Techniques d'enquête, juin 1991
Vol. 17, n° 1, pp. 47–61
Statistique Canada

Le système Blaise ou comment élaborer un système de traitement intégré des données d'enquête

JELKE G. BETHLEHEM et WOUTER J. KELLER[1]

RÉSUMÉ

Grâce aux progrès récents de l'informatique, les organismes nationaux de statistique sont en mesure de produire efficacement des données de première qualité. Le Bureau central de la statistique des Pays-Bas (BCS) recourt de plus en plus au micro-ordinateur pour les diverses étapes de la production de données statistiques. Dans cet article, nous examinons le rôle que jouent les logiciels et les machines dans la collecte et le contrôle des données ainsi que dans la totalisation et l'analyse. Afin de sensibiliser le lecteur aux conséquences néfastes d'une gestion anarchique d'un système décentralisé de traitement de données, nous faisons ressortir toute l'importance de l'intégration. Celle-ci rend la production de données statistiques plus facile à gérer et plus efficace. Cet article est aussi une occasion de présenter le système Blaise, élaboré par le BCS, comme un outil informatique favorisant l'intégration du traitement des données d'enquête. À partir d'une description du questionnaire d'enquête, ce système peut générer automatiquement divers programmes destinés à la collecte des données (IPAO ou ITAO) ou à la saisie et au contrôle des données (IDAO). Il peut aussi être en liaison avec d'autres progiciels. À cet égard, nous décrivons le lien qui existe entre Blaise et les progiciels "maison" Bascula (pour la pondération) et Abacus (pour la totalisation). Ainsi, le système Blaise contrôle et coordonne, et par conséquent intègre, la majeure partie des opérations d'enquête.

MOTS CLÉS: Intégration; traitement des données d'enquête; IPAO; ITAO; micro-ordinateurs; décentralisation; uniformisation.

1. INTRODUCTION

Le Bureau central de la statistique des Pays-Bas (BCS) utilise de plus en plus des micro-ordinateurs pour le traitement des données d'enquête. L'arrivée du micro-ordinateur a eu des effets considérables sur les méthodes de travail de l'organisme statistique néerlandais. Les statisticiens des divisions spécialisées prennent conscience graduellement des possibilités de ce nouvel appareil et c'est pourquoi ils l'utilisent de plus en plus fréquemment dans le quotidien.

Dans cet article, nous allons examiner le rôle des nouvelles techniques informatiques dans la collecte et le contrôle des données ainsi que dans la totalisation et l'analyse. Nous allons insister sur l'importance de l'uniformisation et de l'intégration. Ces mesures présentent trois avantages concrets: elles nous épargnent les effets néfastes d'une gestion anarchique d'un système décentralisé de traitement de données, elles facilitent la gestion de la production de données statistiques et favorisent une plus grande efficacité.

Nous présentons de plus le système Blaise, une création du BCS, comme le pivot d'un système intégré de traitement des données d'enquête. D'une part, Blaise se caractérise plus particulièrement par la cohérence qu'il introduit dans les diverses étapes de la collecte et du traitement des données. D'autre part, il favorise l'uniformisation des méthodes entre les divisions. Si toutes les divisions utilisent le même logiciel pour traiter les données de leurs enquêtes, tous parleront le même "langage" et l'échange de renseignements entre divisions sera plus simple et moins sujet à l'erreur.

[1] Jelke G. Bethlehem et Wouter J. Keller, Bureau central de la statistique, division de l'informatique, C.P. 959, 2270 AZ Voorburg, Pays-Bas.

2. LA PRODUCTION DE DONNÉES STATISTIQUES

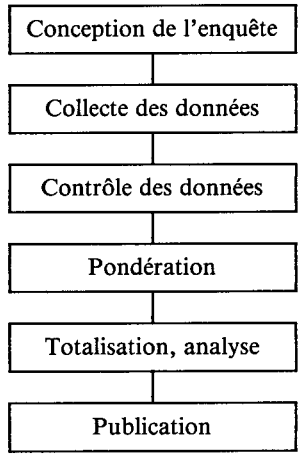
Les organismes nationaux de statistique recueillent des données sur les particuliers, les ménages et les établissements et transforment cette information en des statistiques utiles. La production de données statistiques est un processus souvent complexe, long et coûteux. Cette section vise à décrire les étapes que doit franchir l'organisme de statistique pour mener à bien une enquête, les problèmes qui peuvent se poser en cours de route ainsi que les décisions qui doivent être prises. Le tableau 1 donne un aperçu des diverses étapes d'une enquête.

La première étape est évidemment la conception de l'enquête; le statisticien doit alors déterminer quelle population sera étudiée, quelles données seront recueillies et quelles caractéristiques seront estimées. Comme les organismes statistiques recueillent la plupart de leurs données au moyen d'enquêtes par sondage, il faut élaborer un questionnaire qui contiendra toutes les questions qui seront posées aux répondants. Ce questionnaire est la première description pratique des données à recueillir. De plus, comme il s'agit de sondages, le statisticien doit aussi élaborer un plan d'échantillonnage et veiller à ce que l'échantillon soit prélevé selon les normes.

La deuxième étape du processus est la *collecte des données*. Dans de nombreuses enquêtes, les questionnaires sont remplis généralement à l'occasion d'une interview sur place: un intervieweur se rend à domicile, pose les questions et inscrit les réponses sur un formulaire. La qualité des données recueillies est généralement bonne. Cependant, comme cette méthode nécessite un grand nombre d'intervieweurs, qui peuvent tous être appelés à effectuer de nombreux déplacements, l'interview sur place est une formule qui peut s'avérer coûteuse et chronophage. C'est pourquoi on a recours parfois à l'interview téléphonique. Les intervieweurs appellent les répondants depuis les bureaux de l'organisme; ils n'ont donc plus besoin de se déplacer. En revanche, il n'est pas toujours possible de réaliser des interviews téléphoniques: seules les personnes qui ont le téléphone peuvent être rejointes de cette manière, et le questionnaire ne doit pas être trop long ni trop compliqué. L'enquête postale est une solution encore moins coûteuse car elle n'exige pas les services d'intervieweurs. On envoie un questionnaire aux répondants éventuels et ceux-ci sont priés de le retourner dûment rempli. Bien que l'on dispose de lettres de rappel dans ces conditions, cela n'équivaut pas au pouvoir de persuasion d'un intervieweur; par conséquent, le taux de réponse et la qualité des données recueillies en souffrent.

Tableau 1

La production de données statistiques



Si la collecte des données se fait au moyen de questionnaires, ceux-ci, une fois remplis, doivent faire l'objet d'un traitement approfondi. Pour obtenir des données de première qualité, il est indispensable d'extraire de ces questionnaires toutes les erreurs que l'on peut y trouver. Nous en sommes à l'étape du *contrôle des données*. On y distingue trois types d'erreur: - l'*erreur d'inscription* (lorsque la réponse donnée est en dehors de l'intervalle des réponses possibles, par ex.: âge = 348 ans); - l'*erreur de logique* (lorsqu'il y a incohérence entre des réponses à certaines questions, par ex.: il est raisonnable de croire qu'une personne peut être âgée de 8 ans, et il n'y a rien d'anormal à ce qu'une personne se dise mariée, mais lorsque ces deux caractéristiques se rapportent au même individu, il y a sûrement quelque chose qui cloche, du moins en ce qui concerne les Pays-Bas); - l'*erreur de cheminement* (lorsque l'intervieweur ou le répondant ne respecte pas les instructions relatives à la manière de remplir le questionnaire, c'est-à-dire lorsqu'il ne tient pas compte de l'enchaînement logique des questions: ainsi, des réponses sont données à des questions qui ne concernent pas l'individu, et vice-versa).

Les erreurs qui ont été décelées doivent être corrigées, mais cette tâche peut s'avérer très ardue si elle doit exécutée postérieurement, dans les bureaux de l'organisme statistique. Dans beaucoup de cas, particulièrement en ce qui concerne les enquêtes-ménages, il est impossible de joindre de nouveau le répondant; il faut donc trouver d'autres moyens de résoudre la difficulté. Parfois, on réussira à déterminer une valeur approximative acceptable à l'aide d'une méthode d'imputation mais dans d'autres circonstances, on remplacera la valeur inexacte par un code spécial indiquant que la valeur est "inconnue".

À l'étape du contrôle des données, on procède aussi parfois à ce qu'on appelle le *codage des réponses libres*. Un exemple typique de question à réponse libre est la question relative à la profession du répondant. Il est évidemment plus facile de traiter les questions qui renferment une liste de réponses pré-codées. Toutefois, dans le cas de la profession, cette liste serait interminable et le répondant aurait beaucoup de difficulté à faire le bon choix. Pour éviter ce problème, il suffit de laisser le répondant inscrire en toutes lettres sa profession sur le questionnaire. Ces réponses doivent être ensuite classées si l'on veut être en mesure d'analyser ce genre d'information. La classification des réponses est une opération longue et coûteuse qui doit être confiée à des spécialistes chevronnés.

Le contrôle des données produit un fichier "épuré", c.-à-d. un fichier ne contenant pas d'erreur. Cependant, il est encore trop tôt pour procéder à la totalisation et à l'analyse des données. En premier lieu, l'échantillonnage peut avoir été fait avec des probabilités inégales; par exemple, des établissements ont été échantillonnés avec une probabilité proportionnelle à la taille. Un choix judicieux des probabilités d'échantillonnage permet d'obtenir des estimations plus précises des paramètres de la population mais à la condition que l'on utilise une méthode d'estimation qui tient compte des différences de probabilité. En second lieu, la représentativité de l'échantillon peut être affectée par la non-réponse, c'est-à-dire le phénomène selon lequel il manque des données pour certains éléments de l'échantillon. Si le comportement des non-répondants varie en fonction des caractéristiques démographiques étudiées, les résultats de l'enquête seront biaisés.

Afin de compenser les effets de la non-réponse et de l'inégalité des probabilités d'échantillonnage, on a souvent recours à la *pondération*. À cette étape, on attribue un poids à chaque enregistrement. Les poids sont calculés de telle manière que la distribution empirique pondérée de caractères comme le sexe, l'âge, l'état matrimonial et la région reflète la distribution connue de ces caractères dans la population.

Dans les cas de non-réponse partielle, c'est-à-dire les cas où un enregistrement ne contient pas toutes les réponses voulues, on peut aussi procéder à une *imputation*. En se servant d'un modèle quelconque, on calcule des estimations pour les valeurs manquantes et on les inclut dans l'enregistrement.

À l'issue de cette étape, nous avons un fichier épuré prêt pour l'*analyse*. Celle-ci débute presque toujours par la totalisation des caractéristiques de base. La construction de tableaux n'est pas une opération aussi simple qu'on pourrait le croire. En effet, des facteurs tels la

composition des lignes et des colonnes (construites souvent à partir de plusieurs variables), les quantités figurant dans les cases (effectifs, moyennes, pourcentages), le mode de calcul des pourcentages, le traitement des variables à réponse multiple, la position des totaux et des sous-totaux et beaucoup d'autres facteurs encore peuvent compliquer grandement la tâche des statisticiens.

De nombreux organismes statistiques soumettent aussi leurs données à des analyses qui visent à en révéler la structure fondamentale; ils peuvent ainsi faire une meilleure lecture de ces données. Les résultats de ces analyses peuvent contribuer à perfectionner les enquêtes ultérieures et par conséquent à améliorer la qualité des données ou à réduire les coûts des opérations.

Enfin, les résultats des analyses sont *publiés* sous forme de rapport. Ce rapport renferme normalement des tableaux et des graphiques. Il est essentiel de présenter l'information statistique de manière à ce qu'elle soit comprise de la bonne façon. Les graphiques, les tableaux et les textes doivent être simples et limpides. La construction des graphiques mérite une attention particulière car un graphique équivoque donnera lieu très facilement à de fausses interprétations.

3. NÉCESSITÉ DE L'INTÉGRATION

L'ordinateur a toujours occupé une place importante dans le traitement des données statistiques. À l'origine, il servait uniquement à des opérations comme le tri, le dénombrement et la totalisation. Dans les années soixante et soixante-dix, l'apparition des gros ordinateurs et des progiciels statistiques a permis l'exécution d'analyses plus poussées. L'ordinateur servait aussi de plus en plus au contrôle des données, à la pondération et à l'imputation. Ce n'est que plus tard que l'ordinateur commencera à servir à la collecte des données. La première occasion sera les interviews téléphoniques (ITAO). La dernière décennie a vu naître l'ordinateur portatif; désormais, l'intervieweur peut se rendre au domicile du répondant avec un ordinateur sous le bras. Cette forme d'interview est désignée par le sigle IPAO (interview sur place assistée par ordinateur).

Il ne fait plus de doute que l'ordinateur est utilisé à des fins de plus en plus nombreuses. Il existe du matériel et des logiciels pour presque toutes les étapes du processus de production. En outre, de plus en plus de gens utilisent les outils informatiques. Au début, les ordinateurs étaient réservés à l'usage des spécialistes de l'informatique mais aujourd'hui, les statisticiens et les experts des divisions spécialisées sont rompus aux techniques de l'informatique et réclament avec toujours plus d'intérêt le matériel et les logiciels voulus pour accomplir leur tâche. À l'heure actuelle, le traitement électronique des données peut être confié aux divisions spécialisées pourvu qu'il n'implique pas d'opérations complexes; de fait, les divisions se chargent déjà de cette opération, laissant aux spécialistes de la division de l'informatique l'élaboration et la maintenance des systèmes d'information complexes. Ces changements ont eu pour conséquence de modifier le travail des statisticiens et des experts des divisions spécialisées. Auparavant, ces gens étaient des spécialistes de leur domaine propre, sans plus; mais aujourd'hui, ils ont élargi leur champ de connaissances et diversifié leur expérience en touchant aussi bien à la méthodologie statistique et à l'informatique qu'à des sujets spécialisés. Ainsi, le groupe de spécialistes a fait place à un nouveau groupe qui est plus au fait des divers aspects du traitement des données d'enquête.

Il est bien d'automatiser la production de données statistiques mais il faut être conscient des risques que cela implique. Même si l'utilisation de l'ordinateur est le gage d'une plus grande efficacité et d'une meilleure qualité des données, un usage incontrôlé des nouvelles techniques peut facilement mener à la confusion et, partant, à une productivité moindre. Voici les facteurs qui influent sur l'efficacité de la production de données statistiques:

- **Multiplicité des étapes**
  L'information passe entre plusieurs mains: le répondant remplit le questionnaire, l'agent spécialisé en vérifie le contenu et corrige les erreurs, l'agent de saisie introduit les données dans l'ordinateur et le programmeur élabore des logiciels de contrôle. L'échange de documents entre des personnes ou des divisions peut être une source d'erreur, de confusion et de retard.

- **Multiplicité des systèmes informatiques**
  Les diverses opérations liées au traitement des données peuvent ne pas être exécutées sur le même système. Le transfert de fichiers provoque des retards; en outre, une spécification et une documentation fautives peuvent être à l'origine d'erreurs.

- **Répétition des spécifications relatives aux données**
  La structure des données doit être spécifiée à presque toutes les étapes. Les questions qui reviennent le plus souvent dans les divisions sont les suivantes: Quelle est la signification des variables? Quelles valeurs peuvent-elles prendre? Des conditions s'appliquent-elles au processus de cheminement? Quelles relations entre les variables faut-il vérifier? Bien que les spécifications soient essentiellement les mêmes d'une fois à l'autre, le mode de spécification, lui, peut être très différent. Chaque système a son propre "langage". La première spécification concerne le questionnaire proprement dit. Une deuxième spécification peut être nécessaire pour la saisie des données, et une autre pour le programme de contrôle, la totalisation et l'analyse, *etc.* De toute évidence, ce n'est pas la façon la plus efficace de traiter l'information.

Devant ces difficultés, le BCS propose comme solution l'*intégration*, une intégration sur trois plans: tâches, matériel et logiciel. Examinons tout d'abord l'intégration des tâches.

Dans sa forme classique, le traitement des données consiste en ce que nous appelons des *macro-cycles*. L'ensemble des données d'une enquête passe par des cycles: d'une division à une autre et d'un système informatique à un autre. En premier lieu, la division spécialisée s'occupe d'"épurer" manuellement les questionnaires; ensuite, la section de la saisie des données voit à l'introduction des données de ces questionnaires; en troisième lieu, les fichiers sont transférés dans un gros ordinateur. Un programme soumet les données à une vérification de cohérence; les erreurs détectées sont inscrites sur une liste que l'on envoie à la division spécialisée pour qu'elle fasse les corrections nécessaires. Ce processus de saisie et de contrôle doit être répété un certain nombre de fois avant que l'on puisse dire que les données sont "épurées".

Ce que vise au fond l'intégration des tâches est le remplacement des macro-cycles par des *micro-cycles*. Autrement dit, on devrait songer à faire circuler un seul enregistrement à la fois plutôt que tout le fichier de données. La notion de micro-cycle signifie que toutes les opérations successives se font à l'intérieur du même système informatique et qu'elles sont sous le contrôle d'une seule division. Substituer le micro-cycle au macro-cycle revient à concentrer toutes les opérations de traitement des données dans une seule division, soit la division spécialisée. Comme les statisticiens des divisions spécialisées sont les personnes qui connaissent le mieux le domaine sur lequel porte une enquête, ils sont le plus en mesure de traiter les données, de résoudre les problèmes et de produire des statistiques de première qualité. Pour s'acquitter de ces tâches, il faut, bien sûr, qu'ils aient les outils appropriés, c.-à-d. des systèmes puissants et conviviaux.

L'idée de confier les opérations de traitement des données d'enquête à des spécialistes de l'informatique est révolue. Les statisticiens des divisions spécialisées se rendent compte de plus en plus des possibilités et de l'utilité de l'ordinateur pour leur travail. C'est pourquoi les divisions spécialisées devraient songer à prendre en charge les opérations liées au traitement des données d'enquête, dans la mesure où il s'agit d'opérations peu complexes. Évidemment, il revient à la division de l'informatique de mettre en place l'infrastructure nécessaire. Par ailleurs, cette division demeure responsable de l'élaboration et de la maintenance des systèmes d'information complexes.

Le second volet de la solution du BCS est l'intégration du matériel. Celle-ci vise à faire exécuter le maximum d'opérations sur un seul type d'ordinateur. Compte tenu de ce qu'un grand nombre de statisticiens peu initiés à l'informatique devront se servir d'un ordinateur,

il va de soi que l'on opte pour le micro-ordinateur. On peut ainsi s'offrir la convivialité à un prix relativement raisonnable et avoir accès à toute une série de logiciels utiles.

Sachant que les organismes statistiques traitent régulièrement d'énormes quantités de données, on peut se demander si les micro-ordinateurs peuvent à eux seuls suffire à la tâche et remplacer les gros ordinateurs. Pour être en mesure de répondre à cette question, nous devons distinguer deux types d'opérations liées au traitement des données. Premièrement, les opérations axées sur l'enregistrement, c'est-à-dire celles qui portent sur un seul enregistrement à la fois (par ex.: saisie et contrôle des données). Les opérations de ce genre se prêtent habituellement très bien au traitement interactif. Deuxièmement, les opérations axées sur le fichier, c'est-à-dire celles qui ne peuvent être exécutées convenablement que si l'on dispose de tout le fichier (par ex.: calcul des poids et totalisation). En raison de leur envergure, les opérations axées sur le fichier sont souvent exécutées suivant le principe du traitement par lots.

Il y a quelques années, on était d'avis que les opérations axées sur l'enregistrement pouvaient être faites sur micro-ordinateur mais que celles axées sur le fichier devaient être exécutées sur un gros ordinateur. Comme on fabrique des micro-ordinateurs de plus en plus puissants, l'attention est tournée de plus en plus vers le micro-ordinateur. À l'heure actuelle, la position du BCS est d'utiliser le micro-ordinateur pour toutes les opérations axées sur l'enregistrement et pour un grand nombre des opérations axées sur le fichier (pour des fichiers de moins de 50 méga-octets par exemple). En revanche, nous conservons le gros ordinateur pour le stockage des données et les gros travaux en traitement par lots.

Les personnes qui doivent utiliser un ordinateur devraient être le moins possible mises en contact avec un gros ordinateur. C'est pourquoi le BCS favorise de plus en plus l'utilisation de systèmes frontal et dorsal. Le système frontal est constitué de micro-ordinateurs et c'est ce qu'utilisent les statisticiens pour traiter les problèmes courants. Le système dorsal consiste en un gros ordinateur ou en un mini-ordinateur et sert aux gros travaux; les opérations pertinentes peuvent même se faire à l'insu de l'utilisateur. En particulier, le serveur de base de données est une option très intéressante en ce qui concerne la gestion de bases de données. Selon cette option, les opérations portant sur la base de données s'effectuent sur un mini-ordinateur spécialisé tandis qu'elles sont spécifiées, déclenchées et contrôlées au moyen d'un micro-ordinateur placé sur le bureau de l'utilisateur.

4. UNIFORMISATION

Le BCS utilise de plus en plus le micro-ordinateur (avec système MS-DOS) à beaucoup d'étapes de la production de données statistiques. D'une part cette initiative favorise un traitement plus efficace de l'information, de l'autre elle crée de nouveaux problèmes qu'il faut chercher à résoudre. Si on laisse à chaque division la liberté d'acheter le matériel et le logiciel de son choix, on risque fort de créer une situation chaotique. Les divisions ne parleront plus le même "langage" parce qu'elles n'utiliseront pas les mêmes structures de données ni les mêmes logiciels. De toute évidence, cette perspective commande une politique d'uniformisation vigoureuse. Le BCS a adopté une telle politique et en conséquence, les divisions disposent tout au plus de deux progiciels pour une tâche particulière.

Un autre avantage de l'uniformisation est qu'elle limite le nombre de cours de formation que doivent suivre les utilisateurs. Afin d'assurer la formation de nombreux nouveaux utilisateurs, le BCS organise en moyenne chaque mois 50 cours d'une journée (ce qui suffit à occuper trois salles de cours, dotées de tout le matériel nécessaire, à chaque jour ouvrable).

On doit aussi porter attention à la manière dont sont utilisés les micro-ordinateurs dans l'organisation. La distribution d'un grand nombre de micro-ordinateurs autonomes peut sembler une solution commode mais cela soulève des problèmes qu'il faut chercher à résoudre. Par exemple, il est très facile de copier des fichiers de données (confidentielles) sur le disque dur d'un micro-ordinateur autonome; il y a donc là un problème de sécurité informatique.

De plus, les utilisateurs dans les divisions spécialisées négligent souvent d'effectuer des opérations comme la sauvegarde de fichiers et l'archivage. Par ailleurs, les divisions ne peuvent se partager l'utilisation de fichiers autrement que par l'échange de disquettes. Enfin, pour les grandes organisations qui possèdent beaucoup de micro-ordinateurs autonomes, la distribution des nouvelles versions de progiciels, y compris la documentation pertinente, est une tâche souvent contraignante.

Pour éviter toutes ces difficultés, le BCS a installé quelque 60 réseaux locaux (RL). Chaque division a son propre RL. De dix à soixante micro-ordinateurs sont reliés à un serveur de fichiers haut de gamme articulé sur un microprocesseur 386 et dont la capacité de mémoire peut atteindre 600 méga-octets. Ainsi, près de 2,300 micro-ordinateurs sont reliés de cette manière à un serveur de fichiers et la moitié d'entre eux sont articulés autour du microprocesseur 386SX d'Intel. La sécurité est assurée de trois façons: - protection par mot de passe dans la procédure d'entrée en communication, - chiffrement et - postes de travail sans unité de disquette (seulement 60 des 2,300 micro-ordinateurs sont équipés d'une unité de disquette ou de disque dur). Les opérations d'archivage et de sauvegarde des RL s'effectuent de façon centralisée sous la responsabilité de la division de l'informatique. Chaque nuit, on procède à une sauvegarde complète de plus de 15 giga-octets. Il est clair qu'un tel contexte d'utilisation facilite des opérations comme le contrôle de la version et la mise à jour du logiciel. Avec un RL, la distribution et l'installation des nouvelles versions de logiciels deviennent des tâches simples puisque par une seule commande, on peut télécharger de nouvelles versions dans tous les serveurs de fichiers à la fois. Tous les permis d'utilisation de logiciel prévoient un usage en commun, qui est contrôlé par des logiciels "maison".

Le micro-ordinateur joue un rôle de plus en plus grand dans la production de données statistiques, mais le gros ordinateur ou le mini-ordinateur sont encore indispensables pour certaines applications, comme l'exploitation de grosses bases de données. Dans ce contexte, le BCS a fait de Oracle son système de gestion de bases de données standard. L'élaboration d'un programme de gestion de base de données se fait de préférence sur un micro-ordinateur alors que l'exécution de ce programme se fait sur un mini-ordinateur. Dernièrement, le BCS a élaboré une architecture client-serveur fondée sur un système de gestion de base de données réparties. Les micro-ordinateurs du réseau constituent le système frontal et les mini-ordinateurs, le système dorsal.

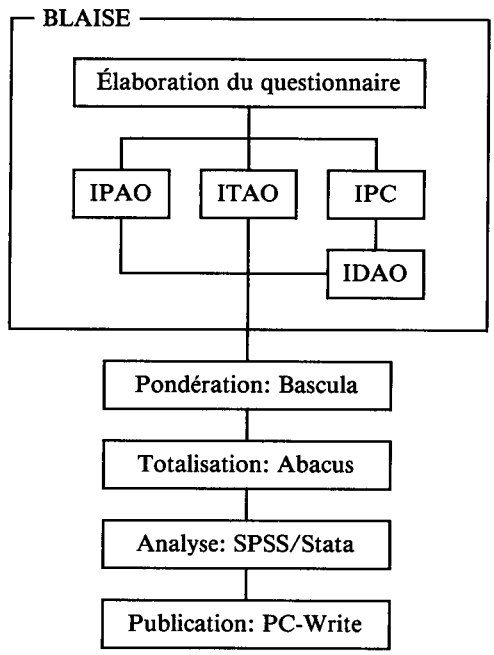
Ainsi, plus les outils informatiques sont à la portée des agents spécialisés des divisions (décentralisation), plus il est nécessaire de centraliser les opérations d'uniformisation et de coordination de l'univers informatique des utilisateurs. Pour plus de détails sur l'infrastructure informatique, voir Keller, Metz et Bethlehem (1990).

5. INTÉGRATION DU TRAITEMENT DES DONNÉES D'ENQUÊTE

Dans la section précédente, nous avons discuté de la nécessité d'une intégration du traitement des données d'enquête. Il a été question plus particulièrement de la concentration des tâches dans les divisions spécialisées et de l'uniformisation du matériel et du logiciel. Or, l'uniformisation du logiciel ne suffit pas. On peut accroître davantage l'efficacité de la production de données statistiques en intégrant les logiciels nécessaires dans un système. Cette section vise à décrire la manière dont le BCS a mis en oeuvre un système intégré de traitement des données d'enquête.

Un système intégré de traitement des données d'enquête doit reposer sur un langage puissant pour la spécification des questionnaires. Cette spécification est la "base de connaissances", qui contient tous les faits relatifs au questionnaire et aux données. Le système intégré doit pouvoir exploiter ces connaissances, c'est-à-dire qu'il doit pouvoir générer automatiquement toutes les applications requises pour le traitement des données, soit d'une part, les logiciels pour la collecte, la saisie et le contrôle des données et d'autre part, les interfaces avec d'autres

Tableau 2
Système intégré de traitement des données d'enquête



logiciels de traitement de données (par ex.: logiciels pour la totalisation et l'analyse). Ainsi, la répétition des spécifications relatives aux données n'a plus raison d'être et la cohérence des données est observée à toutes les étapes du traitement.

Le pivot du système intégré de traitement des données d'enquête mis au point par le BCS est le système Blaise. À l'étape de la conception de l'enquête, le questionnaire est spécifié en langage Blaise et c'est cette spécification qui sert, durant tout le processus, à extraire les données requises pour les diverses étapes du traitement. Le tableau 2 décrit schématiquement le système intégré de traitement des données d'enquête.

Le système Blaise peut générer trois types de programmes: IDAO, IPAO et ITAO (le sigle IDAO signifie "introduction de données assistée par ordinateur"). Ce système intègre les opérations de saisie et de contrôle des données en permettant le traitement interactif des questionnaires. Il peut aussi produire le logiciel nécessaire pour les IPAO ou les ITAO. Le système Blaise est décrit plus en détail dans la section 6.

Quel que soit le mode de collecte des données, nous obtiendrons toujours un fichier "épuré", c.-à-d. un fichier où on ne peut plus trouver d'erreurs. Viendra ensuite le calcul de facteurs de pondération. Le programme Bascula est conçu spécialement à cette fin. Il peut lire directement les fichiers Blaise et extraire de la spécification Blaise l'information relative aux variables, c'est-à-dire les métadonnées. L'exécution de Bascula aura pour effet d'ajouter une variable dans le fichier contenant les facteurs de pondération pour les divers cas. Le programme Bascula est décrit plus en détail dans la section 7.

Le fichier est maintenant prêt pour la totalisation et à cette fin, le système intégré met à notre disposition le programme Abacus. Ce programme peut, lui aussi, lire et comprendre les fichiers créés à l'étape précédente. Voir la section 8 pour plus de détails. La totalisation peut être suivie d'une analyse plus approfondie des données. À cette fin, le système Blaise peut générer des interfaces pour les progiciels statistiques SPSS et Stata. La section 9 contient plus de détails à ce sujet.

Enfin, on préparera une publication à l'aide du logiciel de traitement de texte PC-Write. Comme ce logiciel peut tourner sur le même système que les autres logiciels, il est facile d'importer dans le texte des tableaux et des résultats d'analyses statistiques.

6. LE SYSTÈME BLAISE

Le système Blaise a été élaboré par le BCS et il doit son nom au célèbre théologien et mathématicien français Blaise Pascal (1623-1662). La base de ce système est le langage Blaise, qui sert à spécifier de façon formelle la structure et le contenu du questionnaire d'enquête. Le langage Blaise tire son origine, en majeure partie, du langage de programmation Pascal.

Le système Blaise tourne sur des micro-ordinateurs (ou des réseaux de micro-ordinateurs) dont le système d'exploitation est MS-DOS. Il est le pivot du système intégré de traitement des données d'enquête et comme il est destiné à l'usage des employés des divisions spécialisées, nul n'a besoin d'être informaticien pour l'utiliser. On a conçu le système Blaise dans le but d'offrir aux spécialistes des divisions un outil puissant mais convivial qui leur permette de saisir les données relatives à une enquête et de voir à l'exécution de toutes les étapes subséquentes.

Selon le système Blaise, la réalisation d'une enquête débute par l'élaboration d'un questionnaire en langage Blaise. Un questionnaire élaboré de cette façon contient plus de renseignements qu'un questionnaire classique. On y décrit non seulement les questions, les réponses possibles et les conditions de cheminement dans le questionnaire, mais aussi les relations entre les réponses qui doivent faire l'objet d'un contrôle.

Le système Blaise peut générer des programmes pour l'IDAO ainsi que les IPAO et les ITAO. Un programme IDAO est un système intelligent et interactif destiné à la saisie et au contrôle des données recueillies à l'aide de formulaires. L'agent de la division spécialisée traite un à un les formulaires au moyen d'un micro-ordinateur. Il inscrit les réponses à l'endroit approprié et une fois le questionnaire terminé, il actionne la fonction contrôle pour vérifier s'il n'y aurait pas d'erreur de cheminement ou d'erreur de logique. Les erreurs détectées sont affichées à l'écran avec une explication. L'agent peut corriger ces erreurs soit en se reportant au formulaire ou bien en appelant la personne qui a fourni les renseignements en question. Une fois que toutes les erreurs ont été corrigées, l'enregistrement "épuré" est versé dans un fichier.

Le programme IDAO peut aussi servir à une forme de traitement de données dont on n'a pas fait mention jusqu'à maintenant. En effet, les organismes statistiques ne recueillent pas toujours eux-mêmes les données dont ils ont besoin; ils sont appelés parfois à construire des statistiques à l'aide de fichiers provenant de l'extérieur. Ces données doivent tout de même être vérifiées. Le système Blaise comporte une fonction qui permet d'importer les fichiers de ce genre. Grâce au programme IDAO, on peut soumettre tous les enregistrements à un contrôle intégral par lots. Les enregistrements sont alors classés en deux groupes: "sans erreur" et "avec erreur". Les seconds peuvent être corrigés en mode interactif grâce, une fois de plus, au programme IDAO.

Les programmes IPAO et ITAO peuvent servir aux interviews assistées par ordinateur. Dans ce cas, le questionnaire classique est remplacé par un programme informatique qui contient les questions à poser. Ce programme contrôle entièrement l'interview. Il détermine au fur et à mesure l'ordre dans lequel les questions doivent être posées et vérifie les réponses dès qu'elles ont été saisies. En ce qui concerne l'IPAO, le programme est chargé dans un ordinateur portatif que l'intervieweur apporte au domicile du répondant. Pour ce qui a trait à l'ITAO, le programme est chargé dans un ordinateur de bureau. L'intervieweur appelle le répondant depuis une unité centrale et réalise l'interview par téléphone.

La génération d'un programme IDAO, IPAO ou ITAO de Blaise se fait en un certain nombre d'étapes. Premièrement, on introduit la spécification en langage Blaise du questionnaire au moyen d'un programme de mise en forme de texte; on vérifie ensuite s'il n'y a pas d'erreur de syntaxe. Les erreurs détectées doivent être corrigées et à cette fin, le système rappelle le

Tableau 3

Exemple d'un questionnaire Blaise

```
QUESTIONNAIRE de "l'enquête sur le travail";

QUEST
  NumOrd     "Numéro d'ordre de l'interview?": 1..1000 (CLÉ);
  Age        "Quel âge avez-vous?": 0..99;
  Sex        "Êtes-vous de sexe masculin ou féminin?": (Masculin, Féminin);
  EtaMatr    "Quel est votre état matrimonial?":
               (Marié     "Marié",
               NonMar     "Non marié")
  Emp        "Occupez-vous un emploi?": (Oui, Non);
  DesEmp     "De quel genre d'emploi s'agit-il?": [20] CARACTÈRES;
  Revenu     "Quel est votre revenu annuel?":
               (Inf20     "Moins de 20,000",
               Jusq40     "Entre 20,000 et 40,000",
               Plus40     "Plus de 40,000");
  Deplac     "Quel moyen de transport utilisez-vous habituellement pour aller au travail?":
               SÉRIE [3] DE
               (Pied      "à pieds",
               Bicycle    "à bicyclette",
               Aut        "en voiture ou en motocyclette",
               TransCom   "autobus, tramway, train ou métro",
               Autre      "aucun de ceux-là");
  AutMoyen   "Quel autre moyen de transport utilisez-vous?": [20] CARACTÈRES;

CHEMINEMENT
  NumOrd; Age; Sex; EtaMatr; Emp;
  SI Emp = Oui ALORS
    DesEmp; Revenu; Deplac;
    SI Autre sous Deplac ALORS AutMoyen FIN
  FIN

CONTRÔLE
  SI Age < 15 "Répondant n'a pas encore 15 ans" ALORS
    EtaMatr = NonMar "trop jeune pour être marié!"
  FIN

FINQUESTIONNAIRE.
```

programme de mise en forme de texte et place le curseur à l'endroit où se trouve approximativement l'erreur. Après la correction des erreurs, on soumet la spécification à un nouveau contrôle. Si aucune autre erreur n'est détectée, la spécification est transformée en code source Pascal, lequel est compilé en un programme exécutable.

Le langage Blaise a un double objet quelque peu contradictoire. D'une part, il doit être assez puissant pour pouvoir être appliqué à toutes les grandes enquêtes complexes et d'autre part, les spécifications de questionnaires en langage Blaise doivent être suffisamment lisibles pour que les agents spécialisés puissent s'en servir. De fait, l'information contenue dans un questionnaire Blaise doit être explicite; autrement dit, cette information représente la description fondamentale de l'enquête et est utilisable par tous les intéressés. Le tableau 3 reproduit un questionnaire simple en langage Blaise.

La première partie du questionnaire est désignée par le sigle QUEST; cette section contient la définition de toutes les questions qui peuvent être posées. Une définition se compose d'un symbole d'identification (pour usage interne dans le questionnaire), du libellé de la question telle qu'elle est posée au répondant et d'une énumération des réponses possibles. La deuxième partie du questionnaire correspond à la section CHEMINEMENT. Cette section indique dans quel ordre les questions doivent être posées et à quelles conditions. La section CONTRÔLE sert à définir les vérifications de cohérence.

La description ci-dessus ne reflète pas toute la puissance du langage Blaise. Pour avoir une idée générale de ce langage, prière de se référer à Bethlehem et coll. (1989b) et pour plus de détails, voir Bethlehem et coll. (1989c).

Le système Blaise comprend un module pour le *codage interactif*; ce module permet d'intégrer des opérations de codage à l'étape de la collecte des données ou à l'étape de la saisie et du contrôle. Le module se compose de deux "fonctions". La première met en application une approche hiérarchique du codage. Pour coder une réponse, l'opérateur commence par entrer le premier chiffre du code en choisissant la catégorie pertinente dans un menu. Une fois le premier chiffre saisi, le programme affiche un nouveau menu qui est une description plus détaillée de la catégorie choisie précédemment. Ainsi, on obtient une description de plus en plus détaillée à mesure qu'on approche du dernier chiffre. La seconde "fonction" du module représente une approche alphabétique du codage. Elle vise à trouver une description dans une liste alphabétique. Si cette description n'est pas repérée, le programme affiche la liste à partir de la description la plus près possible de la description recherchée. La liste peut être dressée de manière que l'on puisse y trouver presque toutes les descriptions possibles, y compris les permutations. Cette méthode a l'avantage d'être simple, rapide et contrôlable. Les deux fonctions de codage peuvent être appliquées simultanément.

7. BASCULA

Une fois "épuré", le fichier de données d'enquête produit par le système Blaise n'est pas encore prêt à servir à des inférences concernant la population d'où a été tiré l'échantillon parce que les données disponibles ne constituent pas un échantillon représentatif; il faut donc procéder à certains ajustements.

Afin de tenir compte de la non-réponse et de l'inégalité des probabilités de sélection, il faut souvent calculer des facteurs de pondération. La stratification a posteriori est une méthode de pondération bien connue. Chaque enregistrement reçoit un poids calculé de telle manière que la distribution empirique pondérée de caractères tels le sexe, l'âge, l'état matrimonial et la région reflète la distribution (connue) des mêmes caractères dans la population. Deux facteurs importants peuvent compliquer la stratification a posteriori: l'existence de strates vides et le manque d'information pertinente sur la population. Des recherches ont été faites au BCS afin d'améliorer les méthodes de pondération. Ces recherches ont abouti à la mise en oeuvre d'une nouvelle méthode générale de pondération qui permet de calculer les poids à l'aide d'un modèle linéaire qui établit un rapport entre les variables étudiées dans une enquête et les variables auxiliaires. La stratification a posteriori est un cas particulier de cette méthode. En raison de la généralité de la méthode, on peut appliquer divers schémas de pondération qui exploitent le plus possible l'information existante sur la population sans soulever les problèmes mentionnés plus haut. Voir Bethlehem et Keller (1987) pour plus de détails.

Bascula est un programme général de pondération qui tourne sur des micro-ordinateurs dont le système d'exploitation est MS-DOS. Ce programme réunit plusieurs méthodes de pondération. En premier lieu, il peut exécuter une stratification a posteriori classique. Et si le nombre de strates vides est faible, il peut fondre (c'est-à-dire combiner) ces strates avec des strates avoisinantes. Lorsque le nombre de strates vides est élevé ou qu'il manque d'informations sur la population, Bascula peut exécuter la pondération linéaire décrite plus haut ou effectuer un

ajustement proportionnel itératif (aussi appelé pondération multiplicative ou méthode itérative du quotient). Les poids ainsi calculés peuvent soit être ajoutés au fichier de données ou être stockés dans un fichier à part.

Bascula peut lire directement les fichiers Blaise et extraire de la spécification Blaise l'information pertinente sur les variables. Il revient à l'utilisateur de fournir l'information sur la population. Le programme est à base de menus, donc convivial. Il exécutera une stratification a posteriori complète si cela est possible. Sinon, l'utilisateur doit choisir entre une pondération linéaire ou une pondération multiplicative.

À l'heure actuelle, Bascula ne peut être utilisé que sur un micro-ordinateur. Nous prévoyons mettre au point une version "dorsale" qui tournera sur un gros ordinateur. Bascula a été conçu spécialement pour les enquêtes sociales et démographiques, où la stratification a posteriori se mêle à des méthodes d'estimation relativement simples. En ce qui concerne les enquêtes économiques, il faudra élaborer un logiciel différent qui mettra l'accent sur l'échantillonnage stratifié et des estimateurs plus complexes (estimateur par quotient; estimateur par régression).

8. ABACUS

La totalisation est une des principales étapes de la production de données statistiques et elle a été parmi les premières à être informatisées. Il existe déjà de nombreux progiciels de totalisation dans le monde mais peu sont vraiment conviviaux. Cela s'explique en partie par le fait qu'il faille définir un grand nombre de paramètres pour produire un tableau complexe selon les normes; ces paramètres peuvent être: les variables à utiliser pour les diverses dimensions (ligne, colonne, couche); le choix entre l'enchaînement des variables (c.-à-d. présenter toutes les valeurs d'une variable, puis toutes celles d'une autre variable) ou leur emboîtement (pour chaque valeur d'une variable, présenter toutes les valeurs possibles d'une autre variable) pour une dimension donnée; la quantité figurant dans les cases (fréquence, pourcentage, total, moyenne); le choix de présenter ou non les totaux et les sous-totaux; et de nombreuses caractéristiques de présentation. Pour pouvoir maîtriser tous ces paramètres, les progiciels classiques renferment des langages de commande pour définir les tableaux et ces langages ne sont pas très faciles à apprendre ni à utiliser.

Abacus est un progiciel de totalisation qui tourne sur des micro-ordinateurs dont le système d'exploitation est MS-DOS. Bien qu'on puisse le voir comme un progiciel de totalisation parmi tant d'autres, Abacus a été élaboré de façon très particulière. Premièrement, les tableaux ne sont pas définis au moyen de langages de commande. Le programme est plutôt à base de menus. L'utilisateur conçoit son tableau d'une manière simple et intuitive en mode interactif sans avoir à connaître aucun langage de commande. Deuxièmement, Abacus peut lire directement les fichiers créés par le système Blaise, de même que les fichiers ASCII. Les métadonnées, c.-à-d. l'information relative aux variables dans le fichier, peuvent être produites par le système Blaise ou encore (dans le cas de fichiers ASCII indépendants) être introduites en mode interactif par l'utilisateur. Troisièmement, le programme peut produire des tableaux prêts à la photo.

Une autre caractéristique remarquable d'Abacus est sa vitesse. Il prend environ 6 secondes pour produire un tableau que SPSS produit en 3 minutes (sur le même micro-ordinateur de type 386SX). Si Abacus est aussi rapide, c'est à cause de sa taille relativement modeste; il peut donc utiliser une grande partie de la mémoire comme zone de travail, et ainsi produire des tableaux pouvant contenir jusqu'à 90,000 cases.

Les tableaux générés par Abacus peuvent avoir jusqu'à trois dimensions (couches, lignes et colonnes). Le nombre de variables pour chaque dimension peut atteindre 10; ces variables peuvent être emboîtées ou enchaînées. Dans le tableau 4 par exemple, les variables "emploi" et "sexe" (colonnes) sont présentées selon le mode enchaînement tandis que les variables "région" et "ville" (lignes) sont présentées selon le mode emboîtement. Ce tableau ne comporte pas une troisième dimension (couches).

Tableau 4

Exemple d'un tableau à deux dimensions

Nombre d'enregistrements	Population de Samplonie				
	Total	Emploi		Sexe	
		Oui	Non	M	F
Total	1,000	341	659	511	489
Agria	293	121	172	145	148
Wheaton	144	60	84	70	74
Greenham	94	38	56	44	50
Newbay	55	23	32	31	24
Induston	707	220	487	366	341
Oakdale	61	26	35	36	25
Crowdon	244	73	171	128	116
Smokeley	147	49	98	80	67
Mudwater	255	72	183	122	133

Source: Bureau statistique de Samplonie.

Le tableau ci-dessus contient de simples fréquences, mais Abacus peut calculer aussi des totaux de variables quantitatives et produire des tableaux de pourcentages et des tableaux de moyennes. De plus, les cases d'un tableau produit par Abacus peuvent contenir plus d'un élément à la fois (de fait, elles peuvent en contenir jusqu'à 10). L'utilisateur doit alors décider de la présentation du tableau (c.-à-d. classer les éléments par ligne, par colonne ou par couche). Si des données ont été recueillies au moyen d'une enquête par sondage, Abacus peut être appliqué à des données pondérées; il peut s'agir, par exemple, de poids calculés au moyen du programme Bascula. L'utilisateur n'a plus qu'à définir la variable qui contient les poids.

On accorde beaucoup d'attention à la présentation du tableau car les tableaux produits par Abacus doivent être prêts à la photo. Abacus renferme donc de nombreuses options qui permettent de déterminer la présentation. Ainsi, on peut définir jusqu'à 10 lignes de texte pour l'en-tête et le cartouche d'un tableau et choisir soit des lignes horizontales et verticales (comme dans l'exemple ci-dessus), soit des lignes horizontales seulement ou ni l'une ni l'autre. La disposition du texte dans les en-têtes de colonne et la largeur des colonnes peuvent aussi être modifiées.

On peut procéder à un arrondissement afin de conserver le caractère confidentiel des données du tableau. Non seulement les totaux de case, mais aussi les totaux marginaux sont arrondis au multiple d'une constante donnée, par ex.: 5. Abacus exécute aussi bien l'arrondissement normal que l'arrondissement aléatoire. Si l'utilisateur n'est pas satisfait du tableau obtenu, il peut importer le produit d'Abacus dans le tableur Lotus 123 et pousser plus loin le traitement. Une dernière option qui mérite d'être soulignée ici est la possibilité de créer de nouvelles variables en récrivant les variables existantes (par ex.: âge groupe d'âge). Pour plus de détails sur le programme Abacus, voir Bethlehem et coll. (1989a).

9. ANALYSE

Le BCS n'a pas élaboré de logiciel pour l'analyse des données statistiques principalement parce que le marché compte suffisamment de bons progiciels statistiques. Le BCS utilise lui-même les progiciels SPSS (sur gros ordinateur et sur micro-ordinateur) et Stata (sur micro-ordinateur). Pour que ces progiciels puissent faire partie du système intégré de traitement des

données d'enquête, nous devons trouver des "fonctions" qui nous permettrons d'exporter des données du système Blaise vers ces progiciels. Cela se fait en deux étapes. Premièrement, le fichier de données, en format Blaise, est converti en format ASCII et deuxièmement, l'information relative aux variables, qui est consignée dans le questionnaire Blaise, est traduite de manière à pouvoir être comprise par le progiciel en question. Un fichier provisoire est donc créé. En exécutant ce fichier à partir du progiciel statistique, on se trouve à créer un fichier informatique. Et en chargeant ce dernier, l'utilisateur peut débuter immédiatement son analyse sans se préoccuper de définir les variables, les étiquettes, *etc.*

La procédure que nous venons de décrire ne peut être réalisée qu'avec les progiciels SPSS et Stata et aucun autre. Nous pourrions, bien sûr, élaborer la même procédure pour chaque progiciel statistique connu mais cela exigerait un long travail de programmation. Nous avons choisi une autre solution. Le système Blaise contient un utilitaire spécial, générateur de fichiers provisoires. L'utilisateur "colore" la structure du fichier provisoire dans un logiciel de traitement de texte et exécute le générateur de fichiers provisoires en utilisant comme données d'entrée la description générale qu'il vient de faire du fichier provisoire et le questionnaire Blaise; il obtient ainsi un vrai fichier provisoire. Le générateur de fichiers provisoires permet donc à l'utilisateur de créer des fichiers provisoires pour les progiciels de son choix.

10. CONCLUSION

L'arrivée du micro-ordinateur a eu une incidence considérable sur les opérations des organismes nationaux de statistique. Le statisticien spécialisé recourt de plus en plus au micro-ordinateur et a besoin, pour son travail, d'un système intégré de traitement des données d'enquête comme celui fondé sur le système Blaise. Ce dernier se distingue d'une part par la cohérence qu'il introduit dans les diverses étapes de la collecte et du traitement des données. Il est ainsi plus facile de gérer et de contrôler le processus dans son entier. D'autre part, le système Blaise favorise l'uniformisation entre les diverses divisions. Comme toutes les divisions utilisent le même logiciel pour leurs enquêtes, l'échange de renseignements se fait plus facilement et il y a un moins grand risque d'erreur.

Le système intégré de traitement des données d'enquête est conçu plus spécialement pour des organisations très centralisées, comme le Bureau central de la statistique des Pays-Bas. Dans une organisation de cette envergure, le système intégré de traitement peut amener un accroissement substantiel de l'efficacité. Or, les organismes statistiques n'ont pas tous une structure centralisée. Dans les grands pays en particulier, le traitement des données est souvent décentralisé. Les bureaux régionaux s'occupent du traitement des données dans leurs régions respectives et font parvenir ensuite les fichiers constitués à l'administration centrale. Celle-ci combine les fichiers régionaux en un seul fichier national. Le système intégré de traitement des données d'enquête peut trouver sa place dans un telle structure: l'administration centrale élabore le questionnaire Blaise et envoie à chaque bureau régional des exemplaires du programme de saisie des données ainsi élaboré. De cette manière, on assure la cohérence des opérations de collecte et de contrôle des données dans les diverses régions. Comme les fichiers régionaux auront tous le même format (Blaise), il sera facile de les combiner en un seul fichier national en se servant des fonctions du système Blaise. De plus, comme ces fichiers seront tous "épurés", il ne sera pas nécessaire d'exécuter d'autres contrôles au niveau national. L'administration centrale n'aura plus qu'à analyser les données et les mettre en tableau et à publier les résultats de ses analyses. En outre, les bureaux régionaux ou l'administration centrale pourront publier des données régionales.

Le système Blaise est en usage depuis le milieu de 1986 et sert à un très grand nombre d'enquêtes. On le perfectionne en étroite collaboration avec les utilisateurs et chaque nouvelle version renferme des améliorations. Abacus est en usage depuis plus d'un an et est devenu très populaire auprès des utilisateurs. Quant au programme Bascula, il est encore en voie d'élaboration. On vient juste de lancer le premier prototype.

Nous n'avons pas parlé de la possibilité d'exporter des données et des métadonnées du système Blaise vers le progiciel de base de données Paradox. Nous projetons aussi d'établir un lien entre Blaise et le système de gestion de base de données Oracle. Ainsi, il sera possible de créer une architecture client-serveur pour les utilisateurs de Blaise.

À l'extrémité de la chaîne de production des données statistiques, il y a encore quelques aspects de la publication qui méritent notre attention. En premier lieu, il faudra élaborer un logiciel qui permettra d'évaluer les risques de divulgation d'information confidentielle qui accompagnent la publication de données statistiques. On offrira aussi des "fonctions" qui permettront de mettre les tableaux ou les fichiers à l'abri de ces risques. Enfin, les organismes statistiques se livrent de plus en plus à la publication électronique de données statistiques (par ex.: information sur disquette, sur disque compact-ROM, *etc.*). Pour aider les utilisateurs de ce genre de support à choisir les séries de données dont ils ont besoin, il est nécessaire de mettre à leur disposition des logiciels conviviaux. C'est ce à quoi nous nous occupons présentement.

BIBLIOGRAPHIE

BETHLEHEM, J.G., VAN BUITENEN, A.A.A., HUNDEPOOL, A.J., ROESINGH, M.J., et VAN DE WETERING, A. (1989a). Abacus 1.0, A Tabulation Package, Compact Guide. CBS report, Voorburg: Netherlands Central Bureau of Statistics.

BETHLEHEM, J.G., HUNDEPOOL, A.J., SCHUERHOFF, M.H., et VERMEULEN, L.F.M. (1989b). Blaise 2.0/An Introduction. CBS report, Voorburg: Netherlands Central Bureau of Statistics.

BETHLEHEM, J.G., HUNDEPOOL, A.J., SCHUERHOFF, M.H., et VERMEULEN, L.F.M. (1989c). Blaise 2.0/Language Reference Manual. CBS report, Voorburg: Netherlands Central Bureau of Statistics.

BETHLEHEM, J.G., et KELLER, W.J. (1987). Linear Weighting of Sample Survey Data. *Journal of Offical Statistics* 3, 141-154.

KELLER, W.J., BETHLEHEM, J.G., et METZ, K.J. (1990). The impact of micromputers on survey processing at the Netherlands Central Bureau of Statistics. *Proceedings of 1990 Annual Research Conference, U.S. Bureau of the Census*, 637-645.