

CapiPlus, un système intégré pour la réalisation des enquêtes auprès des ménages

Gilles Luciani, INSEE, France

1. Introduction

Le système CapiPlus développé à l'Insee a pour objectif de proposer aux statisticiens une architecture informatique capable d'accueillir les enquêtes CAPI auprès des ménages en réduisant les coûts de développement informatique à ceux de la programmation du questionnaire de collecte et du programme d'apurement, qui sont spécifiques à l'enquête. Ces programmes, une fois testés sur un échantillon de quelques centaines de ménages, doivent s'intégrer dans le système CapiPlus sans qu'il soit nécessaire de réaliser des développements complémentaires. La possibilité de cette intégration repose sur le respect d'un certain nombre de standards communs à toutes les enquêtes :

- une architecture du programme du questionnaire,
- des blocs de programme de questionnement pour la composition du ménage (le "tronc commun"),
- des variables de gestion,
- des interfaces utilisateurs,
- un carnet de tournée,
- une ergonomie,
- un langage.

J'ai présenté rapidement à la dernière conférence des utilisateurs de Blaise à Helsinki en septembre 1994 ces principales caractéristiques du système CapiPlus, qui était à l'époque en cours de réalisation.

Ce système a depuis été mis en oeuvre pour deux enquêtes :

- « Logement », enquête réalisée auprès de 45000 ménages de novembre 1996 à février 1997,
- « Jeunes et Carrières » en mars et avril 1997 auprès de 60 000 personnes.

La prochaine enquête utilisant CapiPlus sera l'enquête " Patrimoine ", en production en octobre 1997.

2. Résumé

On décrit dans un premier temps l'architecture générale du système CapiPlus : les logiciels utilisés, les matériels, la circulation des données entre les différents sites informatiques.

Il est ensuite expliqué au travers de quelques exemples comment l'utilisation de Blaise dans CapiPlus simplifie la *capisation* d'une enquête, tout en maximisant les deux principaux avantages attendus de la technique CAI :

« *faire plus vite..* » dans la réalisation des programmes de collecte et d'apurement, dans la production de fichiers propres,

« *...et mieux* » dans la qualité des données obtenues à l'issue de l'apurement et dans le confort des utilisateurs du système.

3. L'architecture générale

3.1. Les logiciels

CapiPlus utilise actuellement sur les logiciels et outils suivants :

- Le logiciel Blaise 3 (DEP, Manipula, Manipus, Caméléon) pour la réalisation des questionnaires, les programmes de traitement des données en batch, les interfaces utilisateurs, la production des dictionnaires SAS. Le basculement de la version beta K de Blaise (juin 1996) utilisée pour la production des enquêtes " Logement " et " Jeunes et Carrières " dans la version 1.15 a été réalisé en avril 1997.
- Le logiciel Arbitr (Société Tangram - USA) pour les transmissions des données entre les postes utilisateurs et le site central.
- Des programmes spécifiques écrits en TurboPascal (Société Borland) pour la tabulation des données sur les postes de travail et quelques traitements spécifiques.
- L'utilitaire PkZip (Société PkWare) pour la compression-décompression des données sur les postes de travail.

A l'issue de la collecte, puis de l'apurement des données, le fichier de l'enquête est traité principalement avec le logiciel SAS.

L'applicatif fonctionne sous Dos 5 ou Dos 6, selon les postes.

Les évolutions de cette architecture prévues à court terme sont l'utilisation d'Abacus en batch pour la tabulation des données, l'utilisation de PkUnzip sur site central (IBM, MVS), et l'intégration de CapiPlus sous Windows 3.11 dans une fenêtre Dos. A moyen terme (1998), nous envisageons le remplacement du logiciel arbitr par un logiciel de transmission sous Windows, l'utilisation de Blaise sous Windows.

3.2. Les matériels

CapiPlus est un système informatique réparti sur quatre types de sites informatiques :

Les 900 micros de collecte des enquêteurs : Le parc des micros qui équipent les enquêteurs se répartit actuellement entre des Toshiba T1850 (486 SX RAM4+4 HD80 monochrome) et des Hewlett-Packard Omnibook 5000 CT et Omnibook 800CS (Pentium 100 RAM 16 HD800 couleur). L'objectif est d'homogénéiser dès cette année le parc avec uniquement des Pentium HP.

Une trentaine de postes de travail dans les directions régionales répartis en 18 serveurs.

Les postes de travail du concepteur sur un réseau local unique à la direction générale.

Le suivi de la collecte et l'apurement des données dans les 18 directions régionales qui encadrent les enquêteurs utilise les serveurs sur lesquels sont connectés des micros de bureau de différents types (486 DX et Pentium, RAM16), utilisés pour différentes applications. Toutes les données et programmes Capi sont sur le disque dur du serveur. Les accès concurrents sur les données sont gérés par Blaise. La CPU du serveur n'est pas utilisée.

Enfin le site central par lequel transitent les données entre les enquêteurs et les directions régionales, et sur lequel sont regroupées les données à la fin de l'enquête.

3.3. a circulation des données

On détaille ci-dessous les différentes étapes du cheminement des données depuis la constitution de l'échantillon de l'enquête jusqu'à la restitution des données finales sur le site central.

Avant la collecte :

1. Livraison des échantillons régionaux sur le site central.
2. Intégration de l'enquête sur le serveur régional : importation des programmes.
3. Importation des programmes du poste de collecte, et production de la disquette D1 d'installation de l'enquête sur les postes de collecte.
4. Importation de l'échantillon sur le serveur local en région.
5. Attribution des fiches-adresse (FA) par enquêteur : un échantillon régional avec pour chaque FA un n° d'enquêteur est ainsi constitué.
6. Production de la base des enquêteurs propre à l'enquête. Production de la disquette D2 contenant la base des enquêteurs et permettant la personnalisation des postes de collecte.
7. Eclatement de l'échantillon régional en sous-échantillons " enquêteur ".
8. Conversion de chaque échantillon dans le datamodel de collecte, et production de la disquette D3 des échantillons.
9. Chargement des postes de collecte :
10. - Installation des programmes de l'enquête avec la disquette D1,
11. - Personnalisation du micro de l'enquêteur avec la disquette D2,
12. - Chargement de l'échantillon avec la disquette D3.

Pendant la collecte :

13. Collecte des données par l'enquêteur.
14. Transmission quotidienne via Transpac des questionnaires validés et zippés sur le site central, ou par envoi postal de disquettes pour les enquêteur qui n'ont pas accès à Transpac (3%).
15. Importation quotidienne de ces données sur le serveur de la DR, dézippage et conversion dans le datamodel d'apurement DR.
16. Apurement des données par la DR.
17. Conversion des données dans le datamodel d'apurement Concepteur.
18. Conversion en ASCII et exportation quotidienne des données apurées sur le site central.
19. Importation quotidienne des questionnaires par le concepteur des questionnaires encore erronés.
20. Apurement sur le poste concepteur à la direction générale.
21. Conversion en ASCII et exportation quotidienne des questionnaires apurés par le concepteur sur le site central.

4. L'installation d'une enquête sur les postes de travail

Cette opération est réalisée avec des programmes indépendants de l'enquête. Le gestionnaire indique le code de l'enquête à installer, par exemple PAT_97 pour l'enquête Patrimoine 1997, et l'enregistrement de l'enquête est importé du site central, le fichier des enquêtes disponibles est mis à jour par un manipula ASCII to Blaise.

L'installation de l'enquête, et des éventuelles mises à jour se font par un choix de l'enquêtes grâce à un LOOKUP sur le fichier des enquêtes, puis par l'importation des programmes et fichiers.

Dans l'interface du poste du gestionnaire, le menu primaire regroupe les fonctions communes à toutes les enquêtes, et il y a un menu secondaire pour chaque enquête. Chacun de ces menus sont programmés en Manipus. Le choix du menu de l'enquête que l'on souhaite lancer se fait à toujours par LOOKUP sur le fichier des enquêtes : le code de l'enquête est utilisé comme paramétrage de l'appel de tous les programmes.

Il n'y a pas de limite au nombre d'enquêtes que l'on peut installer sur les postes de travail, sauf celle du disque dur du micro de collecte, ou du serveur.

Le datamodel du fichier des enquêtes :

```
DATAMODEL Enquete
PRIMARY CodEnq
TYPE
    EnqType=(      T1 "Enquête à un seul lot de logements",
                  T2 "Enquête par aires",
                  T3 "Enquête à vagues",
                  T4 "Panels")
FIELDS
    CodEnq          "Code de l'enquête"          : STRING[6]
    NomEnq          "Nom de l'enquête"          : STRING[40]
    DateDeb        "Date de début de l'enquête" : DATETIME
    TypEnq         "Type d'enquête"            : EnqType
    NbQuest        "Nombre de sous-questionnaires" : 1..99
```

L'appel du menu de l'enquête dans le maniplus primaire des postes de travail, en passant les paramètres Code de l'enquête (CodEnq) et numéro de DR (NumDR) :

```
PROCESS AppliDR "Menu primaire CAPI+ - Poste DR"
USES
    Menu          'I:\CAPI\PROGGEN\BLAISE\menu'
    Enquete       'I:\CAPI\PROGGEN\BLAISE\enquete'
INPUTFILE MenuP: Menu ('I:\CAPI\PROGGEN\BLAISE\menudr.asc', ASCII)
                  SETTINGS SEPARATOR='/' CONNECT = NO
INPUTFILE ListEnquete : Enquete(BLAISE3)
                  SETTINGS OPEN=NO ACCESS = SHARED
(...)
DIALOG LookupEnq 'Choix d'une enquête'
    SIZE = (74,19)
    LOOKUP ListEnquete POSITION = (2,1) SIZE = (69,8)
    BUTTON Bouton CAPTION = '~O~K' VALUE = Ok
    BUTTON Bouton CAPTION = '~A~nnuler' VALUE = Nul
(...)
MANIPULATE
(...)
ListEnquete.OPEN('\CAPI\DATAGEN\PARAM\enquete')
    LookupEnq
    CASE Bouton OF
    Ok: CodEnq := ListEnquete.Code
        NomEnquete := ListEnquete.Nom
        DISPLAY ('Vous avez choisi l'enquête: ' + NomEnquete) PAUSE
    ENDCASE
(...)
ListEnquete.CLOSE
ProgN := 'I:\CAPI\PROGGEN\BLAISE\'+CodEnq+'DR /P'+CodEnq+';' + NumDR
Reslt := CALL(ProgN)
```

5. Le menu de l'enquête

Le menu de l'enquête est un setup maniplus qui diffère selon le type de l'enquête. Pour les enquêtes du même type, ces setups ne diffèrent que par les instructions USES qui appellent les datamodels propres aux questionnaires des enquêtes. En effet, le nom du datamodel dans l'instruction USES ne peut être introduit comme paramètre.

Le début du maniplus de l'enquête Logement, et les instructions de lancement du programme d'interview (ordre EDIT) :

```
PROCESS MenuEnquêteur1 "Enquête de type 1 "  
USES  
MenuEnq1      'I:\CAPI\PROGGEN\BLAISE\Menu'  
MetaCollect  'I:\CAPI\LOGT96\PROG\BLAISE\Logt96'  
  
INPUTFILE MenuF : MenuEnq1 ('MENUENQ1.ASC',ASCII)  
                SETTINGS SEPARATOR = '/' CONNECT=NO  
UPDATEFILE      Flux : MetaCollect(BLAISE)  
                SETTINGS OPEN = NO ACCESS = SHARED  
INPUTFILE       FStat : MetaStat(ASCII)  
                SETTINGS OPEN=NO CONNECT = NO      (...)  
MANIPULATE  
  CodEnq := PARAMETER(1) NumDr := PARAMETER(2)  
  PathN:= 'I:\CAPI\'+'CodEnq+'\DATA\  
REPEAT Reslt:= MenuF.MENU('V3.2 04/12/96')  
  CASE Programme OF      (...)  
'Collecte':      (...)  
  FileN := PathN+CodEnq  
  Flux.OPEN(FileN) Reslt:=Flux.EDIT(FileN) Flux.CLOSE      (...)
```

Insistons sur le fait que la seule chose qui diffère entre les programmes des menus des enquêtes de même type est l'instruction USES dans laquelle le nom du fichier du datamodel est en *dur*.

Par conséquent, une fois le datamodel de l'enquête au point, on comprend que son intégration dans un système stabilisé nécessite un minimum d'investissement pour le maintenicien. Il faut tout de même procéder à une compilation de tous les programmes et bien sûr s'assurer par des tests de leur fonctionnement correct.

6. Le chargement de l'échantillon

Dans CapiPlus, le fichier de l'échantillon doit respecter un dessin précis. Dans ce cas, le manipula de conversion de l'échantillon ASCII en échantillon au format Blaise dans la datamodel de l'enquête est toujours le même, avec la même réserve que ci-dessus : le nom du datamodel dans l'instruction USES doit être modifié, et le manipula recompilé.

7. Une qualité accrue dans des délais courts

Dans CapiPlus, les données de collecte sont contrôlées à trois niveaux successifs :

Contrôles de niveau 1 sur le poste de collecte

Le datamodel comprend un grand nombre de contrôles (CHECK et SIGNAL) assurant à la saisie des données correctes. De plus, au cours et à la fin du questionnaire; l'enquêteur valide ou non les parties qu'il vient de saisir. L'état du questionnaire ne pourra être à Bon que si toutes les validations ont été faites, et tous les avertissements supprimés. Quand l'enquêteur décide d'envoyer les questionnaires en DR, un manipula sélectionne tous les questionnaires bons, qui sont zippés dans un fichier. Mais cette technique est insuffisante pour garantir la correction parfaite des données au niveau 1, car après validation, et avant l'envoi,

l'enquêteur peut ouvrir un questionnaire pour le relire (droit au remords), et effacer par erreur des réponses obligatoires, ou rendre incohérent le questionnaire en modifiant des réponses : il suffit pour cela de faire ces modifications en remontant dans le questionnaire, et de quitter par AltX, la sauvegarde étant automatique. C'est pourquoi le manipula de sélection des questionnaires à l'état Bon refait les contrôles grâce à l'instruction CHEKRULES. En cas d'erreurs résiduelles, le questionnaire n'est pas sélectionné pour l'envoi, l'état et la validation finale remis à blanc. Un message prévient l'enquêteur de la dévalidation.

Contrôles de niveau 2 sur le poste de d'apurement en DR

Lorsque les questionnaires reçus en DR sont intégrés dans le fichier d'apurement, ils sont convertis par un manipula dans un autre datamodel (niveau 2), dans lequel on a supprimé des contrôles de niveau 1, rendu plus stricts certains contrôles (fourchettes plus étroites, cohérences plus fines,...), et ajouté de nouveaux contrôles. Afin de ne pas réintroduire à l'apurement des erreurs de niveau 1, il faut être très rigoureux dans la suppression des contrôles de niveau 1, et mettre éventuellement en .SHOW les variables sur lesquelles aucun nouveau contrôle n'est effectué.

Le manipula de conversion dans le niveau 2 fait passer automatiquement à l'état Bon les questionnaires qui n'ont pas d'erreurs de niveau 2.

Le programme produit un fichier des remarques et des questions ouvertes des enquêteurs pour le lot des questionnaires reçus. La production du fichier des remarques oblige à une conversion en ASCII de la totalité du questionnaire, la possibilité n'étant pas offerte aujourd'hui de ne produire que les remarques dans un fichier texte. La consultation de ce fichier permet de se pencher plus aisément sur les questionnaires présentant des problèmes. Les enquêteurs ont pour consigne de motiver systématiquement la suppression d'un avertissement par une remarque.

Le nombre d'avertissements supprimés à la collecte est affiché (SUPPRESSCOUNT). Cet indicateur permet d'attirer l'attention du gestionnaire sur les questionnaires douteux.

Une variable « Arelire » est renseignée à Oui si le nombre d'erreur de niveau 2 est positif.

Manipula EnqToDr

```
USES
  MetaCollect 'I:\CAPI\LOGT96\PROG\BLAISE\Logt96'
  MetaADR     'I:\CAPI\LOGT96\PROG\BLAISE\Logt96F'

INPUTFILE  Collecte : MetaCollect
            ('@I:\CAPI\LOGT96\DATA\ziplist.asc', BLAISE)
OUTPUTFILE ApurtDR   : MetaADR
            ('I:\CAPI\LOGT96\DATA\entree'      , BLAISE)
OUTPUTFILE Rmk      = Collecte ('I:\CAPI\LOGT96\DATA\Remark.asc',
ASCII)
AUXFIELDS  Er1 : INTEGER[4]   Er2 : INTEGER[5]

MANIPULATE
  ApurtDR.MSC := Collecte.SUPPRESSCOUNT
  Er1 := ApurtDR.MSC
  IF Er1 > 0 THEN  ApurtDR.ErrColV1 := Oui
                  ELSE  ApurtDR.ErrColV1 := Non
  ENDIF
  ApurtDR.CHECKRULES
  Er2 := ApurtDR.ERRORCOUNT(SOFT)
  IF ER2 > 0 THEN  ApurtDR.ErrColV2 := Oui
                  ApurtDR.ARelire := Oui
                  ApurtDR.Etat := EMPTY
                  ELSE  ApurtDR.ErrColV2 := Non
                  ApurtDR.VALidDR := Oui
                  ApurtDR.Arelire := Non
                  ApurtDR.Etat := B
  ENDIF
  ApurtDR.Comment.Enq := Collecte.Commentaire
  IF Collecte.Commentaire.Comgen <> Empty THEN
    ApurtDR.RmqEnq := '*'
  ENDIF
  ApurtDR.WRITE
  Rmk.WRITE
```

Quand la DR souhaite envoyer les questionnaires apurés sur le site central, les questionnaires sont à nouveau contrôlés dans le manipula de sélection des questionnaires bons, et éventuellement dévalidés, ils sont convertis dans un datamodel de niveau 3, dans lequel sont programmés des contrôles sophistiqués qui sont traités centralement par des agents très qualifiés appartenant à l'équipe du concepteur. Le manipula place dans un fichier ASCII tous les questionnaires apurés par la DR, et dans un fichier Blaise les questionnaires qui relèvent d'un apurement de niveau 3.

On notera dans le programme l'utilisation de la *wildcard* '@I:\CAPI\LOGT96\DATA\ziplist.asc' dans l'instruction INPUTFILE. Cette façon de faire résout le problème de la concaténation des fichiers de collecte des enquêteurs. A leur importation, un programme Pascal les dézippe chacun dans un répertoire particulier, et crée simultanément le fichier *ziplist.asc* dans lequel se trouve la liste des fichiers importés et dézippés, avec leur chemin complet.

Niveau 3 sur le poste concepteur

Lorsque les questionnaires arrivent sur le poste concepteur, ils ont déjà été convertis dans le datamodel de niveau 3 en sortie d'apurement. Les dernières erreurs sur ces questionnaires *difficiles* sont examinées et corrigées.

Un flux rapide

Dans le cas idéal, un questionnaire gravit les trois échelons de ces contrôles en trois jours :

Envoi sur site central par l'enquêteur le soir du jour J de collecte, importation des questionnaires par le programme de transfert qui se déclenche automatiquement la nuit (on évite les encombrements diurnes des transmissions et on gagne le temps de traitement du manipula de conversion),

Apurement en DR le jour J+1 des questionnaires reçus dans la nuit,

Transfert nocturne des questionnaires apurés sur le site central, une partie étant livrée au concepteur pour les corrections de niveau 3,

Apurement le jour J+2 par le concepteur,

Envoi le jour J+3 du questionnaire définitif sur le site central.

La rapidité de la circulation des données permet à la DR de se retourner vers l'enquêteur, ou au concepteur vers la DR, en cas de doute sur certains questionnaires, ceux-ci étant encore présents dans la mémoire de l'agent.

En terme de volumes, le bilan des contrôles pour l'enquête Logement est :

	Nombre de questionnaires en %
Sans erreurs à la collecte (niveau 1)	100
Aucun avertissement supprimé à la collecte	60
Aucune erreur en entrée d'apurement DR (niveau 2)	62
Aucune erreur en sortie DR (niveau 3)	95

Le nombre d'avertissements supprimés à la collecte est en moyenne de 0,6 par questionnaire.
