

CAPI PLUS et Blaise III : une organisation générale pour les enquêtes de l'Insee

*F. Dussert et G. Luciani, Institut National de la Statistique et des Etudes Economiques (INSEE)
France*

La réussite du passage sur micro-portable de la collecte de l'enquête emploi puis celle de l'enquête budget de famille conduisent l'Insee à généraliser son utilisation pour toutes les enquêtes auprès des ménages.

CAPI PLUS met en place un dispositif général qui doit permettre d'intégrer rapidement tout nouveau questionnaire dans une structure préétablie. Hormis les transmissions, ce dispositif doit intégrer à la fois la gestion de la collecte (case management), le poste de contrôle et d'apurement, définir une ergonomie du poste enquêteur et des normes de développement. Blaise III offre de ce point de vue de grandes possibilités.

1. De l'enquête emploi à CAPI PLUS

Après une implantation progressive, la collecte de l'enquête emploi (Labour Force Survey) est réalisée entièrement sur micro ordinateurs portables depuis mars 1994 : les 900 enquêteurs ont tous utilisé le micro et transmis leur données vers le centre collecteur central à l'aide d'un logiciel spécialisé (Arbiter). En 1994-1995, l'enquête budget de famille (Family Expenditure Survey) était également réalisée avec CAPI, mais les transmissions n'étaient pas assurées : les enquêteurs envoyaient leurs données sur disquette à l'Insee régional pour vérification. A la différence de l'enquête emploi, un poste de vérification (manipula) permettait de vérifier la collecte. En 1995 également, la direction régionale de l'Insee à l'île de La Réunion (dans l'Océan Indien) développait une enquête logement en Blaise 2.4. Chacune de ces applications est réalisée indépendamment des autres.

La réussite en qualité et en délai de ces opérations conduit l'Insee à généraliser progressivement l'utilisation de la saisie sur micro-portable pour toutes ses enquêtes. Restait à créer une véritable infrastructure qui puisse "accueillir" toute nouvelle enquête auprès des ménages.

2. Pourquoi une infrastructure générale ?

D'une enquête à l'autre il a été difficile d'utiliser le travail déjà réalisé ; ainsi le système mis en place pour l'enquête emploi n'a pu être réutilisé pour l'enquête budget de famille notamment pour les transmissions et chaque opération a un coût élevé en conception informatique.

Cette dissociation des opérations conduit à investir de manière importante à chaque nouvelle enquête dans la programmation, l'organisation et la cohabitation sur un même micro de plusieurs

applications. De plus, chaque fois, l'enquêteur doit s'adapter à une nouvelle ergonomie. Par souci d'offrir aux enquêteurs un poste de travail confortable et efficace, il est nécessaire de leur fournir un outil avec un seul mode d'accès quelque soit l'enquête.

Cette difficulté à capitaliser l'expérience est en partie liée à l'organisation de l'Insee.

D'une part chaque "concepteur" d'une enquête en est responsable de sa conception à sa publication : il élabore le questionnaire, rédige les instructions aux enquêteurs puis exploite les résultats et réalise des études à partir du fichier "nettoyé". La durée d'un cycle d'enquête est de trois ans environ. Dans ce cadre, l'apprentissage d'un logiciel de questionnement est une contrainte supplémentaire et il n'est pas souhaitable qu'il investisse dans une programmation relativement complexe.

D'autre part, la sphère informatique est dissociée de la sphère statistique. Dans ce contexte, les rapports entre les uns et les autres doivent être formalisés et les rôles précisés.

Enfin, la responsabilité de la collecte est éloignée géographiquement de la conception. Le réseau de collecte est décentralisé et chaque région Insee est responsable de la gestion, la formation et le contrôle des enquêteurs de sa région. Cette façon de faire nécessite une formalisation et une coordination importante des procédures.

3. Une infrastructure matérielle, organisationnelle et logicielle

L'infrastructure générale de CAPI PLUS vise à mettre à la disposition de tout concepteur désireux de réaliser une enquête en saisie portable l'infrastructure organisationnelle, matérielle et informatique. Elle recouvre à la fois un réseau de collecte équipé et formé à la réalisation d'enquêtes en saisie portable, une infrastructure de communications et de recueils des résultats, un logiciel général de questionnement, un corps de règles et de méthodes. Les différents postes de travail concernés : ceux des enquêteurs, ceux des gestionnaires d'enquêtes en région et le poste du statisticien combinent ces composants.

1. Le poste de travail de l'enquêteur comprend un module de formation, une messagerie et un menu lui permettant d'accéder aux enquêtes à traiter. Pour chacune d'elle, il dispose d'un carnet de tournée pour obtenir, en amont la liste des logements à enquêter et "entrer" dans le questionnaire, en aval l'état d'avancement de son travail et son carnet de rendez-vous.
2. Le poste de gestion de l'enquête en région permet de récupérer l'échantillon, charger les micros des enquêteurs et recevoir les données. Des programmes standards permettent également pour chaque enquête de suivre l'avancement de la collecte de l'enquêteur et vérifier la qualité des questionnaires à partir de contrôles statistiques paramètres. Ce poste doit permettre également de faire les premiers redressements.

3. Le poste de travail du "concepteur-statisticien" auquel le projet CAPI PLUS fournit un ensemble de programmes utilisables sans modification pour toutes les enquêtes. Il constitue en quelque sorte l'ossature du questionnaire que le statisticien va concevoir. Il comprend :
- un ensemble de questions sur l'unité statistique à repérer qui porte sur les caractéristiques du logement et des indications de repérage. Ces indications sont nécessaires pour établir les pondérations ;
 - un ensemble de questions appelé "tronc commun" permet de connaître la composition du ménage et les caractéristiques essentielles de chacune des personnes. Ce tableau de composition du ménage permet de définir la personne de référence. Il est utilisé dans la plupart des enquêtes sur les conditions de vie;
 - des questions de validation en fin de questionnaire qui permettent de dresser un bilan de collecte.

Le questionnaire de l'enquête sera "glisser" dans cette ossature.

Restait à choisir le logiciel à utiliser pour construire cette infrastructure.

4. L'expérience de Blaise II

Nous avons rencontré avec Blaise II (dans les versions 2.39 et 2.4) des limites de trois ordres :

1. L'obligation de scinder des questionnaires volumineux en plusieurs "*setups*" Blaise, les limites de taille étant assez rapidement atteintes (4 *setups* Blaise pour l'enquête Budget de Famille). Cette contrainte nous a conduit à réaliser des développements complémentaires spécifiques pour assurer la cohérence entre les différentes parties du questionnaire ;

2. Le nécessaire détournement de la fonction du "*BLOCK APPOINTMENT*" pour permettre l'affichage de carnet de tournée (liste des logements à enquêter, le choix d'un logement dans la liste provoquant l'ouverture du questionnaire) ;

3. Une programmation informatique complexe qui nécessite une formation assez lourde pour des statisticiens.

5. L'Insee choisit Blaise III.

L'annonce de la version 3 du logiciel Blaise, nous incitait à reconsidérer le choix de Blaise II dans la mesure où une reconversion des enquêtes antérieures était de toute manière nécessaire.

Un premier examen des logiciels du marché nous a conduit à étudier l'un d'eux utilisé par les principaux instituts de sondage français. Il semblait pouvoir supporter sérieusement la comparaison avec Blaise. Nous avons donc fait une investigation poussée sur ce logiciel : réalisation d'un questionnaire regroupant les principales caractéristiques des différentes enquêtes de l'Insee.

La force de ce logiciel réside d'abord dans son interface de développement, puisque les questions sont saisies dans la forme où elles apparaîtront à l'enquêteur (*WySiWyg*). Le mode de programmation des contrôles nous a paru convivial. Il ne semblait pas y avoir de limites pour les gros questionnaires, puisque le programme (<500K) utilise toutes les données du questionnaire comme des données externes (comme le fait aujourd'hui Blaise III). Il fournissait en outre un poste de gestion de la collecte avec des fonctions de contrôles et de tabulations des données sans investissement supplémentaire, conçu plus particulièrement pour CATI.

Mais il n'était pas possible, avec ce logiciel, de remonter dans un questionnaire sans effacer les données saisies. Ce choix est conforme à la pratique des instituts de sondage dans la réalisation des interviews : un questionnaire (souvent court) doit être déroulé sans hésitation du début à la fin, l'enquêteur est très guidé (impératifs de rendement et de rapidité). Dans le cas d'un institut de statistiques, le souci d'obtenir la plus grande qualité possible des réponses au cours de la collecte, pour des questionnements souvent complexes est primordiale. Une plus grande autonomie doit être laissée à nos enquêteurs, et les aller-retours dans le questionnaire sont fréquents et nécessaires.

Blaise III nous a donc finalement paru plus adapté au mode de collecte à l'Insee. D'ailleurs Blaise est utilisé, ce n'est sans doute pas un hasard, par bon nombre d'instituts nationaux de statistiques. Avec Blaise III, les contraintes de taille de questionnaire ont été levées, et la création d'un carnet de tournée est maintenant possible avec Maniplus. De plus à côté de Blaise proprement dit, il existe un ensemble d'outils périphériques qui permet de réaliser l'ensemble des traitements d'une enquête dans le même contexte informatique.

6. Le carnet de tournée en Maniplus :

La version 1.1 de Blaise 3 inclura le nouvel outil Maniplus, dont l'utilité nous paraît essentielle.

Jusqu'à présent, il était nécessaire de réaliser les interfaces des postes de travail avec un outil spécifique : nous avons utilisé Turbopro, puis C, et Clipper. Maniplus offre cette possibilité. Ce nouvel outil nécessite un apprentissage léger, et présente l'avantage d'être évidemment tout à fait

cohérent avec les autres modules de Blaise, et donc de simplifier le développement des applications CAPI, et d'en faciliter la maintenance.

Maniplus nous permet de fournir à l'enquêteur un carnet de tournée qui contient la liste des unités à enquêter. L'enquêteur commence toujours par ouvrir son carnet de tournée: pour chaque unité à enquêter, il visualise l'identifiant du logement, son adresse, des précisions de repérage, un état de son travail (non validé, bon, transmis), un résultat de collecte (logement vacant, résidence occasionnelle, résidence secondaire, impossible à joindre.), un suivi de collecte (refus, abandon, terminé). Ces informations lui permettent de gérer au mieux sa tournée en lui fournissant un aperçu général de l'état de sa collecte. Le choix d'une unité dans ce carnet lui permet d'ouvrir le questionnaire. Maniplus donne une solution tout à fait satisfaisante pour répondre à ce besoin.

Maniplus permet le calcul de variable à l'extérieur du programme de questionnement. Le fonctionnement asynchrone de Blaise rend impossible par exemple le calcul du temps de questionnement (lorsqu'un questionnaire est ouvert plusieurs fois, le temps est recalculé), et le nombre d'ouvertures d'un questionnaire. Maniplus permet cela.

7. Programmer en Blaise : facile ou non

Notre expérience de la programmation en Blaise III est très récente, mais il semble qu'elle reste complexe. Elle nécessite un investissement important en formation, et la mise au point d'un questionnaire est longue et difficile surtout en période d'apprentissage.

Nous avons constaté avec Blaise II deux caractéristiques principales:

1. Le temps d'apprentissage, 2 à 3 mois pour être opérationnel, est lié aux multiples possibilités du logiciel. Les différentes fonctions doivent être programmées dans des langages qui ne sont pas très cohérents (*DEP*, *MANIPULA*, *CAMELEON*, *ABACUS*,...). La séparation des *CHECKS* et des *RULES* complique la tâche par rapport à une programmation classique. Il est obligatoire de faire fonctionner l'analyseur syntaxique pas à pas, ce qui ralentit la mise au point d'un programme. Avec Blaise III, il n'y a pas d'amélioration sensible sur ces points.

2. Le langage est très complet. Si on ne met pas en balance en permanence l'avantage (obtenir un questionnaire convivial) et le coût (temps de programmation nécessaire), on aboutit facilement à un programme volumineux.

Pour conclure, même si des progrès importants ont été faits en Blaise 3 pour faciliter le développement des questionnaires (notamment le mode *Windows*, un paragraphe *RULES* unique, les *INCLUDE*), nous pensons que la maîtrise du logiciel nécessite encore un investissement important.

8. Un pôle de compétence Blaise

Dans ce cadre, plutôt que de confier la programmation du questionnaire à chaque concepteur, il nous paraît préférable de spécialiser une unité de développement à laquelle ils pourront faire appel. Ce pôle a quatre missions :

1. l'expertise des versions successives du logiciel et la formation des concepteurs.
2. le développement des questionnaires et la validation des questionnaires développés par des concepteurs.
3. la maintenance des applications et notamment de l'ossature de questionnaire.
4. l'établissement des normes d'ergonomie, d'écriture, de maintenance et de bonne insertion dans CAPI pour le développement.

Cette centralisation du développement permettra de profiter plus facilement des expériences successives de programmation et de "ramasser" la compétence en Blaise III. Il facilitera l'élaboration de standards de développement et la création progressive d'une bibliothèque de programmes utilisables.

9. La nécessité d'établir des normes

Blaise III est un système très souple qui offre la possibilité de personnaliser son utilisation. Cette souplesse est à double tranchant : laisser chaque concepteur ou programmeur définir la personnalisation induirait pour les enquêteurs et les gestionnaires régionaux des difficultés d'utilisation. C'est dans cet esprit qu'il est recommandé, pour chaque enquête d'utiliser l'architecture de questionnaire proposée sur le poste de travail du statisticien. Dans cet esprit, il nous paraît nécessaire de définir progressivement un corps de normes :

1. Des normes de programmation

Tout développeur de programme de questionnement doit respecter ces normes, définies par le pôle de compétence Blaise.

Par exemple:

- Les mots du langage Blaise sont écrits en majuscules.
- Les noms de *FIELDS* sont composés de 8 caractères au maximum. Cette contrainte est nécessaire pour faciliter la production de dictionnaires SAS.
- Les noms de *FIELDS* doivent être significatifs, dans la mesure où ces noms sont affichés sur l'écran de l'enquêteur devant la zone de saisie.

- Un nom de *BLOCK (TABLE)* commence par B (T) majuscule. On note en commentaire le nom du *block (table)* auquel se rattache une instruction *ENDBLOCK (ENDTABLE)*.
- Un nom de *TYPE* (excepté pour les *BLOCKS* et les *TABLES*, Cf. ci-dessus) commence par Typ, suivi du nom de la variable décrite par ce *TYPE*.
- Règles d'indentation des *IF, THEN, ELSE, ELSIF*.
- Règle d'écriture des *TYPE*, des *CHECKS*.
- Etc..

2. Une norme de présentation de l'écran de saisie

Le paramétrage de la présentation de cet écran s'effectue par des modifications sur le fichier *MODELIB* (et bientôt dans une section *LAYOUT* du programme), et sur le fichier *DEPMENU* (activer ou inhiber certaines fonctions). Un écran standard a été défini, que les enquêtes devront respecter au maximum.

3. Les touches fonction.

Le fichier *DEPMENU* permet également de définir les touches fonction. Notre souci est d'éviter au maximum les combinaisons de touches, et de ne pas bouleverser les habitudes déjà prises.

Un exemple de normalisation :

DESCRIPTION	TOUCHES
Aide à l'utilisation de Blaise	F1
Aide contextuelle (sur les questions)	F12
Mode édition	F2, Backspace, insert.
Afficher la suite d'une question	F8
Se positionner sur une question précise	F5
Confirmer une réponse douteuse	F3
Passer du champ de saisie à la liste des modalités de réponses	F6
Ouvrir une fenêtre remarque	F4
Interrompre un questionnaire, ouvrir un bloc parallèle	CTRL+ENTER
Annuler l'option sélectionnée	Escape
Reconduire une même réponse à une même question dans un tableau	/

Pour codifier "ne sait pas "	?
"refus de réponse"	!
Pour se placer tout au début, à la fin d'un questionnaire	HOME, END
Pour remonter d'une étape dans les menus	ESC
Pour avancer, remonter, par page d'écran	PGDN, PGUP
Pour se déplacer question par question	← ↑ → ↓
Pour effacer le caractère précédent le curseur	BKSPACE
Pour effacer le caractère sur le curseur	DEL
Pour valider la saisie, une option	ENTER
Relancer le micro	CTRL+ALT+DE L

4. Le langage

Blaise est livré en néerlandais et en anglais. Nous devons livrer des interfaces utilisateurs en français. La possibilité de traduire les messages Blaise dans une autre langue ne lève pas toute la difficulté : les problèmes sémantiques subsistent. Il nous faut d'une part adapter les termes induits par la nouvelle technique CAPI à la culture existante en matière d'enquête, et d'autre part trouver les termes adéquats pour désigner les nouveaux concepts. Ce travail de linguiste est délicat, et doit être mené avec un grand soin très tôt, afin de faciliter pour les enquêteurs et les gestionnaires d'enquêtes l'apprentissage de la réalisation des enquêtes par CAPI.

L'infrastructure CAPI avec Blaise III sera mise en place à la fin de l'année 1995 et testée sur les premières versions de l'enquête logement ; la réalisation en vraie grandeur se fera au quatrième trimestre 1996. Parallèlement, l'enquête trimestrielle emploi est développée en Blaise III pour juin 1996. Il n'est pas envisagé de convertir l'enquête emploi en (Blaise II) avant sa refonte en 1999. Les autres enquêtes auprès des ménages passeront progressivement sur micro-portables. La demande est forte de la part des concepteurs qui y voient des gains en qualité et en délais, comme des enquêteurs qui préfèrent, dans une très grande majorité, ce mode de travail. Mais le passage du papier au micro, pour être rentable, nécessite la mise en place de procédures harmonisées pour les informaticiens comme pour le statisticien dans la conception du questionnaire.