

Utilisation de Blaise dans la répartition automatique des fiches adresse entre enquêteurs des enquêtes ménages

Georges Bourdallé, I.N.S.E.E (FRANCE)

1. Présentation

La mise en place de la collecte informatisée sous C.A.P.I d'une enquête auprès des ménages s'accompagne de différentes étapes plus ou moins lourdes, du tirage de l'échantillon national à la répartition en échantillon enquêteurs en passant par la mise à disposition des échantillons régionaux. Trois équipes de personnes sont concernées par celles-ci. Une équipe en charge du tirage de l'échantillon national, une équipe de gestionnaires régionaux en charge du suivi de la collecte et une équipe informatique en charge de la mise à disposition des échantillons enquêteurs. Avant la mise en place d'une répartition automatique des logements ou fiches adresse entre les enquêteurs, la charge du travail entre les équipes n'était pas homogène. Lors de la répartition manuelle de la charge de travail entre les enquêteurs, les gestionnaires d'enquêtes avaient une charge de travail importante sur une période assez courte. En effet, les contraintes faisaient qu'ils n'avaient qu'une semaine pour faire cette répartition. Sur une année courante le nombre de logements enquêtés toutes enquêtes ménages confondues, hors enquête emploi, est de l'ordre de 80000. Au niveau régional, ceci représente un nombre moyen annuel de 3800 logements à répartir. A titre d'exemple, pour l'enquête *Patrimoine* le nombre moyen de fiches adresse à répartir par direction régionale était de 715.

L'objectif de cette répartition automatique de la charge de travail entre enquêteurs est double. Premièrement, diminuer considérablement, de l'ordre de 60%, la charge de travail des gestionnaires d'enquêtes lors de la préparation de la collecte et deuxièmement, une fois le tirage de l'échantillon fait, intégrer la totalité des opérations dans le dispositif C.A.P.I.

1.1 Condition de mise en place d'une répartition automatique

La mise en place d'une répartition automatique des logements enquêtés n'a de sens que si l'on connaît a priori les zones qui seront potentiellement enquêtées. La connaissance de ces zones implique une définition à priori de celles-ci. En France, ceci est rendu possible par la mise en place d'un Echantillon-Maître (E.M) à l'issue du Recensement de la Population (R.P.). L'Echantillon-Maître nous permet d'avoir une liste de communes ou d'îlots (quartiers) parmi lesquelles seront tirés les logements des enquêtes ménages, excepté l'enquête Emploi qui a un échantillon propre. Une fois l'échantillon-Maître constitué, les Directions Régionales procèdent au recrutement de leur réseau enquêteurs en tenant compte de différents critères dont un critère d'éloignement, moins de 30 km, entre la résidence de l'enquêteur et les zones potentiellement enquêtées. Ce réseau enquêteur est donc relativement stable durant la durée de validité de l'Echantillon-Maître. Une fois ce recrutement effectué, les directions régionales sont en mesure de définir les tables de correspondance entre les zones potentiellement enquêtées et les enquêteurs.

1.1.1 Type de logements exclus de cette répartition automatique

Lors du tirage d'une enquête, deux types de logements sont tirés. Les logements issus du dernier Recensement de la Population (R.P. 1999) que nous appellerons les logements R.P et les logements issus de la Base de Sondage des Logements Neufs (B.S.L.N.) que nous appellerons les logements neufs. Ces logements neufs sont situés dans tous les types de ville, aussi bien en milieu rural qu'en milieu urbain. Toutefois, il est à noter que les logements neufs se trouvant dans des villes de plus de 20 000 habitants ne sont pas îlotés, c'est à dire que nous sommes incapables de les localiser au niveau géographique îlot, leur zone d'implantation ne peut donc être affectée a priori à un enquêteur. Dans l'avenir, ceci sera corrigé, l'I.N.S.E.E ayant planifié une opération d'îlotage des logements issus de la B.S.L.N. Les logements neufs représente 7% des logements tirés.

2. Les tables régionales de correspondance entre zones enquêtables et enquêteurs

Du fait de la constitution d'un Echantillon-Maître et d'un réseau d'enquêteur relativement stable les tables de correspondance sont définies une fois pour toutes. Au cours du temps, celles-ci seront mises à jour par les directions régionales à partir d'un datamodel BLAISE. Le taux de pré affectation validé par les directions régionales dépendra du soin apporté lors de la création de ces tables de correspondance.

2.1 Mise à disposition des zones potentiellement enquêtées aux directions régionales et affectation des enquêteurs

Les zones potentiellement enquêtées sont transmises à chaque direction régionale sous forme d'une base de donnée BLAISE. Il est demandé aux directions régionales de compléter cette base de donnée par deux types d'information. Un premier type d'information relatif à l'enquêteur, à savoir le numéro national enquêteur, le nom et prénom de l'enquêteur. Ce premier type d'information permettra la répartition et le paiement de l'enquêteur et un second type d'information relatif aux enquêtes pour lesquelles l'enquêteur est pressenti. Ce second type d'information permettra, en vue de l'appariement avec le fichier échantillonné, la sélection de l'enquête. Les gestionnaires d'enquêtes auront ainsi la possibilité d'affecter une charge de travail à un enquêteur à échéance de 2 ans. La constitution et la mise à jour de ces tables se feront sous forme de questionnaire électronique.

2.1.1 Datamodel de saisie des informations enquêteurs et enquêtes

La saisie par les gestionnaires d'enquêtes des directions régionales I.N.S.E.E des informations relatives aux enquêteurs et aux enquêtes, pour une zone enquêtable donnée, est faite à partir d'un Datamodel développé en Blaise. Quatre types de champs sont proposés aux gestionnaires d'enquêtes. Un premier type de champ relatif aux variables géographiques, ce type de champ indique la zone enquêtable. Un second type de champ relatif au nombre possible d'enquêteurs pouvant agir sur la zone géographique. Un troisième type de champ relatif aux variables d'identification enquêteurs à savoir le numéro d'identification, le nom et prénom de l'enquêteur. Et enfin, un quatrième type de champ relatif aux enquêtes choisies. Concernant ce dernier type de champ, le gestionnaire a la possibilité d'affecter une ou plusieurs enquêtes parmi celles proposées à un enquêteur. Le gestionnaire d'enquête aura la possibilité de saisir les deux derniers types de champ autant de fois que d'enquêteurs sur la zone (voir écran 3 du Datamodel). La liste des enquêtes proposées comprendra toutes les enquêtes planifiées à échéance de deux ans. Seuls les champs relatifs aux variables enquêteurs et enquêtes sont ouverts à la saisie.

2.1.2 Datamodel de mise à jour des zones enquêtées

Trois écrans composent le datamodel de mise à jour. Un premier écran donnant, pour une région, la liste des zones enquêtées. Un second écran définissant le nombre d'enquêteurs pressenti et enfin un troisième écran relatif aux informations enquêteurs et enquêtes.

Ecran1

R22	CODGEO	LIBGEO	STRATE	CL	IL	FIL	NBENG	NSAI	NOM	PRENOM	ENQUETE[1]	ENQUETE[2]	ENQUETE[3]	ENQUETE[4]
R22	60002	Abbecout	0											
R22	60002	Abbecout	0											
R22	60002	Abbecout	0											
R22	80001	Abbeville	2		AM05									
R22	80001	Abbeville	2		AT02									
R22	80001	Abbeville	2		AT06									
R22	80001	Abbeville	2		AT09									
R22	80001	Abbeville	2		AY07									
R22	80001	Abbeville	2		BW02									
R22	80001	Abbeville	2		BX12									
R22	80001	Abbeville	2		XD09									
R22	80001	Abbeville	2		XI02									
R22	80001	Abbeville	2		AL18									
R22	80001	Abbeville	2		AR03									
R22	80001	Abbeville	2		AT12									
R22	80001	Abbeville	2		AY09									
R22	80001	Abbeville	2		BR02									
R22	80001	Abbeville	2		BR11									
R22	80001	Abbeville	2		BV03									
R22	80001	Abbeville	2		XL05									
R22	80001	Abbeville	2		AD02									
R22	80001	Abbeville	2		AK06									
R22	80001	Abbeville	2		BP06									
R22	80001	Abbeville	2		BP07									
R22	80001	Abbeville	2		BX06									
R22	80001	Abbeville	2		BX08									
R22	80001	Abbeville	2		XD07									
R22	80001	Abbeville	2		XL03									
R22	80001	Abbeville	2		BE03									
R22	80001	Abbeville	2		BW03									
R22	80001	Abbeville	2		XM04									
R22	80001	Abbeville	2		XM17									
R22	80001	Abbeville	2		XT04									
R22	80001	Abbeville	2		AT05									

Le gestionnaire d'enquête visualise la liste des zones enquêtées de sa région. Cette liste est triée par département et au sein de chaque département, par nom de commune.

Ecran 2

Blaise Data Entry - C:\REPARTITION_FAVREGION\region22\REGION22

Forms Answer Navigate Options Help

Nombre d'enquêteur sur cette zone

Enter a numeric value between 1 and 10

RGES 22

CODGEO 60002

LIBGEO Abbecourt

STRATE 0

CIL

IL

FIL

NBENQ 4

Old	1/3	Modified by rule	Dirty	Insert	region
-----	-----	------------------	-------	--------	--------

Démarrer Boîte de ré... Air France... NC040023... Explorateur... article5 do... Blaise SAS - (Log... Recherche... Blaise D... 13:46

La première information que le gestionnaire d'enquête doit saisir concerne le nombre d'enquêteurs susceptibles d'être affectés sur la zone concernée. En effet, en fonction de divers paramètres tels que l'étendue de la zone enquêtée, ou bien le type d'enquêtes ou encore des périodes de l'année pendant lesquelles certains enquêteurs ne souhaitent pas faire d'enquêtes, ce nombre peut être supérieur à 1. Il est donné la possibilité aux gestionnaires de définir au plus 10 enquêteurs pour une zone.

Ecran 3

Blaise Data Entry - C:\REPARTITION_FA\REGION\region22\REGION22

Forms Answer Navigate Options Help

Repartition pour 2005. Modifier les attributions de cet enquêteur en saisissant ou cochant les modalités des enquêtes concernées?

1. Loyers et Charges Janvier 2005 13. BDF Vague6
 2. Loyers et Charges Avril 2005 14. BDF Vague7
 3. Loyers et Charges Juillet 2005 15. BDF Vague8
 4. Loyers et Charges Octobre 2005 16. PRCV mai 2005
 5. Epcv Janvier 2005 17. Violence-Santé Nov. 2005
 6. Epcv Janvier 2005 (victimation) 18. GGS Nov. 2005
 7. Epcv Octobre 2005
 8. BDF Vague1
 9. BDF Vague2
 10. BDF Vague3
 11. BDF Vague4
 12. BDF Vague5

Enter at most 18 values

	NSAIG	NOM	PRENOM	ENQUETE
enab(1)	124	MARTIN	PIERRE	1-2-3-4
enab(2)				
enab(3)				
enab(4)				
enab(5)				
enab(6)				
enab(7)				
enab(8)				

Old 2/3 Modified Dirty Navigate region

Démarrer Boîte de ré... Air France... NC040023... Explorateur... article5.do... Blaise SAS - (Log... Recherche... Blaise D... 13:47

Une fois saisi le nombre d'enquêteurs susceptibles d'être affecté sur la zone, le gestionnaire d'enquêtes saisira les informations relatives aux enquêteurs ainsi que leur charge de travail en terme d'enquêtes.

2.1.2 Contrôle de validité des variables d'identification enquêteurs

Afin d'éviter pour le gestionnaire d'enquête lors de la validation de la répartition une charge de travail importante de ré affectation dû à des erreurs de saisie des variables d'identification enquêteur, nous procéderons à un contrôle avec une base enquêteur national appelée « Référentiel enquêteur » dans laquelle sont déclarés tous les enquêteurs en contrat avec l'I.N.S.E.E. Ce « Référentiel enquêteur » étant particulièrement sécurisé, ces accès sont limités. Aussi le contrôle de validité se fera à deux niveaux. Au niveau de l'équipe en charge de la chaîne de tirage, une fois les informations de toutes les directions régionales réceptionnées et à un second niveau, au niveau du poste de gestion, une fois les ré affectations faites et la répartition validée.

2.2 Périodicité des mises à jour

2.2.1 Mise à jour des tables de correspondances

La périodicité de mise à jour des tables de correspondance zones enquêtées/enquêteurs est variable et dépend essentiellement d'imprévus ou contraintes relatives aux enquêteurs (refus de faire l'enquête, maladie, indisponibilité, charge de travail trop importante,...). Pour une facilité de gestion, elle doit rester marginale. La date limite de mise à jour de ces tables permettant une prise en compte pour une enquête donnée est de 30 jours avant le début de la collecte.

2.2.2 Mise à jour du Datamodel de saisie

La mise à jour et la mise à disponibilité du datamodel de saisie est faite dès la planification d'une nouvelle enquête. Sachant qu'à l'I.N.S.E.E, on essaie autant que faire se peut de planifier une enquête au minimum 2 ans avant le début de collecte, les mises à jour du Datamodel se feraient tous les deux ans.

2.3 Restitution des tables de correspondances complétées

Une fois complétées par les informations relatives aux enquêteurs et aux enquêtes, les tables de correspondance régionales sont mises à disposition du responsable du tirage des échantillons d'enquête sur un serveur sécurité FTP.

3. L'affectation des enquêteurs à un échantillon donné

L'opération d'affectation des enquêteurs pour une enquête donnée est faite de façon totalement automatique. Elle consiste en l'appariement d'une table nationale des zones enquêtées/enquêteurs avec le fichier échantillonné. Se faisant en aval du tirage de l'échantillon, cette nouvelle étape n'a aucune incidence sur la chaîne de tirage. Elle n'occasionne aucune charge de travail complémentaire pour l'équipe en charge du tirage.

3.1 Appariement du fichier échantillonné avec les tables de correspondance régionale

Les fichiers régionaux BLAISE des tables de correspondance sont mis au format SAS. A partir des ces tables SAS régionales, nous obtenons une table SAS nationale comprenant les trois types de variables. Les variables de localisation géographiques, les variables d'identification de l'enquêteur et la variable d'identification d'enquête. Une fois la sélection sur la variable d'identification d'enquête faite, le résultat est apparié avec le fichier échantillonné suivant les variables de localisation géographique.

4. Intégration du fichier échantillonné enrichi de la variable « enquêteur » dans le dispositif CAPI

Le fichier échantillonné enrichi de la variable « enquêteur » est fractionné en deux temps. En un premier temps en fichiers régionaux à destination des gestionnaires et en second temps en fichiers enquêteurs à destination des enquêteurs.

4.1 Quelques informations sur le dispositif CAPI à l'I.N.S.E.E

Le dispositif CAPI à l'I.N.S.E.E repose sur trois types de postes ou applicatifs. Un poste de collecte, un poste de gestion et dans certain cas, un poste concepteur. Le poste de collecte est un applicatif installé sur le micro ordinateur portable de l'enquêteur lui permettant de télécharger son échantillon, c'est à dire ses fiches adresses, de faire la collecte et transmettre l'information collectée sur un serveur FTP. Le poste de gestion est un applicatif installé sur le micro ordinateur de bureau du gestionnaire lui permettant actuellement de préparer et gérer la collecte, c'est à dire suivre l'état d'avancement de la collecte, procéder le cas échéant à de l'apurement et enfin valider les enquêtes pour la rémunération des enquêteurs. A terme, avec la mise en place de ce nouveau dispositif de répartition de l'échantillon entre enquêteur, le poste de gestion intégrera une nouvelle fonctionnalité de validation ou modification de la répartition des fiches adresses entre enquêteurs. Le poste concepteur est un applicatif développé uniquement à la demande du concepteur d'enquête. Le

développement d'un tel applicatif est demandé uniquement lorsque la complexité du questionnement impose qu'une partie de l'apurement soit fait par le concepteur de l'enquête.

4.2 Mise à disposition des échantillons régionaux sur le poste de gestion

Les fichiers régionaux d'une enquête sont mis à disposition des gestionnaires des Directions Régionales pour validation directe ou validation après ré-affectation. Cette validation directe ou après ré-affectation se fait à partir du poste de gestion CAPI.

4.3 Mise à disposition des échantillons enquêteurs sur le poste de collecte enquêteur

Une fois validés par les gestionnaires d'enquêtes, les fichiers régionaux sont éclatés en fichiers enquêteurs et mis à disposition de ceux-ci afin de récupération sur le poste de collecte enquêteur à partir d'une connexion téléphonique sur le serveur FTP.

5. Résultats des premières simulations

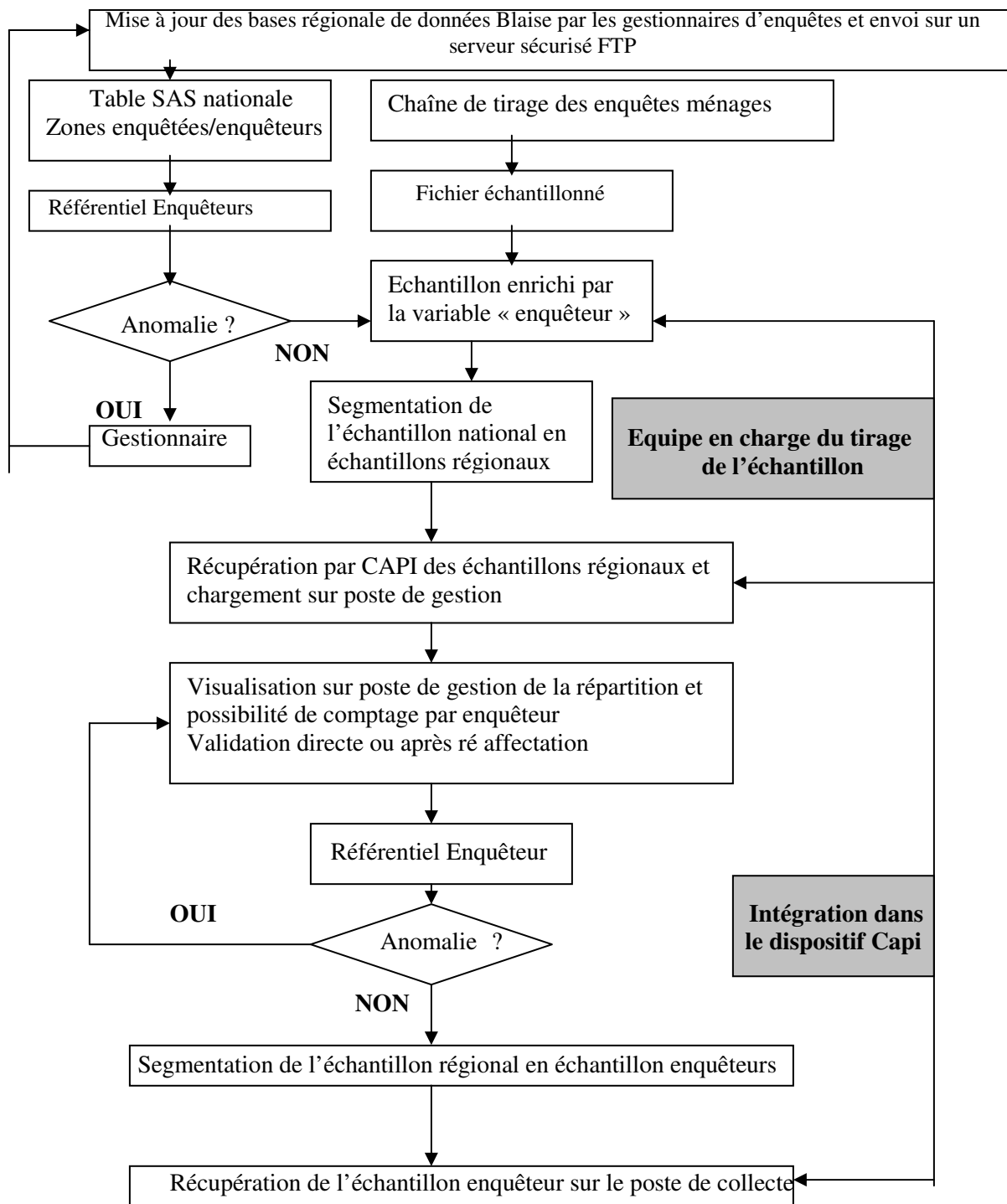
Afin de quantifier le volume des pré affectations, les tests ont été faits auprès de huit directions régionales représentatives en terme de charge d'enquêtes. Ces tests ont portés sur deux types d'enquêtes, un premier type d'enquête, une enquête de grande taille, 15000 logements interrogés, ayant une périodicité de cinq ans, l'enquête PATRIMOINE et un second type d'enquête, des enquêtes de taille moyenne, 8000 logements interrogés, ayant une périodicité de trois fois par an, les enquêtes sur les Conditions de Vie des Ménages (Enquête EPCV). Les logements neufs ne pouvant être îlotés ont été préalablement retirés.

Taux de réussite sur les pré-affectations (en %)

<i>Région</i>	<i>Patrimoine</i>	<i>EPCV Janvier 2003</i>	<i>EPCV Juillet 2003</i>	<i>EPCV Octobre 2003</i>
<i>Centre</i>	64	50	46	39
<i>Nord-Pas-de- Calais</i>	36	65	60	61
<i>Lorraine</i>	42	54	50	44
<i>Alsace</i>	31	33	42	32
<i>Franche-Comté</i>	57	66	54	58
<i>Pays de Loire</i>	40	45	40	42
<i>Rhône-Alpes</i>	57	69	69	52
<i>Languedoc- Roussillon</i>	53	55	60	59
<i>S/Total</i>	48	55	53	48
<i>Total</i>	48		52	

Dans certains cas, 69% des fiches adresses sont affectées a priori correctement. Les différences régionales s'expliquent par le soin et la rigueur apportée à la constitution des tables de correspondance. Ceci nous amène à penser qu'avec un suivi plus rigoureux de ces tables, ces taux pourraient nettement s'améliorer.

6. Schéma d'ensemble de la répartition automatique des fiches adresses



7. Références

Georges Bourdallé, Marc Christine, Laurent Wilms, Echantillons Maître et Emploi, 2000, Actes des journées de méthodologie statistique, Insee Méthodes, N° 100.