques
- être un nom de variable APL : un nom de variable APL est une chaîne de caractères quelconques, ne contenant pas d'opérateurs APL et dont le premier caractère est différent d'un chiffre. (Cette variable APL devra pouvoir contenir les observations de la variable lue).
- être un nom de variable FORTRAN donc une chaîne de six caractères au maximum lettres ou chiffres. (L'utilisateur veut pouvoir travailler sur les sous-routines FORTRAN générées à partir des équations économétriques qui sont écrites en fonction d'identifiants de variables)
- permettre l'accès rapide aux caractéristiques de la variable. On devra pouvoir traduire :
 - un "nom" de variable en un chemin d'accès à la variable
 - un "nom" de classe de variables en une liste de "noms" de variables

On a alors analysé le problème des équations génératrices en imposant à l'utilisateur de donner les nomenclatures des classes d'objets économiques sur lesquelles il veut qu'on génère les équations économétriques.

(1) Lorsque l'on utilise le terme variable, il signifie la variable économique, si l'on fait référence à une variable d'un langage de programmation, on précisera toujours variable APL ou variable FORTRAN.

CHAPITRE II - ETUDE DE L'IDENTIFIANT DES VARIABLES ET
CLASSES DE VARIABLES

2.1 Etude théorique de l'identifiant des variables (1)

Dans le cas d'une relation économique ternaire, la relation "variable" a été définie comme suit :

VAR (nature, nom de relation, objet économique 1,
nomenclature 1, occurrence 1,
objet économique 2, nomenclature 2, occurrence 2,
objet économique 3, nomenclature 3, occurrence 3, I)

où I représente l'ensemble des caractéristiques non identifiantes de la variable.

Dans la suite de l'exposé, on utilisera le symbolisme suivant : N : pour la nature

R : pour le nom de relation

OE_i : pour l'objet économique i

nom_i : pour la nomenclature i

OCC_i : pour l'occurrence i

Dans un fichier de données statistiques, la nature et le nom de relation impliquent automatiquement les objets économiques 1(2,3) et les nomenclatures 1(2,3) suivant que la relation économique est u-naire (binaire, ternaire).

(1) Les concepts utilisés dans ce paragraphe proviennent de "COMPUTER DATA-BASE ORGANISATION" de James Martin - 2e édition et sont issus du modèle relationnel de Codd.

Un nom de relation économique et les objets économiques concernés constituent ce que l'on a appelé un type de relation économique du modèle. Un type de relation économique + les nomenclatures des objets économiques constituent ce que l'on a appelé une classe de relations économiques ou classe de variables. Pour toutes les variables appartenant à la même classe, on aura les mêmes OE1, nom1 (OE2, nom2, OE3, nom3) suivant que la variable est une relation unaire, (binaire, ternaire).

On peut donc décomposer les relations "variable" (ternaires) en deux relations (1) :

- une relation "type de classe de variables" :
type de classe de variables (N, R, OE1, NOM1, OE2, NOM2, OE3, NOM3)
- les relations "variables₁"
variable₁ (NR, OCC1, OCC2, OCC3, I)

La relation "type de classe de variables" est sous la seconde forme normale :

* N,R est la clé de la relation :

- dans chaque n-uple de la relation, la valeur de la clé N,R identifie uniquement ce n-uple
- aucun attribut de la clé ne peut être enlevé sans détruire la 1ère propriété

(1) Les relations binaires pourront être décomposées en 2 relations définies comme suit : une relation 'type de classe de variables 2' (N, R, OE1, NOM1, OE2, NOM2) les relations 'variables 12' (N, R, OCC1, OCC2, I). Les relations unaires pourront être décomposées en 2 relations définies comme suit : une relation 'type de classe de variables 1' (N, R, OE1, NOM1), les relations 'variables₁₁' (N, R, OCC1, I).

Dans la suite du chapitre, on base toute l'étude des identifiants des variables sur les relations ternaires, mais tout peut être facilement transposé pour le cas de relations binaires et unaires.

- * OE1 est en dépendance fonctionnelle totale de N, R car OE1 est en dépendance fonctionnelle de N, R (chaque valeur de N, R n'a pas plus d'une valeur de OE1 qui lui est associée dans la relation) mais OE1 n'est pas en dépendance fonctionnelle ni de R ni de N tout seul.

- * Nom1 est en dépendance fonctionnelle totale de N, R car nom1 est en dépendance fonctionnelle de N, R (chaque valeur de N, R n'a pas plus d'une valeur de nom1 qui lui est associée dans la relation) mais nom1 n'est pas en dépendance fonctionnelle ni de R ni de N seul.

- * de même pour OE2 et OE3, nom2 et nom3.

On définit le domaine "nom variable" comme étant une codification des domaines N, R, OCC1, OCC2, OCC3 (cfr. le paragraphe sur la codification de la clé de la relation "variable 2").

On définit la relation "variable₂" :

variable₂ (nom de variable, I, N, R, OE1, NOM1, OCC1, OE2, NOM2, OCC2, OE3, NOM3, OCC3)

'nom de variable' est la clé de la relation "variable₂", cette clé sera le "nom" de variable demandé par l'utilisateur donc elle sera codifiée en 6 caractères alphanumériques (lettres ou chiffres uniquement, le 1er caractère étant une lettre).

Cette clé étant codifiée, on reprendra chaque élément de la clé en clair dans les attributs non primaires de la relation "variable₂".

Cette relation "variable₂" sera à la base de l'implémentation d'un fichier de données statistiques. Justification des redondances introduites dans la relation "variable₂" :

1. Le nombre de bytes nécessaires pour les attributs redondants de la relation "variable₂" ne sera pas exagéré vu le nombre de bytes nécessaires pour les attributs non redondants de cette relation : une occurrence de variable représentera ± 1000 bytes de données dans le fichier, les 50 ou même 80 bytes nécessaires pour introduire les redondances dans ce même fichier ne seront pas une charge trop lourde pour le système.
2. Les attributs non primaires de la relation "type de classe de variables" (OE1 (2,3), nom1 (2,3)) se retrouvent dans les attributs non primaires de la relation "variable₂".
On pourrait créer un fichier de la relation "type de classe de variables" mais ce fichier ne serait, en fait, quasiment jamais utilisé :

- un utilisateur connaît assez bien les relations économiques qu'il manipule tout au moins celles du fichier projet auquel on s'intéresse ici ;
- dans le cadre d'un fichier projet, l'utilisateur sait que la classe de variables "Exportation en quantité" est construite sur les OE "biens exportés" et sur les "pays importateurs", que ces OE sont vus au travers des nomenclatures nom1, nom2 ;
- ce fichier ne sera d'aucune utilité pour décoder la clé de la relation "variable₂" (cfr. codification de la clé de la relation variable₂, la traduction d'un "nom" de classe de variables").

3. Les attributs primaires de la relation "variable2" se trouveront sous 2 formes dans la relation :

- sous la forme codifiée pour construire la clé de la relation qui sera en même temps le "nom" de variable demandé par l'utilisateur ;
- sous une forme décodée dans les attributs non primaires de la relation.

On ne crée pas de fichiers informatiques de nomenclatures parce qu'à l'heure actuelle, on ne sait pas comment on pourrait l'organiser de façon efficiente. Exemple de problème que l'on rencontrera lors de la création d'un fichier de nomenclatures.

* La nomenclature des "biens importés" contiendra les occurrences : - Biens

- Pétrole

- autres biens que le pétrole

- Service
 - tourisme
 - autres services
- Biens + services

* la nomenclature des "biens exportés" contiendra les occurrences :

- Biens
- Services
 - tourisme
 - autres services
- Biens + services

* Doit-on créer pour cela 2 nomenclatures différentes ?

Ou pourra-t-on, à partir d'une seule et même nomenclature de "biens", retrouver les nomenclatures "biens exportés" et "biens importés".

2.2 La codification de la clé de la relation "variable 2"

La clé 'nom variable' de la relation "variable 2" sera le "nom" de variable demandé par l'utilisateur. Elle sera donc codifiée sur 6 caractères (des lettres et/ou des chiffres, le 1er caractère étant une lettre). Ce que l'on impose quant à la codification du "nom" de variable :

- ce nom devra être unique dans un fichier, être la clé du fichier ;
- pour toutes les variables appartenant à la même classe, l'ordre de présentation de la codification de N, R, OCC1, OCC2, OCC3 devra être le même ;
(cfr. "nom" de classe de variables)
- chaque élément de la clé pourra être codifié sur un nombre variable de caractères mais il faut que dans tous les cas, la longueur du "nom" de variable soit de 6 caractères au maximum.

On laisse cette liberté à l'utilisateur pour qu'il puisse se construire des "noms" de variables aussi significatifs que possible suivant que la relation est u-naire, binaire ou ternaire.

Ex.: Un modèle de la vie économique et sociale peut étudier la classe de relations "exportations en quantité de biens, de Belgique vers d'autres pays" sans s'intéresser aux exportations des autres pays que la Belgique.

On peut remarquer que le pays exportateur "la Belgique" peut être implicite ; pour le modèle, la relation exportation est ici une relation binaire : exportations (biens, pays importateurs).

On pourra codifier :

- la nature des variables de la classe de relations "exportations en quantité" sur 1 caractère par ex: 'Q'
- le nom de relation sur 1 caractère par ex: 'X'
- il reste 4 caractères pour codifier les nomenclatures de biens et de pays importateurs

ex. de codification d'une nomenclature de biens exportés :

code	signification
B	biens
S	service
BS	biens + service

ex. de codification d'une nomenclature de pays importateurs :

code	signification
D	Allemagne
FR	France
NL	Pays-Bas

A partir des codes définis , on peut construire les identifiants de variables suivants :

	QXSD
	QXSFR
	QXSNL
QXBD	QXBSD
QXBFR	QXBSFR
QXBNL	QXBSNL

On ne pourra codifier le pays importateur "USA" par la chaîne de caractères "USA" si l'on garde la même codification pour les occurrences dans la nomenclature "Biens exportés" ; une telle codification construira la chaîne de caractères "QXBSUSA" qui n'est pas un nom de variable.

La relation "type de classe de variables" (1) est sous la 2e forme normale mais pas sous la 3e forme normale : nom1 est transitivement dépendant de N,R qui est la clé de la relation (de même pour nom2 et pour nom3).

On a :

- nom1 est fonctionnellement dépendant de 0E1 (chaque valeur de 0E1 n'a pas plus d'une valeur de nom1 qui lui est associée dans la relation)
- 0E1 est fonctionnellement dépendant de N,R (chaque valeur de N,R n'a pas plus d'une valeur de 0E1 qui lui est associée dans la relation)
- nom1 est fonctionnellement dépendant de N,R (chaque valeur de N,R n'a pas plus d'une valeur de nom1 qui lui est associée dans la relation)

(1) type de classe de var (N, R, 0E1, nom1, 0E2, nom2, 0E3, nom3)

N,R n'est pas fonctionnellement dépendant de OE1 (à une valeur de OE1 peut correspondre plusieurs valeurs de N,R) (de même pour nom2, OE2, et nom3, OE3.)

2.3 Un nom de classe de variables

Un nom de classe de variables sera construit à partir de la relation "type de classe de variables" qui est défini comme suit :

type de classe de variables (N, R, OE1, nom1, OE2, nom2, OE3, nom3)

Pour une classe de variables dans un fichier, la nature, le nom de la relation et les noms d'objets économiques impliquent automatiquement des noms de nomenclatures pour les objets économiques :

pour une occurrence de N,R on a que :

OE1 implique nom1
OE2 " " nom2
OE3 " " nom3

Un objet économique d'une classe de variables dans un fichier est perçu par une et une seule nomenclature.

On peut donc décomposer la relation "type de classe de variables" en 2 relations R1 et R2 telles que :

R1 (N, R, OE1, OE2, OE3)

R2 (N, R, nom1, nom2, nom3)

La relation R2 sera à la base de la construction des noms de classe de variables.

Un nom de classe de variables sera écrit au moyen :

- d'une codification de N,R, utilisant les lettres et les chiffres uniquement

- de une à 3 lettre(s) soulignée(s) pour représenter les références aux nomenclatures (1)

Ex.: le nom de la classe de variables "exportations en quantité" décrite dans le paragraphe sur les noms des variables sera noté :

$\text{QX}\underline{\dot{I}}\underline{\dot{J}}$ { où $\underline{\dot{I}}$ représente la nomenclature des biens exportés
et $\underline{\dot{J}}$ la nomenclature des pays importateurs
QX étant la codification de la nature et du nom de relation.

On a vu que l'utilisateur peut travailler au niveau de classe de variables complète ou de sous-classe de variables.

Un nom de sous-classe de variable sera par ex.: $\text{QX}\underline{\dot{D}}$ où D est une occurrence dans la nomenclature des pays importateurs; $\text{QX}\underline{\dot{D}}$ est le nom de la sous-classe de variables "exportations en quantité vers l'Allemagne.

Une référence à une nomenclature dans un nom de classe de variables se notera par une lettre soulignée ;

- cette lettre soulignée devra être différente des lettres soulignées utilisées dans les opérateurs LEC (cfr. équations génératrices) ;
- on impose de noter une référence à une nomenclature par une seule lettre soulignée parce que :
 - 1) ce sera plus court et plus facile à écrire pour l'utilisateur (surtout lorsque l'on utilise les équations génératrices) ;
 - 2) ce sera plus facile pour la génération des noms de variables

Exemple : la classe de variables "exportations en quantité" sera notée par la chaîne de caractères ' $\text{QX}\underline{\dot{I}}\underline{\dot{J}}$ ' et non par la chaîne de caractères ' $\text{QX } \underline{\text{BIENS}}, \underline{\text{PAYS}}$ '

(1) Au terminal APL, on a la possibilité de travailler avec les lettres et les lettres soulignées ; ce qui étend le jeu de caractères disponibles. Ex.: A est un caractère différent de A.

2.4 Traduction d'un nom de classe de variables

Un nom de classe de variables sera traduit en une liste de noms de variables.

Un nom de variable sera un nom de classe de variables dans lequel on a remplacé les références aux nomenclatures par des occurrences dans les nomenclatures. C'est pourquoi on a imposé pour les noms des variables d'une même classe, d'avoir le même ordre de présentation pour les éléments de la clé. La fonction de traduction d'un nom de classe de variables en une liste de noms de variables est une fonction interactive, elle demandera à l'utilisateur les nomenclatures à partir desquelles on génèrera les noms des variables.

A une telle demande, l'utilisateur peut répondre :

- en donnant un nom de variable APL contenant la liste des occurrences dans la nomenclature ;
- en donnant directement la liste des occurrences dans la nomenclature ;

(dans une étape ultérieure en donnant un nom de nomenclature que l'on pourra aller chercher dans le fichier de nomenclatures).

Cette fonction devra être appelée notamment par la ou les fonctions de lecture de variables : l'utilisateur peut vouloir lire une liste de variables, une liste de classes de variables .

La fonction de traduction d'un nom de classes de variables en une liste de noms de variables GENNAME devra être construite sur le schéma suivant :

appel à la fonction : GENNAME 'nom de classe de variables'

traitement : 1. reconnaître dans la chaîne de caractères
'nom de classe de variables' les références
aux nomenclatures (les lettres soulignées)

2. demander les nomenclatures à l'utilisateur
3. construire le produit cartésien des nomenclatures entrées par l'utilisateur
4. pour chaque n-uple du produit cartésien :
généraler un nom de variable en remplaçant les références aux nomenclatures dans le nom de classe de variables par les occurrences trouvées dans le n-uple courant.

CHAPITRE III - LES EQUATIONS GENERATRICES D'EQUATIONS ECONOMETRIQUES

Une équation génératrice d'équations économétriques sera écrite en LEC et les opérandes auxquels on s'intéresse seront des noms de classes de variables. Une équation génératrice a pour objet de générer les équations économétriques expliquant une classe de variables.

Ex.:

$$\underline{L} \text{ QXIJ} = \underline{C1} \times \underline{L} \text{ QMIJ} + \underline{C2} \times \underline{L} (\text{PXIB} \div \text{PXIM})$$

est l'équation génératrice des équations économétriques expliquant la classe de variables 'exportations en quantité'.

Dans cette équation :

- L est l'opérateur LEC signifiant le logarithme
- C1 et C2 sont les coefficients des équations économétriques
- QXIJ est le nom de la classe de variables 'exportations en quantité' ;

Q = nature de la classe de relation

X = nom de la relation

I = référence à la nomenclature de l'objet économique biens exportés

J = référence à la nomenclature de l'objet économique pays importateur

- QMIJ est le nom de la classe de variables 'importations en quantité'

Q = nature de la classe de relation

M = nom de la relation

I = référence à la nomenclature de l'objet économique biens importés

J = référence à la nomenclature de l'objet économique pays exportateurs

- PXIB est le nom de la classe de variables 'prix des exportations belges' (ex.: prix de l'acier belge à l'exportation)
- PXIM est le nom de la classe de variables 'prix des exportations de biens sur le marché mondial' (ex.: prix de l'acier à l'exportation sur le marché mondial)

Pour que l'on puisse générer des équations économétriques, il faut que la nomenclature I soit la même pour tous les opérandes de l'équation : I est une nomenclature de biens (de même pour J qui est une nomenclature de pays).

Au sein d'une équation générée, on doit avoir la même occurrence dans la nomenclature des biens pour les classes de variables exportations en quantité, importations en quantité, prix des exportations belges et prix des exportations mondiales. (de même pour la nomenclature J)

Si la nomenclature I contient la liste d'occurrences B, S, BS et la nomenclature J la liste D, FR, NL, les équations générées seront :

$$\underline{L} \text{ QXBD} = \underline{C1} \times \underline{L} \text{ QMBD} + \underline{C2} \times \underline{L} (\text{PXBB} \div \text{PXBM})$$

$$\underline{L} \text{ QXSD} = \underline{C1} \times \underline{L} \text{ QMSD} + \underline{C2} \times \underline{L} (\text{PXSb} \div \text{PXSM})$$

$$\underline{L} \text{ QXBSD} = \underline{C1} \times \underline{L} \text{ QMBSD} + \underline{C2} \times \underline{L} (\text{PXBSB} \div \text{PXBSM})$$

$$\underline{L} \text{ QXBFR} = \underline{C1} \times \underline{L} \text{ QMBFR} + \underline{C2} \times \underline{L} (\text{PXBB} \div \text{PXBM})$$

$$\underline{L} \text{ QXSFR} = \underline{C1} \times \underline{L} \text{ QMSFR} + \underline{C2} \times \underline{L} (\text{PXSb} \div \text{PXSM})$$

$$\underline{L} \text{ QXBSFR} = \underline{C1} \times \underline{L} \text{ QMBSFR} + \underline{C2} \times \underline{L} (\text{PXBSB} \div \text{PXBSM})$$

$$\underline{L} \text{ QXBNL} = \underline{C1} \times \underline{L} \text{ QMBNL} + \underline{C2} \times \underline{L} (\text{PXBB} \div \text{PXBM})$$

La fonction de génération d'équations économétriques sera basée sur le même schéma que la fonction de traduction d'un nom de classe de variables en une liste de noms de variables, en entrée on a une équation génératrice, les nomenclatures référencées dans les opérandes de l'équation, en sortie on aura une liste d'équations générées.

appel à la fonction : GENERATE 'équation génératrice'

- traitement :
1. reconnaître dans l'équation génératrice les références aux nomenclatures (qui sont les lettres soulignées différentes des opérateurs LEC (1))
 2. demander à l'utilisateur les nomenclatures
 3. construire le produit cartésien des nomenclatures données par l'utilisateur
 4. pour chaque n-uple du produit cartésien :
généraliser une équation économétrique en remplaçant dans l'équation génératrice les références aux nomenclatures par les occurrences trouvées dans le n-uple courant.

On peut reprendre l'exemple d'équation génératrice donnée en début de chapitre et exécuter la fonction GENERATE en suivant les différentes étapes de la génération :

GENERATE 'L QXIJ = C1 x L QMIJ + C2 x L (PXB ÷ PXM)

1. I et J sont les références aux nomenclatures dans l'équation

2. On demande à l'utilisateur

I =

J =

3. I est une variable APL contenant la liste d'occurrences

B, S, BS; J est une variable APL contenant la liste d'occurrences D, FR, NL

On construit le produit cartésien (I x J) qui est la liste des couples :

(B,D), (B,FR), (B,NL),

(S,D), (S,FR), (S,NL),

(BS,D), (BS,FR), (BS,NL)

(1) Les lettres soulignées utilisées dans les opérateurs LEC sont : A, C, D, E, G, L, M, N, R, V, X.

4. Pour chaque n-uple du produit cartésien :
- donc en particulier pour le 1er (B,D), recopier l'équation génératrice en remplaçant I par B et J par D ce qui donne
- $$\underline{L} \text{ QXBD} = \underline{C1} \times \underline{L} \text{ QMBD} + \underline{C2} \times \underline{L} \text{ (PXBB} \div \text{PXBM)}$$

L'identifiant d'une équation économétrique est constitué par :

- un code équation
- le nom de la variable expliquée (le nom de variable du membre de gauche de l'équation)

On doit donc encore avoir :

- une fonction de génération des identifiants des équations économétriques générées

en entrée on aura :

l'identifiant de l'équation génératrice qui sera constitué d'un code équation suivi du nom de la classe de variables expliquées

en sortie on aura une liste d'identifiants d'équations économétriques

le traitement sera exactement le même que celui de la fonction GENNAME décrite plus haut

- une fonction qui range dans la variable APL "identifiant d'équation économétrique" l'équation économétrique correspondante :

- en entrée on aura : la liste des identifiants des équations économétriques

la liste des équations économétriques

- en sortie : les variables APL "identifiant d'équation économétrique" contiendront les équations économétriques.

CHAPITRE IV - STRUCTURE PHYSIQUE D'UN FICHER PROJET

Une variable d'un fichier projet représente \pm 1000bytes de données, or dans la majorité des traitements, on n'utilise qu'une partie des caractéristiques des variables. On va essayer de regrouper, au sein de fichiers physiques distincts, les caractéristiques des variables suivant leurs utilisations courantes et ce, dans le but de n'occuper qu'un minimum de place en mémoire et de traduire rapidement les noms de variables en chemins d'accès aux caractéristiques des variables.

Pour un fichier projet, on pourra créer les fichiers physiques suivants :

- 1) un fichier de sélection qui devra permettre de traduire les noms de variables en chemins d'accès aux variables. Il contiendra essentiellement la liste des noms de variables du fichier projet exprimés en base 37 (1) ;
- 2) un fichier des observations qui contiendra toutes les informations nécessaires aux manipulations courantes des observations des variables. Ce fichier contiendra pour chaque variable :
 - la liste des observations ;
 - le code de complétion (qui indique s'il manque des observations ou si la liste des observations est complète) ;
 - la localisation temporelle (année et période de la 1ère observation, année et période de la dernière observation et périodicité ;
 - la date de dernière MAJ.

(1) en APL, on peut très facilement traduire une chaîne de 6 caractères alphanumériques en un nombre entier ; la fonction inverse existe aussi : à partir d'un nombre entier, construire une chaîne de caractères.
Si l'on exprime les noms de variables en base 37, c'est pour accélérer la procédure de traduction du nom de variable en chemin d'accès à la variable (cfr. la traduction d'un nom de variable en un numéro de record).

3) un fichier de codification qui contiendra pour chaque variable du fichier :

- l'identifiant de la variable en clair, c'est-à-dire la décodification du nom de variable :
 - . la nature de la relation
 - . le nom de la relation
 - . l'objet économique 1
 - . la nomenclature de l'objet économique 1
 - . l'occurrence dans la nomenclature 1
 - . l'objet économique 2
 - . la nomenclature de l'objet économique 2
 - . l'occurrence dans la nomenclature 2
 - . l'objet économique 3
 - . la nomenclature de l'objet économique 3
 - . l'occurrence dans la nomenclature 3
- le millésime de référence ;
- la localisation dimensionnelle ;
- la localisation géographique ;
- l'équation de définition de la variable ;
- le libellé de la variable ;
- l'origine bibliographique de la variable.

Ce fichier sera utilisé en consultation.

4) un fichier de commentaire qui contiendra pour chaque variable du fichier :

- la date de création de la variable ;
- le responsable de son introduction dans le fichier ;
- le commentaire.

Ce fichier sera utilisé en consultation et très rarement.

La création d'un fichier de codification et d'un fichier de commentaire peut se justifier par le fait que le commentaire d'une variable sera en général assez long, donc la lecture de ce fichier consommera beaucoup de place mémoire (actuellement, on a 560 bytes de commentaires par variable) et il est fort

peu utilisé. Alors que le fichier de codification sera plus souvent utilisé :

- pour l'actualisation d'un fichier projet, on a besoin de l'équation de définition des variables ;
- on peut vouloir décoder le nom de variables, avoir besoin du libellé de la variable...

Le fichier de sélection d'un fichier projet sera accessible en séquentiel, les autres fichiers seront en accès en direct (sur le numéro de record).

Pour accélérer la procédure de traduction d'un nom de variable en un numéro de record dans les autres fichiers, on construira le fichier de sélection de telle sorte que :

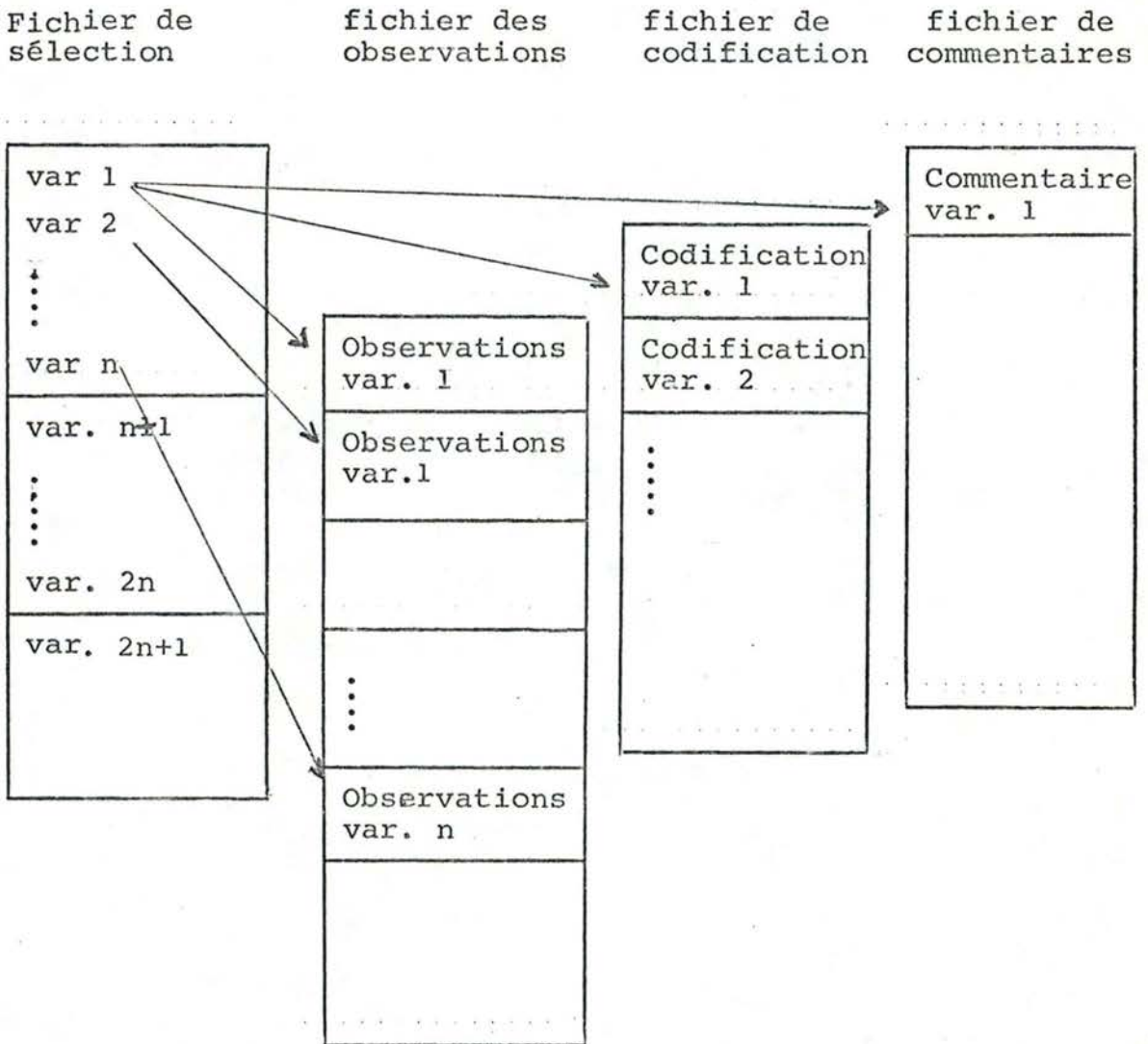
- un record constitue un tableau de sélection pour un ensemble de variables du fichier. Une ligne du tableau correspondra à une variable. Le nombre de lignes du tableau de sélection devra être déterminé de telle sorte que le tableau n'occupe pas trop de place en mémoire ;
- un nom de variable dans le fichier de sélection sera exprimé en base 37, il constituera donc un nombre entier.

La fonction de traduction d'un nom de variable "var 2" en un numéro de record sera construite sur le schéma suivant :

- (1) exprimer le nom de variable "var 2" en base 37 ;
- (2) lire le premier record du fichier de sélection dans un tableau appelé par exemple CLE. Chaque ligne du tableau contiendra un nombre qui est le nom de variable exprimé en base 37 ;
- (3) l'instruction `APL;CLE("var 2 en base 37")` aura pour résultat un indice ; cet indice sera égal à la position de "var 2 en base 37" dans CLE ou sera strictement supérieur au nombre de lignes de CLE si "var 2 en base 37" n'existe pas dans CLE. Dans le premier cas (le nombre de records

lus - 1) x nombre de lignes du tableau CLE + indice = position de la variable dans le fichier qui est aussi le numéro de record dans les autres fichiers. Dans l'autre cas, on lit le record suivant (retour au point 2).

Le schéma de la structure physique d'un fichier projet sera le suivant :



CHAPITRE V - EVALUATION DU TRAVAIL, CONCLUSION ET POSSIBILITES
D'EXTENSION

5.1 Evaluation du travail par rapport aux critiques et souhaits
souhaits des utilisateurs et conclusion

Cette évaluation peut être faite pour les deux niveaux d'analyse que comprend ce travail : pour l'analyse fonctionnelle d'une part, pour l'analyse d'implémentation d'autre part.

Dans l'analyse fonctionnelle, on a modélisé un fichier de données statistiques dans le but de découvrir un identifiant significatif aux variables. L'apport essentiel de cette modélisation est la notion de classe de variables : une classe de variables est un ensemble de variables qui sont toutes construites sur le même type de relation économique défini entre les mêmes types d'objets économiques perçus par des nomenclatures.

La notion de classe de variables a permis d'approcher le problème de la construction d'un modèle économétrique sous un nouvel angle : l'utilisateur construit une classe de variables ou une classe d'équations économétriques plutôt qu'une variable ou une équation économétrique.

Au niveau fonctionnel, on a décrit les procédures de création et de MAJ des différents types de fichiers de données statistiques : création et MAJ des fichiers de base, d'un fichier projet global et du fichier projet utilisateur. On y a décrit aussi les procédures de test de cohérence et d'édicions paramétrables d'un fichier de données statistiques.

On a délibérément réduit l'analyse d'implémentation à l'étape de construction d'un modèle économétrique.

On a commencé une étude de la phase de création interactive du fichier projet utilisateur dont les grandes lignes se trouvent en annexe.

On a alors proposé un type de structure physique pour le fichier projet et décrit des algorithmes de traduction d'un identifiant de variable et de classe de variables en chemins d'accès aux variables, ainsi que l'algorithme de génération d'équations économétriques à partir d'une équation génératrice.

5.2 Possibilités d'extensions

Une première possibilité serait d'adapter le système existant de telle sorte que l'on puisse bénéficier assez directement des notions de classe de variables et classe d'équations. Pour cela, on devra traduire les algorithmes décrits dans l'analyse d'implémentation en fonctions APL, modifier toutes les fonctions qui permettent actuellement de travailler sur une liste de noms de variables ou une liste de noms d'équations de telle sorte que l'utilisateur puisse travailler aussi bien sur les noms de variables que sur les noms de classes de variables ou sur les noms d'équations que sur les noms de classes d'équations. Ces modifications consisteront essentiellement en un appel à la fonction de génération de noms de variables à partir de noms de classes de variables ou à la fonction de génération de noms d'équations à partir de noms de classes d'équations.

L'analyse d'implémentation ne concerne qu'une partie du système décrit dans l'analyse fonctionnelle, une autre extension possible est d'étendre l'analyse d'implémentation à tout le système.

Il serait intéressant de concevoir un fichier de nomenclatures : de définir ce qu'il doit contenir, comment le structurer et comment l'utiliser. Un fichier de nomenclatures peut être conçu fort différemment :

- il pourrait n'être qu'une liste d'occurrences dans les nomenclatures utilisées pour la création d'un fichier de données statistiques ou refléter aussi la structure de chaque nomenclature ;
- on pourrait aussi définir un fichier de nomenclatures dans le but de pouvoir traduire automatiquement les noms de classes de variables en noms de variables ; ce qui impliquerait de définir une nouvelle syntaxe des noms de classes de variables ;
- on pourrait définir un fichier de types de nomenclatures utilisés pour un fichier de données statistiques, pour l'ensemble des fichiers de données statistiques du Bureau du Plan. Un type de nomenclature servirait de référence pour la création d'une classe de variables mais l'utilisateur serait libre de définir les nomenclatures qu'il veut pour la création réelle des classes de variables.

On pourrait aussi définir un ensemble de fonctions qui permette de manipuler les classes de variables et les classes d'équations dans un autre but que celui de retrouver des variables et des équations dont, par exemple, une fonction de description d'un modèle économétrique.

Cette fonction permettrait de retrouver la liste des classes de variables et des classes d'équations définies par le modèle.

Dans cette liste :

- une classe de variables serait décrite essentiellement par son nom, son identifiant complet, l'équation génératrice des équations de définition des variables
- une classe d'équations économétriques serait essentiellement décrite par son nom, l'identifiant de la classe de variables expliquées, l'équation génératrice des équations économétriques.

BIBLIOGRAPHIE

1. BENCI, BODART, BOGAERT, CABANES

"Concepts for the design of a conceptual schema"
Modeling in Data Base Management Systems. G.M. Nyssen
(ed) North Holland Publishing company 1976.

2. BODART

"Problèmes d'organisation et méthodes d'analyse fonctionnelle"
cours donné aux FNDP - Namur (1977-1978)

3. CODD, E.F.

"A Relational Model of Data for Large Shared Data Banks"
comm. ACM, 13, 377-387. (1970)

4. DATE

"An Introduction to Data Base Systems" second edition
ADDISON - WESLEY - 1977

5. Equipe "Grands Fichiers"

"Système de conception et d'exploitation de bases de données"
Première Partie : Modèles et Langages
Projet de recherche CIPS N° I.2/15
Publications de l'Institut d'Informatique de Namur - 1978

6. MARTIN JAMES

"Computer Data-Base Organization" second edition
Prentice Hall

7. POMES

"Etude d'un interface entre un software d'estimation et
de simulation écrit en APL "IESEM" et un software de
simulation écrit en PLI "SIMUL""
Mémoire de fin d'étude - 1978

8. RAYMOND P. POLIVKA, SANDRA PAKIN

"APL : The language and its usage"
Prentice - Hall

9. Publications du Bureau du Plan

- "Dossier Minibank"
référence (78)HJB/cf/1786/1963
- "Dossier SERENA notes techniques"
référence (78)GdA/JPL/HJB-1607/ev/1749
- "Instruction pour la construction de fichiers standards"
référence (76)ThD/TdB/1027/pb/1089
- "TSAR Time Series APL Retrieval"
référence (77)JW/AB/1373/hg/1488
- "SAMI : Système APL de modélisation interactive"
référence (79)DG/HJB/1968/cf/2205

ANNEXE

LE REGROUPEMENT DES FICHIERS DE BASE

La première étape de l'analyse d'implémentation a eu pour objectif de permettre la création interactive du fichier projet utilisateur.

Une variable du fichier projet utilisateur peut être créée à partir de variables définies dans plusieurs fichiers. On a vu que l'APL ne permet d'avoir qu'un nombre limité de fichiers simultanément ouverts; ce nombre est déterminé par le quota de variables partagées fixé pour un utilisateur lors de la génération du système.

Cette étude a pour but de proposer une organisation des fichiers de données statistiques du Bureau du Plan qui permette, pour la phase de construction des variables d'un fichier projet, d'avoir accès aux variables définies dans un nombre quelconque de fichiers.

Cette organisation devra permettre des accès rapides aux variables et ce, en ne consommant pas trop de place en mémoire.

Ce chapitre étudie deux aspects du problème :

1. le principe du regroupement des fichiers de base au sein d'une Banque de Données
2. les accès aux variables de la Banque de Données.

- - - - -

1. LE PRINCIPE DU REGROUPEMENT DES FICHIERS DE BASE AU SEIN D'UNE BANQUE DE DONNEES.

Si l'on veut permettre des accès concurrents à plus de fichiers que ne le permet le quota de variables partagées, on est obligé d'alourdir les procédures d'accès aux variables et ce, tant du point de vue de l'écriture de ces procédures que du point de vue temps d'accès. En effet, avant de pouvoir accéder à une variable, on devra vérifier si cette variable appartient à un fichier actuellement ouvert; si ce n'est pas le cas, on devra fermer éventuellement un fichier, puis ouvrir le fichier voulu, ce qui se traduira par une gestion assez compliquée des variables partagées.

Pour éviter cette gestion complexe des variables partagées, on propose de regrouper les fichiers de base au sein d'une Banque de Données.

Un utilisateur, lors de la phase de construction de variables projet peut avoir besoin :

- d'un fichier projet global
- d'un fichier projet utilisateur
- de 1 à n fichiers de base.

Les fichiers de base ont les avantages :

- d'être communs à tous les utilisateurs du Bureau du Plan
- d'être stables, c'est-à-dire que les MAJ sont périodiques (en général annuelles) et font suite à la parution de nouvelles statistiques officielles.

On peut imaginer plusieurs types de regroupement, par ex :

- un regroupement de tous les fichiers de base
- un regroupement des fichiers de base ayant une même caractéristique, telle que :
 - même périodicité
 - même année de début pour les observations
- un regroupement des fichiers de base utilisés dans la construction d'un fichier projet
- un regroupement des fichiers de base utilisés dans la construction de tous les fichiers projet en cours d'élaboration
- non plus un regroupement des fichiers de base, mais la création de la Banque de Données par sélection de variables dans les fichiers de base.

Le choix d'un type de regroupement plutôt qu'un autre sera basé sur les critères suivants :

- il faut garantir un minimum de fiabilité pour les données; ce qu'on ne pourra faire que si on ne multiplie pas les endroits où sont définies les mêmes variables.
- il faut essayer de regrouper des fichiers dans lesquels le nombre d'observations par variables est approximativement le même (les records des fichiers en accès direct en APL sont de longueur fixe)

-il faut que l'accès aux variables du fichier regroupé soit aisé et surtout rapide.

On propose de regrouper au sein de la Banque de Données tous les fichiers de base intervenant dans la construction de tous les fichiers projet en cours d'élaboration. (1)

Les avantages de ce type de regroupement sont :

- les fichiers projet actuels sont tous annuels et donc construits à partir de fichiers de base annuels. On constituera donc une Banque de Données annuelles.
- les variables des fichiers projet ont un vecteur d'observations qui débute au plus tôt en l'année 1953, ce qui permet d'imposer une longueur unique pour les vecteurs d'observations des variables de la Banque de Données.
- la Banque de Données sera commune à tous les utilisateurs du Bureau du Plan et limite donc la duplication des variables
- la phase de construction de variables projet ne nécessitera plus l'ouverture que de 3 fichiers :
 - la Banque de Données
 - le fichier projet global
 - le fichier projet utilisateur

On analysera plus loin en détail les procédures d'accès à la Banque de Données mais on peut déjà signaler ici que les temps d'accès à la Banque de Données dépendront principalement du nombre de variables définies dans un fichier source et non pas du nombre de fichiers regroupés dans la Banque de Données.

Les inconvénients de cette solution sont :

- les fichiers de base ne participant pas à la Banque de Données doivent pouvoir être accessibles en APL, il faudra essayer de conserver les mêmes procédures

(1) On réserve le terme de "fichier de base" pour les fichiers qui ne font pas partie de la Banque de Données et on utilise le terme de "fichier source" pour un fichier de base dans la Banque de Données.

d'accès aux variables de la Banque de Données et aux variables des fichiers non regroupés

-les fichiers de base que l'on regroupe au sein de la Banque de Données n'ont pas forcément la même année de début des observations, on se verra donc obligé :

-de tronquer les vecteurs d'observations des variables débutant avant 1953. On perd donc de l'information

-de perdre de la place dans la Banque de Données pour les variables dont le vecteur d'observations débute après 1953.

-on devra conserver les fichiers de base sous une forme non regroupée, ce qui crée pour tous les fichiers source une duplication des mêmes informations; et cela pour plusieurs raisons :

a) lors du regroupement des fichiers de base dans la Banque de Données, on est amené à tronquer certains vecteurs d'observations. Or, il peut être intéressant de posséder de plus longues séries d'observations

b) on veut garantir aussi la fiabilité des données de la Banque de Données et pour cela, on propose

-d'interdire toute écriture dans la Banque de Données au cours d'une session APL

-de procéder à la MAJ des fichiers de base sur leur version non regroupée

ce qui permet :

-de travailler normalement sur la Banque de Données même pendant une phase de MAJ

- de n'intégrer dans la Banque de Données que des fichiers ayant été contrôlés par le test de cohérence des fichiers.

2. LES ACCES AUX VARIABLES DE LA BANQUE DE DONNEES

La Banque de Données sera obtenue par le regroupement d'une vingtaine de fichiers de base et représentera \pm 20.000 variables.

La recherche d'une variable parmi 20.000 en APL nécessite la création d'une technique d'accès

En fait, les variables de la Banque de Données sont définies dans un fichier source et on peut essayer de tirer profit de cette situation.

Un accès à une variable de la Banque de Données se fera en deux temps :

- un accès au fichier source
- la recherche de la variable parmi les variables du fichier source, c'est-à-dire d'une variable parmi \pm 1.000

Cette façon de concevoir les accès aux variables de la Banque de Données présente deux avantages importants :

1. que l'on regroupe les fichiers de base au sein d'une Banque de Données ou que l'on garde des fichiers séparés, il faudra accéder à un fichier; seule la procédure de traduction du nom de fichier en accès au fichier sera différente
2. la traduction du nom de variable en accès à la variable se fera dans un contexte restreint à \pm 1.000 variables au lieu de \pm 20.000.

L'accès à une variable de la Banque de Données se traduira par deux opérations :

1. la localisation du fichier source dans la Banque de Données
2. la recherche de la variable dans le fichier source

Pour trouver une variable dans la Banque de Données, on a besoin des seules caractéristiques nom de fichier, nom de variable. Or, une variable complète représente \pm 900 à 1.000 bytes de données; pour diminuer le nombre d'accès physiques et la place mémoire nécessaires à la recherche d'une variable, on propose :

- a) de scinder les caractéristiques des variables en :
- caractéristiques d'accès (nom de fichier et nom de variable)
 - les autres caractéristiques
- b) de constituer un fichier d'accès comprenant uniquement les informations nécessaires à l'accès aux variables.
- c) de construire pour le fichier d'accès des records physiques de records logiques. Chaque record logique contiendra les informations d'accès à une variable d'un fichier source, c'est-à-dire :
- un nom de fichier
 - un nom de variable
 - un numéro du record contenant les autres caractéristiques de la variable

Un record physique sera composé de plusieurs records logiques mais se rapportant uniquement aux variables d'un seul fichier source. On pourra avoir plusieurs records physiques par fichier source (le dernier pourra éventuellement ne pas être rempli).

Le fichier d'accès de la Banque de Données sera accessible en direct sur le nom de fichier source.

Cela impliquera :

1. des records physiques de longueur fixe
2. une méthode de traduction du nom de fichier en un numéro de record dans le fichier d'accès.

1. Des records physiques de longueur fixe

La longueur d'un record physique du fichier d'accès sera constante. Comme on l'a vu plus haut, un record physique contiendra les informations d'accès aux variables d'un seul fichier source. Or, le nombre de variables d'un fichier source à l'autre varie très fort : de 40 à 1.736 avec une moyenne de \pm 1.000 variables par fichier.

On devra déterminer la taille d'un record physique de telle sorte que :

- le nombre de lectures physiques pour retrouver les variables d'un fichier source de taille moyenne soit minimisé

-la place mémoire occupée par le record ne soit pas trop importante.

2. Une méthode de traduction du nom de fichier source.

La traduction d'un nom de fichier source consistera en un numéro de record dans le fichier d'accès de la Banque de Données. Pour cela, on a besoin d'une table de correspondance contenant le nom du fichier source, le numéro du premier record dans le fichier d'accès.

Cette table sera constituée d'une vingtaine d'entrées, une par fichier source. Elle sera construite facilement par la phase de création de la Banque de Données et pourra être sauvée dans un petit fichier qu'il faut lire en début de session APL et qu'on peut conserver en permanence en mémoire.

Un accès à une variable de la Banque de Données se fera selon le schéma suivant :

- 1) traduction du nom de fichier source en un numéro de record dans le fichier d'accès; ce qui se fera par la consultation de la table de correspondance présente en mémoire. Cette première opération donnera le numéro de record à lire dans le fichier d'accès.
- 2) lecture d'un record dans le fichier d'accès. Ce record constitue un tableau pour lequel une ligne correspond aux informations d'accès à une variable, c'est-à-dire le nom de variable et le numéro de record de la variable dans le(s) fichier(s) des autres caractéristiques des variables.

Si l'on trouve une correspondance nom de variable demandée par l'utilisateur, liste de noms de variables du fichier d'accès, on peut accéder aux caractéristiques de la variable grâce au numéro de record correspondant.

Si l'on ne trouve pas de correspondance, c'est que :

- soit la variable demandée n'est pas définie dans le fichier
- soit que l'on doit lire le record suivant dans le fichier d'accès de la Banque de Données. (★)

Si l'on veut encore améliorer les performances du système au point de vue temps d'accès aux variables dans le courant d'une session APL, on peut grouper toutes les déclarations d'accès aux variables en une opération.

On peut procéder comme suit :

a) l'utilisateur donne un ordre de sélection de variables dans la Banque de Données :

```
SELECT 'nom fichier source 1 : var 1, var 2... ;
..... ; nom fichier source n : var 1, var 2...'
```

Cet ordre aura pour effet de consulter le fichier d'accès de la Banque de Données et de construire en mémoire un tableau d'accès pour les variables citées. Chaque ligne de ce tableau sera un record logique du fichier d'accès de la Banque de Données.

b) l'utilisateur pourra lire ces variables au cours de la session APL, ce qui se fera par un ordre :

```
SEARCH 'nom fichier 1;var.1 ; nom fichier 2;var. 2'
```

Cet ordre aura pour effet pour chaque variable citée de consulter le tableau d'accès construit en mémoire par l'ordre SELECT :

Si l'on trouve une correspondance dans le tableau d'accès, en une seule lecture physique, on aura accès aux caractéristiques de la variable.

Sinon, il faudra traduire 'nom de fichier i;var.i' par la procédure normale (points 1 et 2 de la page précédente).

(★) En APL, si l'on codifie les noms de variables en nombres, on pourra, en une seule instruction "C" retrouver l'indice de la ligne concernant la variable dans le tableau d'accès, donc connaître son numéro de record dans le(s) fichier(s) des autres caractéristiques des variables. La codification des noms de variables pourra se faire en traduisant le nom de variable en base 37 et ce, pour les noms de variables dans le fichier d'accès et pour les noms de variables donnés par l'utilisateur.

Cette dernière solution présente plusieurs avantages :

- d'avoir des accès aux caractéristiques des variables en un seul accès physique dans le courant d'une session APL,
- de mieux profiter des records du fichier d'accès de la Banque de Données. En effet, l'ordre SELECT permet la sélection d'une liste de variables par fichier source. La lecture d'un record du fichier d'accès de la Banque de Données donne la correspondance nom de variable, numéro de record pour un ensemble de variables du fichier source (et ce, par une seule instruction "C")

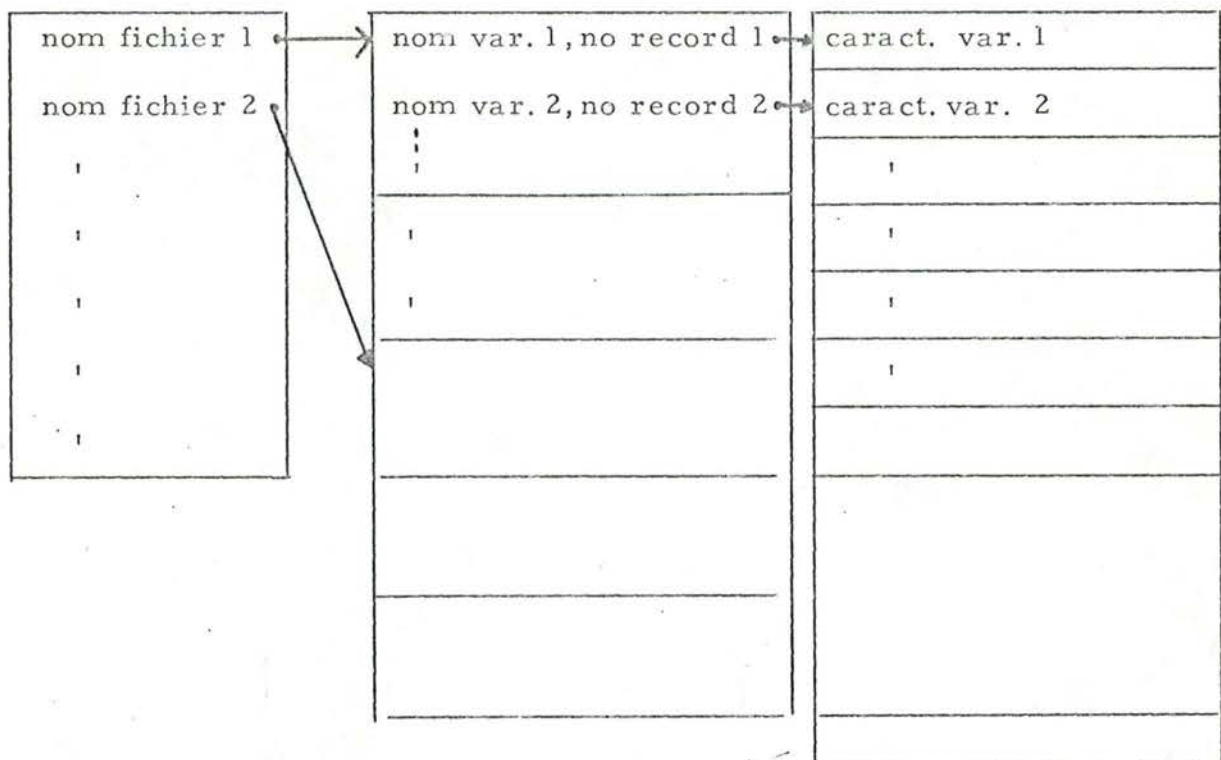
L'inconvénient de cette solution est d'occuper de la place mémoire pour le tableau d'accès.

STRUCTURE DES ACCES AUX VARIABLES DE LA BANQUE DE DONNEES

TABLE 1
en mémoire

FICHER D'ACCES

FICHER DONNEES



BUMP



0 0 3 5 9 0 6 7 2

*FM B16/1980/06

