

Insee

Outils de suivi des trajectoires des personnes en matière sociale et d'emploi

Mylène CHALEIX
Stéfan LOLLIVIER

Juin 2004
N° 98/B010
Class : 1.5.91

Outils de suivi des trajectoires des personnes en matière sociale et d'emploi

Mission « panels »

1. INTRODUCTION	3
2. L'EXISTANT INTERNATIONAL	5
2.1. UN NOYAU DUR ASSEZ REPANDU DANS LA PLUPART DES PAYS DEVELOPPES	5
2.2. DES DEVELOPPEMENTS SPECIFIQUES DANS CERTAINS PAYS	7
3. L'EXISTANT EN FRANCE	9
3.1. BREVE DESCRIPTION DES PANELS EN FRANCE	9
3.2. UNE CULTURE DE COUPE TRANSVERSALE PROBABLEMENT EXCESSIVE	10
3.3. UN MANQUE DE SAVOIR-FAIRE SUR DONNEES LONGITUDINALES	11
4. LES BESOINS	14
4.1. NIVEAU NATIONAL	14
4.2. NIVEAU LOCAL	15
5. LES PISTES POSSIBLES	18
5.1. LES DONNEES ADMINISTRATIVES	18
5.2. LES INITIATIVES EUROPEENNES	19
5.3. LA COMPLEMENTARITE DONNEES - ADMINISTRATIVES ET ENQUETES	21
6. RECOMMANDATIONS	23
6.1. LES OUTILS	23
6.2. LES UTILISATIONS	24
6.3. VALORISATION DES SOURCES	25
6.4. RECOMMANDATIONS STRATEGIQUES	26

1. Introduction

Lors de la préparation du moyen-terme du CNIS (1999-2003), les débats conduits dans les différentes formations sur les statistiques sociales ont mis en évidence que les situations de moyenne et datées ne permettaient pas d'appréhender correctement les trajectoires et histoires individuelles. L'avis insistait sur la nécessité de développer des panels.

L'articulation entre les opérations statistiques et les questions d'évaluation des politiques publiques apparaît moins développée (et peut-être moins facile) dans le domaine de la démographie et des conditions de vie que dans d'autres domaines comme l'emploi, par exemple.

Pourtant les politiques touchant la famille, la pauvreté, la délinquance et les questions de redistribution en général mériteraient d'être mieux fondées sur des données et des analyses quantitatives. De ce point de vue, les analyses longitudinales apparaissent essentielles même si elles sont difficiles à mener avec soin.

Dans ce sens, on insistera en conclusion sur la nécessaire collaboration entre les statisticiens et les chercheurs en sciences sociales et la garantie pour ceux-ci d'avoir un accès privilégié aux sources de données.

(Grandes orientations du moyen terme 1999-2003 - Formation démographie - conditions de vie)

***L'exclusion** - pauvreté - précarité a fait l'objet de plusieurs coups de projecteurs mais ils sont partiels et non coordonnés. Des études plus approfondies pourraient donner des éléments d'informations sur l'intérêt du cumul (activité réduite-allocations) et la santé des chômeurs (exploitation spécifique à prévoir dans la prochaine enquête "Santé"). Les panels sont des outils indispensables pour approcher les phénomènes de précarité. Le nombre de ménages observés dans le panel européen devrait d'ailleurs être sensiblement augmenté pour permettre des études plus fines.*

(Grandes orientations du moyen terme 1999-2003 - Formation santé - protection sociale).

En 2003, lors de la préparation du moyen-terme 2004-2008, le constat est renouvelé. Mais si le diagnostic est inchangé (insuffisance des travaux sur les trajectoires individuelles), le discours sur les panels a évolué. La séance de l'inter formation « statistiques sociales » du 17 octobre 2003 a abordé ce thème des panels mais à la fois sous l'angle de l'intérêt de l'outil mais aussi de ses coûts et limites.

Le CNIS confirme, lors de sa séance plénière de décembre 2003, sa demande d'une mission sur les outils de suivi des trajectoires des personnes en matière sociale et d'emploi.

Le Conseil enregistre la forte demande de constitution de panels de longue durée d'individus afin de suivre les parcours professionnels et sociaux. Il recommande qu'une réflexion approfondie soit menée sur l'intérêt et les limites de l'utilisation de panels à la lumière des expériences déjà réalisées tant sur le plan national qu'international.

(Avis général adopté le 18 décembre 2003)

Cette mission a été confiée par Jean-Michel CHARPIN, directeur général de l'INSEE en décembre 2003 à Stéfán LOLLIVIER, directeur de l'ENSAE et Mylène CHALEIX, membre de l'Inspection générale de l'INSEE (Cf. lettre de mission en annexe 1).

Le diagnostic porté par certains acteurs sur le dispositif actuel de la statistique publique concernant les trajectoires individuelles est sévère. Une illustration récente de cette lacune est développée dans le rapport du CERC (Conseil de l'Emploi, des Revenus et de la Cohésion sociale) sur les enfants pauvres en France.

La France est en retard, en ce domaine d'observation et d'analyse, sur nombre de pays européens (et encore plus sur les pays d'Amérique du Nord, États-Unis et Canada).

Tout d'abord, quand l'information existe, elle est très dispersée ...

En second lieu, l'articulation entre collecte et analyse statistique, d'une part, observations et analyses sociologiques, d'autre part, est ... insuffisante en France.

Une place à part doit être faite à l'amélioration de la connaissance des situations les plus défavorisées ...

De manière plus particulière, le manque d'observation longitudinale (constitution de panels ou suivi de cohortes) crée une zone d'ombre sur les conséquences, à terme, pour les enfants des situations de pauvreté subies par les familles : nous sommes très en retard dans ce domaine.

(Extraits de la synthèse du rapport «Enfants pauvres en France, Progrès nécessaires de l'observation et de l'analyse »)

Pour parachever la critique du dispositif actuel, citons le témoignage d'un chercheur britannique de haut niveau, venu travailler dans un laboratoire parisien :

L'absence de « vraies » données de panel françaises est dommage, et freine la recherche appliquée en France, à mon avis.

Dans cet esprit, le rapport débute par un état des lieux de l'existant et des pratiques au niveau international, puis fait le point sur les données disponibles au niveau français. Il décrit ensuite les besoins exprimés lors des entretiens menés par les missionnaires. Il présente enfin des pistes possibles et énumère un certain nombre de recommandations susceptibles d'être mises en œuvre. Parmi celles-ci figurent des recommandations de court terme, qui peuvent être mises en place rapidement pour un coût modéré. D'autres sont plus structurantes, et renvoient à la place des sources administratives dans le dispositif statistique sur les individus. Leur mise en œuvre nécessitera, notamment avec la CNIL, un examen approfondi des procédures afin de garantir les règles de confidentialité et de respect des personnes.

2. L'existant international

2.1. Un noyau dur assez répandu dans la plupart des pays développés

Des panels généralistes utilisés dans de très nombreux pays

La plupart des grands pays disposent d'un panel généraliste construit sur le modèle du PANEL STUDY OF INCOME DYNAMICS (PSID) américain. Ce dernier, lancé en 1968, assure le suivi longitudinal d'un échantillon représentatif d'individus vivant aux États-Unis, ainsi que des familles auxquelles ils appartiennent. Il s'intéresse en priorité à la dynamique des comportements économiques et démographiques, même s'il aborde aussi des questions sociologiques ou psychologiques. A l'issue de la collecte de 2003, le PSID avait recueilli de l'information sur 65 000 individus et 36 années de cycle de vie. L'enquête est réalisée par le Survey Research Center, Institute for Social Research, de l'Université du Michigan ; sa réalisation a été rendue possible grâce aux généreux soutiens financiers d'agences gouvernementales, de fondations, et d'autres organisations. De nombreuses enquêtes spécifiques ont été collectées sur la base de l'échantillon du PSID :

Active Saving File, 1984-1989 and 1989-1994
Childbirth and Adoption History File, 1985-2001
Death File, 1968-1999
Ego-Alter File, 1985
Estimating Risk Tolerance for the 1996 PSID
Family 'Income Plus' Files, 1994-2001
Geocode Match File
Health Care Burden File, 1993
Hours of Work and Wage Files, 1994-2001
Marriage History File, 1985-2001
Medicare File
OFUM Income Detail File, 1993
Parent Health Supplement, 1991
Parent Identifier File, 1968-2001
Relationship File, 1968- 1985
Retrospective Occupation-Industry Files, 1968-1980
Self-Administered Health Supplement, 1990
Telephone Health Supplement, 1990
Time and Money Transfers File, 1988
Wealth File, 1984, 1989, 1994, 1999, 2001
Work-History File, 1984-1985; 1984-1986; 1984-1987

L'exploitation du PSID a conduit à de très nombreuses publications : le nombre annuel est passé d'environ 200 dans les années 70 à environ 400 à la fin des années 90.

Proche du PSID, le BRITISH HOUSEHOLD PANEL SURVEY (BHPS) a été lancé en 1991 après d'environ 5 500 ménages et 10 300 individus dans 250 aires de la Grande Bretagne. Réalisé par l'Université d'Essex, l'exploitation de ce panel a conduit à plusieurs centaines de publications. Dans le même esprit, le panel socio-économique allemand (SOEP) fournit de l'information depuis 1984 sur les membres de ménages ouest-allemands puis est-allemands

En 2002, plus de 12 000 ménages, soit 24 000 personnes étaient interrogées. Le SOEP a conduit lui aussi à de très nombreuses publications.

De nombreux autres pays, tels la Suisse, le Luxembourg ou la Russie disposent d'un tel panel. EUROSTAT a conduit de 1994 à 2001 le panel européen, d'inspiration analogue, dans les pays de l'union économique de l'époque. Le projet SILC devrait poursuivre la mise en place de tels panels au sein de l'Union Européenne.

Un dispositif sur les enfants

Les pays anglo-saxons ont depuis de nombreuses années organisé de vastes cohortes constituées d'enfants suivis depuis la naissance (ou depuis la grossesse) jusqu'à l'âge adulte, sur la base d'échantillons représentatifs au plan national. Les buts des suivis de ce type sont divers : ils incluent presque toujours l'étude du « développement de l'enfant », appréhendé dans un contexte familial, social, économique, à travers son histoire familiale et son parcours éducatif. La dimension santé est presque toujours présente. La plus ancienne de ces cohortes a été constituée en Grande-Bretagne en 1946. Commencée par une étude de périnatalité auprès de 17 000 nouveau-nés, elle s'est poursuivie sur un sous-échantillon de 5 400 enfants qui ont fait l'objet de 20 enquêtes depuis 1946 (NSHD). Un second échantillon de même taille a été constitué en 1958 (NCDS), et un autre en 1970 (BCS70), cette fois suivis intégralement. Toutes ces cohortes sont encore « actives » : plus de 70 % des enfants de ces cohortes sont encore en observation. Tout récemment, les Anglais ont lancé en 2000 la « MILLENIUM COHORT STUDY » (MCS) : plus de 21 000 enfants nés en 2000-2001, dont une partie des enfants nés de la cohorte 1970...

Aux États-Unis, les premiers travaux ont surtout porté sur les relations entre scolarité et emploi, avec des données complémentaires sur la santé et la consommation de tabac, d'alcool et de drogues. Les enquêtes NLSY79 et NLSY97 n'ont inclue que des adolescents en milieu scolaire, mais la première a été prolongée par une étude sur les enfants nés des mères appartenant à la cohorte : 5 000 enfants ont fait l'objet d'un suivi annuel à partir de 1986. L'étude canadienne (NLSC en anglais, ELNEJ en français) porte sur 22 800 enfants âgés de 0 à 11 ans en 1994, suivis jusqu'à leur 25ème anniversaire. Les pays d'Europe du Nord ont lancé récemment des programmes ambitieux. En particulier, au Danemark, le projet DNBC (DANISH NATIONAL BIRTH COHORT) vise à inclure 100 000 femmes enceintes.

De nombreuses initiatives sur la santé et le vieillissement

Aux États-Unis, l'Université du Michigan conduit le HEALTH AND RETIREMENT STUDY (HRS) depuis 1994 en interrogeant tous les deux ans plus de 22 000 américains de plus de 50 ans. Financé par le National Institute on Aging, l'enquête s'intéresse aux effets du vieillissement sur la santé physique et mentale, ainsi qu'à des aspects économiques tels la couverture sociale, les aspects financiers, l'état sur le marché du travail et la préparation de la retraite. L'enquête est très exploitée puisque utilisée dans 60 chapitres de livre, 346 articles, 261 monographies,...

En Grande Bretagne, l'ENGLISH LONGITUDINAL SURVEY OF AGING (ELSA), inspiré du HRS, est en train de débiter. Il devrait servir de modèle au projet européen SHARE au sein de onze pays, dont la France.

2.2. Des développements spécifiques dans certains pays

Une culture de l'évaluation des dispositifs publics aux États-Unis

Aux États-Unis, la réforme du WELFARE qui s'est déroulée durant les années quatre-vingt dix a insisté sur le retour à l'emploi (Welfare to Work), en transférant une responsabilité prépondérante aux états dans la définition et la gestion des programmes¹. En contrepartie, les états sont tenus de mettre en place un dispositif de suivi des bénéficiaires permettant au niveau fédéral d'assurer un contrôle du respect des règles de décentralisation financière et de gestion. Des pénalités sont imposées aux états qui, soit ne respectent pas certaines règles de la loi, soit ont un défaut de « *reporting* » (pénalités consistant en la réduction de la subvention fédérale).

Ceci s'est traduit par la mise en place de sources nécessaires aux évaluations correspondant à différents domaines. Parmi ces sources se trouvent les enquêtes statistiques à couverture large mais non panélisées comme les vagues de mars du CURRENT POPULATION SURVEY, réalisées par le CENSUS BUREAU ou le NATIONAL SURVEY OF AMERICAN FAMILIES réalisé par WESTAT pour l'Urban Institute, ou des enquêtes sociologiques de terrain, mais également des panels longitudinaux. Le CENSUS BUREAU conduit, depuis 1984, des enquêtes suivant les personnes sur deux ans et demi à quatre ans, les SURVEY OF INCOME AND PROGRAM PARTICIPATION SIPP. Le SIPP 96 pouvait suivre les effets de la réforme sur quatre ans, mais il ne fournissait pas d'éléments d'analyse des trajectoires et des caractéristiques des personnes sur une période antérieure à la loi. Aussi, le CENSUS BUREAU a, conformément à la décision du Congrès, décidé de construire un panel long, le SURVEY OF PROGRAM DYNAMICS, en suivant à nouveau des personnes des SIPP 92 et 93. De très nombreux autres panels longitudinaux de moindre ampleur se sont développés, souvent localement, pour suivre des sous-populations concernées par tel ou tel programme.

Une des caractéristiques remarquables du système de recherche américain est que la grande majorité des données d'enquêtes (au niveau des données individuelles anonymisées) est mise à la disposition de l'ensemble des utilisateurs - documentées et le plus souvent gratuitement - et en ligne ; et ce, qu'il s'agisse d'enquêtes instantanées ou de panels longitudinaux, qu'ils soient réalisés par des organismes publics ou privés.

Une utilisation des sources administratives

Dans les pays d'Europe du Nord (Danemark, Norvège, Suède, Finlande), les données statistiques proviennent fréquemment de sources administratives qui peuvent autoriser un suivi longitudinal (« registers »). En Finlande, le projet CHINEX, a montré notamment que la collecte des revenus du panel européen pouvait provenir de sources administratives plutôt que déclaratives. Cependant, même si les deux origines fournissent la même image de la population dans son ensemble, des différences apparaissent aux deux extrêmes de la distribution des revenus : la conclusion est davantage un enrichissement mutuel des deux types de données qu'une véritable substitution. En outre, la mise à disposition de telles données se heurte à des règles de confidentialité sévères qui réduisent fortement les possibilités d'exploitation.

A partir des sources administratives, le Canada a mis en place un outil analogue au PSID et aux grands panels généralistes, la banque de DONNEES ADMINISTRATIVES LONGITUDINALES (DAL), outil de recherche sur le revenu et la démographie. Le fichier longitudinal contient de nombreuses variables démographiques annuelles au sujet des personnes représentées, de même que des renseignements sur le revenu annuel, à la fois pour le particulier et sa famille, au sens du recensement, pour l'année. Ce sont surtout les ministères gouvernementaux qui se servent des données pour évaluer les programmes et soutenir les recommandations

¹ Cf. Michel DOLLE : « Processus de développement de politiques publiques : les enseignements de la réforme du WELFARE américain », LES PAPIERS DU CERC, N° 2002 – 02. Décembre 2002

stratégiques. Mais les universitaires, les conseillers privés et les chercheurs de Statistique Canada utilisent également ces données pour analyser la situation socioéconomique.

Encadré 1 : Quelques éléments de vocabulaire pour les études longitudinales

Différents types de données sont mobilisables pour conduire une analyse longitudinale. On va essayer ici de préciser un vocabulaire parfois ambigu, ou du moins d'explicitier le vocabulaire adopté au sein du rapport :

↳ *Les cohortes sont des outils utilisés généralement par les démographes ou les épidémiologistes. Une cohorte regroupe des individus ayant un ou plusieurs facteurs en commun, en général leur année de naissance. Ces individus sont suivis au cours du temps avec une périodicité régulière. En général, le suivi s'opère sur une durée longue, dix ou vingt ans, les phénomènes démographiques se manifestant avec une périodicité longue. Une cohorte vieillissant au cours du temps, il est souvent nécessaire de lancer régulièrement de nouvelles cohortes si l'on souhaite examiner les tranches d'âge particulières (notamment l'enfance).*

↳ *Les panels portent la plupart du temps sur des populations moins spécifiques que les cohortes, voire sur la population entière. Alors que le nombre d'individus reste constant dans une cohorte (si l'on fait abstraction de l'attrition), il peut évoluer à l'intérieur d'un panel. Par exemple, en cas de mariage ou de naissance, les conjoints ou les enfants des individus du panel peuvent devenir eux-mêmes des individus panels. En général, les panels s'intéressent plutôt aux aspects socio-économiques et la durée de suivi est moindre que pour les cohortes.*

↳ *L'échantillon de certaines enquêtes répétitives n'est pas renouvelé complètement à chaque opération de collecte de l'enquête, en général pour réduire la variance des évolutions temporelles de macro-évolutions de variables (taux de chômage, opinion,...). C'était par exemple le cas de l'enquête annuelle sur l'emploi dont l'échantillon de logements était renouvelé par tiers, ou de l'enquête sur l'emploi en continu, ou de l'enquête sur les loyers et charges. De ce fait, un suivi longitudinal des individus qui habitent durablement dans le même logement est possible. En général, on ne suit pas les individus qui déménagent, ce qui réduit la représentativité, mais réduit les coûts de collecte. On parle parfois de « panels de logements ».*

↳ *Les enquêtes répétitives peuvent être utilisées pour assurer un suivi des macro-agents même si les échantillons sont renouvelés. Par exemple, chacune des enquêtes répétées assure une représentativité statistique des strates d'individus définies par le sexe et la date de naissance. Cette technique dite des pseudo-panels permet à moindre coût de suivre au cours du temps le comportement moyen des macro-agents, mais pas des individus.*

↳ *Enfin, une enquête en coupe transversale permet au travers de questions rétrospectives de reconstituer l'histoire des individus. Cette pratique en apparence séduisante se heurte à de lourdes difficultés méthodologiques :*

- pour obtenir des réponses à des questions réputées objectives (calendriers d'activité,...), il faut être très précis, ce qui conduit à un questionnement lourd et demande un effort de mémoire assez considérable à l'individu. Un questionnement léger (du type « jeunes et carrières », ou similaire) entraîne de nombreuses incohérences ou des trous dans les calendriers, qui compromettent l'exploitation.

- les questions plus subjectives sont à proscrire : l'expérience montre que les individus reconstruisent le passé, et de nombreux utilisateurs estiment que la quasi-totalité des réponses correspond à cette reconstruction.

3. L'existant en France

3.1. Brève description des panels en France

Les différents entretiens ont permis de dresser une liste des données longitudinales portant sur les personnes disponibles en France (annexe 4). Même si celle-ci ne prétend pas à l'exhaustivité, elle montre que la situation est plus nuancée que certaines critiques radicales le laisseraient entendre ; de fait, les données longitudinales ne sont pas absentes du paysage statistique national, qu'elles soient obtenues par collecte, ou au moyen de l'exploitation de sources administratives. L'exploitation de certaines d'entre elles, comme l'enquête sur les sorties du RMI ou le panel issu des DADS a d'ailleurs largement contribué à l'amélioration de la compréhension d'un domaine, ou l'évaluation de dispositifs.

Un certain nombre de caractéristiques communes apparaissent néanmoins à la lecture de la liste² :

- ? La plupart des sources ne porte pas sur la population dans son ensemble, mais sur un champ plus restreint. Ceci peut s'expliquer par le fait que l'on s'intéresse effectivement à une population particulière (enfants, actifs,...). Mais fréquemment, et en particulier pour les sources administratives, la restriction de champ est « subie », et s'explique par la nature du dispositif. Ainsi, les DADS collectent de l'information sur les salariés des secteurs privé et semi-public, ce qui limite par construction le champ du panel.
- ? La difficulté est que le mode de sélection du champ coïncide fréquemment avec les règles d'attribution de la prestation administrative dont on cherche précisément à évaluer l'effet (demandeurs d'emploi, allocataires d'allocation,...). Ceci pose des problèmes d'exploitation liés à cette sélection endogène du champ, que l'on cherche à résoudre au moins partiellement en mettant en place des échantillons témoins.
- ? Certaines de ces sources ne sont connues que d'un nombre réduit d'initiés. L'accès aux données lui-même est de toutes façons fortement restreint pour des raisons de confidentialité liées aux données administratives.

Le panel européen est l'une des rares sources en population générale dont l'accès est libre. Il n'est cependant pas véritablement documenté, et ne comporte que huit vagues, la collecte ayant été interrompue brutalement.

Des initiatives récentes de la DARES et de la DREES méritent l'attention :

- la DARES a lancé en 2001 une troisième génération de panels visant à évaluer les bénéfices des politiques d'emploi. Un certain nombre d'innovations méthodologiques ont été introduites, notamment la mise en place de groupes témoins afin d'éliminer les biais de sélection. Elle présente surtout la particularité de s'appuyer sur différentes sources d'information complémentaires afin de tirer le meilleur parti et pallier les insuffisances de chacune d'entre elles. La population est tirée dans des sources administratives. Mais elle procède ensuite par entretien aussi bien auprès des individus que des entreprises qui les ont accueillies, afin de comprendre dans les processus d'insertion ce qui peut s'expliquer par le comportement des entreprises et l'évolution de la conjoncture économique.

² Pour mémoire, l'INSERM dispose d'une myriade de panels épidémiologiques portant sur les traitements de patients atteints de différentes affections. Si l'on excepte la cohorte EDEN, leur mode de constitution comme leur finalité obéissent à une logique médicale davantage que statistique.

- l'échantillon national inter-régime d'allocataires de minima sociaux élaboré par la DREES vise à regrouper dans une même base de données des informations issues des différentes sources administratives sur les minima sociaux³. L'opération porte sur un échantillon de personnes âgées de 16 à 64 ans. Chaque année, les traitements statistiques consistent à rechercher dans les fichiers administratifs si ces personnes sont bénéficiaires d'un des quatre minima sociaux considérés (RMI, AAH, API, ASS), ou répertoriés dans les fichiers de l'UNEDIC comme chômeur indemnisé ou non. L'objectif de l'opération est d'étudier les caractéristiques des bénéficiaires, mais aussi de disposer d'une base de sondage qui pourra être utilisée pour la réalisation d'enquêtes spécifiques portant par exemple sur l'insertion sociale. Dans le même esprit d'enrichissement mutuel de différentes sources administratives, mais pour obtenir de l'information visant à améliorer la compréhension des modes de constitution des retraites, la DREES met en place un échantillon inter-régime de cotisants parallèlement à son échantillon inter-régimes de retraites.

Au-delà de la constitution de bases de données statistiques, l'intérêt de ces opérations est l'enrichissement mutuel de différentes sources d'information, qu'il s'agisse de sources administratives multiples dans le cas de la DREES, ou de sources administratives et déclaratives dans le cas de la DARES. Ceci réduit les limites liées au nombre restreint de variables de chacune des sources. En outre, dans les deux cas, les problèmes de sélection de champ sont résolus par un mode de tirage en population générale (minima sociaux), ou par la mise en place d'échantillons témoins (politiques d'emploi).

3.2. Une culture de coupe transversale probablement excessive

Même si les données longitudinales ne sont pas absentes du paysage statistique national, force est de reconnaître que la quasi-totalité du dispositif repose sur des enquêtes en coupe transversales destinées à répondre à un problème spécifique. La raison généralement mise en avant est la question des coûts : l'investissement est moindre au démarrage ; il n'est pas non plus nécessaire de s'engager pour une période de temps longue, ce qui peut présenter un avantage dans un contexte d'incertitude budgétaire⁴. La conséquence de cette « préférence pour le présent » est néanmoins une certaine inefficacité sur le moyen terme :

- Chaque opération étant conduite sur un échantillon séparé, on ne peut capitaliser sur la durée les informations portant sur les mêmes individus sur un sujet donné, ou sur des thèmes connexes. Les rapprochements ne peuvent porter que sur des strates homogènes, ce qui restreint le champ des études de corrélations. En particulier, les méthodes de contrôle des variables inobservables et de l'endogénéité sont souvent insatisfaisantes (manque d'instruments,...).
- Le corollaire est une lourdeur inévitable des enquêtes multithèmes : puisque toute l'information doit être collectée en une seule opération, il faut recourir à des questionnements longs ou des visites multiples.
- L'absence d'information sur le passé des individus conduit les responsables d'enquête à recourir de façon trop systématique à des questions rétrospectives, sans qu'il leur soit possible de valider même partiellement la pertinence des réponses à ces questions (voir en encadré 1 les problèmes posés par ce type de pratique).

³ Ce dispositif complète l'échantillon interrégime de retraités, mis en place depuis plus de dix ans.

⁴ Une autre raison, apparue récemment lors de la mise en place de la collecte de l'enquête SILC, est peut-être une acceptabilité plus difficile des panels de la part des ménages. Dans SILC, la lettre-avis mentionne le nombre de visites prévues alors que pour le panel européen, cet aspect n'était évoqué qu'à la fin du premier questionnement, une fois le contact noué. D'après les enquêteurs, la nouvelle formule dissuade certains ménages de procéder au premier entretien. Il pourrait être utile de s'interroger sur le meilleur protocole de mise en œuvre du concept de « collecte loyale », afin de respecter les droits du citoyen tout en permettant à l'institut de communiquer de façon plus pédagogique avec celui-ci.

- Le dispositif dans son ensemble donne une certaine impression d'éparpillement. En caricaturant quelque peu, le réflexe naturel est de mettre en place une enquête spécifique dès lors que l'on cherche à répondre de façon urgente à une question nouvelle, même si les formations du CNIS et le Comité du Label font leur possible pour éviter les redondances en matière de collecte. Cette démarche se justifie si la réactivité est indispensable, par exemple pour répondre à des demandes ministérielles. Elle privilégie néanmoins la minimisation des coûts à court terme. Lorsque l'on peut planifier le programme d'enquêtes, on peut se demander si la dispersion des moyens au sein du système statistique qui résulte d'un recours presque exclusif aux coupes instantanées n'est pas source d'inefficacité sur le moyen terme. Un exemple est peut-être le dispositif des enquêtes PCV, qui répond bien à son objectif de réponse à la demande de court terme. Celui-ci peut cependant sembler coûteux en termes de moyens, même si cette impression doit être tempérée compte-tenu du nombre d'enquêtes réalisées ; certains lui reprochent en outre une approche trop superficielle liée à la trop petite taille du questionnaire par rapport à la complexité des thèmes que la demande sociale conduit à aborder aujourd'hui.

3.3. Un manque de savoir-faire sur données longitudinales

Au travers des différents entretiens, le manque de savoir-faire méthodologique a souvent été évoqué comme un frein au développement des panels.

- En matière de collecte et de suivi longitudinal, peu d'équipes ont une culture des méthodes de tirage, de renouvellement, de suivi ou de fidélisation des échantillons. Il semblerait que seuls l'INED et l'INSEE, avec l'expérience du pôle de Nancy, disposent des ressources méthodologiques nécessaires à la mise en place d'un panel de moyen terme en population générale. Puisque l'accumulation de compétences nécessaires est plus intense pour les données longitudinales que pour les opérations ponctuelles, une concentration accrue des ressources autour d'un ou plusieurs centres de compétences est sans doute souhaitable. Ces centres pourraient par exemple comparer utilement la méthodologie actuellement en place avec les pratiques du secteur privé⁵. Le corollaire en matière d'organisation est que les centres de production isolés puissent bénéficier d'une façon ou d'une autre des compétences disponibles de façon concentrée.
- Le redressement des non-réponses ou de l'attrition part le plus souvent du principe qu'en se référant aux caractéristiques socio-économiques habituelles et observables, les individus non-répondants peuvent être remplacés par des répondants (hot-deck, repondération,...). La validation de ces a priori pour certaines variables pourrait faire l'objet d'opérations-qualité. Celles-ci pourraient prendre la forme de ré-interrogations de non-répondants⁶, ou encore d'enquêtes méthodologiques complémentaires au recensement portant sur des variables clés, celui-ci étant aujourd'hui annualisé. L'INSEE a procédé dans le passé de la sorte en utilisant pour des études méthodologiques les enquêtes « Famille » (associées aux recensements de la population de 1954 à 1999) ou VQS (Vie Quotidienne ou Santé, enquête associée au recensement de la population de 1999).
- De la même façon, l'exploitation des données longitudinales inquiète les chargés d'études peu habitués à ce genre de sources. Il est ressorti des entretiens que les panels étaient souvent utilisés, en France, comme des coupes successives, ou pour l'étude des transitions entre T et T+1. Un effort de formation est sans doute à

⁵ Certains opérateurs privés n'hésitent pas à rémunérer dans certains cas les enquêtés pour améliorer l'acceptabilité des enquêtes, alors que les gratifications offertes par les acteurs publics sont de l'ordre du symbolique. Au-delà des questions de principe, et même si de telles pratiques ne constituent pas une panacée, il pourrait être instructif de réaliser à leur sujet un bilan coût-bénéfice.

⁶ Notamment sur le thème du divorce et le suivi des personnes après la séparation.

entreprendre. Comme pour la collecte, la mise en place d'un centre de ressources méthodologiques permettrait de rassurer et conseiller les chargés d'études au sein du système statistique. Celui-ci pourrait s'appuyer sur le savoir-faire de l'institut du longitudinal, au sein du LASMAS, qui organise régulièrement des séminaires ou des colloques portant sur différents aspects méthodologiques de l'exploitation des panels.

Encadré 2 : Avantages et inconvénients du longitudinal

Les données longitudinales, qu'elles soient ou non panélistées, sont absolument indispensables pour comprendre les effets de l'âge ou du vieillissement. En coupe transversale, il est en effet impossible de démêler les effets de l'âge de ceux de l'année de naissance si l'on cherche à comprendre n'importe quel phénomène économique ou social. Le suivi des individus par un panel n'est, dans ce cas, pas indispensable, et empiler des coupes transversales portant sur les mêmes variables peut suffire.

En revanche, le suivi des individus est incontournable si l'on recherche à analyser les comportements individuels dans la durée.

- ? *Ainsi, l'analyse des transitions entre différents états requiert le suivi des individus pendant au moins deux dates ; or, ces transitions sont des indicateurs précieux, voire incontournables, si l'on s'intéresse par exemple à la dynamique du marché du travail, ou aux entrées-sorties de pauvreté. Plus généralement, des indicateurs aussi élémentaires que la durée moyenne du chômage ne peuvent être calculés à partir des coupes transversales, et nécessitent un suivi longitudinal des individus.*
- ? *Les trajectoires elles-mêmes présentent un intérêt dans l'étude des comportements ou l'évaluation des politiques publiques. La compréhension des carrières salariales, de l'insertion des jeunes ou l'analyse de la rentabilité des formations nécessite un suivi des individus dans la durée. Des comportements qui comportent des allers et retours comme l'immigration conduisent à des mesures biaisées si l'on se limite à des analyses en coupe.*
- ? *Le suivi longitudinal permet en outre de modéliser une hétérogénéité inobservable des caractéristiques individuelles. La prise en compte de cette dernière dans l'analyse des comportements permet de réduire les biais d'estimation dans l'analyse des comportements ou la mesure des causalités.*
- ? *Le caractère répété des visites permet d'annoncer à l'avance le thème des prochaines interrogations, et permet au ménage de s'y préparer psychologiquement ou matériellement.*
- ? *Le suivi longitudinal permet enfin de réaliser des contrôles de qualité que l'on ne peut conduire en coupe, en comparant les réponses des enquêtés à deux périodes successives, et en leur demandant d'en valider la cohérence temporelle. On peut alors corriger des réponses qui seraient erronées en coupe instantanée. Cette pratique de validation temporelle pose néanmoins problème en terme de respect de la confidentialité des réponses au sein d'une même famille lorsque le protocole de collecte permet à des tiers de répondre à la place de l'enquêté au sein du ménage. La solution adoptée par SILC consiste à demander l'autorisation à l'enquêté de faire part si besoin de ses réponses au répondant à la vague suivante pour assurer le contrôle de cohérence.*

Un dernier avantage des panels ou des cohortes est la capitalisation d'informations sur les individus. Même si l'on se limite à une exploitation en coupe transversale, on dispose de toute l'information accumulée par le passé sur les individus. Cette propriété est largement utilisée par le PSID américain, au travers de nombreuses enquêtes ponctuelles réalisées sur l'échantillon panelisé.

A côté de ces avantages, un certain nombre d'inconvénients sont reprochés aux panels ;

- ? *Le premier d'entre eux est le problème de l'attrition : au fur et à mesure des visites, le nombre de répondants diminue, et la question de la représentativité des « survivants » se pose. Ce constat indiscutable doit cependant être relativisé dans deux directions. En premier lieu, la première cause de non représentativité est la non-réponse dès la première visite, tout aussi observable avec des*

enquêtes en coupe. Ces dernières, moyennant des redressements adéquats, sont cependant jugées exploitables, faute de quoi aucun indicateur statistique ne serait disponible. Au-delà, la plupart des expériences tendent à montrer que l'attrition est surtout à l'œuvre les premières années, et qu'au-delà la participation tend à se stabiliser autour d'un noyau dur. Celui-ci est d'autant plus large que le soin apporté à communiquer avec les enquêtés pour les fidéliser est important. Ceci étant, dans les pays qui disposent de panels, l'attrition n'est pas un obstacle aux publications, y compris aux publications de haut niveau.

- ? La répétition des mêmes questions posées à plusieurs reprises à un même individu peut entraîner un inconvénient symétrique à l'avantage qui consiste à disposer de plusieurs réponses consécutives. Les entretiens se répétant, l'individu apprend progressivement la logique du questionnaire, et repère les questions qui donnent lieu à développement. Dans un souci de raccourcir le temps de collecte, l'enquêté peut être conduit à répondre que « rien ne s'est passé depuis la dernière fois », ce qui conduit à un lissage excessif des réponses par rapport à la réalité.
- ? À cause du suivi longitudinal, les coûts de collecte et de maintenance sont plus élevés pour les panels que pour les coupes transversales. Outre les efforts à consentir pour limiter l'attrition des personnes qui restent à une adresse fixe, une énergie importante doit être déployée pour suivre les personnes qui déménagent ou quittent leur foyer d'origine.
- ? Pour sa part, la méthodologie à mettre en œuvre pour la collecte et l'exploitation est plus lourde que dans le cas des enquêtes en coupe. Comme on l'a vu, le suivi des adresses requiert des efforts spécifiques ; de la même façon, le redressement des non-réponses dues à l'attrition et l'exploitation des trajectoires met en œuvre une méthodologie statistique beaucoup plus aboutie que pour les coupes.
- ? Enfin, un soin tout particulier doit être apporté en ce qui concerne les questions de confidentialité et de respect de l'anonymat des personnes, que ce soit les individus interrogés ou les personnes-relais qui permettent de retrouver l'enquêté en cas de déménagement (cette dernière doit obligatoirement être avertie). Le suivi longitudinal entraîne une accumulation d'information qui rend beaucoup plus facilement repérables les personnes que dans le cas des coupes. À l'exception des États-Unis, rares sont les panels dont l'accès intégral est possible sans restriction.

4. Les besoins

4.1. Niveau national

Lors des entretiens, nos interlocuteurs ont ré-affirmé la nécessité de disposer de données sur les trajectoires individuelles, le plus souvent accompagnées de données sur la famille et le ménage.

Des données longitudinales pour quels besoins ?

Les besoins, en terme de finalités, peuvent être regroupés en deux grandes familles :

- Connaissance de phénomènes : modélisation de théories économiques, explications et prédictions;
- Évaluation des politiques publiques.

Sur le premier besoin, celui de la connaissance, certains phénomènes nécessitent une approche longitudinale pour avoir du sens, dès que la transition d'un état vers un autre n'est pas définitive ou quand les allers-retours entre les deux états existent et viennent « fausser » les résultats issus de données en coupe.

On peut citer dans le domaine de l'Éducation des thèmes comme le retard scolaire ou le choix de l'enseignement privé par exemple (ce dernier passe de 20% en coupe à 4 élèves sur 10 fréquentant à un moment de leur scolarité l'enseignement privé). Un autre exemple emblématique du besoin de suivi individuel est celui de l'immigration.

Certains thèmes se sont développés depuis une quinzaine d'année mais ils ne couvrent pas toujours l'ensemble des sous-populations. Par exemple, si les approches sur la dynamique de la pauvreté ont fait l'objet de nombreuses publications depuis une dizaine d'années (Cf. RAY-JEANDIDIER, communication aux Xè journées du longitudinal, Caen, Mai 2003), les besoins restent importants sur certaines sous-populations comme les enfants par exemple (Cf. rapport du CERC sur les enfants pauvres). De façon générale, la connaissance sur les jeunes générations est relativement faible en dehors de la scolarité proprement dite.

Toujours en matière de connaissance des phénomènes et plus particulièrement sur les transitions entre deux états, les obligations communautaires ont fait leur apparition dans deux domaines :

- ? L'emploi avec les matrices de transition sur le marché du travail : le consensus s'est fait sur la notion de trajectoire lors des travaux sur la qualité de l'emploi en 2001. Ces indicateurs font partie de l'Eurobaromètre.
- ? La pauvreté, avec les entrées/sorties.

Ces indicateurs étaient obtenus avec le panel européen.

Les politiques publiques, en matière d'emploi en particulier, ont évolué d'une problématique de « stocks » à un intérêt porté aux flux et aux trajectoires : discontinuité des trajectoires, récurrence de certaines situations (Cf. Communication de la DARES aux Xè journées du longitudinal, Caen, Mai 2003). De nouveaux outils d'analyse se sont mis en place. D'une analyse descriptive des bénéficiaires, les travaux ont évolué vers l'étude des différentes étapes des trajectoires individuelles.

Pour évaluer les politiques publiques, l'utilisation des seules données administratives pose des difficultés : appréhension des trajectoires et sorties du système limitées aux seuls

bénéficiaires de ces politiques et aux seules données recueillies (pas de population « témoin » ou éligible, pas ou peu de données sur l'environnement de l'individu, causes de la sortie ...). La demande est souvent double : disposer rapidement d'éléments pour évaluer une politique, mais aussi des besoins sur des évaluations des politiques sur plus long terme (suivi, évaluation, efficacité).

Le système statistique public français, contrairement aux autres pays européens, ne tisse pas de liens forts avec les milieux universitaires : peu d'échanges, une mise à disposition des données qui commencent seulement à progresser. La crédibilité du système français face aux décideurs communautaires s'en trouve amoindrie.

Il est cependant indispensable d'améliorer le système statistique public en s'inspirant des « bonnes pratiques » européennes et étrangères dans le domaine de l'utilisation des panels.

Les domaines

Pratiquement tous les domaines sont concernés :

- Démographie, immigration ;
- Social : inégalités, pauvreté, conditions de vie, insécurité ;
- Enfants : données socio-démographiques, éducation, santé ... ;
- Santé ;
- Retraite et vieillissement ;
- Emploi : insertion sur le marché du travail, chômage ;
- Entreprises.

Chacun de ces domaines requiert peu ou prou des données longitudinales. La démographie, qui s'intéresse à des évolutions portant sur un horizon traditionnellement long, est sans doute la discipline où l'on a construit le plus tôt des données longitudinales, comme les cohortes ou les registres de population. Pour leur part, les professionnels de la santé utilisent fréquemment des données longitudinales pour leurs études épidémiologiques. Les statisticiens accusent un certain retard dans la mise en place de ces outils. Les travaux sur l'emploi ont recouru plus récemment à de telles données ; celles-ci sont aujourd'hui incontournables pour comprendre les trajectoires d'insertion, ou encore l'évaluation des dispositifs publics. L'usage des données longitudinales en ce qui concerne les conditions de vie est plus difficile, car l'unité statistique est le ménage, qui s'accommode mal d'une analyse dans la durée contrairement à l'individu. Enfin, les panels de salariés ont largement concouru par le passé à mieux connaître les comportements des entreprises, comme par exemple en ce qui concerne l'impact des nouvelles technologies sur les qualifications et les coûts, ou l'effet des baisses de charges sur l'emploi.

4.2. Niveau local

Les besoins actuels en matière d'études longitudinales portent en premier sur la connaissance de phénomènes et en particulier sur le lien entre géographie et processus/cycle de vie et ceci, plus particulièrement sur les domaines de la formation et de la Politique de la Ville.

On peut citer les exemples suivants : lien entre le lieu d'études et les marchés locaux du travail ; ou encore chercher à répondre à des questions du type : quel marquage individuel du passage dans un quartier donné, les quartiers sont-ils stigmatisants ?.

Les outils permettant d'aborder les phénomènes locaux sont de deux types :

- ? Outils nationaux : la présence de localisation et variables géographiques (résidence, résidences antérieures, lieux de travail ...) dans des outils nationaux permet des études sur des territoires particuliers (ex : entrées/sorties des ZUS à partir de l'ECHANTILLON DEMOGRAPHIQUE PERMANENT, travail en cours à l'INSEE, au département de l'Action Régionale).
- ? Construction d'outils spécifiques pour des travaux sur la population ou une sous-population d'un territoire donné (exemple du PANEL LORRAIN), la plupart des outils nationaux ne peuvent répondre à ce besoin.

La plus grande difficulté, quand on souhaite travailler à un niveau local est la construction d'indicateurs simples résumant des phénomènes longitudinaux. Si on prend l'exemple d'un indicateur comme l'espérance de vie, cet indicateur montre ses limites quand on descend au niveau local car il ne prend pas en compte les déformations des pyramides des âges dans le temps, déformation beaucoup plus lente à un niveau national.

Par contre, la demande dans le domaine de l'évaluation de politiques publiques qui seraient de la responsabilité locale (Région, Département, voire la commune) ne s'exprime pas encore en dehors du lien emploi/formation. Elle a cependant été évoquée en interne à l'Insee. L'idée sous-jacente est qu'il est impensable de monter autant de dispositifs d'observation localement. Une proposition émise par la DSDS serait de déterminer une typologie des dispositifs à partir d'une enquête d'observation des politiques locales, qui permettrait de construire des modèles et ainsi « mutualiser » les données.

Ce problème doit être étudié, avec soin, en effet la nouvelle étape de la décentralisation. Après une première vague de décentralisation (lois Defferre 1982-1983), la décentralisation s'accélère. Cinq domaines sont ou vont progressivement être délégués essentiellement aux départements ou aux régions :

- ? Développement économique-emploi (aides aux entreprises, formation professionnelle, tourisme).
- ? Infrastructures et aménagement du territoire (routes, ports, aéroports..., fonds structurels européens).
- ? Solidarité, santé (RMI/RMA, formation des travailleurs sociaux, logement social, personnes âgées).
- ? Éducation, culture, sport (élaboration des cartes scolaires, établissements scolaires, ATOS, médecine scolaire, CIO).
- ? Protection de l'environnement (eau et déchets).

Cette décentralisation s'accompagnera, tôt ou tard, de demandes d'évaluation locale sur telle ou telle politique ou d'évaluation sur le dispositif même de la décentralisation. La statistique publique devra se préparer à y apporter des réponses.

On sait déjà que les dispositifs locaux peuvent brouiller la vision nationale. Ainsi, des travaux⁷ monographiques ont montré que les prestations versées localement aux ménages démunis se superposaient fréquemment aux prestations versées dans le cadre des dispositifs nationaux. Le problème des trappes à inactivité s'en trouve exacerbé, les incitations au retour à l'emploi pouvant dans certains cas disparaître. Les méthodes d'évaluation ne peuvent s'affranchir de tels dispositifs. S'inspirer des pratiques américaines pour construire un processus d'évaluation pourrait constituer un point de départ intéressant pour l'évaluation des politiques locales.

⁷ D. Anne et Y. L'Horty (2002 : « Transferts sociaux locaux et retour à l'emploi », Economie et Statistique, n° 357-358, 49-71.

Encadré 3 : l'échantillon Démographique Permanent

L'échantillon démographique permanent a été mis en place dans les années soixante-dix. Il répond à la volonté de tirer parti d'informations sur les mêmes individus collectées au fil du temps dans les sources démographiques, recensements et état civil. L'échantillon est constitué par les personnes nées quatre jours déterminés dans l'année, quelle que soit l'année. L'échantillon se renouvelle à partir des naissances ou par la venue de nouveaux immigrés révélée par les recensements ou par un événement enregistré dans un acte de l'état civil. A tout moment, la base est proche d'un échantillon au 1/100 (4/365) de la population résidant en métropole.

Deux sources alimentent le fichier. Les recensements de population décrivent les individus à des dates fixes. L'état civil recueille l'information à l'occasion d'événements majeurs de la vie familiale. Pour chaque individu, le fichier rassemble les informations du bulletin individuel rempli par les intéressés lors des cinq derniers **recensements** de population : 1968, 1975, 1982, 1990 et 1999, s'il a été recensé à ces dates.

Les événements démographiques majeurs faisant l'objet d'un enregistrement à l'état civil sont pris en compte dans l'EDP :

- la naissance d'un individu du champ EDP (si elle survient l'un des quatre jours de référence),
- son éventuelle reconnaissance par un parent,
- le mariage de parents susceptible de le légitimer,
- son ou ses propres mariages,
- la naissance de ses enfants,
- la reconnaissance de ses enfants,
- son décès.

Ces bulletins relatent les circonstances de l'événement (lieu et date) et fournissent des caractéristiques de la personne (sexe, date et lieu de naissance, commune, département et pays de domicile, état matrimonial, catégorie sociale, nationalité). Les événements de l'état civil sont pris en compte jusqu'à l'année 2002. L'année 2003 est en cours d'introduction. La base d'études rassemble désormais 886 000 enregistrements individuels, dont 617 000 concernent des personnes présentes au recensement de la population de 1999. Sur l'ensemble de la base, près de 752 000 personnes sont nées en France, soit 85 % du total, et 135 000 à l'étranger.

Quelques chiffres					
Effectifs EDP présents au ...		et également présents aux recensements successifs de ...			
		1990	1982	1975	1968
RP99	617 195	497 699	399 915	327 705	260 605
%	100	80,6	64,8	53,1	42,2
RP90	588 428		470 808	387 580	309 839
%	100		80,0	65,9	52,7
RP82	551 211			450 965	362 682
%	100			81,8	65,8
RP75	526 014				418 025
%	100				79,5
RP68	485 009				

5. Les pistes possibles

5.1. Les données administratives

L'avantage des données administratives est qu'il s'agit en général de fichiers comportant un grand nombre d'individus, et permettant un suivi sur une longue période sans risque de lasser les personnes. L'exemple le plus emblématique est sans doute celui du panel des Déclarations de Données Sociales (DADS)⁸ qui autorise à l'heure actuelle le suivi annuel des salariés du secteur privé et semi-public depuis 1967, soit pour certains salariés la quasi-totalité de leur cycle d'activité. Cette source véritablement exceptionnelle a même été imitée aux États-Unis puisque le CENSUS BUREAU est en train de mettre en place une base de données administratives analogue. Mais la source est également révélatrice à plus d'un titre des difficultés inhérentes à ce type d'exercice :

- ? Elle a longtemps été réalisée et exploitée hors des programmes de travail habituels : issue à l'origine d'une collaboration avec C. BAUDELLOT, elle a dû sa survie pendant les années 80 à des bonnes volontés locales, l'intérêt des universitaires, et son très faible coût de constitution à partir de l'exploitation standard des DADS. En témoignent les stigmates de 1981, 1983 et 1990, où les données ne sont pas disponibles car les moyens manuels de l'époque ne permettaient pas une exploitation des DADS.
- ? Le nombre de variables disponibles est limité par le contenu de la déclaration. Pour disposer du diplôme des individus, il faut appairer cette source avec L'ÉCHANTILLON DEMOGRAPHIQUE PERMANENT⁹, ce qui n'est pas trop difficile techniquement, mais pose des problèmes de confidentialité, de sorte que les travaux sont restés à l'état de notes internes faute de cadre juridique légitimant cette approche.
- ? Le champ est restreint aux seuls salariés du secteur privé et semi-public. Ceci a longtemps interdit le suivi des individus sortant du champ, notamment en raison des transitions vers et hors du chômage. Les appariements avec des sources complémentaires sont eux-aussi restreints faute d'un cadre réglementaire adapté.
- ? Enfin, l'accès à ces données directement nominatives ne peut être autorisé sans contrôle. Jusqu'ici, les seules personnes autorisées à les utiliser étaient les membres du système statistique public, ou les chercheurs venant travailler au sein des équipes du système statistique, après d'être engagé par écrit à respecter des clauses restrictives de confidentialité et de non-dissémination des données. Cette situation est susceptible d'évoluer dans un futur prévisible avec la mise en place de centres d'accès sécurisés et de procédures visant à assermenter les utilisateurs venant travailler dans ces centres.

On l'a vu, le système statistique dispose en France d'un ensemble de données administratives très riche. Pourtant, chacun de ses éléments pris séparément souffre des restrictions décrites précédemment : faible nombre de variables disponibles, champ restreint, et accès difficile. S'ajoute dans certains cas un coût de mise en place de l'information administrative de base plus élevé que dans le cas du panel des DADS, où le travail est réalisé lors de l'exploitation courante des DADS.

⁸ Cf. annexe 4.

⁹ Cf. annexe 4.

Une solution extrême pourrait consister à appairer entre eux tous ces panels administratifs, en étendant les expériences actuellement à l'œuvre à la CNAM pour des besoins administratifs à l'ensemble des sources disponibles. Cette approche ne nous paraît cependant pas souhaitable pour différentes raisons :

- ? Elle reviendrait à constituer un répertoire dynamique de l'ensemble de la population, et donc de fait l'équivalent des « registres » au sens des pays nordiques. Il se poserait de graves questions de respect des libertés individuelles, le fichier consistant à suivre la totalité de la population française.
- ? Quand bien même ces difficultés liées à la confidentialité pourraient être surmontées, un appariement exhaustif de toutes les sources administratives représenterait un coût considérable en termes de moyens. La réussite d'une opération d'une telle envergure serait compromise de par sa lourdeur même.
- ? Ce coût ne serait pas justifié au regard des besoins, au moins des besoins nationaux et européens répertoriés.

C'est la raison pour laquelle l'orientation suivie par Statistique Canada paraît justifiée ; celle-ci consiste à constituer un échantillon, en l'occurrence de 20%, et à en assurer le suivi longitudinal. En France, la colonne vertébrale d'un dispositif administratif pourrait être utilement constituée par L'ÉCHANTILLON DEMOGRAPHIQUE PERMANENT. Celui-ci serait élargi dans deux dimensions, d'où le nom d'EDP++ :

- ? Appariement avec les sources administratives disponibles. Celui-ci pourrait être mis en place de façon progressive, et pourrait ne retenir que les sources les moins sensibles au vu du respect des libertés individuelles.
- ? Évolution du mode de sondage, de façon à équilibrer la saisonnalité des naissances pour répondre aux préoccupations des démographes et des éducateurs, mais aussi éviter une « prédétermination » statistique à certaines dates de naissance (Taille optimale de l'ordre du 1/20^{ème}).

5.2. Les initiatives européennes

A l'heure actuelle, le système statistique français dispose d'opportunités excellentes pour mettre en place à moindre coût un dispositif longitudinal aux normes internationales. Il suffit pour cela de se contenter de profiter d'initiatives soit européennes, soit inspirées par les expériences étrangères :

- ? Le programme SILC, porté par EUROSTAT, permet de mettre en place un panel analogue au PSID américain, au BHPS britannique ou au SOEP allemand.
- ? L'initiative SHARE est l'équivalent européen du HRS. Portant sur les personnes âgées, il vise à documenter et éclairer la réflexion sur les conséquences économiques et sociales du vieillissement. Se priver en France d'un tel instrument alors que le vieillissement de la population est avéré ne paraît pas raisonnable.
- ? Le projet de cohorte nationale « enfants » actuellement porté par l'INED s'inspire des cohortes d'enfants britanniques et des NLSY américains, et permettrait à terme de fournir des statistiques enfin satisfaisantes sur le devenir des enfants, et notamment des enfants pauvres.

Mais pour véritablement profiter de cette opportunité, il conviendrait que le dispositif statistique adhère à une culture longitudinale, et qu'il cesse de considérer ces opérations comme des greffons inopportuns, comme ce fut le cas lors de l'expérience du panel européen.

En effet,

- ? Le panel européen s'est purement et simplement ajouté au programme statistique existant fondé sur les coupes transversales, sans que les synergies ou les économies d'échelle n'aient été explorées. Ceci peut en partie s'expliquer par le fait qu'il s'agissait du premier panel véritablement mis en place, et donc qu'il pouvait être considéré comme une expérience. D'ailleurs, lorsque la pression communautaire s'est relâchée, l'expérience a cessé, quand bien même elle était techniquement concluante.
- ? À ce jour, aucune documentation digne de ce nom n'est disponible, et les données ne sont accessibles qu'à un cercle d'initiés (la source n'est pas répertoriée). Il n'est donc guère surprenant qu'elle soit moins exploitée que ses consœurs à l'étranger, et il est même étonnant qu'elle le soit autant.

Si l'on cherche à pérenniser un dispositif longitudinal, et à en rentabiliser les coûts, il sera sans doute nécessaire de repenser le système d'enquêtes autour des trois sources évoquées précédemment. Dans cette optique, il serait logique que toute enquête spécifique lancée au sein du système statistique se voit offrir la possibilité d'utiliser l'échantillon de l'une des trois sources longitudinales. Cette possibilité ne constituerait pas une obligation mais le concepteur serait conduit à expliquer les raisons pour lesquelles il souhaite recourir à une enquête isolée¹⁰ :

- ? Un tel système permettrait de capitaliser de l'information sur les ménages et les individus et de réduire la durée des entretiens grâce à des économies d'échelle (non nécessité du tronc commun par exemple). Même pour les enquêtes ponctuelles, on dispose d'un historique des individus. L'expérience étrangère a montré que plus cette information est riche, plus le recours au panel comme base de sondage est demandé.
- ? Les nombreuses publications issues des sources longitudinales à l'étranger montrent que l'attrition n'est pas un frein à l'exploitation, sauf à supposer que les chercheurs et statisticiens étrangers sont moins compétents ou moins rigoureux que leurs homologues français. Quand on considère la place des études provenant de données longitudinales dans les meilleures revues scientifiques internationales, il est permis d'en douter.

Par conséquent, un réexamen de l'articulation entre données longitudinales et coupes ponctuelles paraît nécessaire si l'on cherche à développer les panels en France. Dans cette optique, la coexistence actuelle de SILC et du dispositif PCV peut poser un problème, comme l'a relevé le Comité des Investissements¹¹. Mais la remarque porte tout autant sur le système statistique dans son ensemble, où l'économie d'ensemble du dispositif est rarement évaluée, sauf au coup par coup par les formations du CNIS et le Comité du Label.

¹⁰ Par exemple, une enquête sur un sujet sensible justifie sans doute une distanciation forte entre l'enquêteur et l'enquêté. Celle-ci paraît contradictoire avec la constitution des bonnes relations nécessaires pour les interrogations répétées.

¹¹ A contrario, il faut être vigilant à ne pas poser au même ménage un questionnement excessif, que ce soit lors d'un entretien ponctuel ou dans la durée.

5.3. La complémentarité données - administratives et enquêtes

Contrairement à celle des données longitudinales, la culture de l'enrichissement mutuel des données administratives et des enquêtes est bien établie au sein du système statistique public, avec quelques expériences emblématiques, comme :

- ? L'enquête sur les revenus fiscaux, très utilisée sur les questions d'inégalité et de pauvreté, résulte depuis les années 60 de l'appariement de la source fiscale avec une source déclarative (recensement, puis enquêtes sur l'emploi).
- ? L'enquête « trajectoire des demandeurs d'emploi » porte pour sa part sur une cohorte de nouveaux inscrits à l'ANPE tirés dans les fichiers administratifs, et interrogés tous les ans par entretien.

On pourrait trouver d'autres exemples, mais les deux précédents sont révélateurs de deux types de pratique qui peuvent être appliquées aux opérations évoquées précédemment :

- ? les panels par entretien peuvent être enrichis grâce aux sources administratives. Comme l'est l'enquête sur les revenus fiscaux, des données complémentaires peuvent être mobilisées, ce qui permet de réduire les durées de questionnement ou les coûts de collecte. Mais les sources administratives permettent également de prolonger le suivi temporel au-delà de la période de collecte proprement dite. Par exemple, la source DGI serait en mesure de renseigner sur les ressources et l'environnement démographique des individus pendant un certain temps à l'issue des six ou huit années de collecte dans SILC. De la même façon, le suivi salarial des cohortes d'enfants au-delà de l'insertion sur le marché du travail pourrait être entrepris grâce aux fichiers des DADS, ou des fichiers de paie de la fonction publique.

Il se pose cependant à ce stade une question de déontologie. Des relations de confiance très étroites se nouent entre les enquêtés et le système statistique au cours d'un questionnement longitudinal, bien davantage que lors de la réalisation de coupes. Il est dans ces conditions difficile d'apparier les réponses avec des sources administratives pour des ménages chez qui on sera de nouveau accueilli, sans les informer et/ou sans leur fournir de sérieuses garanties. Ce point devra être résolu d'une façon ou d'une autre si on adopte dans la durée une pratique d'appariement entre réponses et données administratives.

- ? les sources administratives, transversales ou longitudinales, peuvent bénéficier d'un complément d'information provenant d'un questionnement direct portant sur une question ponctuelle. L'utilisateur de ce dernier dispose d'un historique lui donnant les moyens de mieux contrôler l'hétérogénéité individuelle et d'éclairer les mécanismes de sélection lorsque l'on cherche à évaluer une mesure. Dans de nombreux cas, l'EDP++ pourrait servir de base de sondage pour une enquête ponctuelle dans les travaux sur l'évaluation (RMI, dispositif d'insertion, sortie du chômage,...).

Pour terminer cette section, le plus révélateur est de citer la conclusion d'une équipe de chercheurs au questionnaire posé aux interlocuteurs :

Des panels :

- ✓ *un peu plus courts¹² (mais pas trop court, par exemple certaines analyses sur l'enquête sorties du RMI semblaient montrer qu'une vague ou deux de plus auraient permis d'affiner avantageusement les résultats),*
- ✓ *de taille suffisante pour pouvoir travailler sur des sous-groupes,*
- ✓ *bien ciblés sur des populations spécifiques (au sens de choisies en référence à une démarche d'évaluation d'une politique sociale, ce qui signifie le plus souvent que la collecte commence avant la mise en place de la politique en question de manière à pouvoir observer des changements au cours du temps),*
- ✓ *couplés à des informations administratives issues de fichiers de gestion (limiter le temps d'enquête et assurer une meilleure fiabilité de certaines informations),*
- ✓ *et dont l'échantillonnage intégrerait un groupe de contrôle, constituerait alors sans doute une option plus réaliste pour la Statistique Publique en France.*

¹² que le PSID (hors citation)

6. Recommandations

6.1. Les outils

Recentrer le dispositif sur les sources administratives

Recommandation 1 : mettre en place un panel s'appuyant sur des sources administratives

Recommandation 1a : créer un tel panel à partir de l'EDP élargi (EDP++) à la fois en terme de taille, mais surtout en l'enrichissant à partir de sources administratives (fiscales et sociales) tout en respectant l'anonymat des individus et les règles de confidentialité.

Recommandation 1b : mettre en place un Centre d'Appariement Sécurisé (CAPS) unique pour permettre et sécuriser cet appariement.

Un tel outil répond à un certain besoin de centralisation du système, afin de compenser par la multiplicité des sources les insuffisances de chacune d'elles (champ partiel, population spécifique,...). Une telle proposition est réaliste car elle reprend en l'élargissant la démarche entreprise par la DREES dans la confection de l'échantillon national inter régimes de minima sociaux. Elle permet de mutualiser les coûts de développement d'instruments séparés dans différents services du système statistique public. Le statut juridique de ce centre devra être élaboré en accord avec la CNIL.

S'appuyer sur les initiatives internationales et compléter le dispositif par trois panels plus courts obtenus par collecte

Recommandation 2 : Pérenniser un panel en population générale

Il s'agit ici de se mettre au niveau des grands partenaires, qui disposent d'un panel du type PSID. Il convient d'éviter les opérations de stop-and-go, comme celle qui consiste à stopper un panel quand celui-ci commence à obtenir la taille requise, et repartir de zéro deux ou trois ans plus tard. L'intérêt de SILC est qu'il s'agit d'un panel communautaire, et qu'il est donc possible d'empiler les fichiers des différents pays si l'on s'intéresse à des comportements « rares » (santé,..), ou des expériences naturelles (traitement du chômage,...). Dans ce dernier cas, il convient cependant d'avoir une connaissance très précise des réglementations indigènes.

Recommandation 3 : Mettre en place un dispositif longitudinal sur les enfants

Un effort tout particulier doit être consacré aux enfants, sur lesquels le dispositif actuel est faible. Le projet porté par l'INED et visant à mobiliser le système statistique public doit dans cette perspective être recommandé.

Les recommandations ne portent pas sur le dispositif SHARE, celui-ci étant piloté hors du système statistique. Compte-tenu des enjeux, il est néanmoins essentiel de s'y associer ainsi qu'à ses différents prolongements et de réfléchir à la mise en place d'un outil adapté pour étudier le vieillissement.

6.2. Les utilisations

Optimiser la constitution des systèmes d'informations nécessaires à l'évaluation

Recommandation 4: organiser, dans la mesure du possible, les systèmes d'informations nécessaires à l'évaluation autour du dispositif administratif en utilisant l'EDP++ comme base de sondage pour les opérations de collecte visant à enrichir par entretien les sources administratives, à chaque fois que possible. Chaque opération de collecte sera soumise à la CNIL.

Il s'agit à nouveau d'une généralisation de la formule adoptée par l'enquête trajectoire d'emploi à la DARES, ou de l'enquête RMI. L'avantage est que la base de sondage fournit de l'information homogène sur une population-témoin, et donc que l'enquête ne se limite pas aux seuls bénéficiaires. Une telle utilisation du fichier nécessitera un toilettage du statut juridique de l'EDP, qui devra être réalisé en concertation avec la CNIL. Bien évidemment, tel ou tel groupe social peut être sur représenté selon les finalités de l'opération.

L'outil est bien adapté pour la plupart des problématiques se rapportant à l'échelon national. Il atteint néanmoins ses limites en terme de taille lorsque l'on s'intéresse à des problématiques locales, comme l'évaluation de telle ou telle mesure prise au sein d'une collectivité territoriale (et notamment les communes). On l'a vu, mettre en place un fichier exhaustif n'est ni souhaitable, ni raisonnable en terme de coût. Ce que l'on peut suggérer pour répondre à des demandes localisées est de conserver au sein des services producteurs la matière première exhaustive en termes de données administratives pendant une certaine période (5 ans). Si une demande locale apparaît, et qu'elle est jugée recevable par les CRIES (ou leurs équivalents locaux), elle pourra bénéficier de l'infrastructure offerte par le CAPS pour fabriquer une base de sondage à partir des sources administratives locales. Chaque opération de cette nature sera soumise à la CNIL dans son intégralité (utilisation de sources administratives et collecte).

Repenser le dispositif des enquêtes ponctuelles en fonction des panels

Recommandation 5: Articuler dans la mesure du possible les enquêtes légères en coupe autour du dispositif longitudinal

Cette proposition prend le contre-pied de la situation actuelle, dans laquelle chaque enquête est lancée le plus souvent indépendamment des fichiers existants. Au contraire, on se propose ici de capitaliser de l'information sur les ménages en multipliant les visites. La critique selon laquelle la charge d'entretien repose toujours sur les mêmes unités doit être relativisée. D'une part, l'expérience tend à montrer qu'au bout d'un certain temps une « identité panel » se crée au sein des ménages, sous réserve que les responsables entretiennent un tel sentiment au travers d'une politique de communication adéquate¹³. En outre, il n'est pas forcément nécessaire de prévoir des périodes de suivi aussi longues que dans le cas du PSID, dans la mesure où les données administratives pourraient permettre de prolonger un certain temps sans entretien la durée de suivi, même si l'information est moins riche. Bien évidemment, le format en coupe instantanée doit être conservé si le thème ou le mode de questionnement le justifie (urgence, questions sensibles, enquêtes à carnets...).

¹³ Certains responsables organisent même des rassemblements amicaux des individus panel. Sans atteindre cet extrême, un retour d'information soigné (journal du panel,..) est la condition minimale pour que le sentiment d'attachement se développe.

Concrètement et sur le court terme,

Recommandation 6 : Articuler les parties variables des enquêtes PCV autour d'un panel SILC reconfiguré

Cette recommandation renvoie à la recommandation précédente, qui vise à utiliser le plus possible les panels permanents pour mettre en place des questionnements ponctuels. Elle est relativement aisée à mettre en œuvre puisqu'il s'agit d'opérations réalisées au sein du même service. Ici, malgré les efforts entrepris lors de la refonte du dispositif PCV, les risques de recouvrement entre les deux opérations demeurent. Une intégration des deux opérations serait source d'économies budgétaires, tout en permettant une accumulation de l'information sur différents thèmes économiques et sociaux¹⁴.

Recommandation 7 : Utiliser le Centre d'Appariement sécurisé pour compléter le dispositif statistique existant en coupe transversale.

La méthodologie développée dans le centre pour assurer les appariements entre fichiers peut être appliquée indépendamment de la dimension longitudinale. Un des avantages des fichiers administratifs étant leur exhaustivité, des appariements complets entre certains d'entre eux peuvent donner une image de la population dans son ensemble. Par exemple, l'appariement entre les fichiers fiscaux et sociaux est de nature à fournir une information annuelle sur la démographie des micro-zones géographiques, utile pour compléter les indicateurs issus du RRP ; un tel appariement peut également servir à connaître les inégalités de revenu et de niveau de vie à des niveaux géographiques fins (IRIS/ZUS), répondant ainsi à certains besoins de l'action régionale.

6.3. Valorisation des sources

Encourager les initiatives

Recommandation 8 : Encourager les initiatives existantes, en y associant le plus possible les acteurs du système statistique.

Il faut éviter que se reproduise la mésaventure connue par l'enquête « éducation » au début des années 90. Alors que tout était prévu pour procéder à une réinterrogation au bout de deux ans pour évaluer les efforts d'éducation des enfants, l'opération n'a pas été renouvelée faute de volonté (et de responsable). A l'heure actuelle, en ce qui concerne le dispositif statistique portant sur les enfants, il faut, comme on l'a vu, soutenir le projet porté par l'INED. Mais celui-ci ne portera cependant ses fruits qu'au bout de plusieurs années. Sur le court terme, la DEP dispose d'un panel d'élèves et d'étudiants, et cherche à procéder à des visites permettant de décrire le début de leur vie active. Procéder à une enquête par entretien permettrait d'évaluer pour la première fois la réussite des étudiants à la sorti du système scolaire en connaissant très finement leur parcours¹⁵.

¹⁴ Il faut cependant s'assurer au préalable, par exemple grâce à des tests, que les ménages sont prêts à accepter une telle charge de collecte. Le questionnement sur les revenus dans SILC pourrait néanmoins être allégé si on utilisait de une information de type administratif (fiscal, social...), plutôt que de demander au ménage de recopier ses propres documents.

¹⁵ Depuis la consultation des correspondants, un protocole de suivi a été élaboré avec le LASMAS. Celui-ci devrait être mis en place prochainement

Rendre accessible l'information sur les sources

Recommandation 9: Généraliser la constitution d'une documentation normalisée sur les sources (enquêtes,...) afin que celle-ci soit disponible en ligne.

Cette proposition devrait aller de soi indépendamment de toute dimension longitudinale. Force est néanmoins de constater qu'un interlocuteur étranger, même francophone, voire un interlocuteur français étranger au système statistique ne peut avoir connaissance des sources en ligne¹⁶, ce qui est le mode de recherche actuel. On pourrait utilement s'inspirer du modèle que représente le site de Statistique Canada.

Recommandation 10: A court terme, améliorer la documentation des enquêtes existantes, en commençant par les sources principales, et la mettre en ligne.

On peut penser ici au panel européen, aux enquêtes sur les budgets familiaux, ou aux enquêtes sur l'emploi. Les travaux actuellement entrepris dans le cadre du projet de répertoire des sources statistiques pourraient contribuer à la mise en œuvre de cette recommandation.

Recommandation 11: Effectuer une évaluation des enquêtes rétrospectives portant sur un horizon long.

Ces enquêtes nécessitent un effort de mémoire important de la part des enquêtés, et certaines réponses sont susceptibles de biais. Il serait nécessaire de confronter les résultats avec ceux que pourraient apporter d'autres sources pour s'assurer de la qualité des réponses, ou de réaliser des tests méthodologiques comparant les réponses avec ce que pourrait apporter une source administrative. Par exemple, dans le cas de l'enquête FQP, l'image qu'a une génération ancienne de son début de son cycle d'activité pourrait être comparée aux réponses que cette même génération fournissait à l'enquête sur l'emploi de l'époque. Ou encore, une enquête méthodologique pourrait être posée à une cohorte ancienne du panel DADS afin de comparer les réponses rétrospectives à la réalité du début des années 70.

On notera que ces dernières recommandations ne nécessitent pas un effort financier important de la part de l'Institut. Il s'agit davantage d'une rationalisation et d'une valorisation du dispositif existant.

6.4. Recommandations stratégiques

Recommandation 12: Améliorer la coordination entre producteurs de sources au sein du SSP.

La dissémination des services statistiques au sein des ministères présente des avantages certains en matière de positionnement de la statistique dans la fonction publique. La contrepartie est une multiplication de centres de décisions. Même si le CNIS joue bien son rôle de coordination, notamment au travers des interformations, les contacts paraissent insuffisants entre les producteurs de sources. Des relations bilatérales ont bien lieu entre les responsables, mais ceux-ci se réunissent rarement pour discuter des orientations générales. Un moyen d'améliorer la coordination serait de s'inspirer du fonctionnement du système statistique sur les entreprises (Comité plénier SSE et bureau) en créant d'une part un lieu d'échanges d'information et d'autre part une instance plus politique avec les hauts responsables.

¹⁶ A l'exception de l'INED, du CEREQ, et de quelques cas particuliers

Recommandation 13 : Avoir une stratégie commune des différents acteurs nationaux vis-à-vis des interlocuteurs communautaires

En matière de panels d'individus, et sans doute plus généralement, la France ne dispose pas d'une force de lobby équivalente à celle des pays anglo-saxons, voire d'Europe du Nord. Même si la France n'a pas cette culture, force est de reconnaître que les acteurs nationaux, statisticiens publics et chercheurs, ne présentent pas un front uni comme c'est notamment le cas en Grande-Bretagne. Or, l'INSEE à elle seule n'a sans doute pas la légitimité scientifique (nombre et qualité des publications,..) pour prétendre faire contrepoids au système anglo-saxons. Il serait sans doute opportun qu'un rapprochement plus étroit se mette en place, quitte pour l'Institut à mandater des universitaires pour représenter le système statistique, notamment en répondant de façon plus volontariste aux appels d'offre communautaires visant à faire évoluer le système. On peut ainsi s'étonner que l'ADEPS, qui a longtemps collaboré sur les panels avec la DR de Lorraine, ait été absente des discussions visant à mettre en place le système SILC.

Recommandation 14 : Développer une méthodologie du longitudinal

Lors des consultations, de nombreux interlocuteurs publics ont indiqué avoir souffert d'insuffisances méthodologiques en matière de constitution de panels, mais également de difficultés liées à l'exploitation. Ceci s'explique d'une part par la faible culture statistique française en matière de données longitudinales, mais aussi par les difficultés intrinsèques de mise en place et d'exploitation de telles sources : l'investissement méthodologique est plus intense que pour les coupes transversales. Si l'on envisage en France de donner aux panels une place équivalente à celle qu'ils occupent dans les autres pays, il faut effectivement se doter d'une structure qui capitalise le savoir-faire spécifique à ce type de source. La constitution d'un Centre de ressources méthodologique pour le système statistique public serait une solution.

Annexes

	Pages
Annexe 1 : Lettre de mission.....	29
Annexe 2 : Liste des personnes rencontrées.....	31
Annexe 3 : Schéma théorique du CAPS	(document séparé)
Annexe 4 : Données longitudinales disponibles en France.....	32
Annexe 5 : Données longitudinales à l'étranger.....	35

Lettre de mission

Dossier suivi par :
S. Lollivier
Tél. : 0141175155
Fax : 0141173853

Paris, le 11 décembre 2003
N° 486/F001

Objet : **Mission sur « les outils de suivi des trajectoires des personnes en matière sociale et d'emploi »**

Dans les domaines sociaux comme dans le domaine de l'emploi, les besoins de connaissance des trajectoires des personnes sont importants et vont vraisemblablement être appelés à croître dans les années à venir. Ces besoins sont exprimés tant par les acteurs sociaux que par les personnes en charge de l'évaluation des politiques publiques ou encore par le monde des études et de la recherche (statisticiens, démographes, économistes, sociologues, responsables de la santé...).

Les outils actuellement disponibles pour répondre à ces besoins de connaissance ont des limites, qui peuvent être jugées relativement importantes. Ces limites conduisent parfois à prôner la constitution en France d'un ou de plusieurs panels, permettant de suivre les trajectoires des personnes sur une longue durée. La constitution et le suivi au cours du temps de panels est cependant une opération lourde, tant d'un point de vue humain que financier. Il nécessite notamment un engagement fort dans la durée.

Le Conseil National de l'Information Statistique a commencé à organiser la réflexion sur l'intérêt et les limites des panels de personnes. Dans ce cadre, l'inter-formation « Statistiques sociales » du 17 octobre 2003 a demandé à ce qu'une mission d'experts puisse être constituée. C'est pourquoi je confie à Stéfan Lollivier, directeur de l'Ensaë, et à Mylène Chaleix, membre de l'Inspection Générale de l'Insee, une mission sur les outils de suivi des trajectoires des personnes en matière sociale et d'emploi.

Cette mission devra tout d'abord réaliser un **état des demandes** connues relatives au suivi des trajectoires des personnes. Elle devra préciser quels sont les thèmes pour lesquels existent de fortes demandes : l'enfance préscolaire ; la période de scolarité ; la fin des études et l'insertion professionnelle ; les trajectoires d'emploi ; les carrières, salariales et non salariales ; les revenus et les conditions de vie ; la santé ; le départ à la retraite ; la fin de vie, etc. Elle devra souligner si ces demandes relèvent du progrès général de la connaissance ou

sont directement liées à l'évaluation des politiques publiques. Elle devra analyser si elles sont circonstanciées ou pérennes, ainsi que leur degré d'urgence.

Cette mission devra ensuite **décrire les outils existants** à l'Insee, dans les services statistiques des ministères et plus généralement dans l'ensemble de la sphère publique (par exemple à l'Ined, au Cereq, à l'Inserm, dans les organismes de recherche). Ces outils peuvent être des panels mais aussi des fichiers qui ne sont pas des panels mais dans lesquels on dispose d'une information rétrospective. Ils peuvent être d'origine administrative ou issus d'enquêtes spécifiques. En s'appuyant sur les travaux existants, la mission comparera les outils disponibles en France avec ceux utilisés dans certains grands pays de l'Ocde. Elle décrira en particulier les outils internationaux dont une composante est française (tels que le projet Share).

La mission devra faire **un bilan de l'intérêt et des limites** des outils actuellement existants. Ce bilan prendra particulièrement en considération la qualité des données recueillies ainsi que leur précision et le délai d'obtention des résultats. Il tiendra compte aussi de l'intérêt des travaux déjà réalisés à l'aide de ces outils. Plus généralement, la mission pourra préciser les intérêts respectifs des approches en termes de cohorte ou de panel, les risques de biais dus à l'attrition, la taille et la durée optimale de suivi (selon les objectifs...).

S'il apparaît à la mission que les réponses du système statistique public aux besoins sont inadaptées, elle devra alors **faire des propositions** permettant d'améliorer les outils existants (éventuellement en mieux les valorisant) ou visant à les supprimer. Si elle le juge utile, elle pourra proposer la création de nouveaux outils. Dans ce cas, elle s'attachera à traiter les questions statistiques (qualité, précision, délais) ainsi que les questions juridiques, financières et administratives (qui conçoit l'opération, qui la réalise, quels moyens financiers et humains doivent être mobilisés, qui a accès aux résultats, etc.).

Pour cela la mission prendra tous les contacts qu'elle juge utile, tant à l'intérieur qu'à l'extérieur de l'Insee. Elle devra notamment rencontrer des chercheurs ayant à travailler sur les questions de trajectoires ainsi que des personnes ayant en charge la mise en œuvre ou l'évaluation des politiques publiques, et les responsables des bases de données statistiques. Elle pourra s'appuyer sur des contributions demandées aux services compétents de l'INSEE.

Le rapport me sera remis début juin 2004. Il sera présenté au Conseil National de l'Information Statistique.

Le directeur général de l'INSEE

Signé : Jean-Michel Charpin

Diffusion :

- Mme et MM. les membres du Comité de direction,
- M. le chef de l'Inspection générale,
- Mme et MM. les chefs de département de la DSDS,
- M. le directeur de l'Ensaef,
- Mme Chaleix.

Annexe 2 : liste des personnes rencontrées

- **INSEE-IG** : P. Ralle
- **DREES** : M. Elbaum, C.Cases, P. Cunéo, P. Pétour, N. Roth, P. Stroebel
- **DARES** : A. Fouquet, F.Brunet, C. Even, B. Sédillot, F. Vennat
- **INED** : F.Héran, H. Léridon, L. Toulemon
- **INSEE-DSDS** : M. Glaude, G. Desplanques, O. Marchand, F. Maurel
- **CERC** : M. Dollé
- **IRESO** : R. Silberman
- **INSEE-CREST** : S. Grégoir, B. Crépon, F. Kramatz, J.-M. Robin, S. Roux
- **INSEE-DDAR** : J-F. Royer
- **CAE** : L. Bloch
- **INSEE-CNIS** : J-P. Puig, C. Saglietti
- **INSEE-DSCRI** : J-P. Le Gléau, M. Isnard, G. Lang
- **DEP** : F. Rosenwald, J-P. Caille, S. Lemaire, M. Tomasini
- **CNAF** : V. Madelin, R. Mahieu
- **INSEE-D3E** : D. Blanchet

Contributions

- **CNRS - Nancy 2** : B. Jeandidier, J-C. Ray
- **GAINS** : Y. Guillotin, F. Langot, P-Y. Steunou
- **INSEE-DSDS** : D. Goux
- **Delta** : A. Clark, Sevestre
- **Ministère des Sports** : L. Muller
-

Annexe 4: données longitudinales disponibles en France

- OBTENUES PAR COLLECTE

- « PANEL DE MENAGES EUROPEEN » (EUROPEAN COMMUNITY HOUSEHOLD PANEL, ECHP) DANS SA VERSION EUROPEENNE (EUROSTAT) OU DANS SA VERSION FRANÇAISE (INSEE)
De 1994-2001, 8 vagues
- ENQUETE « SORTIES DU RMI » DE L'INSEE, LA DREES, LA CNAF, LA DIRMI, LE CSERC ET LA DARES
Une des sources les plus valorisées.
- ENQUETE « EMPLOI » DE L'INSEE (PANELISEE PAR TIERS SUR TROIS ANS), AVEC L'EXTENSION REVENUS FISCAUX.
Jusqu'en 2001.
- ENQUÊTE « EMPLOI EN CONTINU » DE L'INSEE (PANÉLISÉE SUR DIX-HUIT MOIS) DEPUIS 2002.
- ENQUÊTE « SUIVI DES CHÔMEURS » DE L'INSEE
- ENQUETE « TRAJECTOIRE DES DEMANDEURS D'EMPLOI » DE LA DARES : TDE-MLT 3 INTERROGATIONS ENTRE 1996 ET 1998
- ENQUÊTE « PERMANENTE AUPRÈS DES MÉNAGES SUR LES CONDITIONS DE VIE, PCV » (INSEE) (POUR SON TRONC COMMUN)
- ENQUÊTES « GÉNÉRATION » DU CEREQ
Suivi de jeunes sortant du système scolaire
- COHORTE EDEN DE L'INSERM
Suivi de femmes de la grossesse aux 5 ans de l'enfant.
- « PANEL DE MÉNAGES LORRAINS » (ENQUÊTE SOCIO-ÉCONOMIQUE AUPRÈS DES MÉNAGES LORRAINS, ESEML) DE LA DR INSEE LORRAINE ET L'ADEPS (UNIVERSITÉ NANCY 2 ET CNRS)
6 vagues de 1985 à 1990.

- **PANELS DES BENEFICIAIRES DE POLITIQUE D'EMPLOI (DARES)**
Panel 1ère génération en 1990 : panel téléphonique de jeunes sortant du système scolaire et panel « Chômeurs de longue durée »
Panel 2 dit « loi quinquennal » entre 1993 et 1995
Panel 3 actuel : 4 vagues d'enquêtes entre 2002 et 2005.
- **PANELS ÉDUCATION (DPD)**
Une longue tradition de panels d'élève, ces panels sont suivis jusqu'à la sortie du champ. Ils mélangent des données administratives et des enquêtes socio-démographiques dont les thèmes sont variables.
Sur les collégiens : le plus ancien a été mis en place en 1972, suivi de trois autres panels lancés en 1980, 1989 et 1995 (taille initiale de l'ordre de 20 000)
Sur les écoliers : en 1978 et 1997
D'autres travaux sur les étudiants (mise en place du système d'information SISE) ou sur les enseignants.
- **PROJET SHARE (SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE) - CREDES POUR LA FRANCE**
Projet multi-disciplinaire (santé, revenus d'activité et de retraite, consommation, épargne et patrimoine...). Cette enquête longitudinale pérenne sur un échantillon de personnes de plus de 50 ans se met en place dans 9 pays européens. (Enquête pilote en 2004)
- **PROJET GGS (GENERATION AND GENDER) - INED**
Enquêtes et données rétrospectives (2 ou 3 interrogations au minimum), projet porté par les Nations Unies, rassemble environ 25 pays.
- **PROJET PANEL D'ENFANTS PORTÉ PAR L'INED ET L'INSERM**
Réflexion en cours.

- OBTENUES PAR SOURCES ADMINISTRATIVES

- **PANEL ISSU DES DADS (INSEE)**
Depuis 1967, les données des Déclarations Annuelles de Données Sociales (DADS) des individus nés en octobre d'une année paire (1/25ème) sont conservées. Les données empilées depuis 1976, date de l'introduction du SIREN comme identifiant d'entreprise, ont permis de constituer un fichier organisant cette information sous forme de période d'emploi. Plus récemment, l'échantillon est passé au 1/12ème avec le traitement des années paires et impaires.
- **PANEL D'ALLOCATAIRES DE LA CNAF**
Construit en chaînant les échantillons nationaux des allocataires de la CNAF.
- **PANEL DES BENEFICIAIRES ET EX-BENEFICIAIRES DE L'ALLOCATION DE PARENT ISOLE DE MEURTHE-ET-MOSELLE DE L'ADEPS (UNIVERSITE NANCY 2 ET CNRS) ET LA CAF-54.**
Depuis 1991, données administratives de la CAF - 54 enrichies par une enquête postale pour une des cohortes.

- RELEVÉS DE MISSION DE TRAVAIL TEMPORAIRE (DARES)
([HTTP://WWW.TRAVAIL.GOUV.FR/PUBLICATIONS/PICTS/TITRES/TITRE2096/INTEGRAL/2004.02-08.1.PDF](http://www.travail.gouv.fr/publications/picts/titres/titre2096/integral/2004.02-08.1.pdf))
- DMMO-EMMO (DARES-INSEE)
([HTTP://WWW.TRAVAIL.GOUV.FR/ETUDES/PDF/DE72.PDF](http://www.travail.gouv.fr/etudes/pdf/de72.pdf))
- FICHER HISTORIQUE ANPE DES DEMANDEURS D'EMPLOI
- ÉCHANTILLON INTER REGIMES D'ALLOCATAIRES DE MINIMA SOCIAUX (DREES) : ENIAMS
Début 2002, données administratives issues de l'Insee (État-civil), de la MSA, de la CNAF et de l'UNEDIC, périodicité : annuelle, base de sondage de l'enquête auprès des bénéficiaires réalisée en 2003.
- ÉCHANTILLON INTER RÉGIMES DE COTISANTS (DREES) : EIC
Début : en cours, données administratives issues de l'Insee et des différents organismes d'assurance vieillesse et autres partenaires concernés, PERIODICITE : ANNUEL ?
- ÉCHANTILLON INTER-RÉGIMES DE RETRAITÉS (DREES) : EIR
Début : 1988, périodicité : tous les 4 ans.
- FICHER FAMEU (FICHER ANNUEL DES BENEFICIAIRES DES MESURES DE POLITIQUE D'EMPLOI ET DES ETABLISSEMENTS UTILISATEURS) –DARES :
Pourrait être utilisé de façon longitudinale.
- ÉCHANTILLON DÉMOGRAPHIQUE PERMANENT (EDP)-INSEE
Début : 1968, données administratives Insee (État-civil) enrichi par les données des recensements de la population (1968 à 1999), périodicité : permanente pour l'état civil.
- ENQUÊTE .DANS LES REGISTRES D'ÉTAT CIVIL (INED)
À partir des mentions en marge relevées dans les registres d'état civil.
- ENQUÊTE « TRA »
Comporte des données notariales et militaires.
POUR LES ENQUETES DE L'INED SE REPORTER AU SITE DU SERVICE DES ENQUETES DE L'INED ([WWW.INED.FR](http://www.ined.fr))
- PANEL DES MINEURS SUIVIS EN JUSTICE (MINISTÈRE DE LA JUSTICE)
Opération devant débuter en 2005, elle a pour but d'étudier les trajectoires judiciaires des mineurs.
- ECHANTILLON PERMANENT DES ASSURES SOCIAUX (EPAS) - CNAM ET CREDES
Échantillon au 1/6000è de la population couverte par le régime général de l'assurance maladie. Projet SNIIRAM extension de l'EPAS complété par des données d'enquête.

Annexe 5: données longitudinales à l'étranger

Allemagne

- GERMAN SOCIO-E CONOMIC PANEL –SOEP-
(<http://www.diw.de/english/sop/index.html>)

Australie

- LONGITUDINAL STUDY OF AUSTRALIAN CHILDREN
(<http://www.aifs.org.au/lisac/home.html>)
- AUSTRALIAN LONGITUDINAL SURVEY
(<http://assda.anu.edu.au/studies/D0491.html>)

Canada

- Banque de Données **administratives** longitudinales –DAL-, démographie et revenus (http://www.statcan.ca/francais/sdds/4107_f.htm)
- ENQUÊTE LONGITUDINALE NATIONALE SUR LES ENFANTS ET LES JEUNES –ELNEJ-
(http://www.statcan.ca/francais/sdds/4450_f.htm)
- Enquête nationale sur la santé de la population
(http://www.statcan.ca/francais/sdds/3236_f.htm)
- National Longitudinal Survey of Children and Youth –NLSCY -
(<http://www.hrdc-drhc.gc.ca/sp-ps/arb-dgra/conferences/nlscyconf/flyer-x.shtml>
)
- Étude longitudinale du développement des enfants du Québec - Eldeq -
(<http://www.grip.umontreal.ca/fr/programme/etudes/eldeq.html>)

Danemark

- DANISH NATIONAL BIRTH COHORT (<http://www.bsmb.dk/>)

États-Unis

- PANEL STUDY OF INCOME DYNAMICS (PSID ;
(<http://psidonline.isr.umich.edu/Guide/>)
- NATIONAL LONGITUDINAL SURVEY (NLSY 79, NLSY 97 ;
(<http://www.bls.gov/nls/home.htm>)
- WISCONSIN LONGITUDINAL STUDY (<http://dpls.dacc.wisc.edu/wls/index.html>)
- ADOLESCENT HEALTH STUDY (<http://www.cpc.unc.edu/projects/addhealth>)
- EARLY CHILDHOOD LONGITUDINAL STUDY, BIRTH COHORT
(<http://nces.ed.gov/ecls/>)
- EDUCATION LONGITUDINAL STUDY OF 2002 (<http://nces.ed.gov/surveys/els2002/>)
- NATIONAL EDUCATION LONGITUDINAL STUDY (<http://nces.ed.gov/surveys/nels88/>)

- NATIONAL LONGITUDINAL STUDY OF HIGH SCHOOL CLASS
(<http://nces.ed.gov/surveys/nls72/>)
- BEGINNING POSTSECONDARY STUDENTS (BPS) LONGITUDINAL STUDY
(<http://nces.ed.gov/surveys/bps/>)
- HOMELESS FAMILIES AND CHILDREN: A LONGITUDINAL STUDY
(<http://www.hmcnet.harvard.edu/psych/redbook/66.htm>)
- NATIONAL LONG-TERM CARE SURVEY (<http://nlctcs.cds.duke.edu/index.htm>)
- HEALTH AND RETIREMENT SURVEY (<http://hrsonline.isr.umich.edu/>)
- SURVEY OF INCOME AND PROGRAM PARTICIPATION
(<http://www.sipp.census.gov/sipp/>)
- NATIONAL SURVEY OF FAMILIES AND HOUSEHOLDS
(<http://www.ssc.wisc.edu/nsfh/home.htm>)
- CONSUMER EXPENDITURE SURVEY (CEX ; <http://www.bls.gov/cex/csxovr.htm>)

Luxembourg

- PANEL SOCIO-ÉCONOMIQUE "LIEWEN ZU LËTZEBUERG"
(<http://www.ceps.lu/psell/pselpres.htm>)

Norvège

- Mother and Child Cohort Study (MOBA)
(<http://www.fhi.no/tema/morogbarn/kortproteng.html>)

Pays-Bas

- LS-SEHD (Eindhoven) GLOBE Study
(<http://www.eur.nl/fgg/mgz/mgzprojects/prjct1tm54/project011.html>)
- DNB HOUSEHOLD SURVEY (<http://www.uvt.nl/centerdata/dhs/>)

Royaume-Uni

- BRITISH HOUSEHOLD PANEL SURVEY –BHPS-
(<http://iserwww.essex.ac.uk/ulsc/bhps/>)
- ENGLISH LONGITUDINAL SURVEY OF AGING (<http://www.natcen.ac.uk/elsa>)
- NATIONAL CHILD DEVELOPMENT STUDY (COHORTES 1958 ET 1970) :
<http://www.data-archive.ac.uk/findingData/bcsAbstract.asp>,
- THE MILLENIUM COHORT STUDY
(<http://www.cls.ioe.ac.uk/Cohort/MCS/mcsmain.htm>)
- ELSPAC (Avon) et ELSPAC (European) Study
(<http://www.alspac.bris.ac.uk/alspacext/MainProtocol/>)

Russie

- RUSSIA LONGITUDINAL MONITORING SURVEY (RLMS ;
<http://www.cpc.unc.edu/projects/rlms/project.html>)

Suisse

- PANEL SUISSE DE MÉNAGES (PSM ; <http://www.swisspanel.ch/>)

Union européenne

- ECHP : European Community Household Panel

- European Labour Force Survey (en panel dans tous les pays, mais pas avec les mêmes règles de rotation des échantillons dans tous les pays).
- Projet CHER (HBS for Poland, HHPS for Hungary, PSBH for Belgium ; <http://www.ceps.lu/CherPaco/CherPaco.htm>)