

CAHIER DE ReCHERCHE

AOUT 1997



N° 100

L'IMPENSÉ REBELLE :

L'IDENTIFICATION DES FACTEURS
D'INCERTITUDE DANS LES ENQUÊTES
SUR FICHIERS

**PATRICK DUBÉCHOT
MARIE-ODILE SIMON**

Département " Évaluation des Politiques Sociales "

avec la collaboration de

**Sonia EUGÈNE
du Service des Enquêtes**

Crédoc - Cahier de recherche. N°
0100. Août 1997.

CREDOC•Bibliothèque



CRÉDOC

ENTREPRISE DE RECHERCHE

x

CREDOC

L'IMPENSÉ REBELLE :

L'IDENTIFICATION DES FACTEURS D'INCERTITUDE
DANS LES ENQUÊTES SUR FICHIERS

Patrick DUBÉCHOT
Marie-Odile SIMON

en collaboration avec

Sonia EUGÈNE
du service des Enquêtes

Département "Évaluations des Politiques Sociales"

Mai 1997

142, rue du Chevaleret
7 5 0 1 3 - P A R I S

SOMMAIRE

INTRODUCTION.....	4
PREMIERE PARTIE : LA CONSTITUTION DU FICHIER	9
1. L'UTILISATION DE FICHIERS EXISTANTS.....	12
1.1. Un fichier exhaustif à partir d'une seule source.	12
1.2. Un fichier élaboré à partir de plusieurs fichiers existants et harmonisés.....	13
1.3. Un fichier élaboré à partir de plusieurs listes ou fichiers existants et non harmonisés	14
1.4. Un fichier construit à partir d'un fichier initial existant, incomplet ou inadéquat.....	15
2. LA CONSTRUCTION DE FICHIERS QUI N'EXISTAIENT PAS AVANT L'ETUDE.....	20
2.1. La constitution des listes par un intermédiaire	20
2.2. La sélection des enquêtés directement par l'intermédiaire.....	22
3. LES FREINS A LA CONSTITUTION DE FICHIERS.....	25
3.1. Les freins institutionnels : les conditions exigées par la CNIL.....	25
3.2. La négociation avec les intermédiaires.....	27
DEUXIEME PARTIE : LA CONSTRUCTION DE L'ECHANTILLON.....	29
1. LA BASE DE SONDAGE IDEALE : LE FICHIER EXHAUSTIF.....	31
1.1. Les biais d'un fichier exhaustif.....	32
1.1.1. <i>Les doubles comptes</i>	32
1.1.2. <i>Les omissions</i>	32
1.2. Le tirage des unités statistiques.....	33
2. TIRAGE SUR UN FICHIER PARTIEL : ECHANTILLONNAGE A PLUSIEURS DEGRES	34
2.1. Les intérêts d'une méthode d'échantillonnage en plusieurs étapes.....	35
2.1.1. <i>Limiter la taille et la connaissance du fichier</i>	35
2.1.2. <i>Limiter le temps de constitution du fichier</i>	36
2.1.3. <i>Limiter le coût</i>	36
2.2. Quelle conséquence cela entraîne-t-il sur la précision des résultats?	37

2.3. Comment choisir les unités primaires ?	37
2.3.1. La nature de l'unité primaire.....	37
3. LES BIAIS A L'ECHANTILLONNAGE : LES REFUS OU LES NON-REPONSE.....	40
3.1. Un fichier parfois incomplet	40
3.2. Les fichiers anciens.....	42
TROISIEME PARTIE : INFLUENCE DU TYPE DE FICHER SUR LA PRISE DE CONTACT AVEC LES ENQUETES	43
1. LES IMPLICATIONS DU MODE DE COMMUNICATION RETENU POUR UNE PRISE DE CONTACT ET POUR LA PASSATION D'UN QUESTIONNAIRE.....	46
1.1. Le courrier.....	46
1.2. Le téléphone.....	48
1.3. Le face-à-face	48
2. LA PRISE DE CONTACT A PARTIR D'UN FICHER ANCIEN.....	50
2.1. Influence de la formulation du coupon sur le taux de refus.....	50
2.2. Retrouver les personnes « injoignables » : une nécessité lorsque l'on enquête sur la base d'un fichier ancien.....	51
2.3. Les recherches complémentaires.....	52
2.3.1. Des adresses obsolètes à réactualiser.....	53
2.3.2. L'extension des recherches aux réseaux de l'action sociale et aux employeurs.....	54
2.3.3. Les démarches de terrain	55
3. LA PRESENTATION DE L'ETUDE.....	58
3.1. L'homogénéité de l'information	58
3.2. Une source d'information uniforme.....	58
3.3. Des sources d'informations multiples.....	60
3.3.1. Une présentation de l'étude hétérogène.....	60
3.3.2. Un rapport à l'information différenciée	61
4. LE TEMPS COMME FACTEUR DE REUSSITE.....	62
CONCLUSION.....	64
ANNEXES	73
annexe 1 : bibliographie.....	74
annexe 2 : fiches études.....	77

INTRODUCTION

Décidée dans son principe dès l'été 1995, la mise en oeuvre par le CRÉDOC d'une étude sur le devenir de personnes passées, plusieurs années auparavant, par des structures d'insertion par l'activité économique n'a été effective qu'à l'automne. Elle a débuté par la construction d'un fichier à partir de listes fournies directement par ces structures d'insertion¹. Ce fichier a servi de base à la sélection d'un échantillon. La recherche des personnes et la réalisation des entretiens ont débuté à partir du mois de décembre 1995. Les derniers ont été réalisés au début du mois de juillet. L'exploitation des 452 entretiens et la rédaction du rapport se sont achevées en septembre 1996. Ce délai d'un an, plus long qu'il n'était prévu, entre la conception de l'étude et la fin de la réalisation de l'enquête sur le terrain, témoigne d'une réelle complexité des problèmes à résoudre pour faire coïncider le souci déontologique, la méthodologie liée à aux enquêtes sur fichier et la possibilité de produire des résultats pertinents et valides.

Les difficultés rencontrées dans cette enquête font écho aux problèmes auxquels le CRÉDOC a été confronté dans l'étude ASE/PJJ² menée en 1992, proche dans son principe, puisqu'il s'agissait de retrouver des personnes sorties, cinq années auparavant, des dispositifs et mesures de l'Aide sociale à l'enfance et de la Protection Judiciaire de la Jeunesse. Ces deux enquêtes, ainsi que beaucoup d'enquêtes réalisées par le département Évaluation des Politiques Sociales, s'appuient sur un fichier nominatif. Les situations sont évidemment très différentes d'une enquête à l'autre. Parfois il existe un fichier unique, mais la plupart du temps, il est nécessaire d'agréger plusieurs sources existantes ou même de constituer des listes originelles pour construire ex-nihilo le fichier qui servira de base à l'échantillon.

Depuis 1993, le département Évaluation des Politiques Sociales a, à plusieurs reprises, publié des cahiers de recherche visant à s'interroger sur les modalités concrètes de recueil de l'information. Cette démarche s'inscrit dans une réflexion générale sur la qualité dans les enquêtes, réflexion qui participe à un mouvement plus vaste dans le domaine des études et de la recherche³. Cette attention portée à la qualité apparaît d'autant plus primordiale que le département Évaluation des Politiques Sociales enquête auprès de populations « hors-

¹ Conformément aux articles 15 et 16 de la Loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, nous avons dû obtenir l'accord de la Commission Nationale Informatique et Liberté pour réaliser cette enquête. La procédure obligatoire et ses implications seront évoquées aux différents chapitres de ce cahier de recherche.

² ASE : Aide Sociale à l'Enfance (service du Département) et PJJ : Protection Judiciaire de la Jeunesse (services de l'État, ministère de la Justice).

³ « *La qualité de l'information dans les enquêtes* » recueil des communications des journées d'étude des 13-15 juin 1991 de l'ASU, du CNRS et de l'ENST.

normes », pour lesquelles les cadres théoriques habituels sont mal adaptés. Il faut donc reconstruire de nouvelles méthodologies, dont on souhaite mesurer la pertinence.

Parce que l'interrogation sur la qualité de l'enquête est présente au lancement de chaque nouvelle étude, il nous a semblé nécessaire de développer une réflexion plus globale, qui analyse et capitalise les enseignements de chaque expérience passée.

Un premier rapport a été publié sur la qualité des enquêtes auprès des populations en difficulté d'insertion sociale.¹ Il abordait les conditions de déroulement de ce type d'enquête, et plus particulièrement les effets des caractéristiques socio-démographiques des enquêteurs sur les réponses, ainsi que les conséquences du niveau d'insertion sociale des personnes interrogées sur le déroulement des entretiens et sur la qualité des réponses obtenues. Ce premier travail avait permis d'isoler des pistes pour améliorer la qualité dans le recueil de l'information auprès de ces populations, en particulier au moment du déroulement des entretiens. Le deuxième cahier de recherche² s'est intéressé aux spécificités méthodologiques des enquêtes auprès des publics marginaux. Il poursuivait la réflexion du premier rapport, en l'étendant à d'autres populations particulières, encore peu objet d'enquêtes quantitatives. Ce travail avait montré la variété des méthodologies possibles, et l'absence de modèles théoriques. A partir d'exemples d'enquêtes du CRÉDOC, il reprenait les différentes étapes de la conception d'une étude -l'échantillonnage, le repérage des populations, le déroulement de l'entretien et même l'exploitation-, en pointant les problèmes spécifiques que posent de telles populations, les choix adoptés par le CRÉDOC pour les résoudre et le bilan de ces options méthodologiques.

Ce nouveau cahier de recherche s'inscrit dans cette même démarche de réflexion sur les possibilités d'amélioration de la qualité des enquêtes dans trois registres : la construction des fichiers, l'échantillonnage et la prise de contact avec les enquêtés. Il considère cette question sous un angle particulier, qui est celui des enquêtes sur fichier nominatif, en situation d'incertitude. Par situation d'incertitude, il faut entendre des enquêtes à réaliser auprès de populations difficilement repérables, accessibles : fichier à constituer, incomplet, inadéquat, de mauvaise qualité ou très ancien... Dans les études du département Évaluation des Politiques sociales, l'utilisation de fichiers nominatifs est fréquente pour procéder à une Évaluation d'un dispositif, d'une mesure destinée à une population ou pour analyser l'efficacité de l'intervention d'un organisme ou d'une institution. Il s'agit d'enquêtes pour lesquelles il est nécessaire de repérer des individus en situation particulière, qui représentent une faible part de la population, et qu'il faut avoir identifiés au préalable. Il est difficile en effet d'envisager une recherche au

¹ Patrick DUBÉCHOT, Michel LEGROS, *La qualité des enquêtes auprès de populations en difficulté d'insertion sociale*, Coll. des Cahiers de Recherche du CRÉDOC, n°47, avril 1993

² Marie-Odile GILLES-SIMON, *Les spécificités des enquêtes quantitatives auprès de populations socialement marginales*, Coll. des Cahiers de Recherche du CRÉDOC, n° 68, avril 1995.

hasard parmi l'ensemble de la population, par l'intermédiaire de questions filtres qui permettent de sélectionner les individus concernés. C'est pourquoi, la première phase de l'enquête est destinée à construire la liste des personnes dans le champ de l'étude.

Cependant l'utilisation de listes de noms ou de fichiers nominatifs pose la question du respect de la liberté des individus et de la garantie de leur anonymat. La Commission Nationale Informatique et Liberté est vigilante sur cette question et intervient directement, par un accord ou un refus du protocole de l'enquête, sur la façon dont sont utilisés les fichiers nominatifs déjà existants ou constitués, et dans les modalités de mise en oeuvre. Il faut cependant préciser que dans les enquêtes du CRÉDOC, le nom est simplement utilisé pour repérer les personnes et les coordonnées, mais n'est pas conservé dans l'analyse qui n'est pas personnalisée. Il n'y a évidemment pas de traitement nominatif des données.

Le premier objectif de ce document est d'analyser et de capitaliser l'ensemble des expériences accumulées par le Département Évaluation des Politiques Sociales. Le Service des Enquêtes du CRÉDOC est associé à cette réflexion puisqu'il a participé à l'élaboration et à la réalisation des nombreuses enquêtes qui alimentent ce cahier de recherche. Le deuxième objectif vise à formaliser et à élaborer une problématique de construction d'une population-mère, formalisée par un fichier nominatif, puis d'un échantillon représentatif de cette population-mère, ou si cela s'avère impossible de construire un échantillon « diversifié ». Indiquer les méthodes d'investigation, rappeler les contraintes et analyser les difficultés rencontrées dans l'utilisation de fichiers *ad-hoc*, tel est le double objectif de ce cahier de recherche.

Chemin faisant, une analyse des procédures mises en place dans le cadre d'enquêtes réalisées à partir de fichiers, permet de mettre à jour un espace d'incertitude, aussi bien dans la constitution de l'échantillon que dans la recherche des personnes interrogées, qui échappe à toute conceptualisation. Si en effet, les règles de construction d'un échantillon sont bien connues et constituent le cadre de l'intervention du département Évaluation des Politiques Sociales du CRÉDOC, il est clair qu'un effort particulier d'adaptation ou de « traduction » doit être réalisé dans les enquêtes sur fichier pour que ce cadre maîtrisé puisse coïncider avec une réalité rebelle par nature aux catégorisations et aux simplifications. L'une des ambitions de ce cahier de recherche vise à prendre en compte cet « impensé » qui relève le plus souvent de la pratique ou de l'expérience, dont dépendent pourtant les résultats définitifs de l'étude, et propose plusieurs manières d'en réduire l'aléa.

La première partie de ce document présente les différents types de situations auxquelles nous avons été confrontés au départ d'une enquête. Un premier type de situation concerne l'utilisation de fichiers existants, élaborés avec un objectif différent que celui de réaliser une enquête ; ce cas de figure présente des variantes car on peut disposer d'un fichier unique, ou bien de fichiers existants mais qu'il est nécessaire d'agrèger, ou encore d'un fichier incomplet. Enfin, on peut être amené à construire un fichier ex-nihilo, pour l'enquête. Ces différentes situations seront dépeintes et illustrées avec les enquêtes récentes réalisées par le département Évaluation des Politiques Sociales. Les difficultés et les obstacles que nous avons pu rencontrer lors de la constitution de ces fichiers seront également étudiés.

La seconde partie s'attache à souligner les contraintes qui pèsent sur la construction de l'échantillon et qu'entraînent la possession d'un fichier nominatif contenant différentes imperfections. Dans le plan de ce rapport, la réflexion sur la construction de l'échantillon intervient après l'examen de la constitution du fichier. En réalité, ces deux aspects sont le plus souvent traités simultanément : dans la chronologie de l'étude, la méthode d'échantillonnage est choisie en fonction des possibilités de recueil de données, et le fichier est constitué en respectant le plan d'échantillonnage. Cependant, il est possible d'optimiser la méthode choisie de manière à limiter les erreurs d'échantillonnage qui se répercutent ensuite sur les résultats de l'étude.

La troisième partie analyse l'influence du type de fichier sur un moment particulièrement délicat de la phase de l'enquête : la prise de contact avec les individus sélectionnés. Cette phase de prise de contact sera étudiée en fonction des paramètres déterminés par les types de fichiers tels que les différents modes de communication de l'enquête auprès des enquêtés, les démarches complémentaires de recherche, l'homogénéité des informations transmises à l'enquêté.

Ce cahier de recherche ne considère pas le cas particulier des panels¹. Cependant, un certain nombre de problèmes évoqués dans ce rapport rejoignent des difficultés rencontrées dans la mise en oeuvre d'outils de mesure longitudinaux. En particulier, dans un panel, la succession de vagues d'interrogation à intervalles répétés amène à travailler sur un fichier ancien, dont les coordonnées ne sont pas réactualisées. Cependant, des solutions spécifiques peuvent alors être mises en place, qui ne sont pas applicables à une enquête non renouvelée ou à une première vague d'interrogation, que nous n'évoquerons pas ici².

¹ Echantillon de personnes qui sont enquêtées plusieurs fois d'affilée.

² Anne-Lise AUCOUTURIER, Marie-Odile GILLES-SIMON, Michel LEGROS, *intérêt et faisabilité d'un panel d'allocataires du revenu Minimum d'Insertion*, juin 1994.

PREMIÈRE PARTIE

LA CONSTITUTION DU FICHIER

Les études pour lesquelles le département Évaluation des politiques sociales est sollicité ont fréquemment pour objectif d'analyser les évolutions affectant des publics, d'évaluer des mesures ou des dispositifs mis en oeuvre dans différents registres des politiques d'accompagnement des problèmes socio-économiques. Ces travaux s'appuient sur des enquêtes auprès des responsables d'entreprises ou d'établissements concernés, auprès des professionnels chargés de la mise en oeuvre des mesures ou directement auprès des bénéficiaires.

Pour réaliser ce type d'étude auprès d'un échantillon d'individus ou auprès d'institutions, dans la mesure où le sujet concerne une partie seulement de la population ou une catégorie administrative, il est difficile « d'aller au hasard », parce que les personnes recherchées sont soit en faible nombre, soit réparties de façon très particulières dans l'ensemble de la population. L'idéal est donc de posséder un fichier nominatif unique, à l'échelle de l'étude : nationale, régionale, départementale, cantonale, communale. Cependant, cette situation initiale est certainement la moins fréquente et plus le champ de l'étude couvre un territoire important - comme le plan national-, plus elle devient rare. Cette situation de fichier unique constitue certainement un modèle de référence, un « idéal-type » vers lequel il faut tendre ou dont il faut se rapprocher. Dans ce sens, lorsque ce fichier idéal n'existe pas, la démarche consiste à essayer de le construire à partir d'une agrégation de fichiers existants. En effet, la source d'information unique -la liste des personnes présentes dans les structures ou bénéficiaires d'une mesure sur un territoire donné- n'existe pas toujours, il s'agit alors de rassembler l'information dans l'état où elle se trouve et de la mettre en forme. Dans ce cas, il est fait appel à diverses sources disponibles dans les services, structures et institutions concernées par la population enquêtée.

Enfin, le problème de la construction des listes doit tenir compte de l'objectif recherché. Le fichier recouvre une population ciblée -bénéficiaires d'une politique ou d'un dispositif, acteurs de la mise en oeuvre d'une politique ou d'un dispositif- et, dans ce sens sa constitution, sa forme et son contenu ne peuvent être envisagés sans tenir compte des objectifs de l'étude. S'ils exigent une certaine représentativité de l'échantillon, dans l'hypothèse où l'approche est plutôt quantitative, le fichier de base devra être exhaustif ou recenser un nombre suffisamment important d'individus. Si les objectifs de l'étude sont plus proches d'une analyse organisationnelle, qualitative, la question de la représentativité est moins déterminante et dans ce cas le fichier sera construit différemment. En particulier, la recherche de l'exhaustivité apparaîtra comme moins fondamentale.

Par ailleurs, la phase de constitution du fichier est aussi importante pour recueillir des informations sur la population ou les organismes enquêtés. Au-delà d'une liste de noms, on essaye dans la mesure du possible, d'obtenir d'autres informations qui servent soit à construire l'échantillon, soit à mieux connaître la population enquêtée -variables socio-démographiques, renseignements sur l'organisme-, et éventuellement à prévoir une analyse croisée des informations déjà recueillies par voie d'enquête et des informations recueillies dans le fichier.¹

Cette première partie présente les différentes situations possibles, en les illustrant par des études menées par le département Évaluation des Politiques Sociales. Certaines d'entre elles s'appuient sur des fichiers existants, qui couvrent de manière exhaustive ou non la population à enquêter. Dans d'autres situations le CRÉDOC a été amené à construire un fichier unique à partir de plusieurs sources et quelquefois à partir de listes élaborées spécialement pour l'enquête.

¹ Ce type d'analyse, qui permet de croiser des informations recueillies au cours d'une enquête et des informations administratives est souvent riche d'enseignement car ces deux sources de données peuvent être complémentaires. Cependant, pour rendre possible ces analyses, il est nécessaire de prendre au préalable un certain nombre de précautions, en particulier en harmonisant les nomenclatures utilisées dans les deux sources. Ce croisement d'informations nécessite un accord préalable de la CNIL.

1. L'utilisation de fichiers existants

1.1. Un fichier exhaustif à partir d'une seule source.

Dans la mesure du possible, pour des raisons de commodité et à cause des contraintes budgétaires et de délais, il est préférable de prendre appui sur ce qui existe déjà. Pour l'organisme d'études, puiser dans une source unique pour construire l'échantillon est une situation idéale, surtout si la source ou l'échantillon couvre de manière exhaustive le public concerné. Le fichier existe alors comme tel, et n'a pas à subir d'aménagements, de transformation pour pouvoir être utilisé dans le cadre de l'enquête. Si le fichier unique constitue un idéal, il n'est pas toujours sans poser de problèmes. En particulier, il est difficile de déterminer si le fichier est véritablement exhaustif, c'est-à-dire sans omissions ni doubles comptes.

Le département Évaluation des Politiques Sociales a rarement connu une telle situation. Il est possible de citer à titre d'exemple de fichier unique, national et exhaustif, le fichier SIREN qui contient toutes les entreprises (avec une réactualisation régulière intégrant les entreprises nouvellement créées et éliminant les entreprises ayant déposé leur bilan) ou pour la population, l'échantillon démographique permanent de l'INSEE¹. Deux études récentes du département Évaluation des Politiques Sociales ont utilisé des fichiers qui se rapprochent de cet idéal. Elles s'inscrivaient dans une démarche d'Évaluation, d'analyse de la mise en oeuvre et de l'application des mesures et de leurs effets. La première concerne l'allègement des charges sociales sur les bas salaires, la seconde l'Allocation de Remplacement pour l'Emploi (ARPE).

Pour réaliser l'étude sur l'allègement des charges sociales sur les bas salaires (cf. annexe fiche n°1)², environ 500 entreprises ont été extraites du fichier SIREN par l'INSEE, en respectant des critères de taille -nombre de salariés- et de salaires. Cet échantillon de départ a permis d'interroger autant de responsables d'entreprises dans huit régions différentes.

¹ 1% de la population est suivie longitudinalement, en conservant à la fois les données issues du recensement et les informations provenant d'actes civils (naissance, mariage, décès...).

² En annexe figurent des fiches de présentation des études auxquelles il sera fait référence dans ce cahier de recherche. Ces fiches récapitulatives permettront aux lecteurs de s'y reporter pour connaître de manière plus approfondie les objectifs de ces études et les méthodes utilisées.

Dans la seconde étude citée, l'ARPE, (cf. annexe fiche n°2), c'est un fichier fourni par l'UNEDIC qui a servi de base à l'interrogation des responsables du personnel ou des chefs d'entreprise et d'établissement dont au moins un salarié était parti en préretraite en bénéficiant de ce nouveau dispositif. Cependant, il n'existait pas d'emblée de fichier des établissements correspondant à ce critère. Il a été tout d'abord nécessaire de consulter le Fichier National des Allocataires (FNA) qui recense les allocataires de prestations versées par les ASSEDIC, y compris l'ARPE. Dans ce fichier apparaît pour chaque allocataire le numéro de SIRET de l'établissement qui l'employait. A partir de cette information, par croisement avec le fichier SIREN, il a été possible de constituer un fichier des établissements.

1.2. Un fichier élaboré à partir de plusieurs fichiers existants et harmonisés

Souvent le fichier doit être constitué par regroupement de fichiers locaux dont l'intérêt réside dans le fait qu'ils sont tous construits sur le même modèle, donc harmonisés. Dans ces conditions, l'agrégation des différents fichiers ne présente pas de grandes difficultés. Ainsi, pour mieux connaître les nouveaux arrivants dans le dispositif RMI, la Délégation Interministérielle au RMI a choisi d'interroger un échantillon représentatif de 900 allocataires entrés récemment dans le dispositif (cf. annexe fiche n°3). La Caisse Nationale d'Allocations Familiales dispose de fichiers statistiques nationaux d'individus mais ces fichiers sont anonymes. Il a donc été nécessaire de s'adresser aux Caisses d'Allocations Familiales (CAF) pour constituer l'échantillon puisque chacune d'elles conserve des fichiers nominatifs des bénéficiaires des prestations. Cette source est quasiment exhaustive, puisque les CAF gèrent les dossiers de 97,5 % des allocataires du RMI de métropole, et la Mutualité Sociale Agricole, les dossiers restants. Il est d'ailleurs envisageable, en particulier dans le cas d'évaluations locales dans des départements ruraux, de combiner les deux sources (CAF et MSA) pour constituer la liste complète des allocataires.

Ces fichiers, qui sont en fait des fichiers de gestion, présentent l'intérêt d'être régulièrement mis à jour. Cependant, ils servent uniquement à calculer le montant des prestations et par conséquent, ils contiennent peu d'éléments descriptifs des individus et les variables fiables sont très limitées : ce sont essentiellement celles qui servent effectivement au calcul des prestations -dans le cas du RMI, il s'agit de la situation familiale, de l'âge, des autres ressources, par exemple-. Le deuxième intérêt de ces fichiers réside dans leur harmonisation qui facilite leur agrégation. Les informations stockées sont basées sur les mêmes nomenclatures d'un département à l'autre, elles sont même saisies à l'aide du même logiciel informatique.

Cependant ces situations où les fichiers sont informatisés et couvrent tout un territoire -le département pour les CAF- sont assez rares. De nombreux organismes ou associations travaillent encore sur la base de dossiers papiers. C'est un élément à vérifier au moment même de la conception de l'enquête, car l'absence de fichier informatique peut considérablement compliquer la procédure. Ainsi, de nombreuses Commissions Locales d'Insertion (CLI), qui suivent les allocataires du RMI dans leur parcours d'insertion, ne sont pas informatisées.

Par ailleurs, dans le cadre de l'étude RMI, le recueil de l'information pose peu de problèmes dans la mesure où dans le comité de pilotage de l'étude figure une instance hiérarchiquement influente -la CNAF- sur les sites enquêtés -les CAF-. En fait, la situation la plus favorable est celle d'une structure mère qui a un ascendant hiérarchique direct sur ses structures ou services déconcentrés. Parfois, les études sont décidées par un organisme fédérateur qui n'a pas de pouvoir hiérarchique sur les structures locales. Cet organisme fédérateur ne peut que demander et souhaiter la participation du local. Dans ce cas, cette participation est facilitée lorsque la fédération s'implique dans la démarche de l'étude. Quelquefois, ce type d'étude prend la forme d'étude-action qui permet de créer une dynamique interne au secteur concerné. Malgré tout, les expériences que nous avons pu connaître dans d'autres enquêtes montrent que l'appui hiérarchique ou la dynamique interne par étude-action pour obtenir des listes ou des fichiers ne fonctionne pas toujours, en particulier dans les administrations. Le poids d'un ministère se révèle peu influent sur le comportement des responsables des structures ou des services locaux. Parfois, pour ceux-ci, le refus est même un moyen de « régler quelques comptes » avec le ministère ou l'autorité de tutelle. Les conflits locaux, ou bien les « conflits entre le national et le local » dont nous n'avons pas connaissance au départ, se révèlent au fur et à mesure de l'enquête...

1.3. Un fichier élaboré à partir de plusieurs listes ou fichiers existants et non harmonisés

Les fichiers non harmonisés constituent une source importante de problèmes dans la construction du fichier de la population mère. Notamment, se pose le problème de l'harmonisation des informations, des critères ou des nomenclatures à partir desquels ils sont construits. Pour un même dispositif, pour une même mesure, les critères ou les nomenclatures retenus d'un fichier à l'autre peuvent être différents : des fichiers prennent en compte les bénéficiaires potentiels, d'autres les bénéficiaires effectifs d'une mesure ; une partie repose sur la date de signature d'une convention, le reste sur la date d'effectivité de la convention ; enfin la taille des entreprises peut être regroupée par tranches de salariés différentes d'une entreprise à l'autre et qui ne se recoupent pas (1 à 99 ; 1 à 49 ; 50 à 199...etc.). Le niveau de détail des

coordonnées des individus est aussi très variable : l'adresse peut-être plus ou moins détaillée, avec un numéro de téléphone ou non.

Le CRÉDOC a effectué plusieurs enquêtes sur les dispositifs de préretraite auprès d'entreprises ayant utilisé ces mesures (cf. annexe fiche n°1 et 2). Si la DARES possède des informations au plan national sur les bénéficiaires, elle ne possède pas de fichiers nominatifs. Pour arriver à constituer une liste des entreprises concernées, le CRÉDOC a pris contact avec les Directions Départementales du Travail et de l'Emploi qui négocient au plan départemental les conventions de préretraite avec les entreprises. Les services sont en cours d'informatisation et les listes anciennes sont souvent manuscrites. Par ailleurs les informations recensées ne sont pas toujours les mêmes d'un département à l'autre. Pour ces différentes raisons, elles ne sont pas forcément directement comparables. Agglomérer ces listes oblige donc à travailler sur les différentes nomenclatures utilisées. Ce travail d'harmonisation est très lourd et n'est souvent possible que sur peu de variables communes à tous les fichiers.

1.4. Un fichier construit à partir d'un fichier initial existant, incomplet ou inadéquat

La réalisation d'une enquête à partir d'un fichier initial existant, mais qui s'avère incomplet ou inadéquat, constitue un quatrième type de situations rencontrées dans les enquêtes du département Évaluation des Politiques Sociales. Fréquemment, dans ce cas, l'étude se propose d'analyser une population ou un dispositif pour lequel il existe une source principale, centralisée, qui doit servir de base à la construction de l'échantillon interrogé. Cependant, il se peut que le fichier couvre un champ trop large par rapport à la cible fixée. Il est alors **inadéquat** -la population enquêtée est « perdue » au milieu d'autres populations qui n'ont rien à voir avec le champ de l'étude-. Dans ce cas, on tente d'isoler la population recherchée en sélectionnant les individus concernés, dans la mesure où les variables qui permettent de faire cette sélection sont présentes dans le fichier. Ainsi, si l'on souhaite interroger des entreprises qui ont mis en place un dispositif particulier, il est possible de les isoler rapidement parmi une liste plus vaste d'entreprises à condition qu'une variable booléenne¹ révèle l'adoption de la mesure. En l'absence de cette variable (existant à l'origine ou éventuellement reconstruite), le repérage est plus difficile et ne peut avoir lieu sans prendre contact avec les entreprises, démarche très souvent longue et coûteuse.

¹ Variable binaire de type OUI/NON.

Le fichier peut aussi être incomplet - toute la population ciblée par l'étude n'y figure pas-, les deux situations -fichier incomplet et inadéquat- étant parfois réunies. Dans ces conditions, le travail du chargé d'études prend deux directions : d'une part, un travail de tri, d'aménagement du fichier existant de manière à isoler la population ciblée, d'autre part une recherche de sources différentes susceptibles de compléter les données existantes. Il ne s'agit plus de mettre bout à bout des fichiers obtenus par des voies différentes comme cela a été développé dans les deux paragraphes précédents, mais de recomposer le premier fichier en y agrégeant les nouvelles sources. Parfois, l'utilisation de cette source principale peut être une exigence du commanditaire. L'un des objectifs de l'étude peut être alors d'actualiser le fichier central ou de proposer des aménagements de manière à le rendre plus fiable, plus manipulable. Cette perspective conduit à décider avec le propriétaire du fichier et le commanditaire de l'étude des variables systématiquement saisies.

Dans le cadre de la phase exploratoire de la première partie du programme de recherche financé par la MIRE sur les emplois de l'intervention sociale (cf. annexe fiche n°4), l'objectif était de recenser les employeurs d'intervenants sociaux dans le département du Val-de-Marne. A cette occasion, le commanditaire fournissait un fichier existant : le fichier FINESS géré par le service statistique du Ministère des Affaires Sociales et de la Solidarité. Ce fichier national recense les établissements sanitaires et sociaux. Les données disponibles dans la nomenclature établie dans ce fichier permettent d'identifier l'établissement : la dénomination, la localisation, le statut, la catégorie de l'établissement et sa mission.

Un premier tri a permis de constituer un sous-fichier des établissements et services implantés dans le Val-de-Marne (environ 1 200 établissements, structures ou services). Une seconde opération de sélection a permis de « sortir » de ce fichier les organismes qui ne correspondaient pas au champ de l'étude (SARL du domaine médical : clinique, centres médicaux...etc.). Cette première opération a réduit le nombre d'établissements à considérer. Cependant, les premiers tests entrepris par téléphone ont montré qu'il était nécessaire de procéder à une vérification plus fine de l'inscription des établissements dans le champ de l'intervention sociale : les établissements du type maison de retraite, aide et service de soins à domicile, services hospitaliers, etc., emploient parfois des intervenants sociaux. Si l'employeur déclarait employé un ou des intervenants sociaux, son organisme était ajouté à la liste ; dans le cas où il n'avait pas ce type de salarié dans ses effectifs, il était retiré du fichier. On est exactement dans le cas où il n'existe pas de moyen simple et systématique d'effectuer un tri : il faut presque réaliser une sélection « au cas par cas ».

Tableau n°1
Nomenclature du fichier FINESS (*)

NOM DE LA DONNEE	LIBELLE	EXEMPLE
NUMET	n° d'établissement (ET)	940000052
NUMEJ	n° de l'entité juridique(EJ)	940721384
RSAEJ	Raison sociale de l'EJ	A.F.A.S.E.R. de Champigny
RSAET	Raison sociale de l'ET	Foyer d'hébergement adultes handicapés
ADRESS1	adresse de l'EJ	<i>pas d'information</i>
ADRESS2	adresse de l'ET	19 Villa Corse
CODPOST	code postal	94430
COMMUNE	commune	Chennevières-sur-Marne
TELEPHONE	téléphone	45-94-23-22
STATUT	statut juridique	60 (association)
CAT	catégorie d'établissement	252 (foyer d'hébergement pour adultes handicapés)
DIS	disciplines	920 (social : hébergement ouvert en établissement ouvert pour adultes et familles)
SANIT	lits sanitaires	<i>pas d'information</i>
TASA	Type d'activité	<i>pas d'information</i>
SOC	lits sociaux	0 (obligatoires)
TASO	type d'activité	12 (hébergement de nuit en structure groupée)
ENSEIG	lits enseignement	<i>pas d'information</i>
TAE	type d'activité	<i>pas d'information</i>

(*) Fichier national des établissements sanitaires et sociaux

Ce fichier incomplet ne permet pas d'établir une liste exhaustive des structures employant des intervenants sociaux dans le département. Par exemple, les Centres Communaux d'Action Sociale, les missions locales, qui relèvent de la compétence municipale ou associative, ne figurent pas dans le fichier. Il ne contient pas d'informations sur des structures ou des services qui se sont mis en place plus récemment à l'occasion du développement de la politique de la Ville. De ce fait, il a été nécessaire de vérifier par téléphone auprès des établissements et services « limites » au champ d'investigation -qui ne figuraient pas dans le fichier-, la présence dans leur personnel d'intervenants sociaux.

Pour compléter ce fichier il s'est avéré nécessaire de chercher d'autres sources d'informations : listes de structures ou services d'organismes financeurs, annuaires,

observatoires (cf. encadré suivant). Les sources sont nombreuses mais chacune d'entre-elles apporte des informations sectorielles et partielles, avec parfois des zones de recoupement.

Encadré n°1

Programme MIRE : Observer les emplois et les métiers du social

Exemples de sources complémentaires possibles

- Un repérage des employeurs à partir de listes de structures ou services obtenues auprès des administrations et organismes financeurs : DDTFP, DDJS, CAF, URSSAF, URIOPSS, les actions financées dans le cadre de Contrats de Ville.

- Pour repérer des professions telles que les conseillères en économie sociale et familiale, il est possible de se référer à l'Union HLM qui publie chaque année un annuaire national établissant la liste des organismes HLM, par département.

- Différents annuaires nationaux recensent des établissements à vocation sociale, par département. On peut citer par exemple : l'annuaire de la FEHAP (Fédération des Etablissements Hospitaliers et de l'Assistance Privée); l'annuaire du CREAM (Centre Régional d'Ile-de-France pour l'Enfance et l'Adolescence Inadaptée) ; l'annuaire de la FNARS (Fédération Nationale des Associations d'Accueil et de Réadaptation Sociale). A titre d'exemple, l'annuaire de la FNARS recense 17 établissements dans le Val-de-Marne. Des informations sur la capacité d'accueil de la structure et sur le personnel socio-éducatif sont fournies dans ces documents.

- Des organismes tels que l'ORS Ile-de-France ou le CEDIAS publient également des annuaires qui recensent, département par département, l'ensemble des organismes sociaux et sanitaires. Par exemple, le CEDIAS dispose de 14 annuaires thématiques portant sur des domaines tels que la santé mentale des enfants et adultes, les établissements d'accueil pour enfants et adolescents ou les maisons de retraite.

D'autres pistes plus empiriques ont été explorées.

- le MINITEL comme outil immédiatement disponible,

Le MINITEL permet d'obtenir la dénomination de la structure ou du service, l'adresse et le numéro de téléphone (davantage s'il existe un serveur -3615-). Cette source d'information posait plusieurs problèmes dans le cadre de l'étude :

- l'information ne nous renseignait pas sur les personnes qui travaillaient dans les structures ou des services, sur le statut de ces personnes.

- plus la dénomination renvoie à un terme générique "foyer", "association", plus la dispersion est grande, et la structure ou le service peut être hors du champ de l'étude.

Pour ces deux raisons, la vérification de l'information est indispensable. A titre d'exemple nous avons réalisé un test sur les 24 foyers répertoriés et en fait, 13 foyers étaient dans le champ de l'étude.

- L'information recueillie à partir des BUDGETS PREVISIONNELS des communes

Dans ces budgets figure la liste des associations, des organismes et autres comités locaux que la commune subventionne. Ces documents fournissent également l'état du personnel communal. Ces documents peuvent être consultés au Service Financier des Préfectures ou dans les communes.

- Les ANNONCES dans les journaux spécialisés

Dans les revues spécialisées -par exemple, "Actualités sociales hebdomadaires"-, les annonces sont suffisamment nombreuses pour qu'à chaque numéro quelques annonces concernent un service ou une structure d'un département de l'Ile-de-france. Le contenu de ces annonces présente plusieurs intérêts. Tout d'abord, l'intitulé du poste recherché figure au centre du texte et permet, par exemple, de repérer qu'une structure d'insertion par l'activité économique recherche un travailleur social. Cette information permet immédiatement de déterminer si ce service peut figurer dans le fichier ou pas. Ces annonces fournissent les coordonnées du service ou de la structure, parfois celles de l'organisme gestionnaire, le nom du responsable. Enfin, souvent le texte renseigne sur les diplômes recherchés, les compétences nécessaires, l'expérience requise pour occuper l'emploi. Autant d'informations intéressantes pour la constitution du fichier, notamment lorsqu'il y a un doute sur la nature des emplois dans des structures.

La méthode par recoupements de fichiers, de listes diverses, est indispensable dans ce type de construction de fichier avec un objectif d'exhaustivité. Cependant, elle prend du temps et nécessite de la méthode, pour éviter les trop grandes zones d'ombre ou les recoupements. Les sources redondantes doivent être évitées. Il est nécessaire de bien connaître les sources statistiques pour savoir lesquelles utiliser. Pour cela, on peut se référer à un certain nombre de recueils ou de bases de données qui existent et qui permettent de faire le tour des sources existantes, en particulier dans le système statistique public¹. Mais il convient aussi d'imaginer des sources spécifiques à l'objet de l'étude, comme dans l'exemple précédent les annonces dans les journaux spécialisés.

¹INSEE méthodes : *la clé des sources* Tomes 1 et 2.

Philippe MOATI : *Les sources pour les études sectorielles*, Intervention dans le cadre du Séminaire de formation CRÉDOC organisé par le Département Dynamique des Marchés.

2. La construction de fichiers qui n'existaient pas avant l'étude

2.1. La constitution des listes par un intermédiaire

Certaines études ne peuvent bénéficier d'un fichier existant au départ. En l'absence de fichier national, départemental ou de fichier constitué par une administration ou un organisme, l'enquête ne peut être mise en oeuvre qu'à la condition de construire entièrement un fichier de la population ciblée. Pour cela, des listes d'individus doivent être constituées par des intermédiaires extérieurs au CRÉDOC, « informateurs institués »¹ comme par exemple des responsables de structures ou de services qui sont en contact et encadrent la population ciblée par l'étude. Ce type de démarche s'adresse, en général, à un plus grand nombre de structures et de services différents que dans les situations précédemment étudiées. L'étude ne peut aboutir qu'avec leur accord et leur collaboration. Dans ces conditions, il faut veiller à obtenir une information homogène qui peut s'agréger facilement. Pour cela, une grille type à remplir est envoyée à ce tiers. Communiquée à toutes les structures ou services, elle permet d'obtenir cette homogénéité. Malgré cette précaution, il faut veiller à ce que les personnes indiquées sur les listes correspondent aux critères requis, et s'attacher à vérifier la façon dont les listes ont été complétées.

La deuxième étape de la démarche consiste à agréger les différentes listes d'individus obtenues de manière à obtenir un fichier unique. De ce fichier sera extrait l'échantillon à enquêter.

L'étude sur le devenir des personnes passées par les structures d'insertion par l'activité économique (cf. annexe fiche n°5) constitue un bon exemple de ce cas de figure. En l'absence de fichier national général au secteur d'activité ou de fichiers par types de dispositifs (association intermédiaire, entreprise d'insertion ou régie de quartier), l'enquête ne pouvait se mettre en place que sur la base de listes constituées par les responsables des structures retenues dans l'échantillon. Des rencontres dans chaque site ont eu lieu afin de présenter les objectifs de l'étude et les modalités de réalisation de l'enquête de terrain. Le département Évaluation des Politiques Sociales a élaboré une grille qui a été distribuée aux responsables. Elle reprenait des rubriques correspondant à des indications permettant de retrouver la personne -le nom et le prénom de la personne, la dernière adresse connue- et des données correspondant à des critères d'échantillonnage -les dates d'entrée et de sortie de la structure, l'âge...etc.-. Le responsable

¹ Jean-Claude COMBESSIE : *La méthode en sociologie*, Coll. Repères, Éd. La Découverte. n° 194, 1996.

devait remplir cette grille en y faisant figurer toutes les personnes salariées par sa structure aux postes d'insertion -donc pas les encadrants techniques permanents- dans la période considérée (premier semestre 1993).

Une autre difficulté se manifeste par le non-respect des critères de sélection fournis par le CRÉDOC. Cette situation conduit à réduire la liste de l'échantillon initial : par exemple des listes élaborées par les responsables des structures d'insertion par l'activité économique, il a été nécessaire de rayer des personnes (189) qui ne correspondaient pas aux critères requis. Les causes sont diverses : l'inattention de la personne qui constitue la liste ou une mauvaise compréhension de ce qui est demandé -dans l'étude prise en exemple ci-dessus, certaines listes correspondaient au deuxième semestre 1993 alors que nous demandions la liste des personnes présentes au premier semestre-. Parfois, les personnes renvoient des listes informatisées déjà saisies ou des listes papiers déjà constituées mais qui ne reprennent pas toutes les informations souhaitées ou sur lesquelles il manque des informations -toujours dans l'enquête SIAE, certaines listes ne contenaient pas l'âge des personnes ou la durée du contrat-. Ces situations posent le problème de l'harmonisation des données pour pouvoir les intégrer dans le fichier de base.

Dans cette situation la plus éloignée de l'idéal évoqué au début de ce chapitre, les difficultés sont nombreuses. Les délais de réponses ont été très variables. Ils ont été totalement dépendants de l'intérêt des responsables des structures d'insertion par l'activité économique pour cette étude, « une de plus » diront certains. Il a fallu en convaincre un bon nombre, parler et parlementer de longs moments avec eux au téléphone pour les convaincre de consacrer de l'énergie et du temps pour cette étude. Les résistances et parfois les refus des responsables de ces structures ont été alimentés pour certains par leur rancœur vis-à-vis des pouvoirs publics -commanditaire d'une étude d'évaluation- qui par ailleurs menaient une politique de désinvestissement -selon ces responsables- qui se manifestait par une réduction des subventions ou des postes d'insertion. Dans ce type de situation, le commanditaire -un ministère- qui n'a aucun lien hiérarchique avec les structures n'est pas d'une grande utilité pour convaincre les intermédiaires à coopérer. Parfois, l'origine de l'étude peut être même un facteur défavorable à la participation de ces tiers lorsqu'on leur demande une collaboration effective et lourde en temps et en moyens.

Enfin, dans ces fichiers élaborés par un tiers, il est impossible de vérifier s'il y a eu des « oublis » d'individus. Ces « omissions » peuvent être dues à de l'inattention ou encore à une mauvaise tenue des fichiers, mais la personne en charge de la constitution de la liste a pu aussi sélectionner les individus de la liste, et volontairement éliminer des noms. Les contacts par téléphone et parfois directs -lorsque cela s'avère nécessaire- avec les responsables des structures

permettent au chargé d'études de percevoir la nature des difficultés, des résistances ou des blocages et de comprendre leur origine. Ce contact et cette connaissance favorise fréquemment la résolution des difficultés, voire permet de les anticiper.

2.2. La sélection des enquêtés directement par l'intermédiaire

Comme dans la situation précédente, si l'enquête ne peut s'appuyer ou ne s'appuie pas volontairement sur un fichier existant (le fichier existe, mais un autre mode d'échantillonnage a été choisi) parce qu'elle cible des structures réparties sur un ou plusieurs sites¹ précis, on peut être amené à demander à un tiers de faire une sélection. Dans certaines études, l'objectif est d'analyser des comportements, des attitudes dans un espace ou dans un temps donné. La démarche se rapproche davantage de l'analyse organisationnelle et de l'analyse institutionnelle d'un type de public dans un lieu ou un espace circonscrit. Il importe alors de posséder un échantillon représentatif des lieux et ensuite de cibler les personnes pertinentes à interviewer pour étudier des phénomènes ou des comportements dans ces lieux.

Par exemple, dans l'étude sur la mobilité des personnels dans les hôpitaux publics (cf. annexe fiche n°6), le but n'était pas de travailler sur un échantillon représentatif du personnel hospitalier, mais de pouvoir effectuer des comparaisons entre les différents types de mobilité (interne et externe à l'établissement), la mobilité des différentes catégories de personnel (administratifs, infirmiers, praticiens), les raisons de ces mobilités. Dans la mesure où l'objectif n'était pas de travailler sur un échantillon représentatif, il n'était pas nécessaire de partir d'un fichier existant du personnel. Par ailleurs, les contraintes de temps liées à cette étude ne permettait pas de rechercher et d'adapter des fichiers qui pourraient exister (fichier des salaires). Une quarantaine de directeurs des ressources humaines et de directeurs des affaires médicales ont alors été interviewés au cours d'entretiens semi-directifs. Ils ont proposé et sélectionné des agents et des praticiens (400), en respectant un certain nombre de critères qui leur étaient proposés. Cependant, on a pu constater que la mobilité des personnels interrogés était plus forte que la moyenne du personnel hospitalier. Les tiers ont sélectionné d'eux-mêmes des personnes dont le parcours était « intéressant » par rapport au thème de l'étude et n'ont pas joué le jeu qui consiste à choisir des salariés au hasard.

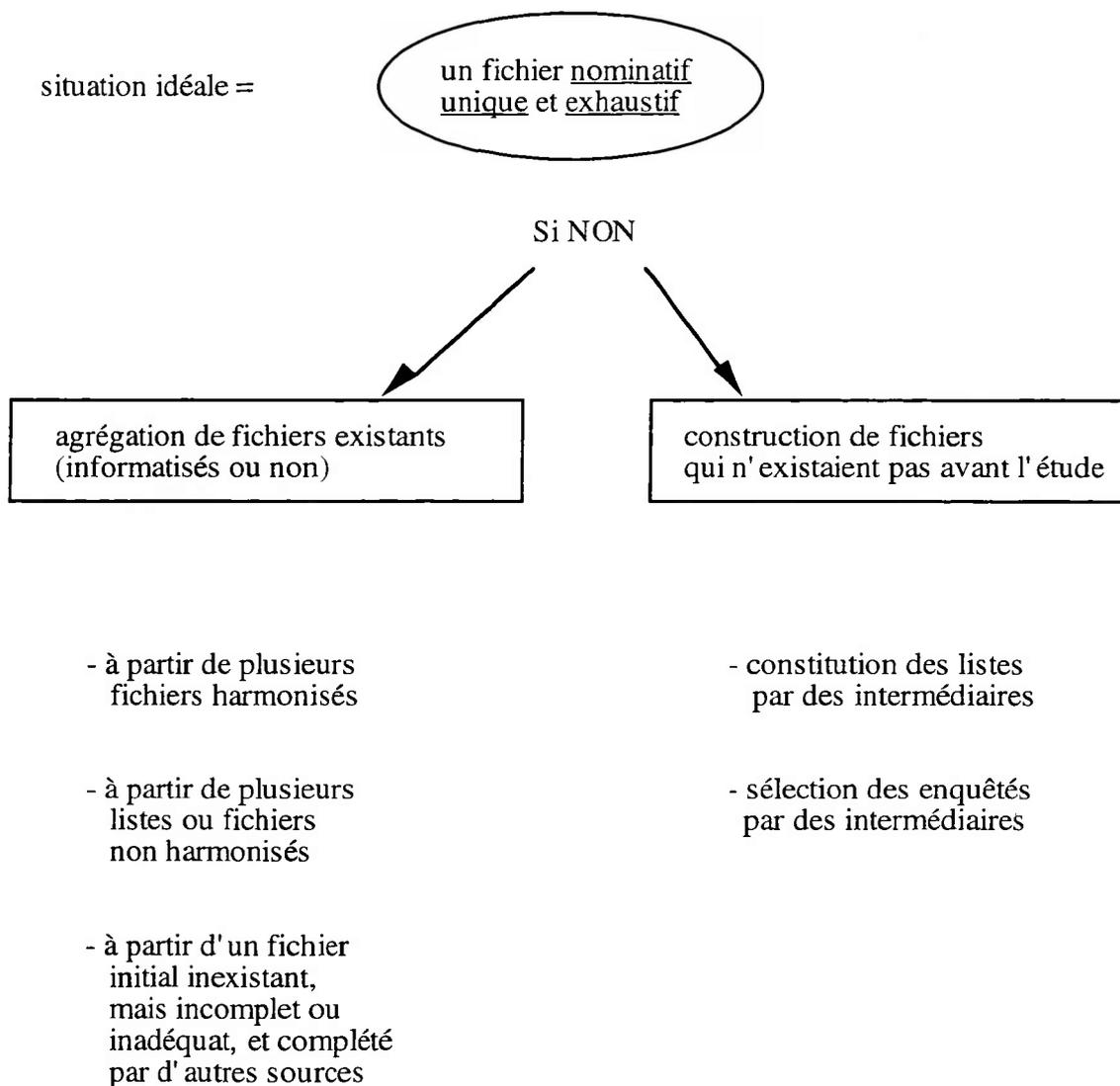
¹ Patrick DUBÉCHOT, Marie-Odile GILLES-SIMON, Michel LEGROS, *L'aire et la manière, construction et usage de la notion de site dans les procédures d'enquête*, coll. des Cahiers de Recherche du CRÉDOC, septembre 1995, n°78.

Dans l'étude réalisée pour PROMOFAP sur les compétences des éducateurs en prévention spécialisée (cf. annexe fiche n°7), l'objectif était de construire des référentiels d'activités et de compétences. La représentativité de l'échantillon n'est pas une donnée fondamentale dans ce type d'étude où il est plus important de s'assurer de la plus grande diversité possible des profils. Malgré tout, les équipes avaient été sélectionnées par le CRÉDOC en fonction de critères relatifs à l'analyse de l'action des équipes de Prévention spécialisée réalisée dans une phase précédente et les directeurs avaient des consignes et des critères précis pour sélectionner les personnes à interviewer dans leur équipe : âge, durée dans la fonction, durée dans le secteur d'activité, etc.

Ce type de méthode n'est envisageable que dans des études qualitatives, qui portent éventuellement sur un grand nombre d'entretiens, où l'on ne souhaite pas construire des échantillons représentatifs. En effet, malgré toutes les précautions, on ne peut empêcher des biais importants au moment de l'échantillonnage, et même dans le déroulement de l'entretien. En particulier parce qu'il y a souvent un rapport hiérarchique entre les personnes qui sélectionnent et les personnes sélectionnées. Cependant, même dans ce type de situation, le chargé d'études doit malgré tout maîtriser l'échantillonnage. Il n'est pas envisageable de laisser à la seule initiative des directeurs ou responsables des services la sélection des personnes à enquêter. Pour cela, il faut bien connaître le secteur de l'étude soit par des travaux précédents, soit par une étude lors d'une phase précédente. Il faut également veiller à la diversité des profils des personnes choisies en regard de cette connaissance préalable.

Schéma n°1 : LES DIFFERENTES SITUATIONS RENCONTREES

La population étudiée est couverte par ...



3. Les freins à la constitution de fichiers

3.1. Les freins institutionnels : les conditions exigées par la CNIL

Beaucoup de fichiers nominatifs -ceux de la CAF, par exemple- sont constitués à des fins administratives et non dans l'optique de les utiliser pour réaliser des enquêtes. Des structures ou des services possèdent des listes des personnes passées dans leurs établissements ou dans leurs locaux là encore dans une logique de gestion -prix de journée, de salaires, etc.- dans une logique d'analyse et d'évaluation de leur action -fiches individuelles concernant les personnes suivies, listes de participants aux activités, etc.-. Quelque soit le cas de figure, l'utilisation de telles informations exige un certain nombre de précautions.

Dans la plupart des enquêtes présentées dans ce document, l'autorisation de la Commission Nationale Informatique et Liberté (CNIL) a été nécessaire pour pouvoir procéder à l'étude. Cette autorisation est obligatoire pour accéder aux fichiers nominatifs gérés par des institutions extérieures et pour les utiliser dans un autre but que celui pour lequel ils ont été mis en place initialement, ou encore pour créer un nouveau fichier. Cette contrainte, tout en constituant une garantie de la protection des individus et qui est tout à fait compréhensible d'un point de vue déontologique, constitue certainement une difficulté supplémentaire lors de la constitution des fichiers de base nécessaires à l'étude et dans la mise en oeuvre des études. La procédure CNIL, même lorsqu'elle est prévue dans le calendrier de réalisation de l'étude et l'échéancier, réserve toujours des « surprises », crée une zone d'incertitudes qui fréquemment s'étend au-delà des prévisions et allonge le temps de réalisation. Enfin, outre l'accord nécessaire pour créer le fichier, la CNIL exige que l'on obtienne l'accord de chaque personne concernée.

LOI DU 6 JANVIER 1978

relative à l'informatique, aux fichiers et aux libertés

L'informatique doit être au service de chaque citoyen. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques. (article 1er)

...

Est dénommé traitement informatisé d'informations nominatives au sens de la présente loi tout ensemble d'opérations réalisées par les moyens automatiques, relatif à la collecte, l'enregistrement, l'élaboration, la modification, la conservation et la destruction d'informations nominatives ainsi que tout ensemble d'opérations de même nature se rapportant à l'exploitation de fichiers ou bases de données et notamment les interconnexions ou rapprochements, consultations ou communications d'informations nominatives (article 5).

...

Hormis les cas où ils doivent être autorisés par la loi, les traitements automatisés d'informations nominatives opérés pour le compte de l'État, d'un établissement public ou d'une collectivité territoriale, ou d'une personne morale de droit privé gérant un service public, sont décidés par un acte réglementaire pris après avis motivé de la CNIL.

Si au terme d'un délai de deux mois renouvelable une seule fois sur décision du président, l'avis de la commission n'est pas notifié, il est réputé favorable (article 15).

...

Toute personne physique a le droit de s'opposer, pour des raisons légitimes, à ce que des informations nominatives la concernant fassent l'objet d'un traitement (article 26).

...

Les personnes auprès desquelles sont recueillies des informations nominatives doivent être informées :

- du caractère obligatoire ou facultatif des réponses ;
- des conséquences à leur égard d'un défaut de réponse ;
- des personnes physiques ou morales destinataires des informations ;
- de l'existence d'un droit d'accès et de rectification.

Lorsque de telles informations sont recueillies par voie de questionnaires, ceux-ci doivent porter mention de ces prescriptions (article 27).

...

Sauf dispositions législatives contraires, les informations ne doivent pas être conservées sous forme nominative au-delà de la durée prévue à la demande d'avis ou à la déclaration, à moins que leur conservation ne soit autorisée par la commission (article 28).

(Journal officiel du 7 janvier 1978 et rectificatif au J.O. du 25 janvier 1978)

3.2. La négociation avec les intermédiaires

Dans les enquêtes pour lesquelles il n'existe pas de fichiers existants et qui nous conduisent à demander aux responsables d'organisme ou de structure de nous communiquer une liste de personnes, les réticences sont fréquentes. Elles sont motivées par un souci déontologique de ces responsables de ne pas transmettre les coordonnées de leurs usagers ou même de leur personnel à un organisme d'étude sans leur avoir demandé leur avis. Ce souci est tout à l'honneur des professionnels de terrain et les chercheurs du CRÉDOC le comprennent et l'acceptent.

Dès les réunions de présentation de l'étude sur l'insertion par l'activité économique, plusieurs administrateurs et responsables de structures ont manifesté leur réticence à communiquer des noms et des adresses. Quelques uns ont refusé de les communiquer, quelle que soit la procédure, et malgré l'accord de la Commission Nationale Informatique et Liberté. Les professionnels de l'intervention sociale ont parfois un attitude très protectrice vis-à-vis des personnes qu'ils prennent en charge. Certains pensent que l'interview peut nuire à la personne, ou au travail éducatif mis en oeuvre et demandent à connaître le guide d'entretien avant de se prononcer. Dans ce cas, sans leur laisser le guide, il est possible de leur préciser les grands thèmes sur lesquels va porter l'entretien. D'autres estiment que les personnes enquêtées ne sauront pas répondre car elles sont trop « déstructurées » ou d'un niveau scolaire trop faible (situation d'illettrisme, par exemple) et dans ces conditions, parfois, demandent à assister à l'entretien.

Pour vaincre ces résistances des professionnels de l'intervention sociale, il est important de prendre du temps pour les rencontrer afin de leur présenter l'enquête et d'évoquer avec eux les modalités de déroulement. Il faut également donner des contreparties comme la promesse d'avoir les résultats de l'étude soit par une note de synthèse, soit par une restitution orale.

Les enquêtes pour lesquelles les responsables sont sollicités pour constituer des listes de personnes à interviewer prennent souvent du retard. Evidemment ces professionnels sont aussi très occupés par leurs tâches quotidiennes et relèguent au second plan la demande de l'organisme d'étude. La perte de temps est encore plus importante lorsque les organismes n'ont pas de listes établies. Dans le cas de l'étude sur l'insertion par l'activité économique, la majorité des structures n'avaient pas de liste établie des personnes passées dans leurs services. Il a donc été nécessaire de les constituer, et pour cela les responsables ou leurs collaborateurs ont dû se replonger dans leurs archives. La rapidité de constitution de ces listes était très liée à la disponibilité en temps, en personnel, en équipement matériel (informatique notamment).

Plusieurs structures ont invité les enquêteurs du CRÉDOC à venir faire ce travail, sous leur contrôle, faute de temps et de moyens.

En règle générale, si les études sont plutôt bien accueillies, en revanche, les contraintes matérielles qu'elles engendrent sont sources de difficultés, voire d'opposition. Il faut alors négocier, discuter, se déplacer pour obtenir après plusieurs semaines, voire plusieurs mois, la liste des personnes.

Un fois l'accord des responsables obtenu, la vigilance s'impose. Pris dans leur activité quotidienne, ils peuvent oublier la tâche qui leur a été confiée. Il faut alors les relancer. Dans cette situation, le contact par téléphone est le plus efficace. Certains échanges téléphoniques peuvent durer un quart d'heure ou une demi-heure. Cette situation révèle un jeu subtil de connaissance mutuelle et de reconnaissance de l'un et de l'autre. Ce moment peut traduire un désir de connaissance mutuelle, mais il peut signifier également un jeu stratégique. L'interlocuteur « occupe le temps » du chercheur, le contraint à l'écouter en échange du temps que lui va perdre à répondre à la demande du chercheur de constituer une liste. Dans les enquêtes du département Évaluation des Politiques Sociales, les intermédiaires, souvent des intervenants sociaux, possèdent une culture de l'oral et ne considèrent pas le temps passé à négocier au téléphone ou en direct comme un temps perdu. Ce moment consacré aux relances est aussi un temps précieux d'appréciation mutuelle qu'il ne faut pas négliger pour le bon déroulement de l'enquête et par respect pour nos interlocuteurs, sans qui l'enquête ne pourrait avoir lieu.

DEUXIÈME PARTIE
LA CONSTRUCTION DE L'ÉCHANTILLON

L'acquisition des informations est une étape essentielle de la démarche d'étude. Pour des questions de temps et de coût, l'observation porte le plus souvent sur un nombre limité d'unités (individu, ménage, entreprise, ...) qui représentent une fraction de la population étudiée et non la totalité¹. Cependant, on souhaite pouvoir inférer les résultats obtenus sur cette partie de la population à l'ensemble.

Le recueil des informations résulte d'un processus assimilable à une chaîne contenant plusieurs maillons, qui sont déterminés en fonction des objectifs de l'étude : la conception du plan d'échantillonnage, la construction de la base de sondage, la conception du questionnaire, la collecte des données... Toute cette chaîne précède l'analyse et conditionne la précision et la validité des résultats obtenus. Or, chaque étape peut être source d'erreurs qui invalident les conclusions de l'étude. Jacques DESABIE², fait apparaître deux composantes à ce qu'il qualifie « d'erreur totale » : « l'erreur d'échantillonnage », provenant de ce que l'on observe un échantillon d'individus et non l'ensemble de la population, et « l'erreur d'observation », provenant de ce que l'on ne dispose pas d'une observation exacte pour chacune des unités de l'échantillon. Dans ce rapport, c'est plutôt à la première composante de l'erreur que nous nous intéresserons. Nous laisserons donc de côté les effets enquêteur, les considérations sur la conception du questionnaire et sa formulation, pour nous intéresser davantage aux moyens de limiter l'erreur d'échantillonnage, dans le cas particulier d'enquêtes sur fichiers nominatifs.

Dans cette deuxième partie, nous allons détailler en fonction des modes de constitution des fichiers qui ont été vues précédemment les méthodes d'échantillonnage optimales, les problèmes rencontrés et les moyens d'y remédier.

Les méthodes de tirage d'un échantillon appartiennent à deux grandes familles : les méthodes empiriques, ou à choix raisonné, la plus connue étant la méthode des quotas ; les méthodes aléatoires, la plus simple étant le tirage aléatoire simple à probabilités égales. Lorsque l'on possède une base de sondage clairement identifiée, dans le cas qui nous intéresse formalisée par le fichier, ce sont les deuxièmes méthodes qui sont privilégiées. Elles permettent en effet de mieux maîtriser la précision des résultats, en particulier parce que l'on peut mathématiquement calculer les intervalles de confiance associés aux estimateurs. On se limitera donc à l'étude des méthodes aléatoires.

¹ Si c'est la totalité de la population qui est observée, on parle alors de recensement et non plus de sondage.

² Jacques DESABIE, *Théorie et pratique des sondages*, 1966

1. La base de sondage idéale : le fichier exhaustif

L'objectif est de construire un échantillon qui soit « fidèle » à la population étudiée, appelée aussi population-mère, c'est-à-dire qui soit statistiquement représentatif. Il sera alors possible d'extrapoler les résultats obtenus sur l'échantillon à l'ensemble de la population-mère. Dans un certain nombre d'enquêtes du CRÉDOC, il n'a pas été envisageable de construire un échantillon représentatif, parce que la population-mère était trop mal définie, ou mal identifiée. En particulier, c'est le problème que nous avons rencontré dans l'enquête sur la grande pauvreté (cf. annexe fiche n°8), réalisée en 1995 pour le Conseil Économique et Social¹. On essaye alors de concevoir un échantillon qui soit le plus diversifié possible, de manière à bien repérer les différentes situations qui peuvent exister dans la population-mère. On peut qualifier ces études « d'enquêtes qualitatives sur un grand nombre d'observations, sélectionnées par une méthode raisonnée ». Si ces enquêtes apportent des éléments de caractérisation qui permettent une meilleure connaissance des populations, elles possèdent une contrainte majeure : il n'est pas envisageable d'inférer les résultats obtenus sur l'échantillon à la population mère. Les enquêtes regroupées dans la partie précédente sous la rubrique des fichiers sélectionnés par un tiers appartiennent le plus souvent à cette catégorie. Ce type d'enquête ne sera pas pris en compte dans cette partie, où l'on privilégie la construction d'un échantillon représentatif.

L'échantillon représentatif est tiré à partir d'une base de sondage, ici le fichier. Selon la qualité de cette source, c'est un objectif plus ou moins facile à atteindre. La situation idéale suppose que le fichier de départ soit totalement exhaustif, c'est-à-dire qu'il inclue l'intégralité de la population-mère. En effet, pour que le tirage soit représentatif, tous les éléments de la population visée doivent avoir une « chance », c'est-à-dire une probabilité non nulle, de figurer dans l'échantillon. Les deux risques principaux lorsque l'on travaille à partir d'un fichier exhaustif sont les omissions et les doubles-comptes.

¹ Marie-Odile GILLES-SIMON, Michel LEGROS, « Politiques sociales : l'épreuve de la pauvreté », collection des rapports du CRÉDOC, n°159, avril 1995

1.1. Les biais d'un fichier exhaustif

1.1.1. Les doubles comptes

Les unités figurant en double compte dans la base de sondage ont une probabilité plus grande de figurer dans l'échantillon que les autres. Les unités de ce type risquent d'être sur-représentées dans l'échantillon, ce qui peut amener au moment de l'analyse des données à amplifier certains résultats ou certains types de comportement.

En particulier, le risque de double compte est plus élevé lorsque le fichier est obtenu par regroupement de plusieurs listes. Il faut donc mettre en place un certain nombre de tests pour repérer les individus ayant des caractéristiques totalement identiques, ou posséder une variable qui joue le rôle d'identifiant, c'est à dire une variable qui ne prend jamais la même valeur pour deux observations différentes. Ce numéro d'identifiant correspond soit à un identifiant officiel, comme le numéro de sécurité sociale pour les individus ou le numéro SIRET pour les entreprises, soit à un numéro incrémenté spécifiquement pour l'enquête. On peut parfois être tenté d'utiliser comme identifiant un numéro d'inscription ou numéro de dossier, en particulier lorsque l'on cherche à interroger les bénéficiaires d'une prestation. Ce type de variables n'amène pas toujours à bien différencier les individus. En effet, il se peut qu'après une interruption des droits, une personne soit réinscrite avec un nouveau numéro, alors qu'elle figure toujours dans le fichier sous son ancien numéro.

Ainsi, pour l'enquête sur l'insertion par l'activité économique, il s'est avéré possible qu'une même personne figure sur plusieurs listes correspondant à des associations intermédiaires ou entreprises d'insertion différentes, car pendant la période de référence, elle avait pu travailler pour plusieurs structures d'insertion par l'activité économique. Il a donc été nécessaire de vérifier que chaque personne ne figurait bien qu'une seule fois sur l'ensemble des listes.

1.1.2. Les omissions

Si la base de sondage est incomplète ou n'est pas à jour, les unités non mentionnées auront une probabilité nulle d'être tirées. Il peut s'en suivre une sous-représentation d'un type d'unités dans l'échantillon, qui peut par conséquent amener à ignorer ou sous-évaluer certains phénomènes.

Ce type de lacune du fichier est plus difficile à repérer et à corriger que les doubles comptes. Il est clair que l'ancienneté du fichier a tendance à accentuer le risque d'omettre des personnes ou des entreprises. Différents moyens pour réactualiser ou compléter un fichier peuvent être mis en place pour y remédier. En particulier, le fichier des employeurs d'intervenants sociaux constitué par le CRÉDOC pour la MIRE en est un bon exemple : nous sommes partis d'un fichier unique, mais très incomplet, qui a ensuite été mis à jour en croisant de nombreuses autres sources. Cela a été possible car l'unité géographique était assez réduite, un département, et la population à repérer bien connue des chargés d'études du CRÉDOC. Cependant, il est difficile d'être assuré que le fichier a été complété en totalité, à moins de connaître précisément le nombre d'observations contenu dans la population-mère, et des « oublis » peuvent subsister. Compléter le fichier est un travail particulièrement long, qui peut aussi s'avérer très coûteux.

Ce problème des omissions est peu important lorsque l'échantillonnage est réalisé à partir de fichiers de gestion, fichiers de paiement en particulier, comme les fichiers d'allocataires du RMI des CAF, qui sont en pratique actualisés, ou encore le Fichier National des Allocataires de l'UNEDIC.

1.2. Le tirage des unités statistiques

A partir d'un fichier exhaustif, toutes les méthodes de tirage aléatoire sont envisageables, des plus simples, comme les tirages aléatoires à probabilités égales ou inégales, aux plus complexes, comme les tirages stratifiés ou à plusieurs degrés.

Cependant, si l'on connaît peu les caractéristiques de la population étudiée, ou encore si l'on souhaite réaliser une enquête qui mesure des variables de nature très différentes¹, on privilégie plutôt le tirage aléatoire simple, avec probabilité égale de tirage. Chaque unité a alors la même probabilité d'être sélectionnée. Ce mode de tirage n'est pas forcément optimal, dans le sens où les intervalles de confiance ou les marges d'erreurs associés aux estimateurs ne sont pas les plus faibles, mais en l'absence d'une connaissance préalable des caractéristiques de la population observée, c'est celui qui assure un niveau moyen d'erreur. En plus, il est particulièrement simple informatiquement à mettre en oeuvre.

¹La qualité d'un échantillonnage se mesure en comparant l'estimateur à la variable estimée. Dans un questionnaire, plusieurs variables sont estimées. Pour chacune d'entre elles, la précision associée peut être différente.

2. Tirage sur un fichier partiel : échantillonnage à plusieurs degrés

Dans la pratique, le fichier que détient le CRÉDOC est rarement un fichier exhaustif. En particulier, on a vu que lorsqu'il s'agit d'enquêtes nationales, les fichiers déjà existants sont rarement nominatifs : la CNAF possède des informations au niveau national sur l'ensemble des allocataires du RMI ou d'une autre prestation, mais elle ne centralise pas un fichier individuel et nominatif. De même, la DARES connaît bien globalement les entreprises qui ont recours aux différents dispositifs d'aide à l'emploi, mais ne les recense pas au niveau national. Il faut donc revenir à un niveau local, régional ou départemental, pour avoir accès aux informations nominatives.

Lorsqu'il n'existe pas déjà, le fichier complet serait en général trop volumineux et trop coûteux à constituer. Le fichier qui sert de base de sondage résulte donc d'une sélection, qui constitue en fait la première étape d'un échantillonnage à plusieurs degrés. C'est pourquoi l'échantillonnage ne peut pas être dissocié dans la logique d'étude à la constitution du fichier. Ces deux aspects sont le plus souvent intimement liés : la méthode d'échantillonnage est choisie en fonction des possibilités de constitution de fichier, qui est ensuite élaboré en respectant ce plan d'échantillonnage.

La méthode d'échantillonnage à plusieurs degrés s'apparente à la construction d'un échantillon-maître : il s'agit de constituer un échantillon représentatif de la base de sondage, en quelque sorte une sous-base de sondage, moins lourde à manier car moins volumineuse, et dans lequel on puisera l'échantillon définitif. C'est le modèle qu'utilise l'INSEE pour construire l'ensemble de ses échantillons à la suite de chaque recensement, il constitue une réserve de logements dans laquelle il puisera les échantillons des enquêtes au cours de la période intercensitaire. Cette réserve représente un échantillon représentatif au 1/20 de l'ensemble des logements recensés.

Le nombre de degrés dans l'échantillonnage est en général de deux : après un premier tirage d'unités, un deuxième tirage dans chacune de ces unités amène à constituer un échantillon d'individus, de ménages ou encore d'entreprises. Mais il peut éventuellement y avoir plus de degrés selon la configuration (par exemple, sélection de départements, puis sélection d'établissements ou d'associations et enfin sélection de professionnels travaillant dans ces établissements). On distingue les différents niveaux de l'échantillonnage en parlant d'unités primaires pour le premier degré, d'unités secondaires pour le second degré et ainsi de suite,...

A chaque niveau, la méthode de sélection est basée sur un tirage aléatoire simple avec probabilité de tirage égale ou inégale. Dans le choix de la méthode, il est préférable de respecter un certain nombre de règles, de manière à pouvoir dépouiller les résultats de l'enquête comme un recensement, c'est à dire avec un poids de un pour chaque observation.

L'échantillonnage stratifié et l'échantillonnage par grappe sont en fait deux cas particuliers d'un échantillonnage à deux degrés : le premier correspond à une situation où toutes les unités primaires sont considérées (probabilité de tirage *au premier degré* de 1), le deuxième correspond à une situation où ce sont toutes les unités secondaires qui sont prises en compte dans l'échantillon (probabilité de tirage *au second degré* de 1).

Parmi les enquêtes du CRÉDOC décrites précédemment, beaucoup correspondent à des échantillonnages à deux degrés : pour l'interrogation des allocataires du RMI, le fichier a été constitué en sélectionnant 20 départements, puis chaque Caisse d'Allocations Familiales sollicitée a effectué un tirage d'allocataires. Dans l'enquête sur les préretraites financées par le Fonds National pour l'Emploi, une vingtaine de départements ont été choisis, et chaque Direction Départementale du Travail et de l'Emploi a transmis une liste exhaustive des entreprises ayant signé une convention départementale de préretraite sur une période de référence. L'enquête sur l'aménagement des rythmes scolaires a porté sur une trentaine d'établissements expérimentaux, qui ont ensuite pris en compte trois classes d'enfants (cf. annexe fiche n°9).

2.1. Les intérêts d'une méthode d'échantillonnage en plusieurs étapes

Cette méthode présente plusieurs intérêts : elle permet de limiter le volume du fichier à constituer, mais surtout elle réduit le coût et le temps passé pour la constitution. Elle fournit aussi le moyen de construire un échantillon sur une population qui n'est pas complètement connue.

2.1.1. Limiter la taille et la connaissance du fichier

Le tirage de l'échantillon a lieu sur un fichier moins volumineux, ne représentant qu'une partie de la population étudiée, et donc plus maniable. Ainsi, faire un tirage de 900 allocataires du RMI à partir d'une base de sondage recensant le million d'allocataires du RMI nécessite des moyens informatiques disproportionnés par rapport à la taille de l'échantillon qui va ensuite être considéré.

De plus, et c'est un des principaux intérêts de cette méthode d'échantillonnage, on n'a pas besoin pour mettre en oeuvre ce type d'échantillonnage de posséder la liste exhaustive de la population-mère, et il est même possible d'y avoir recours si l'on n'en possède pas les caractéristiques complètes. En d'autres termes, on n'est pas obligé d'avoir des informations précises sur le contenu des unités, primaires ou secondaires, qui ne sont pas sélectionnées. En particulier, il n'est pas nécessaire de connaître le nombre d'observations contenues dans ces unités, même si, comme on le verra par la suite, cette connaissance permet d'améliorer la précision du sondage.

Ceci permet de concevoir des plans de sondage raisonnés y compris sur des populations mal ou partiellement connues.

2.1.2. Limiter le temps de constitution du fichier

Les intermédiaires sollicités pour la constitution du fichier voient cette mission se rajouter à leur charge de travail habituelle, et n'ont pas toujours beaucoup de temps à y consacrer. Dans la mesure du possible, on cherche à limiter le nombre de ces intermédiaires et alléger le travail qu'on leur demande. Ainsi, la transmission du fichier d'allocataires des Caisses d'Allocations Familiales au CRÉDOC leur prend du temps, puisqu'il faut qu'elles fassent une extraction de fichier, que conformément à la procédure imposée par la CNIL elles informent les allocataires de l'enquête afin de leur permettre de refuser que leurs coordonnées soient transmises au CRÉDOC... Limiter le nombre de CAF et le nombre d'allocataires par CAF qui figurent dans l'échantillon permet de réduire la part de travail demandé à des tiers.

2.1.3. Limiter le coût

Cet argument est fortement lié au précédent : l'allégement de la constitution du fichier constitue à la fois une économie de temps et d'argent. Le gain financier est réalisé au moment de la constitution de l'échantillon, mais aussi au moment de la collecte, s'il s'agit d'une enquête en face-à-face et si l'un des degrés du plan d'échantillonnage est constitué par un découpage géographique. Le travail des enquêteurs sur le terrain est alors facilité, les temps de transport réduits et donc le coût de l'enquête diminué.

2.2. Quelle conséquence cela entraîne-t-il sur la précision des résultats ?

Ce mode d'échantillonnage n'apporte pas forcément une plus grande source d'erreur dans les résultats de l'enquête. Si l'on respecte quelques règles simples de construction et sous réserve de détenir quelques informations préalables sur la population-mère et sur la variable étudiée, l'échantillonnage à plusieurs degrés peut même apporter des résultats d'une plus grande précision qu'un tirage aléatoire simple.

Les unités primaires constituent des classes d'unités statistiques, parmi lesquelles sera effectué un nouveau tirage. On montre mathématiquement¹ que pour optimiser le tirage de l'échantillon, on doit s'attacher à ce que la variance interclasse des unités primaires par rapport à la variable étudiée de l'échantillon soit la plus forte possible, et que la variance intra-classe soit la plus faible possible. En d'autres termes, il faut que les unités primaires aient des caractéristiques clairement différenciées, et que chaque classe soit « homogène », c'est-à-dire qu'elle soit constituée d'unités statistiques ayant des positionnements proches par rapport aux variables étudiées. Les unités primaires doivent aussi être en nombre suffisant et il faut dans la mesure du possible qu'elles soient de taille comparables.

Si cette règle est bien respectée, le tirage donne de meilleurs résultats que le tirage aléatoire simple à partir de la population-mère. Cette méthode qui amène à procéder par degrés peut donc être préférable au tirage aléatoire simple. Le choix des unités primaires est donc particulièrement important, aussi bien en ce qui concerne sa nature, que le nombre à sélectionner et la méthode de tirage considérée.

2.3. Comment choisir les unités primaires ?

2.3.1. La nature de l'unité primaire

L'unité primaire la plus fréquente correspond à un découpage géographique. Bien qu'il soit préférable de privilégier le niveau de découpage le plus fin possible (logement, quartier, bassin d'emploi, commune, canton), pour des contraintes de faisabilité, et parce que les fichiers accessibles contiennent bien souvent des informations à cette échelle, le CRÉDOC choisit le plus

¹ Jean-Marie GROSBRAS, *Méthodes statistiques des sondages*, Économica, 1987

fréquemment l'unité départementale. Il nous est néanmoins arrivé d'envisager d'autres unités primaires, comme des écoles pour l'enquête sur les aménagements des rythmes scolaires.

Dans cette hypothèse, les départements sont le plus souvent choisis sur la base d'une typologie, afin de constituer des classes homogènes de départements par rapport au critère observé. Les départements constituent en effet des unités primaires qui sont des entités peu nombreuses et de taille importante, il est donc préférable de stratifier leur sélection de manière à représenter les diversités.

En fait, on a vu que les enquêtes ne s'intéressent pas à une variable unique, mais à un grand nombre de variables à la fois. Comme on recherche une bonne précision de tous les estimateurs calculés à partir de l'enquête, il faudrait pouvoir vérifier l'homogénéité des classes par rapport à l'ensemble des variables observées, ce qui est impossible. Cependant, le questionnaire reste centré sur un thème (l'insertion sociale et professionnelle, la pauvreté, la préretraite,...). Il est donc possible d'essayer de définir une homogénéité par rapport au thème. Les variables qui servent de base à la typologie sont choisies dans ce sens.

Ainsi, une typologie des départements pour les bénéficiaires du RMI a-t-elle pris en compte les informations sur le nombre et les caractéristiques des allocataires, mais aussi des informations plus générales sur la situation économique du département (taux de chômage, taux de chômage longue durée, taux de chômage chez les moins de 25 ans, taux de scolarisation) ou encore sur le taux d'urbanisation du département.

Sur les préretraites, cette typologie était fondée sur la répartition des entreprises par secteur d'activité, le nombre de départs en préretraite selon les différents dispositifs existants,...

La typologie est effectuée par une classification ascendante hiérarchique à partir d'une analyse en composante principale. Cette typologie permet de regrouper les départements en un nombre limité de classes. Une même classe contient des départements ayant des caractéristiques proches sur les variables étudiées, c'est-à-dire proches par rapport au thème de l'étude.

Plusieurs départements sont ensuite sélectionnés dans ces classes. Une vingtaine de départements semble représenter un nombre suffisant pour tenir compte de la diversité des situations, et opérationnel pour ensuite gérer les prises de contacts ultérieures. Le choix des départements est effectué selon un échantillonnage stratifié, de préférence proportionnellement au poids que représentent ces classes par rapport à l'ensemble (en termes de population, de salariés, de bénéficiaires d'un dispositif,...). Ce tirage à probabilité inégale est préférable lorsque les unités statistiques sont de tailles très différentes, ce qui est le cas des départements.

Cependant, elle nécessite de connaître une information supplémentaire qui est le nombre d'unités contenues dans chaque département.

Le fichier sera ensuite constitué uniquement sur ces unités primaires. On tire le même nombre d'unités dans chaque département sélectionné. A ce dernier niveau, on privilégie plutôt un tirage d'un nombre équivalent d'individus par unité.

3. Les biais à l'échantillonnage : les refus ou les non-réponse

Il s'agit ici de s'intéresser aux refus et aux non-réponses à un niveau global, c'est-à-dire sur la participation de l'enquête, et non sur des questions particulières au moment du questionnement.

La question des refus et des non-réponses et des conséquences que cela entraîne sur la qualité de l'échantillonnage se pose dans toutes les enquêtes. Cependant, elle se pose avec encore plus de force lorsque l'on est en situation d'incertitude. Cet aspect a déjà en partie été abordé dans un cahier de recherche précédent¹. Nous reviendrons ici sur deux problèmes particuliers.

3.1. Un fichier parfois incomplet

Que le fichier soit unique ou soit issu d'un regroupement de fichiers, lorsque l'on travaille sur des fichiers d'individus qui sont transmis par un ou plusieurs organismes extérieurs, il est nécessaire de respecter la procédure de transmission de fichier imposée par la CNIL.

En particulier, cette procédure implique que les organismes sollicités aient eux-mêmes le fichier avant de le transmettre au CRÉDOC en éliminant les personnes qui ont signifié qu'elles ne souhaitent pas être interrogées.

Le fichier que possède le CRÉDOC n'est donc plus un fichier exhaustif. Un certain nombre d'individus ont été éliminés et ont une probabilité nulle d'être interrogés. Ceci peut entraîner des biais dans la constitution de l'échantillon, d'autant plus que le nombre de personnes qui ont ainsi été supprimées peut être important. En moyenne, il est de l'ordre de 25%, mais il peut parfois avoisiner 50%, suivant le sujet de l'étude ou la population visée.

Si les caractéristiques de la population mère sont connues, il est alors possible d'envisager un redressement, aussi appelé stratification *a posteriori*, de l'échantillon : on accorde un poids plus important aux catégories de la population qui sont sous-estimées dans l'échantillon par

¹ Marie-Odile GILLES-SIMON, *Les spécificités des enquêtes quantitatives auprès de populations socialement marginales*, op.cit.

rapport à leur représentation dans la population-mère¹. Cependant, si les caractéristiques de cette dernière sont mal connues, ce qui est souvent le cas, les redressements sont impossibles.

De plus, dans le cas d'un tirage à deux degrés, il se peut que les refus se répartissent de façon inégale entre les différentes unités primaires. Cela peut amener des distorsions au moment de la sélection des unités secondaires, à moins de connaître pour chaque unité primaire le nombre d'unités statistiques qu'elles contenaient avant épuration du fichier des refus et d'en tenir compte dans la probabilité de tirage appliquée dans chacune d'entre-elles. C'est une information que l'on ne connaît pas forcément, lorsque le tirage est fait par un organisme extérieur.

Par exemple, dans l'étude qu'a réalisé le CRÉDOC sur les aménagements des rythmes scolaires, il semble que les parents d'origine étrangère aient plus souvent refusé que les parents français que leurs coordonnées soient communiquées au CRÉDOC. Ils ont mal compris la lettre d'information qui leur a été diffusée, et ont préféré refuser de répondre à l'enquête. Ce phénomène a probablement entraîné plus de refus dans les Zones d'Éducation Prioritaires qu'ailleurs. En l'absence d'informations précises sur le nombre d'enfants d'origine étrangère impliqués dans l'expérimentation, il est difficile de vérifier si cela a effectivement été le cas, et d'envisager un redressement a posteriori.

De plus, il faut qu'un minimum de personnes ayant la caractéristique recherchée soient présentes dans l'échantillon pour envisager un redressement. Sinon, on prend le risque d'accorder trop d'importance à un nombre faible de réponses. Les redressements ne sont donc pas toujours possibles et ne constituent pas la solution à tous les biais d'échantillonnage.

Le fait que les personnes aient la possibilité de refuser avant le démarrage de l'étude ne signifie pas qu'elles seront toutes d'accord pour participer lorsqu'elles seront contactées par le CRÉDOC. Il y aura donc une seconde vague de refus. Cependant, cette deuxième série de refus est en général assez faible, et ces refus sont statistiquement plus faciles à maîtriser, car les informations nécessaires au redressement peuvent être demandées au moment du contact avec la personne qui exprime son refus.

¹ Pour une description détaillée des différentes méthodes de redressement d'échantillon pour tenir compte de la non-réponse totale, se référer à l'article de Jean-Claude DEVILLE et Françoise DUPONT, *non-réponse : principe et méthodes*, actes des journées de méthodologie statistique du 15 et 16 décembre 1993, INSEE méthodes, n°56-57-58.

3.2. Les fichiers anciens

Si l'échantillonnage s'effectue sur un fichier ancien, non réactualisé, les coordonnées d'un certain nombre de personnes, d'entreprises ou d'organismes ne seront plus à jour. Même si, on le verra par la suite, il est possible de réactualiser en partie ce fichier, cette ancienneté amène à modifier la probabilité que certaines personnes se retrouvent dans l'échantillon. Les personnes qui ont souvent changé de logement et, tout spécialement celles qui n'ont pas de logement stable, donc les plus difficiles à localiser, auront moins de chance d'être intégrées dans l'échantillon. Or ceci peut entraîner des biais importants selon le sujet de l'étude. En particulier, dans les études sur les parcours d'insertion sociale et professionnelle, les personnes les plus difficiles à retrouver peuvent correspondre à celles qui ont les parcours les plus chaotiques. En l'absence de données de cadrage sur cet aspect de la population de référence, il est impossible d'envisager un redressement *a posteriori* de l'échantillon. C'est pourquoi, afin d'éviter une trop forte sous-représentation des personnes les plus difficiles à joindre dans l'échantillon, on s'oblige sur ce type de fichier à retrouver un certain pourcentage de l'échantillon initial¹.

¹ Cette méthode, et les conséquences sur les résultats de l'étude, ont été détaillées dans le cahier de recherche du CREDOC N°68 *Les spécificités des enquêtes quantitatives auprès de populations socialement marginales*, Marie-Odile GILLES, Avril 1995.

TROISIÈME PARTIE :

INFLUENCE DU TYPE DE FICHIER

SUR LA PRISE DE CONTACT AVEC LES ENQUETÉS

Une fois le fichier constitué et l'échantillonnage réalisé, vient la dernière phase précédant la passation du questionnaire : la prise de contact avec l'individu sélectionné afin d'obtenir son accord pour sa participation à l'étude.

Ce contact permet de présenter l'étude, l'enquêteur et le CRÉDOC à la personne sollicitée, lui exposer les objectifs de l'étude et la motivation du commanditaire. Il a lieu par courrier, au téléphone ou en face à face. Dans le cas d'un envoi par courrier le délai entre la lettre de présentation et la demande de rendez-vous est de plusieurs jours, en revanche, pour les deux autres modes de contact, celui-ci a lieu en général juste avant la passation du questionnaire, à moins que la personne accepte de participer mais ne soit pas disponible dans l'immédiat et qu'il faille convenir d'un rendez-vous.

Pour l'enquêté, ce contact peut être vécu comme une intrusion. L'enquête étant basée sur un fichier nominatif, il réalise qu'il a été sélectionné spécifiquement et, éventuellement recherché et retrouvé. Sa motivation à accepter ou à refuser sa participation à l'enquête peut dépendre de plusieurs facteurs :

- du sujet de l'étude (la personne se sent-elle concernée ? a-t-elle l'impression de savoir quelque chose à ce sujet ou d'avoir des choses à dire ?)
- du support utilisé (courrier, téléphone, face à face, contact direct précédé ou non d'un contact indirect ...) pour les transmettre.

Si l'on se replace dans le cas d'un fichier exhaustif, unique et actualisé, tous les modes de communication sont imaginables, en considération du thème et du public. Dès que le fichier, de par sa construction, l'hétérogénéité des sources mobilisées ou encore des informations collectées, peut présenter des erreurs, la méthodologie à adopter pour contacter les individus doit respecter un certain nombre de règles : il s'agira en effet de limiter l'introduction de biais spécifiques à cette nouvelle étape du recueil de l'information, en particulier en travaillant sur la réduction de taux de refus ou de non-réponse.

Il n'est pas dans notre propos de développer ici les problèmes posés par le choix d'un mode de contact et de passation du questionnaire par rapport à un autre. Il est cependant important de les avoir à l'esprit car à ces éléments méthodologiques s'ajoutent les contraintes imposées par le type de fichier.

Après avoir rappelé brièvement les conséquences du choix d'une modalité pour la prise de contact, nous orienterons notre réflexion sur les implications liées au type de fichier, en nous intéressant plus particulièrement au cas du fichier ancien. Nous développerons ensuite les difficultés inhérentes à la présentation de l'étude.

1. Les implications du mode de communication retenu pour une prise de contact et pour la passation d'un questionnaire¹.

Trois modes principaux de communication sont utilisés : le courrier, le téléphone et le face-à-face. D'autres techniques sont envisageables comme la télécopie, le serveur minitel ou internet, mais ils ne seront pas abordés dans ce chapitre à cause de leur faible fréquence d'utilisation pour le moment. L'analyse de l'influence des modes de communication est traitée par rapport aux publics et aux sujets d'enquête sur lesquels travaille le département Évaluation des Politiques sociales. Seront pris en considération plus particulièrement les éléments qui favorisent ou au contraire diminuent le taux de participation à l'enquête.

1.1. Le courrier

Le courrier est le mode de prise de contact le plus impersonnel, le plus systématique, et qui est facilement, et à moindre coût, utilisable à grande échelle. Par courrier on peut envoyer un questionnaire, solliciter un accord de principe pour la participation à l'étude ou tout simplement annoncer une enquête. Nous ne parlerons pas ici de l'interrogation par voie postale, peu utilisée au CRÉDOC et réservée à des enquêtes très spécifiques. Seuls les deux derniers aspects du recours au courrier retiendront notre attention.

En dehors de la procédure CNIL, selon laquelle l'envoi d'un courrier est obligatoire pour obtenir un accord, cette technique est plus volontiers utilisée pour annoncer une enquête, pour officialiser une démarche. C'est le cas dans les enquêtes réalisées par des commanditaires en rapport avec la population à interroger (par exemple une caisse de retraite souhaitant interroger ses adhérents ou un bailleur social faisant une enquête auprès de ses locataires), ou encore dans des enquêtes auprès de professionnels. Dans ce cas, le courrier a un effet d'annonce, il permet d'aider à l'identification de l'interlocuteur (dans une entreprise ou une administration) ou de préparer une réflexion (les thèmes de l'enquête sont annoncés par courrier au préalable ainsi qu'un certain nombre d'informations qui seront demandées lors de l'interrogation téléphonique). Dans ce cas, l'enquêteur qui appellera ou qui passera quelques jours plus tard s'adressera à des personnes déjà informées et déjà sensibilisées à l'étude.

¹ Pour en savoir plus : "La qualité de l'information dans les enquêtes", op. cit.

Cependant en dehors du public des professionnels, le courrier reste un mode de contact très sensible, et de ce fait plus rarement utilisé. En effet, s'il officialise une démarche, il la dépersonnalise d'autant. Lors d'enquêtes avec une première prise de contact par courrier dans le cadre d'une procédure CNIL, les taux de refus sont supérieurs à ceux constatés dans des enquêtes équivalentes où le premier contact a lieu par téléphone ou en face à face. A la réception d'un courrier adressé par un organisme inconnu, l'individu peut s'inquiéter de figurer sur un fichier. La lettre, souvent à double en-tête CRÉDOC et commanditaire, inquiète car elle est assimilée à un courrier « officiel ». D'ailleurs, il peut parfois arriver qu'une confusion s'installe entre la démarche d'étude, qui vise seulement à recueillir des informations et qui n'a aucune conséquence directe pour la personne, et la mission de l'organisme commanditaire qui délivre une prestation ou permet d'accéder à un droit. Ainsi, lorsque le CRÉDOC a réalisé l'enquête sur la prestation expérimentale dépendance (cf. annexe fiche n°10), un certain nombre de personnes ont contacté la Caisse Nationale d'Assurance Vieillesse avec inquiétude, afin de s'assurer que leur prestation n'allait pas être supprimée.

Pour avoir imparfaitement compris le contenu de la lettre, l'individu préférera parfois refuser de participer à l'enquête, de peur des conséquences de ses réponses. Lors de prises de contact directes qui suivront l'envoi d'un courrier, l'enquêteur s'entendra dire *"je n'ai pas su quoi faire, je n'ai rien compris à votre lettre"*. Même les courriers les plus simples peuvent être mal interprétés. Ce n'est pas ici une question de rédaction car le plus grand soin est apporté aux informations qui sont données et aux termes utilisés, en particulier en cherchant à simplifier le plus possible pour que la lettre soit comprise par tous. C'est davantage un rapport à l'écrit qui est mis en cause : les particuliers reçoivent peu de lettres, sont peu habitués à en écrire. Ce phénomène est encore plus accentué dans les milieux modestes. Parfois, ce sont les enfants qui nous appellent de la part des parents non francophones pour comprendre le fin mot de l'histoire ou tel autre fera appeler son voisin ou l'assistante sociale. Dans le cadre d'une enquête auprès d'un public en difficultés (population étrangère, population en situation d'illettrisme, adultes de faible niveau de qualification) l'annonce de l'étude par courrier peut multiplier les refus, qui ne seraient pas formulés lors de contacts directs.

Dans le cas des courriers envoyés dans le cadre d'une procédure CNIL, la lettre doit impérativement contenir un minimum d'informations, qui sont rappelées dans le souci de préserver la liberté des personnes. Ces éléments inquiètent parfois plus les lecteurs qu'ils ne les rassurent, car ils soulèvent des questions qu'ils ne se posaient pas. En particulier, la lettre insiste sur le respect de l'anonymat des réponses, et sur le fait que les informations données par la personne interrogée ne seront pas transmises à d'autres organismes.

1.2. Le téléphone

Le téléphone¹ est devenu l'outil d'enquête par excellence. Il semblerait que tout puisse se dire par téléphone, on y parle de pratiques sexuelles² ou de produit financiers, on peut joindre les personnes âgées ou les enfants³.

Une prise de contact par téléphone suscite moins de réticences qu'un courrier, trop impersonnel, ou qu'un contact en face-à-face, plus impliquant. Bien sûr, les enquêteurs se heurteront toujours aux réfractaires à tout "sondage" et à ceux qui perçoivent tout appel téléphonique de la part d'un étranger comme une violation de leur intimité. Le téléphone est aussi un mode de contact qui permet de renouveler les essais, pour les personnes difficiles à joindre, ce qui est plus difficile à organiser en face-à-face.

En fait, la principale préoccupation des personnes contactées par téléphone porte sur la façon dont l'enquêteur a obtenu leurs coordonnées (quand la prise de contact est nominative). Après une explication, où il est toujours possible de faire passer davantage d'informations que par courrier, les personnes sont rassurées et les entretiens se déroulent le plus souvent sans encombre.

Pendant nombre des sujets de prédilection du département Évaluation des Politiques Sociales du CRÉDOC se prêtent mal à la conversation téléphonique. En effet le téléphone est davantage réservé aux questionnaires dits quantitatifs ou du moins comportant majoritairement des questions fermées.

1.3. Le face-à-face

Le questionnement en face à face est le mode de passation réservé aux entretiens qualitatifs, sur la base de questionnements semi-directifs, mais il s'impose aussi dans d'autres cas. On y a recours dès que le temps consacré à l'entretien est jugé trop long pour être passé au téléphone (30 minutes pour les individus, 45 minutes pour les professionnels sont les limites fixées), ou que la population visée est peu équipée en téléphone - la proportion d'allocataires du

¹ Le taux d'équipement en téléphone est stable depuis plusieurs années, et atteint 96%. Tout le monde ou presque est équipé d'un téléphone, même si toutes les catégories de population ne sont pas équipées avec des taux aussi élevés.

² *Enquête nationale sur les pratiques sexuelles des français*, INED, Revue Population, Septembre/Octobre 1993.

³ Denise BAUER, Marie-Odile GILLES-SIMON, les 9-14 ans : de l'encadrement vers l'autonomie, coll. des rapports du CRÉDOC, n°148, avril 1994.

RMI équipés du téléphone avait été estimé à 70% dans le panel RMI du CRÉDOC- ou encore qu'elle ne sera pas à l'aise avec l'outil -personnes très âgées-. Parfois, on s'adresse à une population qui a certes accès à un téléphone, mais dans des conditions qui ne sont pas favorables à la réponse à un questionnaire (personne hébergées ou en foyer).

On opte également pour le face à face lorsque le travail de réflexion demandé au cours de l'entretien nécessite un va-et-vient entre interviewé et enquêteur, comme dans le cas de reconstructions de trajectoire de vie qui obligent à se souvenir d'informations anciennes et à les confronter entre elles. De même, dès que l'intervention d'une tierce personne est envisagée pour compléter les informations demandées, le face à face semble être la méthode la plus opérationnelle.

La prise de contact en face à face peut-être directe ou précédée d'un courrier ou d'un appel téléphonique (ou encore des deux). Le contact direct, même suivi d'un refus apporte déjà des informations (type d'habitat, éventuellement taille approximative du ménage, âge de la personne contactée) et les refus sont plus souvent motivés. Même si l'on cherche à obtenir des explications, les informations pour comprendre les refus manquent cruellement suite à un courrier, mais aussi à un appel.

Le mode de prise de contact dépend des objectifs de l'étude et de la méthodologie choisie, du public ciblé, de l'objet de l'étude. Cependant, les caractéristiques du fichier dont on dispose ne sont pas sans influence sur ces modes de communication. En effet, si le fichier est ancien ou incomplet, le contact avec les enquêtés sera plus difficile et cette phase de l'enquête demandera plus de temps.

2. La prise de contact à partir d'un fichier ancien

Par fichier ancien, on entend ici un fichier incomplet, non actualisé ou sur lequel des doutes sont formulés quant à la fiabilité des coordonnées.

2.1. Influence de la formulation du coupon sur le taux de refus

En présence d'un fichier ancien, le premier réflexe consiste à renoncer à une prise de contact par courrier sans avoir au préalable fait un travail d'actualisation des coordonnées. Cependant, si l'enquête suit une procédure CNIL, cette étape est un passage obligé, puisque le CRÉDOC ne possède aucun fichier. C'est donc par lettre que les individus vont se voir présenter la démarche de l'étude, le sujet, les commanditaires, le mode de passation (un enquêteur va passer à votre domicile ou vous allez être contacté par téléphone...). Selon les recommandations de la CNIL, un coupon réponse de participation ou de refus de participer à l'étude figurera dans le courrier avec une enveloppe affranchie.

Pour les individus contactés par courrier à partir d'un fichier ancien, la principale difficulté résultera de l'interprétation des non-réponses. Tout en respectant scrupuleusement la procédure CNIL, le choix d'accompagner la lettre de présentation d'un coupon de participation à l'étude ("un coupon oui") ou un coupon de refus ("un coupon non") n'est pas sans conséquence. En effet, si le choix se porte sur l'envoi d'une lettre accompagnée d'un coupon de participation à l'enquête, toutes les non-réponses dans les délais impartis devront être interprétées comme des refus et les individus radiés définitivement du fichier transmis au CRÉDOC. Cependant, pour un nombre indéterminé d'individus, qui varie selon la qualité du fichier, ce silence pourra être dû à la non réception du courrier. Les retours des lettres "NPAI" (n'habite pas à l'adresse indiquée) à l'expéditeur dépassant largement le délai de 15 jours généralement imparti dans les procédures CNIL, il sera impossible de faire la part entre refus tacite et non distribution du courrier.

Pour éviter ce biais, le CRÉDOC a généralement adopté la méthodologie inverse face à un fichier ancien : les coupons à renvoyer ne portent que sur les refus (cas des enquêtes sur la prestation dépendance, entrants dans le RMI, insertion par l'économique, ASE/PJJ...). Les personnes qui ne répondent pas sont considérées comme ayant accepté le principe de l'étude. Ces personnes ont de toutes façons la possibilité de refuser de participer lorsque l'enquêteur

prendra directement contact avec lui. Dans ce cas de figure, les refus étant expressément exprimés, il est alors possible de faire la part entre individu refusant de participer et individu non localisé. Le taux de refus a une réelle signification. Ainsi dans l'enquête sur l'insertion par l'activité économique, sur l'ensemble des 800 individus contactés par courrier, moins de 13% ont renvoyé le coupon-réponse négatif et 43% des courriers sont revenus avec la mention "N'habite pas à l'adresse indiquée" (le compte n'a été définitif que plusieurs mois après la fin de l'enquête). Parmi ces « NPAI », certaines personnes n'ont jamais pu être localisées, mais les autres ont été recontactées pour obtenir leur accord ou leur refus de participer à l'enquête. Si le CRÉDOC avait choisi l'option du coupon de participation, nous aurions comptabilisé au moins 56% de refus et nous n'aurions pas pu chercher à localiser les NPAI. Un pourcentage de non-réponse aussi élevé a des conséquences très graves sur la représentativité de l'échantillon.

Autre intérêt du coupon formulé pour identifier les refus : la négligence est assimilée à un consentement à participer, sachant que l'individu a toujours la possibilité de refuser par la suite lors d'une prise de contact directe. Dans l'hypothèse d'un coupon à renvoyer si l'on souhaite participer à l'enquête, la négligence est irrémédiablement un refus. Seuls les plus "dynamiques", les plus concernés par l'étude ou les plus curieux auront fait la démarche de renvoyer un coupon "oui j'accepte de participer", alors que parmi les autres, un fort pourcentage est prêt à être interrogé, mais ne ressent pas la nécessité de faire la démarche de renvoyer le coupon. Cette constatation n'est pas spécifique au travail sur fichier ancien et est valable pour n'importe quelle enquête sur fichier nominatif.

2.2. Retrouver les personnes « injoignables » : une nécessité lorsque l'on enquête sur la base d'un fichier ancien

Dans une seconde étape, il est possible d'entamer des recherches sur les courriers revenus avec la mention « N'habite pas à l'adresse indiquée ». Mais bien que le courrier ne soit pas retourné avec la mention NPAI, cela ne signifie pas que les coordonnées utilisées soient justes. Dans certains logements vétustes, les courriers peuvent "traîner" sur les boîtes sans nom sans être nécessairement retournés par la poste. Ou encore la lettre peut être gardée par un ancien hébergeant au cas où le destinataire passerait. Les personnes destinataires de ces courriers n'ayant pas renvoyé leur coupon de refus, les enquêteurs vont se rendre sur place, à l'adresse connue, pour s'apercevoir que la personne n'y habite plus. La liste des NPAI va ainsi être complétée par d'autres noms, pour constituer la liste des personnes qui ont changé d'adresse. On entreprend alors une recherche approfondie pour les localiser.

Que la procédure d'approche soit le courrier, le téléphone ou le face-à-face, c'est l'interprétation des « injoignables » et l'énergie consacrée à les retrouver qui décideront de la

fiabilité de l'échantillon des répondants. Ce problème est pris en compte pour les enquêtes téléphoniques, où la méthodologie prévoit un certain nombre de rappels pour chaque numéro de téléphone à des plages horaires variées pour arriver à joindre les personnes très souvent absentes de leur domicile, et améliorer le taux de réponse de l'enquête. Ceci est d'autant plus important suivant le sujet de l'étude : par exemple, les enquêtes faites auprès de familles avec des enfants en bas-âge (comme l'enquête sur les aménagements des rythmes scolaires) ne nécessitent pas de prévoir un grand nombre de rappels, car la cible de l'enquête est souvent à son domicile aux heures habituelles d'enquête (18h-21h les soirs de semaine et 9h-14h le samedi). A l'inverse, la méthodologie de l'enquête réalisée par l'INED sur les comportements et pratiques sexuelles des Français, qui cherchait en particulier à repérer des populations ayant des pratiques sexuelles « à risque » (jeunes célibataires) a prévu un maximum de rappels (jusqu'à douze) .

Pour tous les sujets pour lesquels le CRÉDOC a eu à travailler sur des fichiers anciens - recherche de personnes passées par les structures de réinsertion par l'économie, jeunes passés par les structures de l'Aide Sociale à l'Enfance ou de la Protection Judiciaire de la Jeunesse- se limiter aux individus faciles à retrouver introduisait un biais dans le profil des répondants. En effet rechercher après cinq ou dix ans des individus en parcours de réinsertion ou ceux passés par l'ASE ou la PJJ, et se limiter à ceux qui n'ont pas changé d'adresse revient à n'interroger que les plus stables ou, sans préjuger de leur capacité à s'insérer, les moins mobiles géographiquement. Il est donc nécessaire d'engager des recherches complémentaires.

Comme lors de la constitution du fichier ou l'échantillonnage, le souci est d'englober l'ensemble de la population visée. Tous les efforts vont dans le sens d'un rapprochement maximum entre le fichier des localisés et l'échantillon tiré.

2.3. Les recherches complémentaires

Ces recherches visent par tous les moyens à proposer à chaque individu sélectionné une possibilité de participer à l'enquête. Il y aura donc des recherches complémentaires et pas seulement dans le cas d'un fichier datant de plusieurs années. Dans certaines situations, et malgré la volonté de travailler en temps réel, les délais nécessaires pour reconstituer les fichiers et concevoir l'échantillonnage introduisent systématiquement un décalage entre la construction du fichier et l'investigation sur le terrain. Ainsi lors de l'enquête sur les entrants dans le dispositif du RMI, l'interrogation a eu lieu au deuxième trimestre 1996, les bénéficiaires étant entrés dans le dispositif au second semestre 1995. Le temps écoulé entre l'inscription au dispositif et l'interrogation variait donc de cinq à dix mois. Ce sont des délais très longs pour

une population en parcours d'insertion et rapidement les déménagements, les passages d'un département à l'autre ou d'un hébergement à l'autre viennent brouiller les pistes.

2.3.1. Des adresses obsolètes à réactualiser

Les difficultés rencontrées sont de natures très diverses. La première est l'obsolescence des adresses. Travailler sur des coordonnées postales ou téléphoniques anciennes revient à tabler sur une certaine stabilité géographique des populations à interroger. Dès que la population cible est "marginalisée", jeune, en cours d'insertion ou de bas niveau de qualification, les coordonnées sont vite erronées. Ce problème est accru en région parisienne. Ce ne sont pas des populations qui penseront à faire suivre leur courrier, au contraire, leur grande mobilité et la diversité de leurs points de chute rendent extrêmement difficiles le cheminement de leur courrier.

De quels moyens dispose-t-on pour actualiser d'anciennes coordonnées ? Tout dépend du commanditaire, de sa légitimité et de sa capacité à mobiliser des réseaux d'information (par exemple les réseaux de travailleurs sociaux dans les enquêtes qui nous concernent).

La recherche sur minitel est la solution la plus évidente et la plus simple qui nous soit offerte. Cependant ces recherches ne permettent ni d'identifier les hébergés, ni les personnes vivant en concubinage ou les femmes mariées, car dans ces cas l'abonnement téléphonique est souscrit à un autre nom que celui de la personne à identifier. Elles ne permettent pas non plus de repérer les personnes qui sont sur liste rouge, ni bien sûr les personnes qui n'ont pas le téléphone.

Les enquêtés peuvent avoir demandé leur inscription sur liste rouge. Cette démarche volontaire a pour objectif de se protéger ou de ne pas être dérangé par les multiples sollicitations commerciales. La recherche des enquêtés dans ces conditions pose deux types de problèmes. Le premier est d'ordre déontologique : peut-on contourner les protections dont s'entourne un individu ? Le second est d'ordre technique. A partir du Minitel, la recherche se fait par l'identification des homonymes dans l'espoir qu'ils puissent être des membres de la famille. Enfin, si l'adresse correspond à d'anciennes coordonnées la recherche s'orientera vers l'ancien voisinage.

Dans le cas de l'enquête "insertion par l'activité économique" les structures (associations, entreprises d'insertion) avaient des adresses anciennes, datant de 1993. Dans une phase préalable au démarrage des recherches de terrain, pour tous les retours " n'habite pas à l'adresse indiquée " mais aussi pour les courriers pour lesquels nous n'avions pas de réponse au bout de

quinze jours, un important travail de réactualisation a été entrepris par voie de minitel. Il s'agissait de rechercher, sur la commune et ensuite sur l'ensemble du département les personnes sélectionnées. Il a souvent été nécessaire de contacter les homonymes dans un département, en espérant joindre des membres de la même famille qui pouvaient nous indiquer où joindre la personne recherchée. Lorsque ce premier type de recherche ne donnait rien, nous nous sommes tournés vers les personnels d'encadrement des structures d'insertion par l'activité économique.

2.3.2. L'extension des recherches aux réseaux de l'action sociale et aux employeurs

Il s'agit dès lors de mobiliser tous les moyens légaux à notre disposition. Il existe des possibilités pour retrouver des personnes au niveau national, mais qui ne sont pas consultables (fichier d'imposition, fichier d'EDF, fichier France-Télécom,...). Les autres moyens permettent plutôt de retrouver des personnes qui sont restées dans un périmètre géographique assez restreint, comme la commune, le canton ou le département. Ceci pose un problème en région parisienne, et en particulier dans la petite couronne, où le découpage départemental n'a pas tellement de sens, et où les personnes passent facilement d'un département à l'autre sans que les informations suivent.

Passer par le relais d'une structure (entreprise ou autre) n'est pas la garantie d'obtenir des coordonnées fiables. Ainsi dans l'enquête "Insertion par l'activité économique", les structures enquêtées en 1996 ont en règle générale perdu la trace de leurs salariés de 1993, sauf lorsqu'ils sont encore salariés. Cette perte d'information avait été fortement minimisée lors de la phase de préparation de l'étude et elle a également surpris les instances nationales. Au moment de la constitution des listes, plusieurs responsables avaient exigé de faire la démarche de recherche eux-mêmes auprès de leurs anciens salariés. En fait, aucun n'a pu retrouver l'intégralité des personnes indiquées dans sa liste. Comme le soulignait un responsable d'entreprise d'insertion : *"En fait, je pensais que ce serait plus facile. Trois ans c'est long, mais c'est court aussi. On a l'impression que cela ne fait pas longtemps qu'ils sont partis et que les retrouver n'est pas un problème pour nous. J'ai été moi-même étonné par nos difficultés à les retrouver et par la mobilité des gens"*.

Les renseignements donnés par les responsables des structures pour retrouver leurs anciens salariés étaient fréquemment obsolètes. Avec l'accord des structures d'origine, nous avons communiqué les noms des personnes que nous ne parvenions pas à retrouver, par téléphone ou par courrier, aux responsables des autres structures d'un même département.

Cette procédure, si elle nous a permis de prendre contact avec quelques individus, ne s'est pas révélée vraiment fructueuse. Souvent les personnels d'encadrement nous ont dit croiser un tel ou un tel de temps en temps, dans la rue, dans la cité... mais la plupart du temps, ils ne pouvaient que nous indiquer le quartier ou un lieu "où l'on risquait de le rencontrer". Cette difficulté des professionnels eux-mêmes à nous communiquer des informations précises était déjà une indication sur le peu de moyens dont disposent ces structures pour suivre les personnes passées dans leurs structures ; c'était aussi un signe des problèmes que nous allions rencontrer pour retrouver nous-mêmes leurs anciens salariés.

Il a parfois été nécessaire aux enquêteurs du CRÉDOC chargés de retrouver des individus, de prendre contact dans chacun des départements avec le réseau des structures locales et les réseaux des associations ou des services de l'action sociale (les CAF, les DASS ou les CCAS, etc...). Par ailleurs, des contacts sont également pris avec les organismes dont les missions se situent dans le champ de l'exclusion et de la pauvreté (associations caritatives, services RMI, missions locales, etc...). En fait, l'utilisation de ces réseaux a parfois permis de retrouver des personnes, mais plus souvent les intervenants sociaux n'ont pu nous donner que des pistes, la majorité des informations recueillies étaient elles aussi obsolètes.

2.3.3. Les démarches de terrain

Quand les premières procédures de recherche ont été épuisées, il faut se déplacer sur le terrain. Dans ce type d'enquête, le problème est de trouver le bon interlocuteur, celui qui fournira le renseignement déterminant pour retrouver la personne. Parallèlement aux démarches précédentes, les enquêteurs doivent aller à la rencontre des personnes, sur leurs lieux de vie. Ils se trouvent parfois confrontés à des situations difficiles : des immeubles délabrés, des cités immenses, des habitats précaires. Des lieux où la présence d'un "enquêteur" peut être mal perçue. En effet, demander aux gens que l'on rencontre un renseignement sur une personne est une démarche parfois mal comprise et mal accueillie. L'enquêteur ne peut pas dévoiler l'objet de l'étude mais n'en doit pas moins rassurer sur sa démarche (il n'est ni policier, ni huissier, ni un représentant d'un organisme de tutelle par exemple...). Il ne peut pas exposer les motifs de la recherche, ni dans bien des cas, parler du commanditaire. Il s'agit de ne pas stigmatiser l'individu recherché qui a peut-être totalement rompu avec son passé (par exemple dans le cadre de l'étude sur les jeunes passés par les structures de l'ASE ou de la PJJ), ou qui ne souhaite pas être retrouvé.

D'autres difficultés matérielles se sont présentées. La validité de l'adresse est quelque fois impossible à vérifier parce que, par exemple, les boîtes aux lettres sont rares ou peu lisibles,

parce que sur ces boîtes figurent un autre nom que celui que l'enquêteur possède. Ces démarches s'avèrent longues et délicates.

Lors de démarches sur le terrain, les enquêteurs sont parfois confrontés à des anciens lieux d'habitation où l'ancien hébergeant "stocke" le courrier en vue d'un éventuel passage de la personne recherchée, sans savoir ou le joindre. Dans ce cas, il n'y a pas de refus mais il n'y a plus de solutions pour progresser, si ce n'est laisser un message, signalant que le CRÉDOC le recherche en espérant qu'il passera rapidement et nous contactera.

Les recherches complémentaires sur des fichiers anciens mobilisent énormément de temps et d'énergie. Il n'est jamais tout à fait possible de retrouver l'ensemble des individus d'une liste (pour l'enquête ASE/PJJ, 73% des jeunes passés par des structures de l'ASE et de la PJJ en 1988 ont pu être contactés, 65% des personnes ayant travaillé dans des SIAE au cours du premier semestre 1993 ont été retrouvés pour l'enquête Insertion par l'Activité Economique). Pour ces enquêtes les différences de succès d'un département à l'autre sont importantes. Elles sont essentiellement dues au niveau d'implication des réseaux potentiels d'information. Cependant, une autre explication est possible. En effet, pour ce type de démarches et recherches qui relèvent presque du travail de détectives, le profil socioprofessionnel des enquêteurs a son importance. Par exemple, pour l'étude ASE/PJJ les évaluations postérieures effectuées sur les démarches de terrain ont prouvé que certains résultats étaient expliqués par la capacité des enquêteurs à nouer des liens avec les réseaux locaux comme les associations de quartier ou de leur connaissances personnelles des quartiers où enquêter¹.

Dans tous les cas, l'énergie dépensée pour retrouver une personne est parfois inutile, parce que celle-ci est toujours libre de refuser de participer à l'étude.

¹Patrick DUBÉCHOT, Michel LEGROS, *La qualité des enquêtes auprès de populations en difficulté d'insertion sociale*, op.cit.

EXEMPLES DE DIFFICULTÉS RENCONTRÉES AU COURS DES RECHERCHES

Exemple n°1 :

L'adresse fournie par la structure n'était pas bonne, le courrier était revenu avec la mention NPAI. Nous téléphonons au responsable de la structure d'insertion. Il entame de son côté des recherches, se renseigne et nous communique de nouvelles coordonnées. Nouvelle tentative de contact qui n'aboutit pas. Une recherche par le MINITEL nous donne une adresse dans une autre commune. Nous prenons contact avec la personne, mais c'est un homonyme. Nous contactons les autres structures d'IAE. Nous obtenons une adresse en Italie. Un courrier est envoyé. L'enveloppe nous a été encore retournée avec la mention " n'habite pas à l'adresse indiquée ".

Exemple n°2 :

A partir de la liste des personnes présentes dans la structure, nous contactons une personne dans les délais impartis (quinze jours après l'envoi du courrier). La femme qui nous répond affirme qu'elle est un homonyme et qu'elle n'a jamais travaillé dans l'association évoquée. Peu de temps après, nous recevons un appel d'un monsieur qui cherche des renseignements sur une enquête menée par notre organisme. Il pose des questions sur le pourquoi, pour qui, dans quel but ? ...etc. Il dit qu'il est concerné par l'enquête et qu'il va réfléchir avant d'accepter l'entretien. Deux semaines après, la femme rappelle pour nous dire qu'elle se méfie des enquêtes et qu'elle a fait téléphoner un ami pour en savoir plus sur cette enquête. L'entretien a pu se dérouler chez elle sans autre problème.

Exemple n°3 :

L'enquêtrice arrive devant un immeuble de centre ville assez délabré. Le courrier n'étant pas revenu, nous avons supposé que l'adresse était bonne. Elle cherche un numéro d'escalier. L'endroit est plutôt sombre et il n'y a visiblement que des boîtes aux lettres arrachées ou posées par terre. Elle ne trouve pas le nom recherché. Des gens la regardent. Elle va vers eux et leur dit qu'elle souhaiterait rencontrer monsieur T. Silence, réflexion. Il habite bien là, mais ils disent qu'il n'est pas là en ce moment. Que faire ? Continuer à rôder et chercher ? L'enquêtrice choisit de repartir. Deux jours plus tard, elle revient et est attentive à ne pas rencontrer les mêmes personnes. Finalement, après quelques tentatives elle trouve une personne qui lui indique l'appartement de monsieur T. Elle est accueillie assez froidement au début. Mais rapidement l'atmosphère se détend et l'entretien peut se réaliser.

Exemple n°4 :

Après une recherche sur le MINITEL sur des coordonnées retournées comme NPAI, une trentaine d'homonymes sont identifiés sur un département. Un rapide appel à chacun ne permet pas de retrouver la jeune femme recherchée. Quelques jours plus tard, une femme nous appelle d'un autre département : elle avait été contactée par un membre de sa famille lui indiquant que le CRÉDOC la recherchait pour un questionnaire. Elle était mariée (avait donc changé de patronyme) et vivait dans la ville d'à côté mais située sur un autre département.

3. La présentation de l'étude

Une fois la personne à interroger identifiée, ses coordonnées confirmées, il s'agit d'obtenir son accord. Le travail en amont sur la construction des fichiers et la rigueur d'échantillonnage doit être poursuivi et validé par une interrogation homogène, par des taux de refus maîtrisés afin de ne pas biaiser l'échantillon des répondants.

3.1. L'homogénéité de l'information

Le cas de figure idéal consisterait à pouvoir donner à chacun la même information, la même présentation de l'étude afin de les placer devant les mêmes possibilités de refuser ou d'accepter leur participation. Deux conditions doivent pour cela être respectées : contacter chacun des individus, et diffuser dans un second temps autant que possible la même information.

3.2. Une source d'information uniforme

L'information uniforme devrait idéalement être diffusée sur le même support à tous les individus en même temps. Dans la réalité, à part les quelques tentatives de travail avec automates d'appel¹ qui restent limitées dans leurs applications, cette unicité de l'information est impossible, ne serait-ce que parce qu'elle passe par plusieurs enquêteurs et non un seul. Il s'agit alors de s'en approcher.

Le premier écueil provient de la réaction de l'individu, car malgré tous les efforts d'harmonisation, le premier biais irrémédiable reste la réaction de chacun face à la démarche d'enquête. Pour certains, il s'agit d'un refus d'être interrogé, compté, d'"entrer dans les statistiques". D'autres réagissent en fonction de l'objet de l'étude selon ses convictions ou son expérience (le fameux "*ça ne m'intéresse pas*" honni par tous les enquêteurs). Il n'y a souvent rien à faire devant un refus non motivé, ou un refus de principe. La marge de négociation est étroite et il est difficile de faire revenir la personne sur sa décision.

¹Les numéros de téléphone sont composés automatiquement et la marche à suivre est indiquée par un enregistrement, la personne contactée répond en appuyant sur les touches de son combiné.

Le second biais provient du rapport de chacun au support de communication adopté : courrier, téléphone ou contact direct en face-à-face. Ce sont les étrangers, les personnes âgées, les personnes qui ont des difficultés de lecture (illettrisme) qui refuseront ou auront un *a priori* négatif devant une lettre. Au téléphone, les personnes qui ne sont pas familiarisées avec l'outil, ou qui ont des problèmes d'expression orales, les mal entendants, les non francophones hésiteront à accepter. En face-à-face, les individus refuseront l'intrusion physique de l'enquêteur dans leur immeuble ou leur appartement. Ces réticences sont connues du Service des Enquêtes et les enquêteurs sont sensibilisés à ces situations et sont formés à les affronter, à négocier, à argumenter par exemple en mettant en avant telle ou telle information susceptible de valoriser la personne et son apport à l'étude. Indiquer que le témoignage restera anonyme constitue aussi un élément qui soulève les réticences.

Le troisième élément à prendre en compte est "l'effet enquêteur". Au téléphone ou en face à face interviennent les caractéristiques de l'enquêteur. Les personnes contactées peuvent réagir différemment¹ selon que l'enquêteur est un homme ou une femme, un français ou un étranger, un jeune ou moins jeune. Les compétences de l'enquêteur sont également un facteur important : sa capacité à entrer en relation avec les personnes, à engager un dialogue, à négocier.

Il est certain que cet ensemble de facteurs, ne peut assurer de l'uniformité de l'information. Même par courrier, on ne peut la garantir. Dans le cadre de l'enquête sur les entrants dans le dispositif du RMI, telle CAF avait joint des enveloppes affranchies aux coupons réponses, telle autre des enveloppes mais sans les affranchir, une dernière n'y avait pas mis d'enveloppe. Pour l'enquête ASE/PJJ, les jeunes devaient recevoir une lettre des structures par lesquelles ils étaient passés avec une lettre du CRÉDOC les informant de la démarche, dans trois départements sur cinq, les Conseils Généraux ont préféré ne pas utiliser de papier à en-tête afin de ne pas "effrayer" le destinataire, ne pas le troubler quant à l'objet de l'étude.

Cependant le courrier laisse une trace de la présentation de l'étude. Il est toujours possible d'y revenir. La personne contactée garde ainsi les coordonnées du CRÉDOC et peut éventuellement demander conseil à son entourage avant de décider si elle participe ou pas à l'enquête. Lors des passages au domicile des personnes âgées dans le cadre de l'étude CNAV, les enquêteurs ont souvent vu la lettre, précieusement conservée avec les autres documents administratifs.

Dans certains cas, enfin, il est recommandé de renoncer à cette stricte homogénéité de l'information à destination des interviews. En effet si les enquêteurs sont mis en garde contre

¹Cf. *Survey research methods : a reader*, édition Eleanor Singer and Stanler Presser; Chicago, 1989.

tout risque d'écart et d'interprétation dans le fil du questionnaire, ils ont parfois carte blanche pour faire une présentation de l'étude adaptée au degré de curiosité de chaque personne interrogée. Devant des individus qui cherchent à être rassurés, il faudra insister sur l'anonymat des réponses, replacer le témoignage dans la taille de l'échantillon (« *vous serez 900 à être interrogés* »), pour d'autres, insister sur ces points qui leur paraissent aller de soi éveillerait plutôt la suspicion. On a aussi pu constater que la plupart des personnes interrogées demandaient peu d'informations sur le commanditaire et l'objet de l'étude avant la passation du questionnaire, et voulaient surtout connaître les thèmes d'interrogation. A la fin de l'entretien, elles cherchent plus à situer l'étude dans son contexte.

3.3. Des sources d'informations multiples

Dès que nous nous plaçons dans l'hypothèse d'une construction de listes par un tiers, ou de fichiers reconstruits par différentes sources, la diffusion de l'information se fait par un relais. En effet, nous constatons souvent, qu'en dehors de toute procédure CNIL, les structures contactées souhaitent pouvoir demander elle-même l'accord de leurs "ressortissants" avant de communiquer leurs coordonnées au CRÉDOC.

Ce comportement protecteur vient ajouter un écran entre la personne sélectionnée et le CRÉDOC et tout repose sur le bon vouloir d'un interlocuteur, de sa disponibilité, de son niveau d'implication dans l'étude. Il s'agira alors de mobiliser toutes les ressources à notre disposition (lettre du commanditaire, appel ou visite du chargé d'étude) pour impliquer les personnes-relais et les rassurer quant au traitement qui sera fait des informations communiquées.

Cette diffusion de l'information en cascade pose deux problèmes, une présentation de l'étude aux enquêtés potentiels hétérogène et un rapport à l'information différencié.

3.3.1. Une présentation de l'étude hétérogène

La présentation de l'étude est faite par chacune des structures et non par le CRÉDOC, il n'y a donc pas de présentation homogène, le niveau d'information est disparate, mettant en avant des arguments différents. Ces informations sont données soit en face à face, soit au téléphone, soit par courrier. L'information est individuelle ou collective, par voie d'affichage par exemple. Toutes les interprétations sont possibles de la part des personnes sollicitées, celui-ci pense que sa participation est obligatoire, tel autre pense que c'est une réunion de groupe, un autre comprend qu'il participe à un stage.

Le CRÉDOC a le recours d'apporter le maximum de supports écrits aux personnes relais. Si une annonce doit être passée dans les locaux, le CRÉDOC qui propose de la rédiger, de même que tous les courriers diffusés pour sensibiliser les futurs enquêtés. Il s'agit de minimiser autant que possible les biais d'une interprétation par trop subjective. En dernier ressort, c'est l'enquêteur qui devra clarifier la démarche et donner les bonnes informations.

3.3.2. Un rapport à l'information différenciée

Le rapport à "l'informateur" est également hétérogène. Dans l'étude sur la mobilité hospitalière, l'étude a été présentée par quelqu'un qui, le plus souvent, entretient un rapport hiérarchique avec l'enquêté potentiel. Deux types de réactions opposées peuvent alors se manifester : soumission à la hiérarchie avec le biais que cela peut laisser supposer dans les réponses ou au contraire rébellion. Ces réactions peuvent dépendre à la fois de l'atmosphère collective qui règne dans les structures (une restructuration en cours dans l'hôpital par exemple), mais aussi de la relation personnelle qui existe entre le salarié et l'intermédiaire. Cette situation influe sur le taux de réponses, mais aussi sur le déroulement de l'entretien. Il est donc important d'identifier les blocages, de préciser le caractère indépendant de l'organisme chargé de l'enquête. Ceci peut rassurer les salariés et les amener à accepter de participer et à répondre en toute confiance.

4. Le temps comme facteur de réussite

La question du temps est essentielle dans ce type d'enquête. Elle l'est sur deux plans : dans les contacts à établir avec les responsables des structures et services " terrain de l'enquête " ; dans l'estimation que l'on peut faire à priori des difficultés à retrouver les personnes enquêtées.

Il est très important de prendre du temps dans les contacts que l'on établit avec les responsables des structures ou des organismes sollicités. Ceux-ci sont rarement directement commanditaires de l'étude et ne sont pas contraints de participer. Il faut donc les convaincre de l'utilité de l'étude et les inciter à lui consacrer une partie de leur temps. Ceci est d'autant plus important que ces organismes qui prennent en charge des populations en difficulté ou qui gèrent des mesures ou des dispositifs sont souvent sollicités pour des enquêtes sur leur public et leur activité. Les responsables des structures et services concernés sont souvent sceptiques sur l'utilité de ces études car elles leur paraissent parfois déconnectées de leur réalité quotidienne. Ils sont également réticents car les résultats des investigations leur parviennent rarement. Il ne faut donc pas sur-estimer l'intérêt que va susciter notre démarche parmi les personnes sollicitées.

Il faut que l'organisme enquêteur accepte de prendre son temps, éventuellement d'en perdre dans ses contacts avec les responsables des organismes du site enquêté, avec les intervenants sociaux ou avec toute autre personne référente, pour en gagner ensuite. Les intervenants sociaux sont très sensibles à l'attention que l'on peut porter à leur travail, aux personnes concernées par l'enquête. Parfois, il est nécessaire que le chargé d'étude ou l'enquêteur démontre sa connaissance du secteur étudié, du fonctionnement du dispositif ou de la mesure avant que les portes s'ouvrent et que les informations demandées soit communiquées. En fait, le temps passé est aussi un temps d'observation et d'approfondissement de la connaissance du secteur enquêté et de ses acteurs et notamment sur le fonctionnement « informel » des institutions.

Le deuxième aspect autour de cette question du temps se situe dans la phase de recherche des personnes de l'échantillon à interviewer. Là encore, il ne faut pas sous-estimer le temps nécessaire pour repérer des personnes à partir de coordonnées obsolètes. Lorsqu'il faut retrouver des individus plusieurs années après qu'ils soient passés par des dispositifs ou des structures, le temps d'enquête s'allonge. Dans l'étude sur l'insertion par l'activité économique, les démarches se sont avérées plus longues que prévues, malgré les expériences antérieures (comme l'ASE/PJJ). Malgré la conception d'un plan d'enquête adapté, pour chaque étude sur fichier il existe une zone « d'ombre » et d'incertitude qui se situe au début de l'enquête, à partir

de la construction du fichier jusqu'à l'entretien ou le passage du questionnaire. Dans cet espace d'incertitude apparaît toujours quelque difficulté imprévue dans le bel édifice théorique bâti avant le démarrage de l'étude. Cet espace d'incertitude tient principalement de la qualité de l'information recueillie -les caractéristiques de la construction du fichier initial- et de l'attitude des intermédiaires de l'enquête que sont les responsables des services et structures des sites enquêtés.

Et puis au-delà des moyens mis en place pour retrouver les personnes, de tous les efforts déployés pour l'identification des lieux d'habitation, pour la prise de contact, l'établissement d'un contact n'assure pas pour autant la réalisation de l'entretien. La personne peut refuser explicitement de participer, mais il y a aussi bien d'autres façons de refuser. L'individu peut, soit accepter l'entretien au téléphone et ne pas être au rendez-vous fixé avec l'enquêteur, soit adopter une telle stratégie de contournement des questions que l'entretien n'aboutira pas. Dans ces enquêtes sur fichier, il faut beaucoup de patience et du temps. Le temps constitue un paramètre essentiel dans la gestion de l'espace d'incertitude mis en lumière par les expériences successives du CRÉDOC dans les enquêtes sur fichier. Ce temps de réalisation peut être certainement de mieux en mieux appréhendé mais il est peut probable qu'il soit un jour totalement maîtrisé.

CONCLUSION

Les études réalisées par le CRÉDOC s'inscrivent souvent dans une logique quantitative qui implique la construction d'un échantillon représentatif et l'utilisation d'un questionnaire fermé comme mode d'interrogation. Quelques études ont une ambition plus qualitative et s'appuient, le plus souvent, sur un échantillon diversifié et des entretiens semi-directifs. Cependant, les commanditaires du département Évaluation des Politiques Sociales hésitent fréquemment entre ces deux logiques : ils veulent que l'enquête se déroule auprès d'un grand nombre d'individus qui constituent un échantillon représentatif pour pouvoir ensuite déduire des conclusions de l'étude des résultats pour l'ensemble de la population ; par ailleurs, ils souhaitent recueillir une information multidimensionnelle sur des processus complexes et des éléments de nature qualitative, comme les trajectoires d'insertion. L'entretien-questionnaire ou le questionnaire-entretien, qui mêle des parties fermées et des temps d'entretien semi-directif ou de questions ouvertes résulte de ce compromis, mais il demande des techniques d'analyse complexes et longues à mettre en oeuvre : en particulier les techniques d'analyse lexicale ou encore la construction de trajectoires sont des méthodes auxquelles on peut avoir recours¹.

La conception d'une enquête recouvre plusieurs opérations qui s'enchaînent et, souvent, se superposent. Chacune d'elles est déterminante pour la qualité des résultats et implique des options précises. La définition des objectifs de l'étude et de la population-mère, la sélection de l'échantillon, le mode d'accès aux interviewés engagent ainsi l'étude dans sa phase réellement opératoire, et chacune de ces opérations entraîne des conséquences particulières. Si à chaque phase, l'objectif est similaire -réduire les biais qui pourraient conduire à produire des conclusions erronées sur les résultats finaux- les moyens de l'atteindre sont spécifiques.

Définir la population, c'est sélectionner les catégories de personnes que l'on veut interroger, c'est déterminer les acteurs dont on estime qu'ils sont en position de produire des réponses aux questions que l'on se pose et qu'ils se posent eux-mêmes dans le cadre de leur intervention, de leur mission. Cette étape est donc indissociable des objectifs de l'étude et conditionne les étapes suivantes. Les critères de choix de la population à interroger doivent tenir

¹ Plusieurs cahiers de recherche sur ces thèmes ont été publiés par le CRÉDOC ces dernières années. On peut citer parmi les plus récents :

Valérie BEAUDOUIN, F. BROCHET, *Analyse lexicale de corpus en anglais*, Coll. des Cahiers de Recherche du CRÉDOC n°95, sept 1996.

Valérie BEAUDOUIN, *Analyse textuelle et structures narratives de récits*, coll. des cahiers de recherche du CRÉDOC n°82, nov 1995.

Valérie BEAUDOUIN, Anne-Lise AUCOUTURIER, *Parcours d'insertion de jeunes en difficulté*, coll. des cahiers de recherche du CRÉDOC, n°66, sept 1994.

Denise BAUER, *Parcours singuliers, repérer et interpréter les trajectoires atypiques*, coll. des cahiers de recherche du CRÉDOC n°54, octobre 1993.

compte de la faisabilité et peuvent se fonder sur des variables tout à fait simples telles que l'âge (ou une catégorie comme par exemple « les jeunes ») ou plus complexes comme le type de mesure dont a bénéficié l'enquêté. Le département Évaluation des Politiques Sociales du CRÉDOC, de par la nature même de ses axes d'études et de recherche, mais aussi de par ses compétences, est amené à conduire des études dans lesquelles le critère de choix de la population se fonde fréquemment sur des variables complexes : présence dans un dispositif d'insertion, bénéficiaire de mesures sociales ou acteurs de ces dispositifs ou mesures.

Ces études s'appuient en général sur des échantillons représentatifs. Comme les critères qui déterminent la population-mère sont parfois complexes et que les populations étudiées sont difficilement repérables en cherchant « au hasard » dans la population française, il est préférable de travailler à partir d'un fichier nominatif, qui permet de localiser les personnes à interroger. Un fichier qui correspondrait exactement à la population ciblée par l'étude, ce que l'on peut qualifier de fichier « idéal-type », existe rarement. La plupart du temps, il faut le reconstituer, en agrégeant ou en complétant les sources existantes, voir en partant de rien. Le CRÉDOC se heurte parfois à de grandes difficultés lorsqu'il se livre à ce type d'exercice, qui nécessite en particulier d'animer un réseau de partenaires qui détiennent les informations initiales. Ce cahier de recherche a permis d'élaborer une typologie des situations possibles lorsque l'on décide de monter une enquête basée sur un fichier nominatif, de synthétiser les difficultés rencontrées aux différentes étapes qui précèdent la collecte des données, et de proposer des solutions. La typologie, qui recense six possibilités, s'organise par rapport à la situation idéale, qui constitue la référence vers laquelle on cherche à tendre au moment de la construction des fichiers. Les différentes situations existantes s'éloignent plus ou moins de cette situation de référence. Les deux schémas qui suivent présentent les différentes situations en ordre décroissant, en rappelant pour chacune d'entre elles les problèmes principaux et les solutions envisageables.

Les problèmes rencontrés lors de l'agrégation de fichiers EXISTANTS

<u>Type de fichier</u>	<u>Constitution du fichier</u>	<u>Echantillonnage</u>
Fichier unique et exhaustif	Pas de problèmes particuliers <i>contrôler les omissions et les doubles-comptes</i>	→ Toute méthode d'échantillonnage possible
Fichier élaboré à partir de plusieurs fichiers harmonisés	Situation proche de l'idéal-type pas de problèmes particuliers, <i>contrôler les omissions et les doubles-comptes</i>	} Echantillonnage à plusieurs degrés : sélection d'unités primaires, souvent les départements, dans lesquelles on choisit des unités statistiques Construction d'une typologie des départements pour construire des classes homogènes par rapport au thème étudié
Fichier élaboré à partir de plusieurs listes ou fichiers non harmonisés	Problème de l'harmonisation des informations, des critères ou des nomenclatures, <i>harmoniser sur les quelques variables communes, exposer clairement la demande au moment des contacts avec les intermédiaires</i> Fichiers parfois non informatisés, listes manuscrites à saisir	
Fichier construit à partir d'un fichier initial, incomplet ou inadéquat, complété par d'autres sources	Inadéquat : le fichier couvre un champ plus vaste que celui étudié, <i>trier et aménager le fichier de manière à isoler la population ciblée</i> Incomplet : il manque des groupes entiers de la population, <i>croiser avec d'autres sources pour compléter le fichier initial</i>	→ Toute méthode d'échantillonnage possible

Les problèmes rencontrés lors de l'agrégation de fichiers NON-EXISTANTS

<u>Type de situation</u>	<u>Constitution du fichier</u>	<u>Echantillonnage</u>
La constitution des listes par un intermédiaire	Les principales difficultés consistent à obtenir la participation des intermédiaires sollicités, <i>passer du temps à présenter l'étude et à négocier</i> , et à recueillir une information homogène, <i>fournir une grille type et vérifier le remplissage</i>	→ Toute méthode d'échantillonnage possible
Sélection des enquêtés par un intermédiaire	Les intermédiaires sélectionnent des personnes qu'elles estiment intéressantes, <i>il faut leur imposer des critères de sélection</i> , il peut exister des rapports hiérarchiques entre les enquêtés et les intermédiaires qui augmentent les taux de refus ou influent sur le déroulement de l'entretien, <i>il faut mettre en avant l'indépendance du CREDOC et l'anonymat des entretiens</i>	→ Donne lieu à un échantillon diversifié, mais non représentatif

La plus grande difficulté à résoudre lors de la constitution de l'échantillonnage, c'est la connaissance le plus souvent partielle de la population-mère. En l'absence d'informations centralisées, reconstituer le fichier exhaustif de la population serait le plus souvent long et coûteux. Une partie seulement des intermédiaires qui possèdent l'information sont contactés et mis à contribution pour élaborer le fichier. Le recours à des échantillonnages à plusieurs degrés ou par grappe permet, malgré cette connaissance partielle, de construire un échantillon représentatif de la population initiale. Cependant, d'autres éléments interviennent qui peuvent introduire des biais dans la construction de l'échantillon. En particulier, afin d'assurer le respect des libertés et le droit de regard des individus sur les informations détenues dans des fichiers qui les concernent, ce type de procédé d'enquête nécessite l'autorisation de la Commission Informatique et Liberté. Celle-ci demande que les personnes soient averties de la création d'un nouveau fichier et qu'elles puissent refuser d'en faire partie. Le refus de participer est un élément qui est inhérent à toute enquête facultative, mais la particularité de ce processus est que les refus s'expriment avant que le CRÉDOC intervienne. Or, pour assurer la représentativité de l'échantillon, il est nécessaire de posséder un minimum d'informations sur les refus, au moins leur nombre. Ce problème, s'il n'est pas possible de le résoudre entièrement, peut cependant être minoré en prenant quelques précautions, aussi bien au moment de la constitution du fichier, que dans la prise de contact avec les personnes, ou encore au cours de l'analyse.

Le deuxième problème spécifique concerne l'obsolescence des coordonnées qui figurent dans le fichier. Se limiter seulement aux personnes dont les coordonnées sont fiables peut entraîner des biais d'autant plus importants que le fichier est ancien. Par exemple, si l'on recherche des personnes à partir d'un fichier datant de trois ou cinq ans, plus de la moitié des coordonnées peuvent ne plus être valables. De plus, suivant le thème de l'étude, en particulier s'il s'agit d'insertion sociale et professionnelle, les personnes qui n'ont pas changé de domicile ont probablement des trajectoires et des comportements différents de ceux qui ont beaucoup bougé. Il est donc indispensable de se livrer à une réactualisation du fichier, afin que le fichier des personnes repérées soit le plus proche possible de l'échantillon sélectionné. Cette réactualisation est un travail minutieux, qui demande de multiplier les approches.

Enfin, le dernier aspect considéré dans ce cahier de recherche concerne la prise de contact avec les personnes à enquêter. Le plus souvent, cette prise de contact est d'abord indirecte, dans la mesure où elle a lieu par courrier, pour informer les personnes de l'enquête et leur laisser la possibilité que leurs coordonnées ne soient pas transmises au CRÉDOC. De plus, ce premier contact n'est pas directement pris en charge par le CRÉDOC. Ce n'est que dans un deuxième temps que la personne recherchée est en contact direct (en face-à-face ou par téléphone) avec un enquêteur du CRÉDOC. Là encore, cette approche en plusieurs étapes peut être à l'origine de

taux de refus élevé, et avoir des répercussions sur les résultats de l'étude. C'est pourquoi il faut veiller à assurer une forte homogénéité de l'information transmise, pour que chaque enquêté soit dans une situation à peu près semblable au moment du commencement de l'entretien.

Il semble également important d'être attentif à ce que l'on pourrait appeler « l'esthétique de la réception », c'est-à-dire aux conditions dans lesquelles l'enquêté reçoit l'enquête. Cette attention se justifie pour au moins deux raisons. La première concerne la manière dont on informe la personne de sa qualité d'enquêté. Cette « qualité » peut être vécue différemment selon les individus à cause de l'objet de l'étude, leur situation actuelle, la façon dont ils ont été retrouvés. La « qualité » d'enquêté peut être vécue comme stigmatisante. Cette situation fréquente dans les études auprès de publics très en difficulté. Les personnes rencontrées ne souhaitent pas forcément être repérées comme « en difficulté ». Une enquête réalisée en collaboration avec le Groupe Permanent de Lutte contre l'Illettrisme (GPLI) réalisée en 1991¹ est à cet égard pleine d'enseignements sur les conditions à mettre en place pour atteindre les personnes et leur faire accepter l'entretien alors même qu'ils ne se perçoivent pas comme « illettrés ».

L'enjeu dans ce type d'études est d'arriver à ce que chacun puisse trouver une place, sa place, dans l'enquête. Cette dimension constitue la seconde raison pour laquelle il est nécessaire d'être attentif aux conditions dans lesquelles la relation s'établit avec l'organisme d'étude et particulièrement les enquêteurs. Il est certain que l'on ne peut pas gommer l'effet d'imposition de la situation, mais en revanche, il est important de l'atténuer par une attitude d'écoute attentive de l'enquêté. L'enquêteur doit être en permanence conscient de « la violence symbolique » de toute forme d'interrogation à cause de « *la dissymétrie entre les interlocuteurs très inégalement pourvus en capital économique et surtout culturel* »² inhérente à la situation d'enquête, car comme le souligne Alain Blanchet « *l'entretien, comme technique d'enquête, est né de la nécessité d'établir un rapport suffisamment égalitaire entre l'enquêteur et l'enquêté pour que l'enquêté ne se sente pas, comme dans un interrogatoire, contraint de donner des informations* »³. Or, une telle situation de contrainte modifie les rapports entre interviewer et interviewés et par là-même peut modifier aussi la nature de l'information produite. Dans les enquêtes présentées dans ce cahier de recherche, beaucoup concernent l'évaluation d'un dispositif et ses effets sur la situation des individus bénéficiaires, certaines nécessitent une recherche des personnes à interroger plusieurs années après leur passage. Dans ces conditions, l'arrivée d'un enquêteur est appréhendée « comme une situation officielle de perquisition »⁴. Il

¹ BAUER D., DUBÉCHOT P., MARESCA B., CHARRIER O., *Histoires individuelles et situations d'illettrisme*, CRÉDOC, septembre 1991.

² Pierre BOURDIEU et alii : *La misère du monde*, Éditions du Seuil, février 1993.

³ Alain BLANCHET et alii., *L'entretien dans les sciences sociales*, Paris, Dunod, 1985.

⁴ François de SINGLY : *L'enquête et ses méthodes : le questionnaire*, Nathan Université, Paris, 1992.

faut dès lors, que l'organisme d'étude donne des informations sur l'objet de l'étude, son déroulement, le traitement de l'information et sur la façon dont l'anonymat des réponses sera respecté. Il faut absolument que l'enquêté sache comment il est protégé.

Ce cahier de recherche avait pour objectif d'analyser et de capitaliser le savoir-faire acquis au fil des expériences lors d'enquêtes sur fichier. Il avait aussi pour ambition d'identifier et de donner corps à cet espace de manoeuvre situé entre un plan d'enquête -parfois imposé- raisonné et méthodique et la mise en oeuvre des pratiques non prévues, non définies. L'idée était d'identifier cette zone d'incertitude alimentée par l'expérience, nourrie de l'invention et l'ajustement permanent pour répondre au mieux aux problèmes rencontrés dans les enquêtes par fichiers ou listes de personnes.

Les deux temps typiques de cet espace d'incertitude sont certainement la première et la dernière phase, celles de la constitution du fichier et celles de la recherche des personnes ou du contact avec les enquêtés. La collecte des informations, des données ouvre un espace de négociation avec les détenteurs de l'information qui demande des compétences relationnelles, un savoir-faire organisationnel et surtout une capacité à admettre et à gérer l'incertitude. Il faut pouvoir accepter de voir le cadre prédéfini se tendre et parfois se déchirer pour malgré tout atteindre les objectifs fixés : obtenir un fichier suffisant et en adéquation avec l'enquête. La phase de recherche des individus et de prise de contact s'inscrit également dans cet espace d'incertitude et d'invention qui se modifie, évolue, est bouleversé à chaque enquête. Les compétences nécessaires dans cette phase laissent une grande part à l'improvisation, à l'imagination pour trouver les solutions adaptées aux situations rencontrées et qui viennent compléter les procédures théoriques, habituelles, et déjà expérimentées.

Au fond, la tentation de tout scientifique pourrait être de formaliser l'informel. Cette tentation est certainement à l'origine de ce cahier de recherche. L'idée de trouver, face à chacun des problèmes rencontrés au cours des enquêtes sur fichier, une solution adaptée est à l'origine de cette réflexion. L'idée de produire des éléments et des instruments de normalisation, de trouver des procédures reproductibles n'est pas étrangère à notre démarche. Ce type d'approche normalisatrice serait celle que définirait, par exemple, une recherche de certification-qualité (ISO 9002) ; on peut toutefois s'interroger sur la pertinence réelle de cette approche ou sur les conditions même sa faisabilité : peut-on définir des procédures simples et univoques qui puissent répondre à toutes les incertitudes que laisse la réalisation d'une enquête sur des listes ou des fichiers, d'autant plus incertains qu'ils recensent une population marginale ? En revanche, une capitalisation et une formalisation, de l'expérience de ces enquêtes, des savoir-faire des chargés d'études du département Évaluation des Politiques Sociales, du Service des

Enquêtes et des enquêteurs permet de dégager des procédures souples et adaptables aux différentes conditions d'enquêtes sur fichier.

Le projet est ambitieux et peut-être peu raisonnable. Il répond certainement à un souci de capitalisation de l'expérience acquise, du savoir-faire expérientiel¹. Il répond aussi peut-être à l'angoisse d'être face à l'aléatoire et de devoir gérer ce champ d'aléas, d'être contraint à inventer et à vivre l'incertitude alors qu'il faut réussir, obtenir des résultats et atteindre les objectifs définis dans le protocole d'étude. La zone d'incertitude définie dans ce cahier de recherche s'affirme rebelle aux cadres théoriques et conceptuels et laisse une part importante au savoir-être des chargés d'études et des enquêteurs, notamment à leur capacité à gérer l'incertitude, l'informel, à leur capacité d'adaptation et d'initiative dans un cadre cependant de mieux en mieux cerné au fil des expériences.

¹ Guy LE BOTERF : *De la compétence. Essai sur un attracteur étrange*, Les Éditions d'Organisation, 1994.

ANNEXES

Annexe 1 :

BIBLIOGRAPHIE

AUCOUTURIER Anne-Lise, GILLES-SIMON Marie-Odile, LEGROS Michel, *Intérêt et faisabilité d'un panel d'allocataires du revenu Minimum d'Insertion*, juin 1994.

BAUER Denise, GILLES-SIMON Marie-Odile, *Les 9-14 ans : de l'encadrement vers l'autonomie*, Coll. des Rapports du CRÉDOC, n°148, avril 1994.

BEAUDOUIN Valérie, F. BROCHET, *Analyse lexicale de corpus en anglais*, Coll. des Cahiers de Recherche du CRÉDOC n°95, sept 1996.

BLANCHET Alain et alii., *L'entretien dans les sciences sociales*, Paris, Dunod, 1985.

BLANCHET Alain, *Dire et faire dire. L'entretien*, Paris, Colin, 1991.

BOUDON Raymond et François BOURRICAUD, *Dictionnaire critique de la sociologie*, (1982) Paris, PUF, rééd.1993.

BOURDIEU Pierre, *La misère du Monde*, Éditions du Seuil, Paris, février 1993.

COMBESSIE Jean-Claude : *La méthode en sociologie*, Coll. Repères, Éd. La Découverte. n°194, 1996.

CNRS et ENST, *La qualité de l'information dans les enquêtes* » recueil des communications des journées d'étude des 13-15 juin 1991 de l'ASU.

DESABIE Jacques, *Théorie et pratique des sondages*, 1966.

DEVILLE Jean-Claude et Françoise DUPONT Françoise, *Non-réponse : principe et méthodes*, articles dans Actes des journées de méthodologie statistique du 15 et 16 décembre 1993, INSEE méthodes, n°56-57-58.

DUBÉCHOT Patrick, GILLES-SIMON Marie-Odile, LEGROS Michel, *L'aire et la manière, construction et usage de la notion de site dans les procédures d'enquête*, Coll. des Cahiers de Recherche du CRÉDOC, septembre 1995, n°78.

DUBÉCHOT Patrick, LEGROS Michel, *La qualité des enquêtes auprès de populations en difficulté d'insertion sociale*, Coll. des Cahier de Recherche du CRÉDOC, n°47, avril 1993.

DURKHEIM Émile, *Les règles de la méthode sociologique* (1895), PUF, Paris, 1987.

GILLES-SIMON Marie-Odile, *Les spécificités des enquêtes quantitatives auprès de populations socialement marginales*, Coll. des Cahiers de Recherche du CRÉDOC, n° 68, avril 1995.

GILLES-SIMON Marie-Odile, LEGROS Michel, *Politiques sociales : l'épreuve de la pauvreté*, Collection des rapports du CRÉDOC, n°159, avril 1995.

GOFFMAN Erving, *La mise en scène de la vie quotidienne*, (1956), Éd. Minit, Paris, 1973.

GROSBRAS Jean-Marie, *Méthodes statistiques des sondages*, Économica, Paris, 1987.

HÉLAN François, *L'assise statistique de la sociologie*, article dans un numéro spécial de la revue de l'INSEE, Économie et statistique, « Sociologie et statistique », 1984.

INSEE méthodes : *La clé des sources*. Statistiques démographiques et sociales, Tomes 1 et 2, août 1992.

LE BOTERF Guy : *De la compétence. Essai sur un attracteur étrange*, Les Éditions d'Organisation, 1994.

MOATI Philippe : *Les sources pour les études sectorielles*, Intervention dans le cadre du Séminaire de formation CRÉDOC organisé par le Département Dynamique des Marchés.

SINGLY François de, *L'enquête et ses méthodes : le questionnaire*, Éd. Nathan Université, Paris, 1992.

Annexe 2 :
FICHES ÉTUDES

fiche n°1

Intitulé et références du rapport :

« Allègement des charges sociales sur les bas salaires : intégration dans la stratégie des entreprises et effets d'emploi. »

Christine HENRIOT, Juillet 1996 (*)

(*) rapport confidentiel

- Commanditaire de l'étude :** Comité Interministériel d'Évaluation des Politiques Publiques.
- Objectifs :** Évaluation du dispositif : analyser l'intégration des mesures dans la stratégie de l'entreprise ainsi que son effet sur l'emploi à court et moyen terme.
- Méthode de recueil de l'information :** Interrogation de 486 responsables d'entreprises concernées par les mesures dans huit régions. L'interrogation se faisait au cours d'un entretien en face-à-face d'une demi-heure à trois-quarts d'heure. En complément, une cinquantaine d'entretiens ont été réalisés auprès de secrétaires de comités d'entreprise.
- Population-mère :** Des responsables d'entreprises.
- Fichiers de départ :** Fichier SIREN de l'INSEE.
- Méthode d'échantillonnage :** Le tirage de l'échantillon de départ a été effectué par l'INSEE, en respectant des critères de taille et de salaires. La taille de l'entreprise était mesurée selon le nombre de salariés (moins de 9, de 10 à 99, plus de 100). Les entreprises ont été classées en trois catégories : secteurs à bas, moyens et hauts salaires.

Fichier constitué :

L'échantillon a été redressé pour être représentatif de l'échantillon national sur les entreprises à bas et moyens salaires.

Principaux problèmes :

- Le taux de non-réponse des responsables d'entreprises (50%) : pas le temps, déjà trop sollicité, enquête non obligatoire,
- Le taux élevé de non-réponse des secrétaires des comités d'entreprise : méconnaissance du dispositif.

fiche n°2

Intitulé et références du rapport :

« Le dispositif ARPE (*) : principales caractéristiques des bénéficiaires et des personnes réembauchées ».

SIMON Marie-Odile, en collaboration avec RUEFF Delphine

Décembre 1996 (**)

(*) Allocation de Remplacement pour l'Emploi

(**) rapport confidentiel

Commanditaire de l'étude : UNEDIC Union Nationale Interprofessionnelle pour l'Emploi dans l'Industrie et le Commerce.

Objectifs :

- Recueillir des informations sur la réalité des embauches compensatrices après les départs en préretraites avec une Allocation de Remplacement Pour l'Emploi et sur les stratégies d'embauche des entreprises.
- Savoir quelles sont les entreprises et les salariés les plus concernés par ces dispositifs de préretraite,
- Comparer les caractéristiques (sexe, âge, niveau de qualification...) des partants et des personnes réembauchées en compensation.

Méthode de recueil de l'information : Enquête téléphonique.

Population-mère : Entreprises dont un ou plusieurs salariés sont partis en préretraite par le dispositif ARPE.

Fichiers de départ :

Le fichier qui a servi de base a été fourni par l'UNÉDIC. Il a été construit en plusieurs étapes :

- * fichier exhaustif des allocataires de l'ARPE au 31/07/96. Ce fichier reprend le numéro de SIRET de l'établissement qui employait ces allocataires. A partir de là, il est possible de constituer un fichier d'établissements,

- * le fichier d'établissements présumés correctement identifiés (clé du SIRET correcte) a été rapproché du Répertoire National des Employeurs (RNE de l'UNEDIC).

- * sélection des établissements qui appartiennent à la zone géographique des départements retenus.

Méthode d'échantillonnage :

Une vingtaine de départements ont été sélectionnés à partir d'une typologie de l'ensemble des départements. Les variables prises en compte sont : le taux de chômage, le taux de chômage des jeunes, la répartition des salariés par grand secteur d'activité.

Fichier constitué :

Un fichier de près de 5 000 entreprises représentant 8 850 allocataires a ainsi été constitué. 501 entreprises ont été interrogées.

fiche n°2 bis

Intitulé et références du rapport :

« Le dispositif PRP (*) : principales caractéristiques des bénéficiaires et des personnes réembauchées ».

SIMON Marie-Odile, en collaboration avec RUEFF Delphine

Janvier 1997 (**)

(*) Preretraite progressive

(**) rapport confidentiel

- Commanditaire de l'étude :** Direction de la Prévision du Ministère de l'Économie et des finances.
- Objectifs :**
- Recueillir des informations sur la réalité des embauches compensatrices après les départs en PRP sur les stratégies d'embauche des entreprises.
 - Savoir quelles sont les entreprises et les salariés les plus concernés par ces dispositifs de préretraite,
 - Comparer les caractéristiques (sexe, âge, niveau de qualification...) des partants et des personnes réembauchées en compensation.
- Méthode de recueil de l'information :** Enquête téléphonique.
- Population-mère :** Entreprises dont un ou plusieurs salariés sont partis en préretraite par le dispositif PRP au cours des années 1994, 1995 et 1er semestre 1996.

- Fichiers de départ :** Inexistant. Il a été nécessaire de s'adresser aux Directions Départementales du Travail et de l'Emploi qui nous ont fait parvenir une liste des entreprises ayant signé une convention de préretraite départementale au cours des trois années de référence. Ne sont donc pas considérées dans l'échantillon les entreprises qui ont signé des conventions au plan national, avec la délégation à l'emploi.
- Méthode d'échantillonnage :** Une vingtaine de départements ont été sélectionnés à partir d'une typologie de l'ensemble des départements. Les variables prises en compte sont : le taux de chômage, le taux de chômage des jeunes, la répartition des salariés par grand secteur d'activité. 500 entreprises ont été interrogées.
- Fichier constitué :** Un fichier de près de 5 000 entreprises représentant 8 850 allocataires a ainsi été constitué.
- Principaux problèmes :** Les DDTE ont le plus souvent envoyé une photocopie de leurs cahiers d'enregistrement. Les informations qui sont conservées dans ces cahiers ne sont pas identiques d'un département à l'autre, en particulier les nomenclatures peuvent varier (bénéficiaires potentiels, bénéficiaires réels).

fiche n°3

Intitulé et références du rapport :

**« Les nouveaux arrivants au RMI :
profils, parcours antérieurs, rapports à l'emploi et à la famille »**

ALDEGHI Isa

Collection des Rapports du CRÉDOC, n°173,

Novembre 1996

- Commanditaire de l'étude :** Délégation Interministérielle au Revenu Minimum d'Insertion (DIRMI).
- Objectifs :** Analyser les évolutions affectant le public du RMI en interrogeant des allocataires entrés le plus récemment possible dans le dispositif.
- Méthode de recueil de l'information :** Interrogation par questionnaire en face-à-face de 900 allocataires.
- Population-mère :** Les allocataires entrés dans le dispositif au cours du premier semestre 1995.
- Fichiers de départ :** Fichiers des Caisses d'Allocations Familiales.
- Méthode d'échantillonnage :** La CNAF dispose de fichiers statistiques nationaux d'individus, mais ces fichiers sont anonymes. Il a donc été nécessaire de s'adresser à chaque CAF pour constituer la base de sondage. Pour éviter de mobiliser les 118 CAF, une vingtaine ont été sélectionnées. Cette sélection repose sur la construction d'une typologie des CAF en fonction du profil des allocataires du RMI ainsi que sur des indicateurs d'évolution du RMI. Un tirage au sort parmi les allocataires entrés dans le dispositif au cours du premier semestre 1995.

Fichier constitué : Les fichiers des différentes CAF ont été assemblés pour constituer un seul fichier

Principaux problèmes :

- Les fichiers des CAF ont été constitués à des fins administratives et non dans l'optique de les utiliser pour réaliser des enquêtes.
- L'avis de la CNIL a été sollicité avant le tirage de l'échantillon. Ses préconisations ont été scrupuleusement respectées. Chaque personne tirée au sort par les CAF a reçu un avis de la CAF dont elle dépend l'informant des objectifs de l'enquête CRÉDOC et lui donnant la possibilité sous quinze jours de refuser que son nom soit transmis à l'organisme de recherche. Un coupon-réponse pour le refus accompagné d'une enveloppe avec port payé figurait dans chaque envoi. Les CAF ont ensuite communiqué des listes d'adresses au CRÉDOC expurgées des coordonnées des personnes qui avaient manifesté leur refus de participer à l'enquête.
- travail sur fichiers limités dont certaines coordonnées étaient obsolètes.
- travail de recherche et d'identification.

fiche n°4

Intitulé et références du rapport :

« Observer les emplois et les qualifications du social »

Patrick DUBÉCHOT, Pierre LE QUÉAU, Michel MESSU

étude en cours

- Commanditaire de l'étude :** Mission Interministérielle Recherche et Expérimentation (MIRE).
- Objectifs :**
- Première phase : recenser les employeurs du département du Val-de-Marne employant des intervenants sociaux quelque soit leur qualification, leur dénomination de poste.
 - Deuxième phase : analyse des emplois des intervenants sociaux dans un champ particulier - l'insertion sociale et professionnelle des jeunes- et définition de la notion d'intervention sociale et de ses bornes.
- Méthode de recueil de l'information :**
- Première phase : recensement exhaustif des employeurs d'intervenants sociaux dans le Val-de-Marne ainsi qu'une identification quantitative et qualitative de ces employeurs.
 - Deuxième phase : sélection, parmi ces employeurs, de ceux dont les services ou les structures emploient des intervenants sociaux dans le domaine de l'insertion sociale et professionnelle des jeunes.
- Population-mère :**
- Première phase : les employeurs d'intervenants sociaux
 - Deuxième phase : les intervenants sociaux du Val-de-Marne dans le domaine de l'insertion sociale et professionnelle des jeunes.

- Fichiers de départ :** - Première phase : FINESS (Fichier National des Établissements Sanitaires et Sociaux) du Val-de-Marne.
- Méthode d'échantillonnage :** - Première phase : FINESS recouvre des domaines plus larges que celui de l'étude. Il a donc été nécessaire d'extraire les établissements hors-champ. Mais FINESS est aussi incomplet par rapport au domaine observé, il a fallu le compléter.
- Deuxième phase : parmi le fichier exhaustif du Val-de-Marne, ont été sélectionnés les services et structures employant des intervenants sociaux dans le domaine de l'insertion sociale et professionnelle des jeunes.
- Fichier constitué :** - Première phase : fichier exhaustif des employeurs d'intervenants sociaux du Val-de-Marne,
- Deuxième phase : sélection exhaustive des employeurs dans le domaine de l'insertion sociale et professionnelle des jeunes.
- Principaux problèmes :** - Fichier inadéquat et incomplet : FINESS recouvre des domaines plus larges que celui de l'étude. Il a donc été nécessaire d'extraire les établissements hors-champ. Mais FINESS est aussi incomplet par rapport au domaine observé.
- Première phase :
1- vérification par téléphone que les structures ou services à la limite du domaine observé (maison de retraite, services hospitaliers...) emploient effectivement des intervenants sociaux.
2- recherche et exploitation d'autres sources pour compléter le fichier (Minitel, annuaires, budgets prévisionnels des communes, petites annonces dans des revues spécialisées...etc.)

fiche n°5

Intitulé et références du rapport :

« Le devenir des personnes entrées dans les dispositifs d'insertion par l'économique »

Patrick DUBÉCHOT, Christine HENRIOT, Marie-Odile SIMON

Collection des Rapports n°175, Novembre 1996

- Commanditaire de l'étude :** Commissariat Général du Plan.
- Objectifs :** - Retrouver et interroger une fraction significative d'une population sortie des structures d'insertion par l'activité économique (SIAE) dans quatre départements.
- Méthode de recueil de l'information :** Questionnaire en face-à-face avec quelques questions ouvertes.
- Population-mère :** Les individus présents dans l'une des structures entre le 1er janvier et le 30 juin 1993.
- Fichiers de départ :** Inexistant.
- Méthode d'échantillonnage :** Le principe de cette étude étant d'analyser les effets du passage des personnes dans leur trajectoire d'insertion professionnelle, l'Instance d'évaluation a considéré qu'une distance de trois années après la sortie des structures pouvait constituer un délai suffisant pour mesurer ces effets. De ce fait, seules les SIAE existant entre janvier et juin 1993 ont été retenues dans l'échantillon.
- Fichier constitué :** 800 personnes sorties de ces SIAE entre le 1er janvier et le 30 juin 1993.

Principaux problèmes :

- Listes constituées pour l'enquête par les SIAE,
- Délai de réalisation plus long que prévu : démarrages de l'enquête décalés dans le temps en fonction des constitutions des listes dans les départements,
- Questions déontologiques : procédure CNIL pour contacter, puis interroger les personnes,
- Recherche sur fichiers limités,
- Adresses obsolètes,
- Difficultés pour retrouver les personnes.

L'avis de la CNIL a été sollicité avant le tirage de l'échantillon. Ses préconisations ont été scrupuleusement respectées. Chaque personne tirée au sort a reçu un avis l'informant des objectifs de l'enquête CRÉDOC et lui donnant la possibilité sous quinze jours de refuser que son nom soit retenu. Un coupon-réponse pour le refus accompagné d'une enveloppe avec port payé figurait dans chaque envoi. Les listes d'adresses ont été expurgées des coordonnées des personnes qui avaient manifesté leur refus de participer à l'enquête.

fiche n°6

Intitulé et références du rapport :

« La mobilité des personnels dans les hôpitaux publics »

Christine HENRIOT

décembre 1996 (*)

(*) rapport confidentiel

- Commanditaire de l'étude :** Direction de la Prévision du Ministère de l'économie et des finances.
- Objectifs :**
- Déterminer les facteurs déclenchant la mobilité des agents administratifs et soignants, des praticiens hospitaliers,
 - Décrire les dispositifs existant dans les hôpitaux pour accompagner ou favoriser cette mobilité.
- Méthode de recueil de l'information :**
- Interrogation d'une part,
- * des agents administratifs et soignants, des praticiens par questionnaire fermé, administré en face à face et d'une durée d'une demi-heure à une heure et d'autre part,
 - * des directeurs des ressources humaines et des directeurs des affaires médicales par entretien d'une heure à une heure et demi.
- Population-mère :** Professionnels de 20 hôpitaux répartis de façon uniforme entre cinq régions.
- Fichiers de départ :** Inexistant. Chaque établissement a fourni le nombre d'agents, de praticiens, le nombre de départs, de départs en retraite, celui des détachements, des mises en disponibilité, des mutations, des changements de services.

- Méthode d'échantillonnage :** L'échantillon d'agents et de praticien ne visait pas à la représentativité, mais cherchait à couvrir des situations diverses, en termes de responsabilités, de spécialités exercées. Les personnes interrogées ont été désignées par la direction de l'hôpital en fonction des critères fournis par le CRÉDOC sur cette diversité.
- Fichier constitué :** Dans chacun des hôpitaux, ont été rencontrés :
- * 4 aides-soignants, 4 infirmiers, 4 agents administratifs,
 - * entre six et dix praticiens,
 - * les directeurs des RH et des AM.
- Principaux problèmes :** Les directions ont proposé des agents et des praticiens dont le parcours était intéressant par rapport à l'étude, donc des personnes ayant connu une forte mobilité. Les professionnels interrogés entretenaient un rapport hiérarchique avec la personne ressource qui avait aidé à constituer l'échantillon.

fiche n°7

Intitulé et références du rapport :

**« La Prévention spécialisée aujourd'hui :
situations professionnelles et compétences »**

Patrick DUBÉCHOT

Collection des Rapports n°170

septembre 1996

- Commanditaire de l'étude :** PROMOFAF (Fonds d'assurance formation du secteur sanitaire et social).
- Objectifs :**
- Réaliser une photographie de la Prévention spécialisée et analyser l'évolution des pratiques professionnelles.
 - Produire un outil de diagnostic de l'état des compétences actuelles des professionnels de ce secteur.
- Méthode de recueil de l'information :**
- Première phase : entretiens semi-directifs auprès des principaux acteurs « gestionnaires » de la Prévention,
 - Deuxième phase : entretiens semi-directifs auprès des éducateurs
- Population-mère :**
- Première phase : les principaux acteurs « gestionnaires » de la Prévention spécialisée dans les quatre régions choisis par le commanditaire.
 - Deuxième phase : les éducateurs des associations ou services de Prévention spécialisée de ces départements.

- Fichiers de départ :** Inexistant.
- Méthode d'échantillonnage :** Deux éléments ont participé à la sélection des services ou associations enquêtées.
- * sans chercher une représentativité, il a été tenu compte du nombre de services ou associations de PS dans chaque département,
 - * les services ou associations dans lesquels des professionnels de la PS ont été interrogés devaient être adhérents à PROMOFAP.
- Par ailleurs dans chacune des régions un Groupe d'Appui Technique composé d'administrateurs et de professionnels a participé à la construction des deux échantillons :
- première phase : deux critères ont présidé à la construction de l'échantillon : les statuts des personnes interrogées et les lieux d'exercice de leur activité,
 - deuxième phase : la sélection des associations a eu le souci de respecter la diversité de la PS observée dans la première phase. Ce sont les directeurs des services et des équipes qui ont sélectionné les personnes à interviewer.
- Fichier constitué :**
- Première phase : 43 personnes ont été interviewées.
 - Deuxième phase : 33 éducateurs ont été interviewés.
- Principaux problèmes :**
- Le respect de la diversité observée sur un si faible échantillon.
 - La négociation avec les professionnels dans la construction de l'échantillon.

fiche n°8

Intitulé et références du rapport :

« Politiques sociales : l'épreuve de la pauvreté »,

Marie-Odile GILLES-SIMON, Michel LEGROS

collection des rapports du CRÉDOC, n°159,

avril 1995

- Commanditaire de l'étude :** Conseil Économique et Social.
- Objectifs :** Dans le cadre d'un programme d'évaluation des effets des politiques publiques en matière de lutte contre la pauvreté, le CRÉDOC devait observer directement auprès des populations, l'impact des mesures.
- Méthode de recueil de l'information :** Entretiens semi-directifs.
- Population-mère :** Les personnes présentes sur les sites. Ceux-ci se définissaient comme des lieux de regroupement, temporaires ou non, de personnes isolées ou de familles en situation de présomption de grande pauvreté.
- Fichiers de départ :** Inexistant.
- Méthode d'échantillonnage :** Construction de la population à enquêter en trois phases différentes :
- * au cours de la première, 10 zones géographiques ont été sélectionnées,
 - * à la seconde, 48 sites différents ont été retenus,
 - * enfin 756 personnes ont été rencontrées.
- Chacune des deux premières étapes visait à privilégier la sélection des territoires sur lesquels la densité de personnes en situation de pauvreté pourrait être la plus forte.

Fichier constitué : Fichier des sites : une recherche de la diversité par zone géographique.

Principaux problèmes : Aller à la rencontre des personnes (dans les gares, dans le métro, dans les caravanes...etc.) : travail de formation des enquêteurs et test des entretiens particulièrement approfondis.

fiche n°9

Intitulé et références du rapport :

Aménagement des rythmes scolaires

Christine HENRIOT

mai 1997 (*)

(*) rapport confidentiel

- Commanditaire de l'étude :** Ministère de la Jeunesse et des Sports.
- Objectifs :** interroger 800 familles (parent-enfant) sur leur connaissance des modifications intervenues dans l'école.
- Méthode de recueil de l'information :** Interrogation par téléphone. Le questionnaire était d'une durée totale d'environ 30 mn :
* questionnaire d'une quinzaine de minutes à un parent,
* questionnaire d'une dizaine de minutes pour l'enfant.
- Population-mère :** Parents et enfants (9-11 ans) inscrits dans des établissements pilotes.
- Fichiers de départ :** Liste des établissements sélectionnés pour cette expérimentation.
- Méthode d'échantillonnage :** Les écoles contactées ont été déterminées par le ministère sur des critères tenant compte à la fois du nombre d'heures scolaires hebdomadaires et du type d'aménagement mis en place.
- Fichier constitué :** Liste des parents acceptant de participer.

Principaux problèmes :

- Problème déontologique et procédure CNIL : délais entre le contact établissement et le retour des listes contacts des établissements qui diffusent l'information (lettre et coupon de refus) auprès des parents d'élèves des classes concernées. Listes des "acceptants" adressées au CRÉDOC.
- Aucune information sur les refusants (non francophones, désintérêt, peur de révéler la méconnaissance du dispositif...).
- Multiplication des relais de diffusion de l'information : des directeurs d'écoles aux instituteurs, aux enfants, aux parents,
- Accueil du courrier par les parents et l'accord dépend de leur réaction.

fiche n°10

Intitulé et références du rapport :

« La prestation dépendance : expérimentations »

Frédérique MARTIN, Marie-Odile SIMON

août 1996 (*)

(*) rapport confidentiel

- Commanditaire de l'étude :** Caisse Nationale d'Assurance Vieillesse (CNAV)
- Objectifs :** Évaluation des expérimentations de la prestation. Six mois après le démarrage de cette expérimentation, un premier rapport (juillet 1995), faisait un premier bilan statistique sur les demandes déjà traitées et les principales caractéristiques des bénéficiaires. L'objectif de la nouvelle phase d'évaluation de 1996 était de compléter les informations sur la grille AGGIR après un an d'utilisation.
- Méthode de recueil de l'information :**
- Entretiens de groupe auprès des équipes médico-sociales,
 - Entretiens auprès des différents régimes partenaires (Régime Général, MSA...),
 - Questionnaire en face-à-face auprès des bénéficiaires et le plus souvent en présence d'un proche (« utilisé » comme personne ressource lorsque la personne âgée ne connaissait pas les informations demandées ou lorsque son état de santé ne lui permettait pas de répondre.)
- Population-mère :** Les bénéficiaires de la prestation expérimentale dépendance.
- Fichier de départ :** Fichier global de la CNAV (4 691 personnes) constitué à partir des fichiers départementaux transmis par les CRAM.

Méthode d'échantillonnage : L'échantillon construit est proportionnel au nombre de courriers envoyés par département.

Fichier constitué : Après avoir écarté les refus, conformément à la procédure CNIL, un fichier de 3380 bénéficiaires a été constitué concernant les douze départements de l'expérimentation.

Principaux problèmes : - Problème déontologique et procédure CNIL : délais.

CAHIER DE RECHERCHE

Récemment parus :

L'évolution de l'emploi dans l'industrie manufacturière française

Philippe MOATI, Laurent POUQUET - n°92 (1996)

Méthode d'étude sectorielle - Vol. 2

Philippe MOATI - n°93 (1996)

Modélisation des choix alimentaires des ménages

Patrick BABAYOU, Aude COLLIERIE DE BORELY - n°94 (1996)

Analyse lexicale de corpus en anglais

Valérie BEAUDOUIN, Frédéric BROCHET - n°95 (1996)

Évaluer ou l'esprit des méthodes

Michel LEGROS, Guy POQUET - n°96 (1996)

Approche de la structure du paysage associatif dans le domaine de l'environnement

Bruno MARESCA - n°97 (1996)

Les nouvelles logiques productives dans les PMI : déterminants et impact sur les performances

Philippe MOATI, Laurent POUQUET - n°98 (1997)

La consommation en 1997 : vers le cyber consommateur ?

Patrick BABAYOU - n°99 (1997)

Président : Bernard SCHAEFER Directeur : Robert ROCHEFORT
142, rue du Chevaleret, 75013 PARIS - Tél. : 01 40 77 85 01

ISBN : 2-84104-085-2

CRÉDOC

Centre de recherche pour l'Étude et l'Observation des Conditions de Vie