

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Sommaire

Sommaire

Finalité du traitement (annexe 6)	page 2
Fonctions de l'application (annexe 8)	page 4
Sécurité et secrets (annexe 10)	page 9
Catégories de données (annexe 11)	page 15
Catégories de destinataires (annexe 12)	page 17
Mesures prises pour informer les enquêtés de leurs droits (annexe 13) ...	page 20
Exercice des droits des enquêtés (annexe 14)	page 23
Avis de conformité (documentation complémentaire 1)	page 24
Note sur l'échantillonnage (documentation complémentaire 2)	page 27
Projet de convention Insee-Ined (documentation complémentaire 3)	page 36
Sommaire du questionnaire principal (documentation complémentaire 4) (le questionnaire est joint en document autonome)	page 48
Note sur les questionnaires « 15-24 ans » (documentation complémentaire 5) (les questionnaires sont joints en documents autonomes)	page 49

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 6 : Finalité du traitement

Contexte de l'enquête

L'enquête « trajectoires et origines » sur la diversité des populations en France se situe dans la continuité de l'enquête « mobilité géographique et insertion sociale » (MGIS) réalisée par l'Ined en 1992.

La plupart des enquêtes du système statistique comportent déjà des questions permettant de repérer les personnes immigrées et, dans certaines d'entre elles (Formation-Qualification professionnelle, Emploi, Logement), les descendants d'immigrés. Les données ainsi collectées sont toutefois insuffisamment détaillées et portent sur des effectifs trop restreints pour analyser précisément les trajectoires des immigrés et de leurs descendants selon leurs différentes origines.

De plus, les enquêtes existantes n'abordent pas l'ensemble des questions que pose l'intégration : les relations avec les autres, les pratiques sociales. Enfin, le processus d'intégration ne peut se mesurer que dans la durée et nécessite une approche longitudinale, d'où la nécessité d'une enquête spécifique.

La demande d'une nouvelle enquête sur l'immigration et l'intégration a été formulée par le Premier ministre en 2003, réitérée en 2004 par la Cour des comptes, et soutenue depuis plusieurs années par le ministère des Affaires Sociales.

Cette enquête prend place un an après celle de la Drees, ciblée sur les primo-arrivants, et se déroule la même année que le module *ad hoc* 2008 de l'enquête Emploi sur la « situation professionnelle des personnes nées à l'étranger ».

Les utilisateurs potentiels, outre l'Insee et l'Ined, sont multiples et pour la plupart impliqués dans le projet :

- Les délégations, comités interministériels ou organismes de conseil intervenant dans les domaines de l'immigration, de l'intégration ou de la lutte contre les discriminations. Le Haut conseil à l'intégration, la Haute autorité de lutte contre les discriminations et pour l'égalité, la délégation à la ville et au développement social urbain, l'agence nationale de Cohésion sociale ont participé à la mise en place du projet d'enquête ;
- Les administrations centrales de différents ministères, partenaires du projet : la direction de l'Animation de la Recherche et des Études Statistiques (Dares), la direction de la Recherche, des Études, de l'Évaluation et de la Statistique (Drees), la direction de la Population et des Migrations (DPM), la direction générale de l'habitat, de l'urbanisme et de la Construction (DGUHC) ;
- Les agences chargées de l'admission au séjour et de l'accueil des migrants ont également été consultées. L'Agence Nationale pour l'Emploi (ANPE) et l'Agence pour la Formation Professionnelle des Adultes (Afp) font partie du groupe d'élaboration de l'enquête.
- Les agences d'urbanismes, à commencer par l'Institut d'aménagement et d'urbanisme de la région Île-de-France (IAURIF) ;
- Les laboratoires universitaires spécialisés dans les questions d'immigration, d'intégration et de discrimination, qu'ils soient français ou étrangers.

Objectifs de l'enquête

L'enquête « trajectoires et origines » (TeO) porte sur la diversité des populations en France ; elle cherche à appréhender dans quelle mesure les origines (géographiques, nationales ou sociales) sont susceptibles de modifier par elles-mêmes les chances d'accès aux principaux biens qui définissent la place de chacun dans la société : le logement, l'éducation, la maîtrise de la langue, l'emploi, les services publics et prestations sociales, la planification familiale et la santé, les relations sociales, la nationalité, la citoyenneté...

Pour atteindre cet objectif, l'enquête TeO doit permettre, d'une part, d'étudier les conditions de vie des individus au moment de l'enquête et leurs trajectoires sociales et, d'autre part, de recueillir les

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 6 : Finalité du traitement

éléments relatifs à leur « origine » pour étudier ses interactions avec les autres facteurs de différenciation dans la société française (âge, sexe, catégorie sociale et niveau d'instruction, revenu et patrimoine, type d'habitat, etc.).

S'agissant des conditions de vie au moment de l'enquête, l'enquête examine la situation sociale des individus en décrivant leur situation à l'égard de l'école, de l'emploi, du logement, de la santé, de la vie familiale, de la sociabilité. Étudier les conditions de vie permet en effet de rendre compte de l'intégration sociale de chaque individu à la société française.

L'intégration étant conçue comme un processus de long terme, l'enquête s'intéresse également aux trajectoires sociales des individus. Des biographies migratoires et professionnelles sont donc intégralement collectées. Des questions spécifiques portent sur la jeunesse des personnes interrogées, une période potentiellement déterminante pour la trajectoire des personnes. Enfin, le questionnaire détaille certains événements de vie caractéristiques du « passage à l'âge adulte » (1^{ère} union, 1^{er} emploi durable, 1^{er} logement autonome ...).

Dans cette approche rétrospective, l'enquête s'attache à l'étude des stratégies d'ascension sociale, aux obstacles à l'égalité des chances (discriminations) et, de façon corollaire à celle des processus d'identification par soi-même et par autrui.

Les résultats de cette enquête feront l'objet d'une large diffusion, sous forme de publications (notamment dans les collections *Insee-Première* et *Population et Sociétés*) et de fichiers de données individuelles anonymisées à destination des partenaires, des chercheurs et du grand public.

La Direction des Statistiques Démographiques et Sociales (DSDS) de l'Institut National de la Statistique et des Études Économiques (Insee) et l'Institut National d'Études Démographiques (Ined) assurent conjointement la maîtrise d'ouvrage de l'enquête.

La Direction des Statistiques Démographiques et Sociales (DSDS) de l'Institut National de la Statistique et des Études Économiques (Insee) assure en outre la maîtrise d'œuvre.

Cette enquête n'a pas de caractère obligatoire.

Fondements juridiques

- Les données sont confidentielles conformément à la loi n°51-711 du 7 janvier 1951 ;
- Les modalités d'échantillonnage de l'enquête TeO ont fait l'objet d'une déclaration (récépissé Cnil N° 1173603, en date du 6 avril 2007) ;
- L'extension de saisie de données issues de la collecte 2007 du Recensement de la population, nécessaire au tirage de l'échantillon de l'enquête TeO a reçu un avis favorable de la Cnil (N° 874454 - 1^{ère} modification, en date du 24 avril 2007).
- Avis de conformité N° 309/D131 du 20 décembre 2007 modifiant l'avis de conformité N° 243/D131 du 1^{er} Juin 2007, délivré par le Comité du label après l'avis rendu par le Cnis en réunion plénière le 19 décembre 2007 (l'avis de conformité est joint en documentation complémentaire 1).

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 8 : Fonctions de l'application

L'enquête TeO se décompose en trois opérations distinctes :

1. L'enquête principale : une enquête en face-à-face auprès de 24 200 personnes ;
2. L'enquête « jeunes » : une enquête postale auprès des enfants des personnes précédemment interrogées ;
3. un ensemble d'enquêtes qualitatives complémentaires.

1. L'ENQUÊTE PRINCIPALE

1.1. L'échantillon

Le volet principal de l'enquête TeO concerne l'ensemble de la population française d'âge actif résidant en France métropolitaine¹ et vise un échantillon de 24 200 répondants pour 37 000 fiches-adresses (+ 3 000 fiches-adresses de réserve) correspondant à des logements ordinaires. Ce sont cependant des individus spécifiquement identifiés qui sont recherchés au sein de ces logements.

Ce volet de l'enquête portant de façon approfondie sur l'insertion sociale et professionnelle, le champ de la population enquêtée a été réduit aux personnes d'âge actif (18-59 ans), afin d'en privilégier la représentativité².

L'enquête se limite de façon pragmatique aux ménages ordinaires, bien qu'il existe des collectivités spécifiques aux immigrés : la collecte qui avait été réalisée dans les foyers pour l'enquête MGIS s'était avérée très difficile et la plupart des personnes résidant en collectivités n'appartiennent pas aux classes d'âge concernées par l'enquête.

L'échantillon du volet principal de l'enquête est composé de 5 sous-échantillons distinguant :

- les immigrés âgés de 18 à 59 ans (9 600 réponses attendues) ;
- les descendants d'immigrés âgés de 18 à 49 ans (9 600 réponses attendues)³ ;
- les personnes nées dans un département d'Outre-mer âgées de 18 à 59 ans (800 réponses attendues) ;
- les personnes dont au moins un parent est né dans un département d'Outre-mer âgées de 18 à 49 ans (800 réponses attendues) ;
- les personnes n'appartenant à aucun des groupes ci-dessus, échantillon dit « natifs descendants de natifs », âgées de 18 à 59 ans (3 400 réponses attendues).

La logique de l'échantillon est d'assurer une représentation minimale des groupes dont on pense qu'ils peuvent être soumis à des discriminations du fait de leur origine, quelle que soit leur nationalité. C'est la raison pour laquelle les populations originaires des DOM (les immigrants en métropole et leurs descendants nés en métropole) font l'objet d'une surreprésentation spécifique. C'est aussi la raison pour laquelle les descendants de migrants constituent un groupe à part, même s'ils sont très majoritairement Français.

Parmi les immigrés et descendants d'immigrés, il a été choisi de concilier une représentation d'ensemble de toutes les origines et la surreprésentation de certaines origines peu répandues

¹ En raison de ses spécificités, l'immigration dans les collectivités d'Outre-mer (il s'agit d'une immigration récente et d'origine différente de l'immigration en métropole) ne pourrait être abordée que dans le cadre d'enquêtes spécifiques.

² En 2003 a été réalisée une enquête « Passage à la retraite des immigrés » (PRI), qui traite spécifiquement de la situation des immigrés lorsqu'ils vieillissent en France. Une nouvelle investigation sur cette problématique de façon aussi rapprochée est donc inutile.

³ Pour des raisons techniques liées aux sources mobilisées pour l'échantillonnage la tranche d'âge est limitée à 18-49 ans pour les descendants d'immigrés et les descendants de natifs d'Outre-mer

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 8 : Fonctions de l'application

(Vietnam, Laos ou Cambodge, Turquie, Afrique centrale et Golfe de Guinée, Afrique sahélienne occidentale).

L'échantillon est réalisé conformément à la déclaration faite à la Cnil (récépissé N°1173603 en date du 6 avril 2007) à partir de la population recensée en janvier 2007. Pour constituer l'échantillon des descendants d'immigrés et de personnes nées dans un département d'Outre-mer, l'information disponible dans le recensement est insuffisante et doit être croisée avec celle issue d'autres sources, de manière à restreindre l'enquête aux personnes ayant au moins un parent immigré. Les sources mobilisées à cette fin sont l'échantillon démographique permanent et les bulletins de naissance de l'état civil consultés en mairie après autorisation des procureurs de la République.

Rappel des modalités de constitution de l'échantillon

L'échantillon de l'enquête sera constitué à partir des personnes qui ont été recensées en 2007, soit l'ensemble des personnes vivant dans les communes de moins de 10 000 habitants recensées en 2007 (une commune de moins de 10 000 habitants sur cinq) et environ 8% de la population vivant dans des communes de plus de 10 000 habitants.

Par ailleurs, la collecte de l'enquête est limitée aux zones recensées qui font partie de l'échantillon maître, soit la quasi-totalité des zones urbaines mais une partie seulement des zones rurales.

Pour obtenir un échantillon nominatif, une saisie spécifique des noms et prénoms a été effectuée de façon complémentaire à la saisie effectuée dans le cadre de l'échantillon démographique permanent (EDP), conformément à l'avis favorable rendu par la Cnil (N° 874454 1^{ère} modification en date du 24 avril 2007) ; cette saisie couvre près de 300 000 personnes recensées, nées un jour EDP⁴ ou entre le 20 et le 25 de chaque mois, résidant dans une zone couverte par l'enquête, et susceptibles d'entrer dans le champ de l'enquête (c'est-à-dire nées en métropole entre 1958 et 1990 pour celles susceptibles d'être descendant de domiens ou d'immigrés, nées entre 1948 et 1990 pour les immigrés potentiels, nés étrangers à l'étranger). Cette saisie permet de constituer la base de sondage au sein de laquelle l'échantillon de l'enquête sera tiré de façon aléatoire, groupe par groupe.

Parmi les personnes dont le nom et le prénom sont saisis, 55 000 sont immigrées⁵ ; parmi les fiches-adresses correspondant à ces personnes, 12 000 à 13 000 fiches-adresses seront tirées aléatoirement qui constitueront l'échantillon immigré.

De la même manière, l'échantillon de personnes nées dans un département d'Outre-mer peut être constitué directement à partir de cette saisie : on tirera aléatoirement près de 1 000 fiches-adresses parmi les 5 000 natifs d'Outre-mer identifiés⁴.

Par contre, dans l'ensemble des personnes nées en métropole dont le nom et le prénom ont été saisis, soit environ 240 000, une fraction seulement sera constituée de descendants d'immigrés ou de descendants de domiens ; il est impossible de les repérer à travers les seules informations qu'ils ont fournies lors du recensement, aucune information sur le lieu de naissance des parents n'étant renseignée. Il est indispensable pour constituer les échantillons de descendants de consulter les bulletins de naissance de l'état-civil.

Afin de limiter les recherches directes dans les registres des communes de naissance, l'Insee procède à une consultation préalable des bulletins anonymes de naissance qu'il collecte depuis 1968. Pour chacune des 240 000 personnes nées en métropole est recherché l'ensemble de ses correspondants « anonymes » (c'est-à-dire les personnes nées le même jour dans la même commune avec le même sexe). Si aucun de ces correspondants n'est descendant d'immigré ou de domien, il est inutile d'organiser une recherche en mairie. A l'opposé, si tous les correspondants sont descendants d'immigrés ou de domiens, il n'est pas nécessaire d'aller consulter l'état civil. Par ailleurs, cette opération préalable permet de surreprésenter certaines origines. Par exemple, la recherche en mairie sera effectuée systématiquement s'il existe au moins un descendant d'immigré(s) turc(s) parmi les correspondants. Elle ne sera effectuée que pour une fraction des personnes qui ont un correspondant ayant des parents immigrés algériens. Enfin, pour les personnes nées entre 1958 et 1967, aucune information préalable n'est disponible et la consultation des registres est systématique.

On estime à environ 70 000 le nombre de bulletins de naissance à vérifier dans les registres d'état civil. Cette opération sera effectuée par les agents des directions régionales de l'Insee après autorisation des procureurs de la République, conformément au test réalisé à l'été 2007 en Ile-de-France, en Aquitaine et en Lorraine.

Par la suite, l'échantillon sera constitué comme indiqué précédemment par tirage aléatoire au sein de chacun des cinq groupes de la base de sondage, de façon à obtenir le nombre de répondants visés pour chaque groupe.

⁴ Il s'agit des 1^{er}, 2, 3 et 4 de janvier, avril, juillet et octobre

⁵ Estimation fondée sur la proportion de personnes appartenant à cette catégorie dans l'ensemble de la population résidant en France

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 8 : Fonctions de l'application

1.2. Déroulement de l'enquête, information des personnes

L'enquête sera réalisée de septembre à décembre 2008 auprès des personnes constituant l'échantillon.

L'Insee assurera la collecte des données, par enquêteurs, accompagnés - si besoin - d'interprètes (afin de garantir la neutralité du questionnement que ne pourrait garantir l'intervention d'un proche de l'enquêté), en face à face, sur la base de questionnaires informatisés (CAPI)⁶.

La durée de passation du questionnaire est variable selon la maîtrise du français et la nécessité d'un interprète : en moyenne, elle s'établit à 1 heure 20 mais peut atteindre 2h00 et justifier que la réalisation de l'enquête se déroule en deux visites.

L'enquêté est informé de la réalisation de l'enquête par une lettre-avis et par une plaquette d'informations. Les modalités d'information des enquêtés sont précisées en annexe 13.

L'échantillon de l'enquête TeO étant constitué d'individus spécifiés et non de logements. Il est nécessaire d'assurer un suivi d'adresse en cas de déménagement depuis le recensement, afin que toute personne présente dans l'échantillon soit contactée pour réaliser l'enquête. Un protocole de recherche d'adresse est prévu dans les cas où la personne a déménagé. Les enquêteurs pourront s'adresser à la famille, aux voisins, à la mairie et consulter l'annuaire téléphonique. Si l'enquêteur obtient la nouvelle adresse de l'enquêté, il la reporte sur la Fiche-adresse qu'il retourne à sa direction régionale de l'Insee afin qu'elle assure le suivi permettant la réalisation de l'enquête. Pour affiner cette recherche, une consultation complémentaire de fichiers administratifs est envisagée sous réserve de faisabilité technique et d'accord des responsables de ces fichiers ; une confrontation d'adresses pourrait être effectuée a priori pour l'ensemble des enquêtés ; L'appariement - réalisé par l'Insee aurait pour seul objectif de permettre aux enquêteurs de disposer - avant la première prise de contact - des différentes adresses possibles des personnes qu'ils ont à interroger. En aucun cas les responsables des fichiers administratifs n'auraient accès aux résultats de ces appariements, ni détaillés, ni agrégés. La mise en œuvre de cette consultation fera l'objet - le cas échéant - de formalités préalables vis-à-vis de la Cnil en modification au présent dossier.

1.3. Questionnaire face à face

Le groupe de conception du questionnaire, constitué des membres de l'équipe de conception (Insee et Ined) ainsi que des chercheurs venant de différentes institutions ou administrations a élaboré un questionnaire commun aux cinq sous-échantillons de l'enquête. Des filtres intégrés dans le questionnaire permettent cependant d'éviter les questions qui ne seraient pas pertinentes pour certains individus : par exemple, on ne demande qu'aux descendants d'immigrés s'ils sont déjà allés dans le pays d'origine de leurs parents et seulement aux immigrés quel type de logement ils ont occupé à leur arrivée en France.

L'enquête TeO étant centrée sur les trajectoires sociales des individus et sur l'analyse des processus d'insertion, le questionnaire s'intéresse aux trajectoires individuelles et comprend des éléments rétrospectifs sur les parcours scolaires, professionnels, familiaux et résidentiels.

Le recueil des éléments constitutifs de l'origine, aussi bien géographique que sociale, culturelle ou résidentielle fait l'objet d'une attention particulière. Chaque sphère de la vie des enquêtés est approfondie dans un module spécial, qu'il s'agisse de l'éducation, du logement, de l'emploi, de la vie matrimoniale, de la vie citoyenne, des relations sociales.

À vocation multi-thématique, l'enquête TeO aborde donc les principales dimensions de l'intégration, conçue comme l'ensemble des chances d'accès aux diverses ressources de la vie sociale. Il s'agit de repérer les situations où s'expriment les différentiels d'opportunité et de traitement et où peuvent se manifester les stratégies collectives.

⁶ Les données sont chiffrées, l'accès à l'ordinateur s'effectuant par mot de passe et carte à puce.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des populations en France

Annexe 8 : Fonctions de l'application

L'identité, notamment dans sa relation aux origines des enquêtés, est abordée à travers différents aspects : transmission familiale des langues, religion, mais aussi perception par l'enquêté comme dans le regard des autres de son identité.

Le thème des discriminations sera abordé de façon transversale dans différents modules. Enfin, la localisation géographique des individus permettra de constituer une base de données contextuelles (permettant une caractérisation socio-démographique du voisinage) qui sera intégrée dans la base d'exploitation. Elle permettra d'enrichir l'information disponible sans alourdir le questionnaire.

En dehors du tableau des habitants du logement, le questionnaire comporte donc une quinzaine de modules thématiques dont un descriptif des principales variables figure en annexe 10 (catégories de données et destinataires)

Les modules du questionnaire en face à face concernent les thèmes suivants :

- Revenus
- Nationalité et origine des parents
- Langues
- Trajectoires migratoires, rapports au pays d'origine
- Relations familiales et transmissions
- Enfants
- Image de soi et regard des autres
- Éducation
- Vie professionnelle
- Religion
- Vie matrimoniale
- Logement et cadre de vie
- Vie citoyenne
- Santé
- Discriminations
- Relations sociales

La table des matières du questionnaire figure en documentation complémentaire n°4.

Le questionnaire est fourni en complément du présent dossier.

2. L'ENQUÊTE AUPRÈS DES JEUNES

2.1. Le champ et l'échantillon de l'enquête

Les modalités de la transmission familiale des valeurs et les parcours scolaires étant des éléments forts dans la formation des identités et dans l'intégration sociale, un volet complémentaire de l'enquête a été élaboré afin d'approfondir ces questions auprès de jeunes en cours d'études ou à la fin de leurs études.

Ce volet complémentaire concerne les enfants de 15 à 24 ans des personnes interrogées dans le cadre de l'enquête TeO, qui vivent avec leur parent enquêté ; il permettra notamment d'étudier les parcours scolaires des jeunes immigrés et descendants d'immigrés (y compris de troisième génération) en les confrontant aux parcours réalisés par leurs parents.

Ce volet complémentaire a reçu un avis d'opportunité et a fait l'objet d'un examen par le comité du label réuni le 17 décembre 2007 ; son inscription effective dans l'enquête TeO est couverte par l'avis de conformité définitif rendu par le Comité du label suite à la réunion plénière du Cnis du 19 décembre. Cet avis est joint au présent dossier (documentation complémentaire 1).

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 8 : Fonctions de l'application

2.2. Le questionnaire « jeunes »

Le volet complémentaire s'adresse aux enfants de 15 à 24 ans des personnes interrogées dans le cadre du volet principal de l'enquête TeO qui vivent au domicile de leur parent « enquêté TeO » ; deux questionnaires différents ont été élaborés, l'un est soumis aux « Étudiants, collégiens, lycéens », l'autre s'adresse aux jeunes « En emploi ou en recherche d'emploi »).

Il s'agit de questionnaires auto-administrés dont le protocole de mise à disposition est le suivant : si l'enquêté TeO a des enfants de 15 à 24 ans vivant à son domicile, et sous réserve (à l'issue de l'enquête en face à face) de son accord pour l'enquête auprès de ses enfants mineurs (le cas échéant), il lui est demandé de remettre à chacun de ses enfants concernés par l'enquête le questionnaire adéquat sur lequel aura été préalablement reporté le prénom de l'enfant ainsi que le N° de gestion de l'enquête en face à face TeO permettant de procéder ultérieurement au rapprochement des données. Les questionnaires lui sont remis avec une enveloppe T à destination de la direction régionale de l'Insee.

Les enfants majeurs disposent pleinement de tous leurs droits et en sont informés par une mention figurant sur le questionnaire ; l'enquête n'a pas de caractère obligatoire.

Les questionnaires présentés au comité du label du 17 décembre sont fournis en complément du présent dossier. Une note décrivant les modifications qui leur seront apportés conformément aux recommandations du comité du label figure en documentation complémentaire n°5.

3. LES ENQUÊTES QUALITATIVES COMPLÉMENTAIRES

Il est prévu dans un délai maximum de trois années après la fin de collecte de l'enquête TeO la réalisation de 20 post-enquêtes qualitatives non obligatoires maximum, portant chacune sur un échantillon restreint à une cinquantaine de personnes ayant donné leur accord lors de l'enquête TeO 2008 pour être à nouveau contactées.

Ces post-enquêtes auront pour principal objectif de recueillir des précisions nécessaires à l'exploitation de réponses apportées à des questions posées lors de l'enquête TeO. Ces post-enquêtes seront programmées par l'Ined sur appel à projets de recherche, et seront réalisées dans le cadre de conventions par des chercheurs mis à disposition de l'Insee (les projets seront présentés au comité du label du CNIS et feront l'objet de formalités préalables vis-à-vis de la Cnil en modification du présent dossier).

D'un point de vue pratique, les chercheurs indiqueront les critères de sélection des personnes à interroger (en fonction de certaines de leurs réponses à l'enquête). L'Insee sélectionnera des personnes correspondantes aux critères souhaités et transmettra ensuite aux chercheurs, mis à disposition de l'Insee avec signature d'un engagement de confidentialité, les fiches-adresses correspondantes. Le nombre de fiches-adresses transmis excédera le nombre de personnes à contacter pour compenser les pertes liés au refus ou au déménagement de certains enquêtés. Les chercheurs disposeront alors de trois mois à compter de la mise à disposition des fiches-adresses pour prendre contact avec les personnes, les interroger et retranscrire les entretiens. A l'issue, les fiches-adresses seront détruites et les supports conservant le contenu des entretiens seront expurgés des informations permettant une identification directe ou indirecte de la personne enquêtée.

L'ensemble des réponses recueillies aux différents questionnaires (face à face, « jeunes » ou de post-enquête) sont anonymes et confidentielles.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 10 : Sécurités et secrets

1. Mesures générales de sécurité et de confidentialité prises au niveau de l'Insee

La politique de sécurité interne de l'Insee (PSI) est définie dans un « cahier des règles et standards de sécurité » approuvé par le comité de direction et régulièrement mis à jour, notamment pour tenir compte des évolutions de la réglementation, des risques nouveaux pouvant apparaître ou du niveau de nos propres exigences. Les directeurs ont des responsabilités en matière de sécurité clairement définies. Ils s'appuient sur des conseillers locaux de sécurité qui ont suivi une formation spécifique et qui sont chargés de suivre cette mise en place dans ses détails techniques.

Outre les règles générales relatives à la sécurité des locaux et aux conditions d'utilisation des ressources informatiques, des règles spécifiques sont applicables en fonction de la sensibilité des informations traitées.

I - Sécurité

1.1 - Sécurité physique des bâtiments et des locaux

L'Insee, sur l'ensemble des sites dont il a la responsabilité, a mis en place une gestion de la sécurité adaptée à la nature des travaux effectués et orientée vers deux objectifs, la sécurité incendie et la protection contre les intrusions.

Sécurité incendie

Outre la satisfaction aux exigences liées à la réglementation existante, les locaux de l'Insee contenant des matériels informatiques ou des documents sensibles sont dotés de dispositifs de détection incendie. Les salles machines et les bandothèques des centres informatiques ainsi que les salles renfermant les serveurs de la direction générale et des directions régionales sont en outre équipées de systèmes d'extinction automatique.

Sécurité contre les intrusions

Les bâtiments sont placés, en dehors des périodes d'occupation, sous la protection de sociétés spécialisées dans la télésurveillance. Pendant les périodes d'ouverture des locaux, l'organisation des espaces est conçue de façon à permettre un contrôle de l'évolution du public dans les locaux, les zones sensibles sont munies de dispositifs de contrôle d'accès (badges, digicode ou serrures de sûreté).

Un soin particulier est apporté aux contrôles de l'accès aux salles machines. Celles des centres nationaux informatiques sont contrôlées strictement par badges avec journalisation des entrées et des sorties. Pour les salles des serveurs des directions régionales, des normes ont été édictées par le département de l'informatique.

1.2 Sécurité logique

Contrôle de l'accès logique au système informatique

La sécurité logique fait référence à des mesures de protection liées à l'accès aux ordinateurs, aux données et aux logiciels. Dans ce domaine l'Insee a défini dans le cahier des charges déjà cité un ensemble de règles pour se protéger contre l'accès abusif, limiter l'accès aux données et gérer les aspects relatifs à l'intégrité et à la sauvegarde des données.

Le contrôle d'accès sur les machines centrales sous MVS met en œuvre le logiciel ACF2. Les serveurs fonctionnent sous le système Windows Server 2003.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 10 : Sécurités et secrets

La gestion de la sécurité

Les responsables de la sécurité sont les chefs d'établissement. Ils sont assistés par des conseillers locaux de sécurité spécialement formés à ces questions. Chaque système est administré par un administrateur de la sécurité informatique (ASI) dont la liste est contrôlée ainsi que les droits et devoirs attachés à la fonction.

Les administrateurs de la sécurité logique conservent un historique non modifiable et non effaçable des tentatives de violation des règles fixées.

Sécurité de la saisie

L'enquête est réalisée sur le terrain à l'aide d'un questionnaire informatisé, selon la technique Capi (computer-assisted personal interviewing). Les enquêteurs disposent d'un microordinateur fourni par l'Insee leur donnant seulement accès à l'application Capi. Les données saisies sont chiffrées à la volée sur une partition du disque dur. Les enquêteurs disposent pour cela d'un certificat stocké sur une carte à puce protégée par un code confidentiel. Une procédure ad hoc a été définie pour permettre aux responsables désignés de l'Insee de pouvoir recouvrer les données déjà saisies en cas de perte de la carte à puce par l'enquêteur.

Sauvegarde des données

Des dispositions de sauvegarde sont en vigueur sur l'ensemble des centres informatiques et des serveurs. Ils correspondent notamment au besoin de se prémunir d'un sinistre majeur. Dans ce cas les copies de sauvegarde réalisées chaque mois sont stockées dans un lieu distinct de celui de leur production. Dans beaucoup d'autres cas, des sauvegardes dites « applicatives » sont effectuées pour faire face à une défaillance survenant sur une application particulière. Dans tous les cas, les droits d'accès établis pour les fichiers « actifs » s'appliquent pour leurs versions « sauvegardées ».

II - Confidentialité

Les informations individuelles contenues dans les fichiers sont confidentielles conformément à la loi n°51-711 du 7 janvier 1951.

Les règles de base font référence au premier alinéa de l'article 226-13 du code pénal ainsi qu'à l'article 26 de la loi N° 83-634 du 13 juillet 1983 portant droits et obligations des fonctionnaires.

Tous les fonctionnaires de l'Insee sont informés qu'ils sont astreints au secret professionnel conformément aux dispositions de l'article 10 de l'ordonnance N° 59-244 du 4 février 1959.

Les agents contractuels se sont engagés sur l'honneur à ne faire, pendant la durée du contrat comme après son expiration, aucune communication écrite ou verbale concernant les documents ou les résultats de travaux qu'ils auront exécutés, dirigés ou contrôlés sans y avoir été autorisé préalablement par le chef de service lui-même, sous les peines prévues par l'article 226-13 du code pénal.

Les données sont classifiées selon leur sensibilité en trois catégories : non confidentielles (NC), en accès restreint (AR) ou sous haute protection (HP). Les opérations classifiées HP doivent être dotées d'instructions particulières pour assurer la confidentialité. En particulier, les droits d'accès sont donnés individuellement et régulièrement vérifiés ; les données sont stockées sur des machines sécurisées et ne doivent être copiées ni sur un ordinateur portable ni sur un support amovible (à l'exception des copies de sauvegardes qui sont conservées dans des locaux sécurisés). D'autres mesures sont relatives au traitement des documents papier antérieurs ou postérieurs aux traitements informatiques (utilisation des déchiqueteurs, rangement dans des armoires fermées à clef, etc.).

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 10 : Sécurités et secrets

2. Mesures générales de sécurité et de confidentialité prises au niveau de l'Ined

ANNEXE SÉCURITÉS ¹



Déclarant

Numéro de dossier attribué par la CNIL si vous en disposez déjà
NOM ou raison sociale : **Institut National d'Études Démographiques**
N° SIREN ou SIRET : **180037020 00021**
Téléphone **+33 1 56 06 20 00**
Adresse électronique : **lamouche@ined.fr**

Dossier n°
Sigle : **Ined**

L'architecture informatique, les sécurités et sauvegardes

1. Description du système informatique. Il est constitué :

- d'un parc de micro-ordinateurs sans serveur central
- d'un mini/petit serveur d'entreprise
- d'un ensemble de serveurs au sein de l'organisme ou externalisés
- d'un gros ordinateur au sein de l'organisme ou externalisé
- par l'hébergement chez un fournisseur internet.
- Nom de l'hébergeur :
- autre architecture informatique :

Nom(s) du (des) fournisseur(s) et du (des) modèle(s) : **Dell, HP**

Nom(s) du (des) système(s) d'exploitation : **Linux Red Hat, Windows 2003 Server**

2. Nature du réseau informatique permettant les échanges d'informations en interne.

- aucun réseau (par ex. des micro-ordinateurs isolés)
- un réseau local d'entreprise. Nom (ex. Netware) :
- un serveur interne accessible de l'extérieur via internet
- un hébergement externe accessible via internet.
- un extranet mis en œuvre par un Réseau Privé Virtuel (RPV ou VPN en anglais).
Nom du dispositif technique ou du prestataire :
- des lignes privatives louées à un opérateur de télécommunication
- utilisation de technologies sans contact. Nom (ex. WiFi) :
- utilisation de postes de travail nomades (ex. micro-ordinateurs)
- autre type de réseau :

Nombre total de postes de travail : **290**

Éventuellement, nom(s) du (des) logiciel(s) réseau(x) ou du moniteur de télétraitement :

.....

3. En cas d'échanges d'informations avec des partenaires ou organismes extérieurs, préciser le(s) procédé(s) technique(s) utilisé(s) :

- support magnétique ou analogue (disque, bande, cd-rom, clé USB,..) : **CDRom**

Chiffrement : **OUI NON**

¹ Plusieurs cases peuvent être cochées en réponse à une question

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 10 : Sécurités et secrets

- messagerie internet. Chiffrement : OUI NON
- transfert de fichier par internet. Chiffrement : OUI ~~NON~~
- transfert via un réseau privatif. Nom éventuel du réseau :
- Chiffrement : OUI NON
- autre procédé :
- Chiffrement : OUI NON

4. Sécurité (protection) physique des locaux et équipements, sauvegarde du système informatique

- Décrire brièvement les dispositifs/procédures permettant d'assurer la sécurité physique des locaux et équipements informatiques (badge d'accès, gardiennage etc.) :

Le centre d'hébergement dispose d'un onduleur garantissant la continuité en alimentation électrique des serveurs, en cas de défaillance électrique.

Le site est gardé et surveillé 24 heures sur 24. Les accès au site sont contrôlés en permanence par une alarme. La salle informatique n'est accessible, via des badges sécurisés, qu'aux personnes accréditées.

- Mesures assurant la sauvegarde du système informatique
 - Type de support utilisé : **Disque dur, bandes LTO2**
 - Fréquence des sauvegardes : **Quotidienne**
 - Chiffrement des sauvegardes : OUI ~~NON~~
 - Lieu de stockage : **Coffre anti-feu fermé à clef et situé dans une salle indépendante de la salle hébergeant les serveurs**
- Protection supplémentaire du lieu de stockage des supports de sauvegarde. Préciser :
.....

5. Protection contre les intrusions extérieures utilisant le canal des réseaux informatiques.

Procédé(s) technique(s) utilisé(s) :

- un routeur. Nom : **CISCO**
- un pare-feu (firewall). Nom : **boîtier ASA**
- un système complet de détection d'intrusion (IDS). Nom : **SNORT**
- autre procédé :

6. Mesures destinées à assurer la confidentialité des données lors du développement de l'application informatique.

- Le développement de l'application s'effectue dans un environnement informatique distinct de celui de la production (par ex. sur des ordinateurs différents, dans des salles machine différentes)
- Le personnel affecté aux tâches de développement est distinct de celui assurant la gestion ou l'exploitation des équipements informatiques de production
- La mise au point des logiciels s'effectue sur des données fictives et non sur des données réelles
- Autres mesures destinées à protéger la confidentialité des données de production :
.....

7. Mesures destinées à assurer la confidentialité des données lors des opérations de maintenance des équipements informatiques

- Les interventions de maintenance des matériels sont enregistrées dans une main-courante
- Les interventions de maintenance des matériels par un sous-traitant se font en présence d'un informaticien de l'entreprise
- La télé-maintenance des matériels n'est pas autorisée
- Les supports de stockage envoyés à l'extérieur à fin de réparation font l'objet d'une procédure de protection particulière. Si oui, préciser laquelle :
- Les supports de stockage destinés à la destruction font l'objet d'une procédure de protection particulière. Si oui, faire une description

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 10 : Sécurités et secrets

8. Mesures destinées à assurer la confidentialité des données lors des opérations de maintenance des logiciels informatiques

- Les interventions de maintenance des logiciels dans l'environnement de production sont enregistrées dans une main-courante
- Les interventions de maintenance des logiciels de l'environnement de production se font sous le contrôle du chef d'exploitation en respectant une procédure spécifique
- La télé-maintenance des logiciels de l'environnement de production n'est pas autorisée
- Une procédure particulière est mise en œuvre dans le cas où une opération de maintenance logicielle nécessiterait un accès aux fichiers de données nominatives. Si oui, la décrire : **Dban méthode Gutmann**

Le logiciel d'application

9. Il met en œuvre :

- une base de données.(ou un logiciel de gestion d'un entrepôt de données).Nom : **MySQL 5.0**
- un (des) progiciel(s). Nom(s) :
- un infocentre. Nom :
- un logiciel d'analyse de données permettant des statistiques/profilages/segmentations
Nom :
- Autre :

10. Finalités mettant des procédés techniques particuliers

- carte à puce
- biométrie (voir également la rubrique 13)
- RFID (reconnaissance à distance par radio-fréquence)
- vidéo-surveillance
- autre :

11. Authentification/identification des personnes habilitées à accéder à l'application. Le contrôle d'accès se fait-il par :

- un mot de passe.

Préciser :

- s'il a une structure obligatoire (ex. alphanumérique, présence d'un caractère spécial...) **Séquence alphanumérique aléatoire générée par le système mélangeant numérique et alphanumérique**
- sa longueur minimale : **8 caractères**
- sa durée de vie avant changement obligatoire :
- s'il y a interdiction de réutiliser les n précédents mots de passe :
- s'il y a interdiction d'utiliser certains mots de passe (ex. date de naissance, prénom,..) : **Des consignes sont données régulièrement aux utilisateurs**
- s'il y a blocage automatique du terminal d'accès au bout d'un certain nombre d'essais infructueux (si oui, préciser ce nombre)
- des profils d'habilitation définissant pour chaque utilisateur les fonctions autorisées ou les catégories d'informations accessibles
- une carte à puce
- un dispositif biométrique (voir également la rubrique 13)
- autre
- Lors d'une connexion, des informations concernant la précédente connexion s'affichent sur le terminal (par ex. date, heure et identifiant de l'utilisateur)

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 11 : Catégories de données

Cette annexe complète le tableau de la rubrique 11 du formulaire de déclaration.

Elle donne quelques éléments de détail sur les données identifiantes (permettant d'identifier les personnes) et sur les variables appartenant à des domaines concernés par l'article 8.

1. Origine des données

Toutes les données issues de la collecte sont directement obtenues auprès des personnes enquêtées : le fichier de collecte est certes initialisé à partir de données issues du Recensement de la population et de l'État civil⁷, mais en cours d'entretien, ces données font l'objet d'une nouvelle collecte auprès de l'enquêté ; les données issues de cette collecte directe se substituent (normalement à l'identique) aux données d'initialisation.

Le fichier d'analyse sera uniquement enrichi de données contextuelles non identifiantes, relatives aux caractéristiques socio-démographiques du quartier de résidence.

2. Conservation des données

Les fiches-adresses sont détruites dans le mois suivant la validation du fichier détail complet à l'exception des fiches-adresses des personnes ayant accepté de participer à une post-enquête, qui pourront être conservées jusqu'à la réalisation des post-enquêtes, dans la limite de trois années après la réalisation de l'enquête TeO. Aucune donnée nominative n'étant saisie, la destruction des fiches-adresses éteint le droit d'accès et de rectification.

3. Données identifiantes et sensibles

L'enquête TeO traite notamment des informations individuelles suivantes :

A - Données d'identification

- Volet principal : le nom de l'enquêté figure exclusivement sur la fiche-adresse papier et n'est pas saisi lors de la collecte TeO ; sont saisis : le prénom, la date de naissance et le lieu de naissance,
- Volet complémentaire et autres membres du ménage : les prénoms figurant dans le tableau des habitants du logement et sur le questionnaire « jeunes » sont saisis, ainsi que la date de naissance et le lieu de naissance.

F - Adresse et caractéristiques du logement

L'adresse de l'enquêté (et des membres du ménage) figurant sur la fiche-adresse papier n'est pas saisie. Seule la commune de résidence est saisie.

L'adresse antérieure (dans la limite de 5 ans avant l'enquête) sera saisie pour permettre d'étudier la mobilité géographique.

K - données « sensibles » (article 8)

Données sur la religion

⁷ Il s'agit des variables suivantes :

- le prénom, le sexe, l'année de naissance, le numéro de fiche-adresse (qui viennent du Recensement de la population),
- le type d'échantillon selon les 5 sous-groupes, qui vient du croisement des données issues du Recensement de la population ou de l'échantillon démographique permanent avec les données d'État civil.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 11 : Catégories de données

Les questions relatives à la religion répondent à trois objectifs de recherche.

1. Étudier les pratiques religieuses et le sentiment d'appartenance à une religion.

Les questions REL1 à REL5 sont destinées à étudier la place de la religion dans la socialisation familiale des individus, tandis que les questions suivantes permettent d'examiner le rapport que les individus entretiennent actuellement à la religion. La mise en perspective de la place de la religion dans l'enfance de l'enquêté et de son rapport à la religion aujourd'hui permettra de décrire comment les pratiques religieuses se transmettent d'une génération à une autre et de repérer les changements éventuels d'affiliation religieuse qui peuvent survenir au moment d'un mariage par exemple. Possibilité est offerte aux enquêtés de se déclarer « sans religion » et il est demandé aux enquêtés de préciser quelle est leur religion lorsqu'ils en ont une (REL7). La réponse est recueillie « en clair » de façon à laisser les enquêtés définir eux-mêmes leur religion, et les modalités « refuse de répondre » ou « ne sais pas » sont lues, conformément aux recommandations du comité du label. Un recodage des réponses en catégories agrégées sera opérée de façon à empêcher toute identification indirecte des personnes lorsque cela s'avérera nécessaire.

2. Identifier les supports d'une possible discrimination en raison de leur religion

REL10 demande aux individus de préciser s'ils portent des vêtements, bijoux ou autres accessoires permettant d'identifier immédiatement leur religion. Etant entendu que les individus peuvent subir des discriminations en raison de leur religion, spécialement lorsque celle-ci est facilement identifiable, cette information est rendue nécessaire pour confronter les réponses des enquêtés aux questions sur les discriminations, où l'on demande de préciser si la religion est la raison pour laquelle ils ont été discriminés (cf. DIS2).

3. Étudier l'impact de l'importance accordée à la religion sur d'autres pratiques

De nombreux travaux ont montré combien la religion constitue un facteur explicatif à d'autres pratiques. Les choix matrimoniaux (choix du conjoint et décision de se marier ou non), le désir d'enfant, les pratiques contraceptives sont très liées à l'importance que les individus accordent à la religion. En cela les réponses obtenues aux questions sur les pratiques religieuses constituent des déterminants indispensables à l'enquête. L'identification de la religion de l'enquêté et de celle de son conjoint (VM17) permettra d'établir une mesure de la part respective de la mixité et de l'endogamie religieuse au sein des couples.

Données relatives aux opinions politiques

Deux questions relatives aux opinions politiques visent à mesurer l'intérêt des enquêtés pour la politique et les opinions :

La première permet de graduer l'intérêt porté à la politique locale (échelon communal), à la politique nationale et internationale de la France et à la politique du Dom ou du pays d'origine (de l'enquêté ou de ses parents).

La deuxième permet à l'enquêté d'indiquer - s'il le souhaite - s'il a (ou non) une sensibilité politique et le cas échéant de la préciser sur l'échelle « très à gauche », « plutôt à gauche », « au centre », « plutôt à droite », « très à droite ». L'enquêté a la possibilité d'utiliser les modalités « refuse de répondre » ou « ne sait pas ».

M - Santé, données génétiques, vie sexuelle

Le module consacré à la santé aborde brièvement l'état de santé de l'enquêté sans permettre d'identifier aucune pathologie ; il porte surtout sur l'accès et le recours aux soins, ainsi qu'au renoncement à ces soins. Il comporte la question transversale des discriminations ressenties dans le domaine de la santé. Des questions relatives au désir d'enfants et à l'usage de méthodes contraceptives sont incluses. Elles visent à établir s'il existe un besoin d'information des populations migrantes relativement à la conception. Ces données constituent de surcroît des indicateurs d'intégration.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 12 : Catégories des
destinataires

Plusieurs types de fichiers seront issus de l'enquête TeO :

1. Les fichiers bruts de collecte (fichier des résultats et fichier des informations de collecte),
2. Les fichiers de diffusion (fichier « production et recherche », fichiers « d'études » et fichiers « grands publics »)

L'Ined, l'Insee et les Archives de France sont les seuls destinataires des fichiers bruts de collecte de l'enquête.

L'Insee est responsable de la transmission aux Archives de France du fichier brut de collecte et du fichier de production et de recherche communs à l'Insee et l'Ined.

1) Fichiers bruts de collecte

Seuls les dépositaires des fichiers bruts ont accès à ces fichiers bruts, ainsi qu'aux fichiers qui en sont issus en totalité, ou pour partie, dès lors qu'ils comportent des données à caractère personnel couvertes par le secret statistique, désignées comme informations « individuelles ».

Les différentes données collectées seront conservées sous forme de tables, ou fichiers distincts, raccordables par un identifiant « enquêté » non significatif, avec un fichier de centralisation. Chacun de ces fichiers fait l'objet de mesures de conservation (lieu, durée) et de droits d'accès spécifiques

1.1. Fichier échantillon

Destinataires : Insee

Format : fichier au format SAS

Remarque sur les utilisations : accès par la maîtrise d'ouvrage Insee, pour construire les fichiers de diffusion (en particulier enrichissement à partir de l'Iris)

Observation : l'individu

Contenu :

- identifiant non significatif au niveau de l'individu.
- localisation à l'Iris du lieu de résidence
- nom, prénom, date de naissance (variables issues de la saisie spécifique du Recensement de la population, supprimées après impression des fiches-adresses)
- variables issues de la collecte annuelle du Recensement de la population conformément au contenu de l'échantillon maître ; le fichier ayant été constitué avant codification de la CPS, les variables de situation vis-à-vis de l'emploi collectées lors du RP sont intégrées au fichier-échantillon
- pays de naissance des parents (variable issue de la recherche dans l'état civil, supprimée après impression des fiches-adresses)
- pondérations initiales

Lieu de conservation : Insee

Durée de conservation : longue durée

1.2. Fichier des résultats de l'enquête

Destinataires : Insee, Ined

Format : fichier au format SAS

Remarque sur les utilisations : accès par la maîtrise d'ouvrage Insee, Ined pour construire les fichiers de diffusion

Observation : l'individu

Contenu :

- identifiant non significatif au niveau de l'individu.
- Ensemble des variables de l'enquête TeO
- Commune de résidence
- Données contextuelles localisées
- Pondérations

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 12 : Catégories des
destinataires

Lieu de conservation : Insee, Ined
Durée de conservation : longue durée

1.3. Fichier des informations de collecte

Destinataires : Insee, Ined
Format : fichier au format SAS
Remarque sur les utilisations : maîtrise d'ouvrage Insee, Ined
Observation : l'individu
Contenu :

- identifiant non signifiant au niveau de l'individu
- État de la fiche adresse (faite /refus/abandon)
- Nombre de jeunes interrogeables
- Nombre de jeunes interrogés
- Date d'état (date de réalisation si fait, d'abandon ou de refus sinon)
- Nombre de visites
- Heure de début et durée pour chaque visite et durée totale.
- Identifiant non signifiant de l'enquêteur.
- Résultat de collecte (acceptation ou échec et sa cause)
- Une variable commentaire regroupant l'ensemble des commentaires éventuels
- Région de la résidence du ménage.
- Type d'espace de la commune de résidence (zonage en aire urbaine)
- Accord pour post-enquête

Lieu de conservation : Insee, Ined
Durée de conservation : longue durée

2) Fichiers de diffusion

L'Insee et l'Ined sont co-dépositaires des fichiers bruts et responsables de la constitution des fichiers de diffusion au sens de la loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques.

Le contenu précis des fichiers de diffusion ne peut être décrit précisément avant d'avoir expertisé le contenu des fichiers bruts de collecte. Les fichiers de diffusion seront élaborés par un groupe dit de « constitution de fichier », composé de chargés d'études et de chercheurs désignés par la Maîtrise d'ouvrage.

Trois types de fichiers de diffusion seront produits : fichier « production et recherche », fichiers « d'études » et fichiers « grands publics »

2.1. Le fichier de « production et de recherche »

Le fichier de « production et de recherche », destiné aux services statistiques ministériels et aux chercheurs, est construit à partir des fichiers bruts des enquêtes après contrôle des données, recodifications et traitement de la non-réponse. Il ne contient aucune information directement nominative. Il comporte l'ensemble des réponses au questionnaire de l'enquête dès lors qu'elles ne constituent pas des informations individuelles par recoupement de variables indirectement nominatives. Pour cela, des suppressions de variables ou regroupements de modalités sont effectués.

Le fichier de production et de recherche de l'enquête TeO est constitué à partir du fichier brut de l'enquête en face à face, enrichi des données des questionnaires jeunes. Il sera constitué dans un délai de un an après la fin de la collecte. Cette période correspond à la phase d'apurement et d'anonymisation du fichier.

Les modalités d'accès au fichier de production et de recherche par les chercheurs dépendront de la nature (« sensible » ou non) des données mise en œuvre :

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 12 : Catégories des
destinataires

Le fichier de production et de recherche sera rendu disponible aux chercheurs par une habilitation donnée par le réseau Quêtelet, avec lequel chacune des parties a passé une convention. Le réseau Quêtelet s'assurera notamment du statut réel de chercheur de la personne qui demande l'accès au fichier, du bien-fondé du projet de recherche scientifique présenté et de la nécessité d'avoir recours au fichier de production et de recherche pour mener à bien ce projet. Lorsque l'accord est donné, une licence d'usage spécifiant le sujet de la recherche et le fichier de production et de recherche sollicité est signée par le chercheur et visée par le directeur du laboratoire.

L'accès aux données sensibles du fichier de production et de recherche ne sera possible que si le projet de recherche le nécessite et après avis du Conseil scientifique du Comité de concertation pour les données en sciences humaines et sociales.

Destinataires : chargés d'études ou chercheurs à l'Insee et à l'Ined, Services statistiques des ministères, chercheurs sous convention sur projet passé avec l'Insee ou l'Ined et via le Réseau Quêtelet

Format : table(s) au format SAS.

Remarque sur les utilisations : recherches et études

Observation : l'individu

Contenu : Fichier complet des résultats de l'enquête TeO (y compris l'enquête auprès des jeunes), après correction de la non réponse et retraitement (suppression de toute information directement ou indirectement nominative)

Lieu de conservation : Insee, Ined

Durée de conservation : longue durée

2.2. Les fichiers « d'études »

Des fichiers « d'études » seront construits à la demande, à partir du fichier de production et de recherche. Ils comportent les informations utiles aux objectifs d'études. Pour cela, des suppressions de variables ou des regroupements de modalités peuvent être effectués.

Dans le cas présent, les fichiers d'études mis à la disposition des organismes partenaires seront identiques aux fichiers de production et de recherche, le champ des études intéressant chacun des partenaires couvrant un grand nombre des problématiques abordées par l'enquête.

Destinataires : financeurs pour leur besoin de recherche propre, autres demandeurs de fichiers « à façon »

Format : table(s) au format SAS.

Remarque sur les utilisations : recherches et études

Observation : l'individu

Contenu : identique au fichier de production et de recherche pour les organismes financeurs, sur mesure pour les autres demandeurs

Lieu de conservation : Insee, Ined

Durée de conservation : longue durée

2.3. Les fichiers « grand public »

Des fichiers de diffusion « grand public » seront construits à partir du fichier de « production et de recherche ». Les fichiers « grand public » sont obtenus par suppression des informations dont le traitement nécessite des précautions ou des compétences particulières et, éventuellement, par de nouveaux regroupements de modalités. Ces fichiers ne comprendront aucune donnée relevant de l'article 8.

Les fichiers grands publics sont accessibles sur les sites internet de l'Insee, de l'Ined et des organismes partenaires.

Destinataires : grand public

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

**ANNEXE 13 : Mesures prises pour
informer les intéressés de leurs droits**

Les enquêtés seront informés grâce à deux supports distincts :

1. Tous les enquêtés recevront une lettre-avis (ci-jointe) avant la réalisation de l'enquête. Cette lettre contenant l'information légale sur l'enquête est systématiquement accompagnée d'une traduction en 6 langues (allemand, anglais, arabe, chinois, portugais, turc) ;
2. Une plaquette d'informations accompagne cet envoi ou est remise par les enquêteurs au moment de l'entretien (ci-jointe) ;

Lettre-avis

NB : Cette lettre est systématiquement accompagnée d'une traduction en 6 langues (allemand, anglais, arabe, chinois, portugais, turc)

En tête de la direction régionale.

A l'attention de M/Mme.....

Objet : Etude statistique sur la diversité des populations en France

Madame, Mademoiselle, Monsieur,

L'Institut National de la Statistique et des Etudes Economiques (Insee) et l'Institut National des Etudes Démographiques (Ined) préparent une importante étude sur la diversité des populations, notamment en fonction de leurs origines. Pour cela, une enquête statistique est réalisée auprès de personnes résidant en France.

Cette étude vise à étudier l'accès au logement, à la santé, à l'éducation, à l'emploi des personnes selon leur histoire personnelle, leur mobilité géographique, leur origine sociale et celles de leurs parents.

A cet effet, un échantillon de 37 000 personnes a été tiré au hasard sur l'ensemble du territoire métropolitain. Vous êtes l'une d'entre elles. Madame / Monsieur....., enquêtrice / enquêteur de l'Insee, prendra prochainement contact avec vous. Elle / il sera munie / muni d'une carte officielle l'accréditant.

Votre participation à cette enquête est importante car l'échantillon auquel vous appartenez est représentatif de la diversité des situations de la population.

Pour faciliter l'entretien, il serait souhaitable qu'elle / il puisse vous rencontrer seul(e) pour vous interroger sur les thèmes abordés dans l'enquête.

Comme la loi en fait la plus stricte obligation, soyez assuré(e)s que vos réponses resteront strictement anonymes et confidentielles. Elles serviront uniquement à l'établissement de statistiques.

Je suis persuadé que vous aurez à cœur de faciliter la tâche de ma collaboratrice / mon collaborateur et que vous lui réserverez le meilleur accueil.

D'avance je vous remercie pour votre participation et vous prie d'agréer Madame, Mademoiselle, Monsieur, l'expression de ma considération distinguée.

Le Directeur Régional

Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête est reconnue **d'intérêt général et de qualité statistique**, sans avoir de caractère obligatoire.

Label n°243/D131 du Ministre de l'économie, des finances et de l'emploi, valable pour l'année 2008.

En application de la loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques, les réponses à ce questionnaire sont protégées par le secret statistique et destinées à l'Insee.

La loi n° 78-17 du 6 janvier 1978 modifiée, relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites à la présente enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Ce droit peut être exercé auprès des directions régionales de l'Insee.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

ANNEXE 13 : Mesures prises pour
informer les intéressés de leurs droits

Quelles sont
les **missions**
de l'**Ined** et de l'**Insee** ?

● **Les missions de l'Ined**

L'Institut national d'études démographiques (INED) est un établissement public de recherche scientifique. Il étudie la situation des populations. Il suit l'évolution des naissances, des décès, des unions, des séparations, des migrations, etc. et informe le grand public sur les questions de démographie.

● **Les missions de l'Insee**

L'Institut national de la statistique et des études économiques (INSEE) est un organisme public chargé de collecter, traiter, analyser et diffuser l'information statistique à caractère économique, démographique ou social. Pour remplir sa mission statistique en toute indépendance, il dispose d'une large autonomie vis-à-vis des autres administrations.

En aucun cas, l'Insee n'a le droit de communiquer les renseignements individuels obtenus lors d'une enquête statistique.

(loi n° 51-711 du 7 juin 1951, sur l'obligation, la coordination et le secret en matière statistique).

À quoi servent
les informations
collectées par l'Insee et l'Ined ?

Les enquêtes de l'Insee et de l'Ined servent à alimenter des études économiques et sociales qui intéressent un large public.

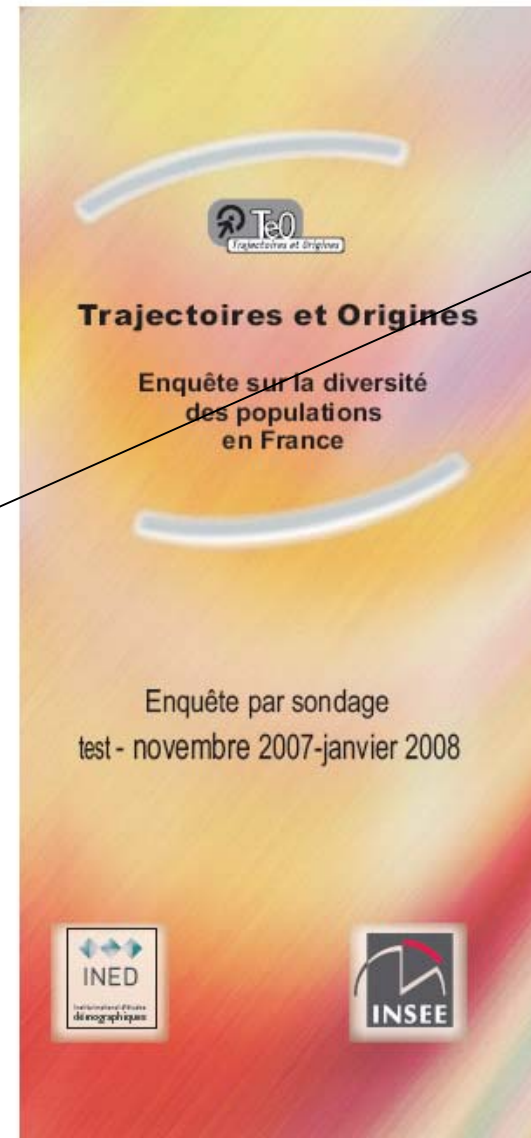
Chaque année, 250 000 questions sont adressées à l'Insee.

Elles émanent d'horizons les plus divers :
journalistes, entreprises,
enseignants, chercheurs,
pouvoirs publics, collectivités locales,
administrations ou particuliers.

Pour toute demande d'information d'ordre économique et social, vous pouvez :

- ▶ **téléphoner** à Insee contact : 0 825 889 452 (0,15€ la minute) du lundi au vendredi de 9h à 17h
- ▶ **déposer un courriel** sur le site internet de l'Insee : www.insee.fr cliquer sur « contact »
- ▶ **consulter** les sites internet de l'Insee : www.insee.fr de l'Ined : www.ined.fr
- ▶ Et pour plus d'informations, vous pouvez aussi consulter le site de l'enquête : <http://iteo.site.ined.fr>

▶ **Merci pour votre participation**



La version définitive de la
plaquette intégrera les
modifications suivantes :

1) dans le paragraphe
concernant le secret
statistique, on indiquera
que « l'Insee et l'Ined
n'ont le droit... »

2) à la suite du paragraphe
précédent sera ajouté le
paragraphe suivant :
« conformément à la loi
relative à l'Informatique,
aux fichiers et aux
libertés, cette enquête à
fait l'objet des formalités
préalables auprès de la
Cnil (loi n° 78-17 du 6
janvier 1978 modifiée) »

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des populations en France

ANNEXE 13 : Mesures prises pour informer les intéressés de leurs droits

● Pourquoi cette enquête ?

Il existe un décalage important entre la place des questions d'intégration et de discrimination dans le débat public et les données statistiques disponibles sur ces questions. L'enquête entend apporter un éclairage scientifique sur ce débat, en décrivant les conditions de vie et les parcours de vie des personnes immigrées, descendantes d'immigrés, des originaires des DOM et descendantes d'originaires des DOM, et en situant leur parcours par rapport à celui de personnes nées en France métropolitaine dont les parents sont aussi nés en France métropolitaine.

Le questionnaire parle de la scolarité, de l'emploi, de la vie familiale, des langues parlées, des conditions de logement, des pratiques religieuses, de la participation civique, de l'état de santé, de l'expérience du racisme, du rapport aux origines et du regard des autres.

● Qui est concerné par l'enquête ?

Toutes les personnes résidant en France métropolitaine et âgées de 18 à 59 ans peuvent répondre à l'enquête. C'est pourquoi celle-ci s'appelle « Trajectoires et Origines » : enquête sur la diversité des populations en France ».

Quels que soient vos origines ou votre lieu de naissance, vous pouvez être interrogé.

● Pourquoi vous ?

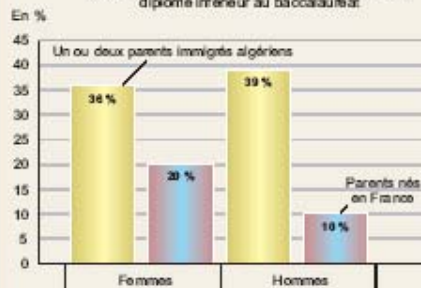
Afin de disposer de résultats représentatifs de l'ensemble de la population, il est inutile d'interroger chaque individu : ce serait coûteux et long. C'est pourquoi l'Insee réalise des sondages. Vous avez été sélectionné au hasard avec 1 000 personnes. Ces personnes représentent l'ensemble de la population.

● Quelques résultats en 1992

En 1992, une enquête similaire, appelée Mobilité géographique et insertion sociale (MGIS) a été conduite par l'Ined et l'Insee. Par exemple, à niveau de diplôme égal, parmi les jeunes hommes âgés de 20 à 29 ans sur le marché du travail, nés en France, ceux de parents algériens, étaient plus souvent au chômage que ceux de parents nés en France. Ce résultat fut l'un des premiers à révéler statistiquement l'existence et l'ampleur d'inégalités de traitement fondées sur l'origine dans l'accès à l'emploi.

Les jeunes de parents immigrés algériens plus touchés par le chômage

Taux de chômage des personnes de 20 à 29 ans, diplôme inférieur au baccalauréat



Source : Enquête MGIS, Ined, 1992

● Comment va se dérouler l'enquête ?

La collecte se déroule de fin novembre 2007 à début janvier 2008. Un enquêteur vous contacte pour obtenir un rendez-vous ou réalise immédiatement l'enquête, si vous êtes disponible lors de sa première visite. Il sera muni d'une carte officielle (spécimen ci-dessous) et tenu au secret professionnel. Il réalise l'enquête sur micro-ordinateur portable.



Si vous souhaitez avoir plus de renseignements, vous pouvez consulter le site internet de l'enquête ou joindre le responsable de l'enquête, dans votre région. Son numéro de téléphone est précisé sur la lettre jointe à ce dépliant.

● Vos réponses sont-elles anonymes ?

Oui. Vos nom et adresse ne servent que pour la collecte et son contrôle. Ils sont détruits quelques semaines après. Les autres informations recueillies sont conservées séparément, sans lien avec vos noms et adresses, de façon parfaitement anonyme. Aucune donnée nominative ne sera jamais transmise par l'Insee à un autre organisme, administration fiscale notamment.

La version définitive de la plaquette intégrera la modification suivante :

dans le paragraphe listant les thèmes abordés par le questionnaire, on parlera des « discriminations ressenties », plutôt que d'« expérience du racisme ».

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Annexe 14 : exercice des droits des
enquêtés

1. Droit d'opposition

L'enquête n'a pas de caractère obligatoire.

En outre, s'agissant des questions « sensibles » portant sur les origines ou sur les opinions religieuses, il est systématiquement rappelé aux enquêtés qu'ils ont la possibilité de refuser de répondre. Ce rappel s'effectue par deux moyens :

- par la lecture d'une phrase introductive annonçant le sujet qui va être abordé et le droit de refuser dont bénéficie l'enquêté, avant même que la question soit posée ;
- par la lecture systématique de la modalité « refuse de répondre » à chaque question « sensible », y compris pour les questions qui appellent des réponses en clair.

Ces mesures exceptionnelles dans une enquête sociodémographique ont été prises à la demande du Comité du label (Cf. avis de conformité, figurant en documentation complémentaire 1).

2. droit d'accès

Le droit d'accès des enquêtés sera exercé auprès des directions régionales de l'Insee.

Le droit d'accès aux données à caractère personnel concernant un mineur peut être exercé par son représentant légal.

Les fiches-adresses étant le seul moyen d'identifier les enquêtés, le droit d'accès s'exerce jusqu'à la date de destruction des fiches-adresses (au plus tard 3 mois après la validation du fichier brut de collecte).

3. droit de rectification

Le droit de rectification des enquêtés sera exercé auprès des directions régionales de l'Insee.

Le droit de rectification sur des données à caractère personnel concernant un mineur peut être exercé par son représentant légal.

Les fiches-adresses étant le seul moyen d'identifier les enquêtés, le droit de rectification s'exerce jusqu'à la date de destruction des fiches-adresses (au plus tard 3 mois après la validation du fichier brut de collecte).

L'Insee transmettra à l'Ined le fichier comportant les modifications, que ce dernier mettra en œuvre sans délai.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 1 :
avis de conformité (label)

Avis du Cnis (réunion plénière du 19 décembre 2007)

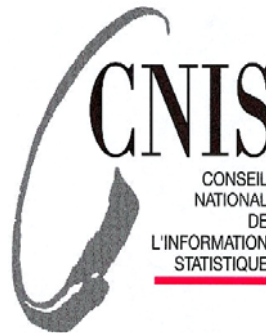
Le Conseil réaffirme **son soutien à la réalisation de l'enquête «Trajectoires et origines des migrants et de leurs descendants » (TeO)** qui permettra de mieux appréhender en France métropolitaine les phénomènes d'intégration mais aussi de discriminations liées à l'origine y compris pour les populations originaires des DOM. Il prend acte de la large concertation organisée sur les questions «sensibles» de l'enquête au sein du Cnis. Dans l'état actuel de la réflexion et de la concertation sur les discriminations dont peuvent souffrir les personnes du fait de leur origine, il juge très inopportune l'élaboration par le système statistique public (SSP) d'une nomenclature ethno- raciale de référence, et recommande de recourir à des questions ouvertes pour appréhender la façon dont les personnes se perçoivent.

Suite à la décision du Conseil Constitutionnel en date du 15 novembre 2007 qui considère que « les traitements nécessaires à la conduite d'études sur la mesure de la diversité des origines des personnes, de la discrimination et de l'intégration (...) ne sauraient (...) reposer sur l'origine ethnique ou la race », il approuve la décision de l'Insee et de l'Ined de retirer du test en cours les deux questions sur l'autodéfinition de la couleur de peau. Par ailleurs, il souhaite que, parmi l'ensemble des causes de discrimination citées dans l'enquête, celles qui sont liées à l'origine soient relativisées. Il demande notamment que la question ID9 du module « image de soi et regard des autres » sur les façons de se définir soit placée en tête du questionnaire.

Le Conseil prend acte que les services producteurs appliqueront des règles déontologiques spécifiques pour la publication de résultats portant sur les variables sensibles de cette enquête : en effet, la mesure des discriminations ne peut se contenter de se fonder sur l'observation des inégalités prises isolément, mais elle doit impérativement faire référence à une comparaison « toutes choses égales par ailleurs » selon les principes adoptés par le conseil scientifique de la revue Économie et Statistique le 28 septembre 2005. Il demande par ailleurs que le Conseil scientifique du Comité de concertation pour les données en sciences humaines et sociales soit consulté systématiquement avant de donner l'accès à ces variables sensibles pour des projets de recherche. Le Cnis sera informé des projets retenus par ce Conseil.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 1 :
avis de conformité (label)



COMITÉ DU LABEL

Paris, le 20 décembre 2007 – N° 309/D131

AVIS DE CONFORMITÉ
Modifiant l'avis n° 243/D131 du 1^{er} juin 2007

Enquête Trajectoires et Origines
(TeO)

Services producteurs : Ined et Insee

Opportunité : avis favorable d'ensemble émis le 15 mai 2006 par la formation Démographie, conditions de vie, complété par une délibération du 12 octobre 2007 de cette même formation sur les questionnements sensibles et l'avis adopté par l'assemblée plénière du 19 décembre 2007

Réunions du Comité du label du 31 mai 2007 et du 17 décembre 2007 (formation Ménages)

Cette enquête a pour objectif d'appréhender dans quelle mesure l'origine est en soi un facteur d'inégalités ou simplement de spécificité dans l'accès aux différentes ressources de la vie sociale (logement, langue et éducation, emploi, loisirs, services publics et prestations sociales, contraception, santé, nationalité, réseau de relations, marché matrimonial...). Elle s'intéressera à l'articulation entre l'origine et les autres catégories de distinction dans la société française (genre, classe, phénotype, âge, quartier...) afin d'analyser les processus d'intégration, de discrimination et de construction identitaire au sein de la société française dans son ensemble.

Les principaux thèmes du questionnaire seront les trajectoires migratoires, le lien avec le pays d'origine ou le pays de naissance des parents, l'éducation, la formation, l'emploi, le cadre de vie, le logement, le quartier, les parcours résidentiels, les pratiques matrimoniales, les langues parlées, la religion, les rapports aux institutions et la santé. Certains thèmes se trouveront de manière transversale dans plusieurs modules : discrimination, pratiques communautaires, relation de genre.

L'enquête portera sur cinq sous-groupes de personnes résidant en France métropolitaine dans des logements ordinaires : 9 600 immigrés âgés de 18 à 59 ans ; 9 600 descendants d'immigrés nés en France métropolitaine âgés de 18 à 49 ans ; 800 personnes nées dans un département d'outre-mer âgées de 18 à 59 ans ; 800 personnes nées en France métropolitaine dont au moins un parent est né dans un département d'outre-mer âgées de 18 à 49 ans ; et enfin, 3 400 personnes n'appartenant à aucun des quatre groupes précédents, qui seront tirées dans un échantillon dit « témoin » ou « population de référence », âgées de 18 à 59 ans.

La collecte sera réalisée en face à face par enquêteur, sous Capi, avec une durée de questionnement d'une heure et trente minutes. Elle se déroulera d'août à fin décembre 2008.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 1 :
avis de conformité (label)

L'enquête interrogera également, sous la forme d'un questionnaire spécifique auto-administré, les enfants des adultes interrogés dans l'enquête principale, dès lors qu'ils ont entre 15 et 24 ans et vivent au domicile parental.

Les premiers résultats seront publiés six mois après la fin de la collecte. Ces publications comprendront des articles pour *Insee première*, *Population et Sociétés*, ainsi que des publications courtes des autres partenaires de l'enquête. Des articles plus approfondis seront réalisés pour des ouvrages ou revues.

Le Comité du label demande qu'un certain nombre d'améliorations ou d'approfondissements techniques soit apporté au dossier.

Pour cela, une note méthodologique globale devra rassembler tous les éléments de compréhension sur le protocole, qui restera néanmoins complexe : utilisation conjointe de l'échantillon de l'EDP et du RP 2007, appariement avec des sources externes, segmentation de l'échantillon en plusieurs sous-populations, surreprésentation ponctuelle de certains groupes, calcul des pondérations, constitution et mobilisation d'une réserve, procédures de redressement.

La lettre-avis devra reprendre les mentions standards et tenir compte des quelques remarques ponctuelles soulevées en séance.

Le questionnaire a fait l'objet d'un certain nombre d'observations en séance. Les experts en communiqueront la liste au service qui est invité à en tenir le plus grand compte.

La nomenclature retenue pour les religions n'est pas visible de l'enquêté qui déclare spontanément sa religion d'appartenance. Le codage qu'en fait l'enquêteur et qui entre ensuite dans les fichiers de données ne doit cependant pas conduire à donner un détail plus grand pour certaines religions que pour d'autres. La nomenclature de codage proposée à l'enquêteur doit donc être homogène quant à son niveau de détail pour les différents groupes de religion, ce qui n'est pas le cas pour l'actuelle nomenclature. Il convient par ailleurs de rajouter la catégorie « animiste » à la nomenclature des religions.

Le Comité du Label demande que la possibilité de ne pas répondre soit rappelée pour chaque questionnement sensible (questions des modules « trajectoires migratoires », « identité et image de soi », « religion » et questions relatives aux discriminations). Lorsqu'une question propose des modalités dont l'une (ou plusieurs) d'entre elles est (sont) relative(s) à des domaines sensibles, il devra être dit explicitement que les modalités « ne sait pas » ou « ne souhaite pas répondre » sont admises. Le questionnaire actuel comporte ces modalités mais demande à l'enquêteur de ne pas les exprimer.

Le titre du module « Identité et image pour les autres » étant mal adapté et susceptible d'incompréhensions, le Comité du label demande au service de proposer un titre en cohérence avec le thème des questions posées.

En ce qui concerne le questionnaire destiné aux enfants de l'enquêté, le Comité du label demande que les formulations concernant la discrimination et l'origine adoptent la même logique que celles de la dernière version du questionnaire principal.

Le Comité du label attribue le label d'intérêt général et de qualité statistique à l'enquête Trajectoires et Origines (TeO).

Cet avis de conformité est valide pour l'année 2008

Le Président du Comité du label

Signé

Jacques Garagnon

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

Ce projet de note actualise la note N° 982/DG75-F402 du 14 avril 2006

Direction des Statistiques Démographiques et Sociales

Département des Prix à la Consommation des Ressources et des Conditions de vie des
ménages

Unité « Méthodes Statistiques »

Dossier suivi par :
Marc CHRISTINE
Tél : 01.41.17.59.74
Fax : 0141.17.62.87
Messagerie : marc.christine@insee.fr

Paris, le 19 décembre 2007
N° XXX/DG75-F402

Objet : Note générale d'information sur les orientations de la DSDS en matière de construction des systèmes d'échantillonnage des enquêtes ménages.

1. Les échantillons issus du RP 1999.

L'Insee a conduit régulièrement un chantier de rénovation des échantillons des enquêtes ménages après chaque RP ponctuel tel qu'il était organisé au cours des décennies précédentes. Ainsi, le RP de 1999 a donné naissance à un nouvel échantillon-maître et un nouvel échantillon Emploi. Ceux-ci ont été tous les deux mis en service courant 2001 :

- pour l'échantillon-maître - EM - : en septembre 2001, avec le tirage de la 1^{ère} enquête qui l'a utilisé : l'enquête Logement 2002, puis, pour la partie « recours aux extensions régionales », mobilisant un EM complémentaire, l'EMEX⁸, en 2002, avec le tirage de l'enquête Santé et de ses extensions.
- au 1^{er} juillet 2001 pour l'enquête Emploi renouvelée.

Ces échantillons ont été calibrés pour durer sur une période de 9 ans environ.

La mise en œuvre de l'EM 1999 a impliqué une procédure, particulière et nouvelle à l'époque, *d'édition des fiches-adresses (FA)*. Celle-ci se fait par récupération des bases-adresses-logements (BAL) constituées à l'issue de la lecture optique massive de l'ensemble des feuilles de logement du RP 1999 et par association, à tout logement tiré dans l'échantillon final, du numéro correspondant de l'image stockée dans la BAL, à fin d'impression du bandeau contenant le nom et l'adresse de l'occupant du logement au RP 1999. Cette information est indispensable pour que les enquêteurs puissent identifier le logement sur le terrain au moment de la collecte et enquêter le bon logement désigné par le tirage aléatoire.

L'ensemble de ces traitements ont été décrits dans le dossier remis à la CNIL le 27 juillet 2000 et ont obtenu un avis favorable de celle-ci le 3 avril 2001.

⁸ *E*chantillon-*M*aître pour les *EX*tensions régionales.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

2. La situation future.

Dans le contexte très différent du nouveau recensement annuel mis en place en 2004, mais avec des finalités analogues en termes d'échantillonnage des enquêtes ménages, l'Insee travaille actuellement à la conception et la mise en œuvre de nouveaux échantillons d'enquêtes ménages issus des bases des enquêtes annuelles de recensement. La méthodologie est complètement revue pour tenir compte du fait que le recensement devient annuel mais ne porte chaque année que sur une partie du territoire.

Ce projet est baptisé OCTOPUSSE⁹.

L'objectif est de mettre en œuvre un nouveau système d'échantillonnage qui doit être opérationnel courant 2009, pour assurer le tirage des enquêtes ayant lieu à compter du 2^{ème} semestre 2009. Une opération spécifique est montée également pour assurer l'augmentation de l'échantillon Emploi à partir du début 2009 et le renouvellement de l'échantillon Emploi actuel à l'horizon 2010.

La chaîne informatique de ce nouveau système sera construite pour une période de 10 ans. Toutefois, les principes généraux sur lesquels ce système est fondé pourraient être reconduits au-delà de cette échéance et, partant, les modalités d'acquisition et de traitement des données, du moins si l'organisation des recensements annuels n'est pas modifiée d'ici là. Seule l'obsolescence technologique naturelle (ainsi que des raisons statistiques, cf. infra, § 2.2) commande d'édicter une durée de vie « raisonnable » pour l'outil informatique mis en œuvre.

La principale nouveauté du projet est la *constitution des bases de sondage* : alors que, dans les recensements décennaux, c'est l'ensemble et la totalité des logements recensés qui servaient de base, le principe général retenu est que *les enquêtes d'une année donnée seront puisées dans la base constituée à partir de la dernière campagne annuelle du recensement disponible*. Contrairement, donc, à l'échantillon-maître de 1999 où les bases de sondage de logements étaient connues et constituées dès 1999, dans le futur système, elles ne seront connues qu'après chaque collecte annuelle du recensement.

Concrètement, pour tenir compte du délai nécessaire entre la mise à disposition des échantillons et le démarrage de la collecte sur le terrain, on admettra que les enquêtes commençant à partir du mois de mai $n + 1$ seront tirées dans la base constituée à partir de la campagne n du RP. En revanche, les enquêtes de début d'année $n + 1$ seront tirées dans la base de l'année antérieure, soit $n - 1$.

Dans tous les cas de figure, on diminue donc considérablement l'intervalle de temps entre la date de constitution de la base et la date de collecte.

Ce décalage était l'un des points faibles des échantillons construits sur des RP quasi-décennaux. Le vieillissement des bases de sondage entraîne en effet une perte de la qualité : destructions ou changements de destination des logements, recomposition urbaine entraînant des difficultés, voire des impossibilités de repérage, nécessité de recourir à une base complémentaire pour couvrir les logements construits après le dernier RP (« BSLN », base de sondage des logements neufs, constituée à partir des flux de permis de construire), impossibilité de sélectionner valablement des échantillons sur population ciblée, du fait de probabilités importantes et augmentant au cours du temps de changement de l'occupant du logement depuis le dernier RP, méthodes de calage ou de traitement de la non-réponse fondées sur les caractéristiques des habitants connues au RP¹⁰ moins pertinentes du fait de ces changements etc... De ce point de vue, le nouveau RP sera un facteur de qualité des nouveaux échantillons.

⁹ Organisation Coordonnée de Tirages Optimisés Pour une Utilisation Statistique des Echantillons.

¹⁰ On notera qu'à l'heure actuelle les enquêtes ménages sont recalées sur les données de l'enquête Emploi annualisée, qui est la plus grosse enquête réalisée en termes de taille d'échantillon. A terme, avec la disponibilité d'estimations structurelles annuelles rendues possibles par la concaténation de 5 campagnes successives annuelles du nouveau RP, il est vraisemblable qu'on cherchera plutôt à recalculer toutes les enquêtes ménages sur ces données issues du recensement.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

Le nouvel échantillon-maître sera fondé, comme les précédents et l'actuel (EM 1999), sur un *système d'unités primaires (UP)*. Rappelons que ce système a pour but de réduire les déplacements des enquêteurs, ce qui a pour effet d'augmenter la qualité, en diminuant les effets de lassitude par rapport à des temps de déplacement qui seraient excessifs, et aussi de mieux maîtriser les coûts de collecte afférents. Le principe est de constituer et de sélectionner aléatoirement des zones géographiques (agrégation de communes en milieu rural, unités urbaines..) qui seront fixes pendant toute la durée de vie de l'EM. Pour chaque enquête ménages, on tire alors un échantillon de logements à l'intérieur de ces zones.

Toutefois, les UP du futur EM seront construites différemment du passé pour s'adapter au principe général de toujours tirer les échantillons de logements dans la partie du territoire recensée l'année précédente : il s'agira de zones géographiques constituées par regroupement de communes de façon que les 5 groupes de rotation du RP (en termes de communes pour les petites¹¹ et en termes d'adresses pour les grandes¹²) soient présentes au sein du regroupement. Comme dans les précédents EM, ceci permettra d'associer à chaque zone un ou plusieurs enquêteurs, habitant a priori dans la zone ou non loin de celle-ci, de façon à minimiser les déplacements (les UP seront de ce fait baptisées « **zones d'action d'enquêteurs** », ZAE), tout en assurant - ce qui est la nouveauté - que, quelle que soit l'année d'enquête, on disposera au sein de la zone d'un « vivier » de logements servant de base de sondage, issu du groupe de rotation recensé l'année précédente.

Ces zones seront fixes pendant 10 ans, ainsi qu'il est expliqué au § 2.2.

Le calibrage de ces zones a été fixé à 300 logements principaux pour chacun des groupes de rotation représenté au sein de la zone (soit au minimum 1500 logements par zone). Ce seuil a été fixé empiriquement de la manière suivante : on considère en année courante un maximum de 10 enquêtes ponctionnant chacune 25 fiches-adresses dans un groupe de rotation donné, plus une marge de sécurité.

Comme en 1999, un échantillon-maître complémentaire pour les extensions régionales d'enquête (EMEX) sera construit, de manière à assurer une meilleure couverture territoriale des échantillons pour les régions bénéficiant d'une extension. Concrètement, cela reviendra à sélectionner des ZAE complémentaires dans chaque région, qui seront activées dès qu'une extension régionale sera décidée pour une enquête donnée et seulement dans ce cas.

La base de sondage pour les enquêtes auprès des ménages constituée à partir de l'échantillon-maître contient - à l'instar de la base de sondage issue de l'échantillon-maître 1999 - les variables issues de la collecte de l'enquête annuelle du recensement de la population utiles à la mesure de la représentativité des échantillons d'enquête.

La liste des variables utilisées pour le tirage d'échantillons dans l'échantillon-maître de 1999, telle que décrite dans le dossier Cnil susvisé, est la suivante :

pour chaque logement, la base de sondage contient les informations suivantes :

- un identifiant FL (Feuille de logement) ;
- la catégorie du logement ;
- le nombre de pièces ;
- le nombre de personnes y habitant ;
- la position socio-professionnelle et l'âge du chef de ménage.

À cette liste de variables, il pourrait s'avérer opportun d'ajouter d'autres variables issues du recensement de la population, utiles notamment pour analyser la non-réponse dans les enquêtes-ménages.

Ce traitement fera l'objet d'une déclaration à la CNIL courant 2008

2. 1 Nouvelle édition des fiches-adresses.

L'édition des fiches-adresses reposera, comme pour l'EM 1999, sur la *mobilisation des images contenant le nom et l'adresse de l'occupant du logement au RP*, tel qu'inscrits à la main par celui-ci sur la feuille de logement lors du recensement. Ceci nécessite d'associer à tout identifiant de logement le n° d'image

¹¹ Moins de 10.000 habitants. Notion de base du RP.

¹² 10.000 habitants ou plus. Notion de base du RP.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

correspondant (opération assurée et déclarée à la CNIL par le RP) et de récupérer ce n°, dans le cadre du processus de tirage, pour aller chercher dans la BAL l'image correspondante.

La description de ce traitement figurera dans le dossier qui sera déposé en 2008 à la Cnil, de façon analogue à celle faite en 2000 pour l'EM 1999. Les conditions d'extraction et d'impression des images sont en cours d'étude par les équipes informatiques en charge du projet. Sans pouvoir les décrire précisément, compte tenu des technologies à définir, on peut toutefois dire qu'elles seront de nature tout à fait analogue à ce qui a été mis en œuvre lors de la réalisation de l'EM 1999. En particulier, après une étude d'opportunité, il est apparu que *la solution d'impression centralisée dans un Centre Informatique de l'Insee, avec des conditions de sécurisation des locaux dédiés, restait la meilleure* vis-à-vis des risques techniques et aussi pour mieux assurer la sécurité des transferts d'information en aval impliqués par cette édition des FA.

2.2 Disjonction inter-temporelle au sein de l'échantillon-maître.

Un même logement ne pourra pas être enquêté la même année dans le cadre de deux enquêtes différentes (sauf en cas d'enquêtes répétées dans le temps) : en effet, tout logement tiré sera « marqué » informatiquement pour ne pas être réutilisé lors d'un tirage ultérieur de la même année.

De surcroît, par construction, un même logement ne pourra pas être tiré dans deux échantillons d'enquêtes ayant lieu deux années différentes au cours d'un cycle de 5 ans. En effet, les bases de sondage seront disjointes, puisqu'elles impacteront des groupes de rotation annuels (de petites communes ou d'adresses dans les grandes communes) différents.

Cas particulier :

En cas de réserve insuffisante dans certaines unités primaires pour une enquête n+1, tirée normalement dans le groupe de rotation du RP n (si, par exemple, une très grosse enquête a lieu une année donnée et que les communes du groupe de rotation impacté l'année considérée aient une taille proche du seuil minimum en termes de nombre de logements, soit 300), on sélectionnera des logements à titre complémentaire parmi les logements non déjà sélectionnés au sein des collectes précédentes du RP appartenant à la même UP.

Dans certains cas où l'intérêt d'une enquête réside principalement en la mesure d'évolutions d'une année sur l'autre, il est envisagé, pour éviter les « biais de rotation » liés au fait qu'on n'impacterait pas les mêmes communes lors de deux années successives, de tirer son échantillon dans l'ensemble des communes des 5 groupes de rotation du RP au sein de chaque UP.

Ces cas de dérogation à la règle générale devraient toutefois rester limités.

Donc, en règle générale, le système d'échantillonnage qui sera construit assurera une disjonction automatique sur une période de 5 ans.

Cette solution est d'ailleurs cohérente avec le principe de conservation des BAL du nouveau RP, qui est fixée à 6 ans : on ne sera jamais dans la situation de devoir mobiliser des campagnes du RP de plus grande ancienneté. La question des enquêtes locales, qui n'est pas instruite à ce jour, pourra toutefois montrer qu'il peut être nécessaire, pour des besoins très spécifiques, de mobiliser également 5 campagnes du RP pour construire une base de sondage ayant la couverture géographique suffisante.

Peut-on aller plus loin ?

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

- dans le passé, la période intercensitaire étant de l'ordre de 10 ans, la disjonction était assurée par nature sur cette période. Ici, on a une période plus courte mais néanmoins suffisante.
- si l'on voulait assurer une disjonction sur une période de 10 ans, cela nécessiterait que l'on soit en mesure de repérer un logement recensé au cours d'une période de 5 ans pour éviter de le tirer à nouveau au cours de la période quinquennale suivante. **Ceci ne serait possible que si l'on disposait d'un identifiant des logements unique et pérenne.** Or cette option, difficile à mettre en œuvre, n'a pas été retenue par les équipes en charge du recensement.

De surcroît, il faudrait augmenter le seuil minimal de taille des nouvelles UP (ZAE), calibrées, comme on l'a vu, de façon à assurer la réalisation de 10 enquêtes annuelles ponctionnant chacune de 25 à 30 FA (soit, pour une durée de vie de 5 ans, une taille minimale de 1500 logements). Si chaque ZAE devait alimenter les enquêtes d'une période décennale, on devrait accroître leur taille, ce qui conduirait à augmenter leur étendue géographique, donc à risquer d'accroître sensiblement les déplacements des enquêteurs qui, chaque année, doivent aller sur la ou les communes du même groupe de rotation au sein de la même zone, ceci entraînant une détérioration de la qualité et une augmentation des coûts.

Une solution alternative serait de constituer le système de ZAE pour une période de 10 ans, mais avec un renouvellement au bout de 5 ans (du moins pour celles qui sont tirées de manière aléatoire, un certain nombre de ZAE en urbain étant retenues d'office dans les zones les plus denses¹³) : mais ceci entraînerait la définition de nouvelles zones d'enquête au bout d'une période de 5 ans et nécessiterait vraisemblablement un renouvellement plus fréquent, bien que partiel, du réseau d'enquêteurs ou aurait pour conséquence possible une augmentation, entre les deux périodes, de leurs déplacements, à réseau inchangé.

- Comparaison avec l'EM 1999.

Dans l'actuel EM, les UP ont été astreintes à respecter une fourchette de 1800-3600 logements principaux. L'EM était calibré sur une période de 6 années, avec 10 enquêtes annuelles ponctionnant chacune 30 FA, ce qui est cohérent avec les hypothèses retenues pour le futur EM en termes de seuil minimum. Dans la pratique, l'EM 1999 durera 8 ans, ce qui suppose que la ponction annuelle soit moins forte que prévu si l'on ne veut pas qu'il s'épuise prématurément. A ce jour, en fait, le nombre de logements principaux tirés dans l'EM 1999 depuis sa mise en service est d'environ 378.000, avec un taux de ponction global de 23 % (alors qu'on a dépassé la moitié de sa durée de vie).

Une étude sur la période 2001-2005 montre toutefois que, sur les 203 UP sélectionnées au titre de l'échantillon-maître 1999 dans le rural et dans le petit urbain, 20% d'entre elles se sont vues ponctionner à plus de 300 résidences principales au moins une année. Notons que ces unités primaires ont toutes dépassé le seuil en 2003, qui a été une année un peu exceptionnelle en termes de nombre de logements sélectionnés.

Mais la différence entre les deux EM - l'actuel et le futur - est que, dans l'actuel EM, une ponction de plus de 300 logements une année donnée dans une UP peut être compensée en sens inverse, par une moindre sélection de logements, l'année suivante. En revanche, dans le futur EM, même si un groupe de rotation d'une année donnée est, à l'intérieur d'une ZAE, très au-dessus du seuil minimum de 300 logements et qu'il n'est ponctionné que partiellement par les tirages d'enquêtes de l'année, il ne peut servir de réserve ou de compensation (sauf les cas exceptionnels mentionnés plus haut) pour les tirages d'une autre année.

De ce point de vue, le mode de constitution des futures ZAE entraîne une petite rigidité qui se traduit par une disjonction nécessitant des principes nouveaux, alors que la taille globale des UP et les hypothèses de tirage restent à peu près les mêmes dans les deux EM (actuel et futur).

¹³ Les 37 plus grandes communes de la France métropolitaine.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

2.3 Disjonction entre l'échantillon-maître et l'échantillon Emploi.

Au stade actuel de la réflexion¹⁴, une augmentation conséquente de la taille de l'échantillon Emploi actuel a été prévue pour une mise en service à partir du début 2009. Par ailleurs, le renouvellement de l'échantillon actuel était prévu à compter de la mi-2010.

La méthodologie de constitution de ces futurs échantillons est en cours de définition et ne peut être indiquée de manière complète dans la présente note. Elle donnera lieu à une note ultérieure dès qu'elle aura été fixée définitivement.

Ces échantillons seront vraisemblablement constitués de « grappes » de logements voisins, se rapprochant de la logique antérieure des « aires ».

Ce système particulier se justifie par le fait que l'enquête Emploi est une enquête relativement simple et rapide mais avec un très gros échantillon (environ 200.000 logements interrogés chaque année) : un même enquêteur a donc la possibilité d'enchaîner plusieurs enquêtes dans une même journée, et cela est d'autant plus nécessaire que la période de collecte, du fait du caractère continu de l'enquête, est très réduite (2 semaines et 2 jours après la semaine de référence couverte par l'enquête).

Les avantages de ce mode d'échantillonnage sont qu'il permet d'assurer un meilleur ratio entre le nombre de questionnaires collectés et le coût de collecte et que, par expérience, les taux de réponse s'avèrent meilleurs dans ce type d'échantillonnage que lorsque celui-ci est dispersé. En contrepartie, outre le fait que ce mode ne serait pas adapté à des enquêtes complexes et longues, où le même enquêteur aurait du mal à enchaîner plusieurs ménages à interroger, l'inconvénient majeur est la détérioration de la précision due à « l'effet de grappe ».

De conception différente de l'échantillon-maître et conçu indépendamment de ce dernier, l'échantillon Emploi devra donc subir une disjonction par rapport à celui-ci, postérieurement à sa constitution.

La méthodologie de cette disjonction sera définie dans le cadre des travaux méthodologiques en cours et définie ultérieurement.

2.4 Disjonction avec les précédents échantillons issus du RP 1999.

Celle-ci ne sera pas possible du fait de l'absence d'un identifiant commun pérenne permettant de suivre les logements entre ces deux RP. De même, aucune disjonction ne sera possible avec des logements qui auraient été tirés de la BSLN sur la période 2001-2008.

Cette situation n'est pas nouvelle : dans le passé, et pour les mêmes raisons, la disjonction avec les échantillons issus du RP précédent ou de la BSLN sur la période intercensitaire, ne pouvait pas non plus être assurée. Il peut donc y avoir un petit risque que des logements interrogés au titre de la précédente période intercensitaire le soient à nouveau dans le cadre du nouveau système de tirage. Ceci étant, la probabilité est sûrement faible : en urbain, les zones les plus denses seront bien entendu impactées dans les nouveaux échantillons mais le vivier de logements disponible est important ; en rural, on procèdera au tirage d'unités primaires, la probabilité de tomber sur des UP déjà tirées en 1999 reste faible. De surcroît, il ne faut pas oublier que, même si les logements sont communs, les ménages peuvent ne plus être les mêmes, du fait de déménagements.

3. Le régime transitoire.

Compte tenu des délais habituels pour la construction d'un outil informatique réalisant les différentes fonctionnalités de tirage mais aussi de calcul des poids de sondage des logements tirés, outil qui se veut adapté à

¹⁴ Faisant suite aux décisions prises par l'Insee après le rapport des Inspections Générales des Finances et des Affaires Sociales sur le devenir de l'enquête Emploi.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

la généralité des enquêtes ménages, le projet OCTOPUSSE est appelé à se terminer courant 2009. Cela veut dire, par corollaire, que toutes les enquêtes ménages réalisées jusqu'au 1^{er} semestre 2009 inclus continueront d'être tirées dans l'EM actuel, enrichi de la BSLN, avec les problèmes de vieillissement signalés ci-dessus inhérents à tout système fondé sur un recensement ponctuel.

Or certaines enquêtes de la période transitoire, du fait de leurs spécificités, nécessitent de tirer leurs échantillons d'ores et déjà dans les nouvelles bases annuelles du RP, avant même que l'outil général de tirage - qui est précisément l'objet du projet OCTOPUSSE - soit opérationnel.

Il s'agit des enquêtes suivantes, ainsi que de leurs tests préalables¹⁵ :

- Trajectoires et Origines (TeO) (collecte de septembre à décembre 2008, 40.000 FA)
- Technologies de l'Information et de la Communication (TIC) (collecte en avril 2008, 4.000 FA)

Dans tous les cas, le tirage dans le nouveau RP et non dans l'EM 1999 se justifie par le fait que les populations concernées sont rares et difficiles à repérer. Il faut donc éviter toute déperdition, ce qui serait le cas si l'on tirait dans une base trop ancienne, ne donnant pas l'assurance que les ménages habitant les logements sélectionnés aient les caractéristiques attendues : tirer dans l'EM 1999 n'assurerait donc pas la qualité adéquate et pourrait engendrer des coûts additionnels.

Pour répondre à ces besoins spécifiques *et sur le champ des enquêtes précitées*, l'UMS met en place un processus de tirage provisoire.

Ce processus est développé en interne pour ce qui est de la programmation informatique des « moteurs » de tirage. Le principe général est une combinaison du principe nouveau (tirer les enquêtes d'une année dans la base de l'enquête annuelle de recensement de l'année précédente) et de la configuration actuelle des UP de l'EM 1999 (pour bénéficier de la localisation actuelle du réseau d'enquêteurs).

Bien entendu, toute la qualité est recherchée et seule la non-automatisation des différentes étapes du processus - objectif visé dans la nouvelle chaîne de traitement qui sera mise en place courant 2009 - reflète le caractère provisoire des méthodes et outils mis en œuvre.

Examinons plus particulièrement quatre points importants.

3.1 Le contenu de la base de sondage.

La liste des variables utilisées pour le tirage de l'échantillon de l'enquête TeO, est légèrement enrichie par rapport à la liste des variables figurant dans les bases de sondage constituée à partir de l'échantillon-maître de 1999 :

pour chaque logement, la base de sondage contient les informations suivantes :

- un identifiant FL (Feuille de logement) ;
- la catégorie du logement ;
- le nombre de pièces ;
- le nombre de personnes y habitant ;
- le nombre d'enfants mineurs y habitant ;

¹⁵ Ont relevé également de ce processus spécifique les enquêtes Mode de garde (collecte en avril-mai 2007, 8.000 FA) et Vie Quotidienne et Santé (140.000 FA), qui sert de filtrage à l'enquête Handicap-Santé (collecte d'avril à juillet 2008, 25.000 FA).

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

- l'ensemble des variables relatives à la position de l'enquêté TeO vis à vis de l'emploi qui permettront de coder sa situation socio-professionnelle (cette codification n'ayant pu être réalisée sur l'ensemble des données de l'enquête annuelle de recensement de 2007 avant le tirage de l'échantillon TeO) ;
- la date de naissance de l'enquêté ;
- le sexe de l'enquêté ;
- le pays de naissance de l'enquêté ;
- la nationalité de l'enquêté ;
- le diplôme de l'enquêté.

L'échantillon étant exceptionnellement un échantillon d'individus et non de logements, les variables se rapportent à l'individu, qu'il soit ou non personne de référence (ex « chef de ménage »). Ces variables seront utilisées à des fins strictement méthodologiques permettant de valider les résultats de l'enquête.

3.2 L'édition des fiches-adresses (FA).

Celle-ci se fait par mobilisation, reconnaissance et lecture des BAL du nouveau recensement au moyen d'une adaptation du processus actuel d'édition opérée par les équipes informatiques de l'Insee pour adapter le processus actuel au format et au contenu des nouvelles BAL, très proches d'ailleurs de ceux du RP 1999.

Le processus d'impression proprement dit se déroulera dans des conditions voisines de celles de l'impression qui se fait en régime de croisière actuellement à partir des BAL du RP 1999, avec des impératifs de sécurité de même nature. L'envoi des FA aux directions régionales de l'Insee sera assuré de manière postale, par Colissimo.

3.3 La disjonction avec les échantillons précédents.

Comme il a été dit plus haut, il ne sera pas possible d'assurer la disjonction avec les échantillons issus du RP 1999 (y compris de l'échantillon Emploi actuel), du fait de l'absence d'un identifiant commun des logements. Toutefois, le risque de retomber sur un logement déjà enquêté est très faible¹⁶. De surcroît, ces enquêtes ayant lieu relativement postérieurement au RP 1999, ce risque est encore amoindri par les probabilités importantes que le ménage ait changé au cours de la période¹⁷.

Dans tous les cas, l'UMS a édicté des principes que les Directions Régionales de l'Insee ont transmis sous forme de consignes aux enquêteurs : les enquêteurs doivent vérifier si un logement nouvellement tiré ne fait pas partie d'un échantillon tiré sur le RP 1999 et, le cas échéant, même s'ils doivent essayer de réaliser l'enquête en cours, doivent faire preuve de circonspection dans leurs relances.

3.4 Disjonction avec les futurs échantillons issus du système de tirage de croisière.

On a vu que, dans le futur système de l'échantillon-maître, le principe d'attaquer les bases issues de la dernière campagne du RP disponible permettait une disjonction automatique sur un cycle de 5 ans. Les premiers échantillons tirés dans le cadre de ce nouveau système le seront au 2^{ème} semestre 2009, à partir des bases du RP 2008.

Donc, des tirages même précoces dans le nouveau RP, mais ne sollicitant que des bases RP antérieurs à 2007 ne devraient pas poser de problèmes vis-à-vis de toutes les enquêtes tirées en régime de croisière à partir des bases 2008 ou 2009 dans l'échantillon-maître. A partir de mai 2011 (enquêtes tirées dans les bases 2010 et au-delà),

¹⁶ Un logement tiré aujourd'hui à taux uniforme parmi l'ensemble des logements principaux aurait une probabilité de l'ordre de 1,5 % de tomber parmi les logements déjà tirés dans l'EM 1999 à ce jour.

¹⁷ D'après l'enquête Permanente sur les Conditions de Vie (PCV) de janvier 2005, on estime à 36,7 % la proportion de logements ayant changé de titulaire entre le RP 1999 et 2005.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 2 :
Note sur l'échantillonnage

on pourrait retomber accidentellement sur des logements déjà tirés dans les bases 2005 à 2009 et enquêtées 5 ans auparavant, à condition toutefois que ces logements appartiennent à des ZAE du nouvel échantillon-maître recoupant des UP de l'EM 1999, ce qui limite la probabilité d'occurrence. Il sera impossible d'assurer une disjonction entre l'ancien système d'UP et le nouveau système de ZAE.

Le cas particulier de la disjonction avec les futurs échantillons Emploi sera traité quand la méthodologie de constitution de ces derniers sera arrêtée.

3.4 Conséquence pratique : déclarations à la CNIL.

La présente note permet d'informer la CNIL des projets de l'Insee en la matière. Elle ne constitue évidemment pas une déclaration de traitement. Comme il a été dit plus haut, la déclaration relative à la constitution de l'ensemble des échantillons à partir du nouveau RP aura lieu courant 2008.

S'agissant des opérations ponctuelles nécessitant un tirage précoce dans le nouveau RP, telles qu'évoquées au §3, la déclaration relative à leur échantillonnage sera incluse dans la déclaration générale de chacune des enquêtes concernées.

Le Chef-adjoint de l'Unité Méthodes statistiques

Marc CHRISTINE

Diffusion :

- M. LOLLIVIER (DSDS)
- MM. DESPLANQUES, LEFEBVRE, LHERITIER, VERGER (DSDS)
- MM. CUNEO, LE GLEAU, LANG, ISNARD (DSCRI)
- MM. LOONIS, FAIVRE (UMS).
- M. JANSOLIN, Mmes MENARD, PIETRI-BESSY, YAOUANCQ (UMS)
- M. BERLEMONT (CNIP)
- Mme GILLES (DAP)

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES
133, boulevard Davout
75980 PARIS CEDEX 20

Ministère de l'économie, des finances
et de l'emploi
INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES
18, boulevard Adolphe Pinard
75675 PARIS CEDEX 14

C O N V E N T I O N

N° **PROJET**

Relative à la réalisation par l'Institut National d'Études Démographiques
et l'Institut National de la Statistique et des Études Économiques de l'enquête

« *Trajectoires et origines, Enquête sur la diversité des populations en France* » dite « **TeO** »

Entre les soussignés :

L'INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES, désigné ci-après par les
initiales Ined, et représenté par son Directeur, François HERAN,

de première part,

et

**L'INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES
ÉCONOMIQUES**, Direction générale du ministère de l'économie, des finances et de l'industrie,
désigné ci-après par les initiales Insee, et représenté par son Directeur général, Jean-Philippe
COTIS,

de deuxième part,

Ensemble désignés par l'expression « les parties »,

Il est convenu ce qui suit :

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Préambule

Le 3 décembre 2003, à l'occasion de l'anniversaire de la « Marche pour l'égalité » (dite « Marche des Beurs »), le Premier Ministre affirmait la nécessité d'améliorer la connaissance des populations immigrées et de leurs descendants : *« Pour nous permettre de bien comprendre les réussites et les échecs, nous manquons d'éléments quantitatifs : nous avons besoin d'une grande enquête statistique et sociologique sur la question de l'intégration et des inégalités, parce qu'il n'y a rien de complet aujourd'hui. Je demande donc au Haut Conseil à l'Intégration, à l'Ined, à l'Insee et au Ministère des Affaires sociales de me proposer une méthode ambitieuse pour une étude qui fera le bilan de l'intégration en France depuis 20 ans, et qui utilisera toutes les ressources permises par la loi de 1978 sur la protection de la vie privée ».*

La Cour des comptes, de son côté, dans son rapport public sur « L'accueil des immigrants et l'intégration des populations issues de l'immigration » (novembre 2004, p. 148), tient pour acquise la volonté de l'Insee, de l'Ined et de la Drees de coopérer à une nouvelle enquête sur le sujet. Le directeur de la Population et des migrations a, pour sa part, confirmé le caractère prioritaire d'une telle enquête lors du conseil d'administration de l'Ined tenu le 15 décembre 2005. Régulièrement, à l'occasion des séances du Conseil scientifique et du Conseil d'administration de l'Ined, la Drees (direction de la recherche, de l'évaluation et des études statistiques au ministère du Travail, des Relations sociales et de la Solidarité et du ministère de la Santé, de la Jeunesse et des Sports) a manifesté son intérêt pour une enquête élargie sur les questions d'intégration et de discrimination.

L'enquête Trajectoire et origine (TeO) qui fait l'objet de la présente convention répond à ces différentes demandes.

Objet de la convention

La présente convention a pour objet de fixer les modalités juridiques, scientifiques, techniques et financières de la collaboration entre l'Ined et l'Insee pour la réalisation de l'enquête TeO.

Toute modification de la présente convention donnera lieu à la rédaction d'un avenant. Elle pourra notamment être modifiée pour se conformer aux dispositions de la délibération de la CNIL auprès de laquelle le dossier de formalités préalables sera déposé selon les modalités prévues à l'article 5.

Durée de la convention

La présente convention est conclue à compter de sa signature pour expirer à sa complète réalisation, au plus tard le 31 décembre 2010.

Elle pourra être résiliée par l'une ou l'autre partie pour un motif dûment justifié ou pour non-respect des dispositions de la présente convention.

La résiliation interviendra à l'issue d'un préavis de deux mois à compter de la notification de la demande de résiliation.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

En cas de résiliation, les deux parties détermineront ensemble, dans le délai de préavis, les suites à donner, compte tenu de l'avancement de l'enquête. Elles feront l'objet d'un relevé de conclusions.

En tout état de cause, la convention conservera ses effets sans limitation de durée en ce qui concerne les dispositions de l'article 6 ci-dessous relativement aux règles de propriété intellectuelle.

Objectifs et responsables du projet de recherche

L'enquête vise à chercher dans quelle mesure l'origine est en soi un facteur d'inégalités ou simplement de spécificité dans l'accès aux différentes ressources de la vie sociale (logement, langue et éducation, emploi, services publics et prestations sociales, contraception, nationalité, réseaux de relations, etc.). Pour ces différents thèmes, l'enquête s'intéressera à l'articulation entre l'origine (migratoire) et les autres catégories de distinction dans la société française (genre, classe, couleur, culture, âge, etc.).

L'objectif principal de l'enquête est donc d'analyser le processus d'intégration de la société française dans son ensemble (en prenant comme analyseur l'accès aux ressources de la vie sociale) afin de tester la pertinence des différents types de clivage et la force des frontières entre groupes constitués. En d'autres termes, la question de l'influence de l'origine sera examinée en cherchant, dans le même temps, à saisir et isoler l'impact des autres formes d'identités et d'identification. Une présentation plus détaillée des objectifs de l'enquête figure dans l'annexe 1 à la présente convention. Le questionnaire utilisé pour le test de novembre-décembre 2007 et programmé informatiquement à cette occasion figure en annexe 2 de la présente convention.

Dès lors, l'enquête ne peut pas seulement porter sur les immigrants et leurs descendants immédiats nés en France, elle doit également porter sur ceux que, faute d'une meilleure expression, les parties appelleront les « natifs descendants de natifs ». Certains groupes de population, a priori exposés à des difficultés d'accès aux ressources de la vie sociale, seront surreprésentés (habitants des quartiers sensibles, originaires des DOM, d'Afrique subsaharienne, etc....).

Le projet est placé sous la responsabilité scientifique des concepteurs de TeO :

- pour l'Ined : Cris Beauchemin, Christelle Hamel, Patrick Simon,
- pour l'Insee : Elisabeth Algava, Catherine Borrel.

Les opérations de collecte sont placées sous la responsabilité de Cécile Ménard (CPOS, Insee).

En cas de défection de ces personnels, les parties s'engagent à assurer la continuité de la fonction dans les meilleurs délais.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Conditions d'exécution

Principe général

L'Insee et l'Ined partagent la maîtrise d'ouvrage de l'enquête TeO. Elle est assurée conjointement par le directeur de l'Ined et par le chef du département de la démographie de l'Insee.

Partage des responsabilités au sein de la maîtrise d'ouvrage

Les deux parties interviennent conjointement à tous les stades de conception, d'exécution et d'exploitation de l'enquête.

L'Ined et l'Insee assurent ensemble la communication externe sur le projet d'enquête (communiqués éventuels). En la matière, l'interlocuteur du directeur de l'Ined est le directeur Général de l'Insee ou son directeur de cabinet. L'interlocuteur du directeur général de l'Insee est le directeur de l'Ined. L'Ined et l'Insee s'informent mutuellement des interventions qu'ils sont amenés à faire dans les médias, dans la mesure du possible avant ces interventions.

L'Ined et l'Insee assurent ensemble la conception du questionnaire. Si nécessaire, les arbitrages sont réalisés en commun par les deux co-maîtres d'ouvrage.

L'Insee est responsable de la préparation et du tirage de l'échantillon. Il s'engage à informer et consulter l'Ined sur les arbitrages méthodologiques en matière d'échantillonnage.

L'Insee est responsable de la préparation et de la réalisation de la collecte de l'enquête sur le terrain. Les concepteurs Ined de l'enquête peuvent cependant contribuer à l'animation des formations des enquêteurs.

L'Insee est responsable de l'élaboration du fichier brut de l'enquête.

L'Ined est responsable du projet scientifique de l'enquête. Il assure l'animation du plan de diffusion de l'enquête.

Calendrier d'exécution

L'enquête se déroulera sur le terrain entre septembre et décembre 2008. La publication des premiers résultats aura lieu au plus tard 14 mois après la collecte. Le calendrier prévisionnel des opérations et son phasage figurent en annexe 3 de la présente convention.

Organisation institutionnelle

La mise en œuvre de l'enquête, de sa conception jusqu'à son exploitation, sera sous la responsabilité des groupes et comités suivants dont la liste des membres figure en annexe 4 de la présente convention :

Le comité d'orientation :

Il est chargé de suivre étape par étape la réalisation de l'enquête et d'arbitrer sur les alternatives qui peuvent se présenter. Il veille à ce que le questionnaire réponde aux objectifs assignés à l'enquête, tels qu'ils figurent dans l'avis d'opportunité du CNIS. Il comprend des représentants des organismes partenaires financiers dans la réalisation de l'opération, ainsi que des chercheurs experts dans les domaines couverts par l'enquête. Il est co-présidé par le directeur de la DSDES et par le directeur de l'Ined.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Le groupe de coordination de la maîtrise d'ouvrage (MOA) :

Il est chargé de coordonner toutes les phases de l'enquête et d'assurer le suivi de l'exécution. Il est composé de représentants de l'Insee et de l'Ined. Il propose des arbitrages à la maîtrise d'ouvrage.

Le groupe de travail et d'exploitation :

Il est chargé de la préparation, de la conception du questionnaire, de l'exploitation de l'enquête et de la production d'un ouvrage collectif. Il est animé par les responsables scientifiques de l'Ined. Les participants sont les membres du groupe de coordination et des experts issus des Services Statistiques Ministériels et du monde scientifique. Les membres de ce groupe ont été désignées *Intuitu personae*. En cas de défection d'un membre, son remplacement sera conjointement décidé par le directeur de l'Ined et le chef du département de la démographie de l'Insee, sur proposition du groupe de coordination de la maîtrise d'ouvrage (MOA).

Le comité de pilotage technique :

Il est chargé du suivi technique de l'enquête : déroulement des différentes étapes, calendrier des tests, solutions techniques. Il est composé de la maîtrise d'ouvrage, des concepteurs, du CPOS, du CDAP et des équipes informatiques (responsables des équipes Blaise, Capi et de l'aval). Il est présidé par le chef du département de la Démographie.

La composition de chaque groupe ou comité peut être modifiée sur proposition de ses membres et sous réserve d'acceptation par la maîtrise d'ouvrage. Après un délai de 2 mois sans refus explicite d'une des parties, la modification sera réputée effective.

Dispositions relatives au respect de la loi n° 78-17 du 6 janvier 1978

Démarches auprès du CNIS

L'Insee et l'Ined ont présenté le projet de l'enquête TeO au CNIS en mai 2006. L'avis d'opportunité a été délivré en date du 21 août 2006 sous le n° 215/D130.

L'Insee a considéré que l'avis d'opportunité accordé à l'enquête le 21 août 2006 portait sur l'ensemble du projet mais était insuffisamment explicite sur les questions sensibles. Le questionnaire de l'enquête a donc été soumis à un avis explicite du CNIS, dans la mesure où il contient des questions sensibles, au sens de la CNIL.

L'enquête a obtenu le label « d'intérêt général et de qualité statistique » le 31 mai 2007, sous réserve d'obtention d'un avis d'opportunité du CNIS sur les questions sensibles.

L'avis du CNIS sur les questions sensibles a été délivré lors de l'assemblée plénière du CNIS du 19 décembre 2007, suite à la réunion de la formation Démographie et Conditions de vie du 12 octobre 2007.

Suite à cela, le comité du Label a émis un avis de conformité définitif le XX décembre 2007.

Les avis d'opportunité et de conformité sont joints dans l'annexe 5.

A compter de la date de signature de la présente convention, il est convenu que toutes les démarches entreprises auprès du CNIS doivent faire l'objet d'un accord préalable entre les parties.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Démarches auprès de la CNIL

L'enquête faisant l'objet de la présente convention contient des données concernées par l'article 8 de la loi n°78-17 du 6 janvier 1978 (appelées données sensibles par la suite) ; à ce titre, elle est donc soumise au régime d'autorisation édicté par cette loi.

La procédure spécifique d'échantillonnage de l'enquête a fait l'objet de démarches auprès de la Cnil. En premier lieu, il était nécessaire de faire saisir les noms et prénoms des individus qui figuraient sur leur bulletin individuel de recensement en 2007. Cette démarche a fait l'objet d'un avis modificatif au dossier de saisie du recensement de la population.

Par ailleurs, la procédure de consultation de l'état civil dans les mairies de naissance des individus pour constituer un échantillon de descendants d'immigrés a fait l'objet d'un test, déclaré à la CNIL qui s'est alors prononcée favorablement sur la mise en œuvre de l'échantillon en vraie grandeur.

Les deux réponses de la Cnil figurent en annexe 6.

Le dossier de formalités préalables auprès de la CNIL a été déposé à la suite de cette assemblée, le XX décembre 2007.

A compter de la date de signature de la présente convention, il est convenu que toutes les démarches entreprises auprès de la CNIL doivent faire l'objet d'un accord préalable entre les parties. L'INSEE et l'INED s'engagent à s'informer mutuellement sur toutes les démarches entreprises auprès de la CNIL.

Les différents fichiers de l'enquête : propriété et diffusion

L'Ined et l'Insee seront considérés conjointement comme service dépositaire des données de l'enquête. Différents fichiers seront issus de l'enquête et chacun répondra à des règles différentes quant à leur diffusion et leur utilisation : le fichier brut de collecte, le fichier de production et de recherche (et les fichiers extraits), le fichier « tous publics ».

Seuls le fichier de production et de recherche d'une part et les fichiers « tous publics » d'autre part, peuvent faire l'objet d'une diffusion hors Ined et Insee. Toute cession de fichier à des tiers devra respecter les règles relatives au secret statistique, conformément aux dispositions en vigueur de la loi n°51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière statistiques.

Le fichier brut de collecte

Services dépositaires des données

Le fichier brut de collecte (sortie de CAPI) est le fichier regroupant l'ensemble des variables collectées auprès de l'échantillon total de l'enquête TeO. Ce fichier comporte les réponses des personnes enquêtées à toutes les questions du questionnaire, ainsi que les variables permettant le calcul des pondérations, issues des réponses au recensement. Toutefois, il ne comporte pas de variables directement nominatives (noms des enquêtés et des enquêteurs, adresse des logements enquêtés). L'Insee s'engage à livrer ce fichier à l'Ined dès que possible et, en tout état de cause, dans un délai maximal de 2 mois après la fin de la collecte.

Ces données complètes directement tirées de l'enquête permettent d'identifier les individus par recoupement de variables indirectement nominatives. L'article 6 de la loi n°51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques dispose que le service dépositaire ne peut communiquer les renseignements individuels à quiconque avant un délai de cent ans. Comme pour toutes les enquêtes statistiques, cette obligation sera appliquée avec la plus grande rigueur. L'Ined et l'Insee, considérés conjointement comme

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

service dépositaire, seront tenus aux obligations de protection des données et de non-communication de celles-ci, conformément à l'article de loi susmentionné.

Ce fichier brut ne sera pas utilisé directement à des fins d'étude mais servira à l'élaboration du fichier de production et de recherche.

Déontologie de diffusion des résultats à partir des données complètes

L'Insee et l'Ined, qui auront accès au fichier le plus détaillé de l'enquête, s'engagent à respecter une déontologie que l'on peut résumer de la façon suivante : la mesure des discriminations ne peut se contenter de se fonder sur l'observation des inégalités prises isolément ; elle doit impérativement faire référence à une comparaison « toutes choses égales par ailleurs ». Cette règle n'exclut cependant pas une analyse descriptive préalable dans l'étude des discriminations.

Cette position a été précisée par S. Lollivier lors d'une intervention qu'il a faite le 3 octobre 2006 devant le Haut Commissariat à l'immigration : « Les comparaisons entre groupes doivent se faire (...) dans la mesure du possible, après avoir contrôlé les autres variables qui contribuent à la construction des différences, qu'elles soient démographiques, sociales ou économiques. (...) Les situations de discrimination (...) ne se résument pas au constat des différences ou des inégalités brutes mais (...) doivent persister « toutes choses égales d'ailleurs ». Certes, l'exigence d'une analyse multivariée s'impose dans tous les secteurs de la statistique sociale, mais elle prend une valeur déontologique particulière en présence de variables aussi chargées que les variables d'origine. »

Le fichier de production et de recherche

Contenu et modalités de constitution

Le contenu du fichier de production et de recherche devra être aussi complet que possible sans pour autant rompre l'anonymat. Ce fichier sera accessible aux services statistiques ministériels (SSM) et aux chercheurs selon des modalités définies à la section b ci-après.

Certaines variables seront par conséquent exclues du fichier de production et de recherche :

- prénoms
- données géographiques à des niveaux commune ou inférieur
- jour de naissance des différentes personnes du logement

D'autres variables seront supprimées et réintroduites sous des formes regroupées (typologies...) :

- le pays de naissance et la nationalité
- toutes les questions ouvertes renseignées en clair

Certaines variables en clair, comme la religion ou la profession, pourront être réintroduites après une suppression des libellés qui risqueraient de rompre l'anonymat (par exemple « préfet »).

Des variables contextuelles externes, décrivant les caractéristiques sociodémographiques du voisinage seront introduites dans le fichier. Les thèmes couverts seront la composition

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

démographique, l'emploi, les revenus, les équipements. Elles seront suffisamment agrégées pour ne pas rompre le principe d'anonymat mais elles seront autant que possible élaborées en utilisant le zonage le plus pertinent selon les sujets : région, département, bassins de vie ou zones d'emploi, commune, zone Iris en ville (sans que ni l'Iris lui-même ni la commune ne figurent dans le fichier).

Par ailleurs, au-delà du respect de l'anonymat, le caractère sensible de certaines variables, au sens de la Cnil, devra être pris en compte lors de leur introduction dans le fichier de production et de recherche et leur diffusion (cf. infra).

Un groupe dit de « constitution de fichier » sera mis en place à l'issue de la collecte. Ce groupe comprendra les membres du groupe de travail et d'exploitation de l'enquête TeO (liste des membres en annexe 3), auxquels s'ajouteront des experts de l'Insee et de l'Ined nommés par les directeurs de ces deux parties. Ce groupe a vocation à formuler les recommandations qui présideront à la constitution du fichier de production et de recherche.

Pour la constitution du fichier de production et de recherche et afin de construire des regroupements ou typologies à intégrer au fichier de production et de recherche, le groupe de constitution du fichier pourra disposer de fichiers extraits du fichier brut de collecte, totalement anonymes, non appariés avec le fichier de production et de recherche.

Suivant ces recommandations, la cellule Statistiques et Études sur l'Immigration de l'Insee constituera un fichier de production et de recherche à partir du fichier brut. Ce fichier prendra en compte les redressements de variables, les recodifications, les pondérations et les variables contextuelles externes.

En cas de désaccord sur la constitution du fichier de production et de recherche, les arbitrages devront être réalisés par le directeur de l'Ined et le chef du département de la démographie de l'Insee,

Un fichier de production et de recherche provisoire sera mis à disposition du groupe de constitution dans un délai maximal de deux mois après la fin de la collecte. Ce fichier sera rendu définitif dans un délai de 12 mois après la fin de la collecte.

Certaines variables, notamment celles issues de données contextuelles et des questionnaires auprès des enfants des personnes enquêtées définis à l'article 11, ne seront sans doute pas disponibles dans leur version définitive un an après la collecte. Dans ce cas, une nouvelle livraison du fichier de production et de recherche, augmenté de nouvelles variables, sera réalisée ultérieurement. A sa réception, l'ensemble des destinataires détruiront leur fichier pour le remplacer par la dernière version.

Calendrier et règles de diffusion du fichier de production et de recherche

Le fichier de production et de recherche est destiné aux services statistiques ministériels (SSM) et aux chercheurs.

La diffusion du fichier de production et de recherche ne pourra pas commencer avant la sortie des publications princeps de l'Insee et de l'Ined (Population et sociétés et Insee Première). Il devra être disponible dans un délai de 12 mois après la fin de la collecte.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Le fichier de production et de recherche sera rendu disponible aux chercheurs par une habilitation donnée par le réseau Quételet, avec lequel chacune des parties a passé une convention. Le réseau Quételet s'assurera notamment du statut réel de chercheur de la personne qui demande l'accès au fichier, du bien-fondé du projet de recherche scientifique présenté et de la nécessité d'avoir recours au fichier de production et de recherche pour mener à bien ce projet. Lorsque l'accord est donné, une licence d'usage spécifiant le sujet de la recherche et le fichier de production et de recherche sollicité est signée par le chercheur et visée par le directeur du laboratoire.

L'accès aux données sensibles du fichier de production et de recherche ne sera possible que si le projet de recherche le nécessite et après avis du Conseil scientifique du Comité de concertation pour les données en sciences humaines et sociales.

Passé les douze premiers mois après la collecte, les chercheurs souhaitant obtenir un fichier totalement anonymisé extrait du fichier brut de collecte pour des besoins d'étude spécifiques pourront faire une demande à la maîtrise d'ouvrage. Les deux parties de la maîtrise d'ouvrage devront signifier leur décision dans un délai de 2 mois. En cas d'autorisation, le fichier devra être élaboré par un des deux services dans un délai de 6 mois.

Le fichier « bases de sondage »

Ce fichier anonyme contient les variables issues du recensement des personnes initialement tirées dans l'échantillon, même si elles n'ont pas été enquêtées, ainsi que l'indication de réalisation de l'enquête ou, en cas d'échec, de la raison de l'échec (refus de l'individu, personne impossible à joindre, autre refus, ménage impossible à joindre, etc). Il sera exclusivement exploité à des fins de calcul des pondérations et d'études de la représentativité et de la qualité des données. Il est la propriété de l'Insee. Il est élaboré dans un délai de deux mois suivant la collecte. Seul l'Insee pourra disposer du fichier « base de sondage » dans le cadre de l'élaboration et de la validation du fichier de production et de recherche. Toutefois, les personnels de l'Ined ou mandatés par l'Ined pourront avoir accès à ce fichier à l'Insee, dans ses locaux ou dans un centre sécurisé, par simple avenant à cette convention précisant leur nom, la durée de leur mise à disposition et le but de la recherche, dans le but unique de travailler sur qualité des données par l'étude des non-réponses et autres échecs de collecte, et pour comparer l'échantillon des personnes sélectionnées à la population cible fixée avant l'enquête. A cette fin l'Insee conserve et archive ce fichier.

Fichier « tous publics »

Ce fichier « tous publics » répond à la politique de l'Insee en matière de mise à disposition d'enquêtes. C'est un fichier anonyme allégé destiné à être mis en libre accès sur internet. Il ne comportera aucune donnée sensible. Son contenu précis sera défini par le groupe « constitution de fichier ».

Le fichier « tous publics » sera mis en libre accès sur internet par l'Insee et l'Ined à compter d'un délai maximum de 12 mois après la fin de la collecte.

Utilisation et valorisation des données

Les données de TeO feront l'objet de plusieurs phases de valorisation ci-dessous explicitées et synthétisées dans l'annexe 5 décrivant le schéma d'accès aux données.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Documentation technique et premières publications

Au terme de la préparation du fichier de production et de recherche (au maximum 12 mois après la fin de la collecte), l'Insee et l'Ined s'engagent à publier les documents suivants :

- un rapport méthodologique, sur les conditions de réalisation de l'enquête et d'utilisation des données, ainsi qu'une documentation du fichier de production et de recherche ;
- une première publication institutionnelle : dans un premier temps, dans les revues Population et sociétés et Insee Première, si possible en sortie simultanée ; et, dans un second temps, dans des publications restreintes (4 pages) des partenaires financiers. Ces publications seront les premières communications publiques des résultats de l'enquête TeO.

Ouvrage collectif

L'Ined s'engage à coordonner la production d'un ouvrage collectif pour valoriser les données de TeO. Cet ouvrage sera produit dans le cadre du groupe de travail et d'exploitation (cf. liste des membres en annexe 3). Dans la mesure du possible, cet ouvrage sera soumis à publication dans un délai de 14 mois à compter de la date d'ouverture de l'accès aux fichiers d'études, c'est-à-dire dans un délai de 26 mois après la fin de la collecte.

Autres publications

Toute autre publication issue de l'exploitation des données de TeO ne pourra pas être soumise à publication avant un délai de 12 mois après la fin de la collecte.

Règle de citation

Toutes les publications qui exploiteront les données de l'enquête TeO devront explicitement la citer sous la forme suivante : « Enquête TeO (Trajectoires et Origines), Ined-Insee, 2008 ».

PROTECTION eT Archivage des données

En tant que co-dépositaires des données et tout particulièrement du fichier brut de collecte, l'Insee et l'Ined s'engagent en matière de protection des données selon les modalités précisées en annexe 6.

Le droit d'accès et de rectification des enquêtés sera exercé auprès des directions régionales de l'Insee. L'Insee transmettra à l'Ined le fichier comportant les modifications, que ce dernier mettra en œuvre sans délai.

L'Insee est responsable de la transmission aux Archives de France du fichier brut de collecte et du fichier de production et de recherche communs à l'Insee et l'Ined. Le fichier base de sondage sera également transmis par l'Insee aux Archives de France.

Budget de l'opération

Le coût estimatif externe total de l'opération est évalué à 1 962 000 euros. Les coûts internes « Insee », comprenant la collecte, l'échantillonnage, la « capisation », sont évalués à 2 606 000 euros. Les coûts internes « Ined », comprenant notamment les frais de personnel et de missions, sont évalués à 662 351 euros.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Modalités financières

Cette partie est à l'étude.

Opérations complémentaires

Les opérations complémentaires décrites ci-dessous font partie intégrante du projet et sont examinées au même titre que l'enquête par le CNIS, le comité du Label et la CNIL. La déclaration à la CNIL prévue à l'article 5 de la présente convention devra tenir compte de la prévision de ces opérations complémentaires.

Enquêtes qualitatives

Il est prévu dans un délai maximum de trois années après la fin de collecte de l'enquête TeO la réalisation de vingt post-enquêtes qualitatives non obligatoires maximum, portant chacune sur un échantillon restreint à une cinquantaine de personnes ayant donné leur accord lors de l'enquête TeO 2008 pour être à nouveau contactées.

Les entretiens seront réalisés selon les modalités suivantes :

Sous condition que leur accord explicite ait été recueilli, les personnes interrogées seront sélectionnées en fonction de leurs réponses à l'enquête quantitative.

L'Ined coordonnera les demandes des différents partenaires désireux de réaliser des entretiens qualitatifs. La maîtrise d'ouvrage devra donner son accord sur les projets nécessitant des entretiens. L'Ined coordonnera l'étape de sélection des personnes à interroger, et fournira à l'Insee la liste des identifiants correspondant aux personnes ainsi sélectionnées.

L'Insee, via le département de la Démographie, s'engage à faire remonter des Directions régionales concernées les fiches adresses correspondant à cette liste d'identifiants. Ces fiches adresse seront conservées par le département de la démographie pour une durée maximale de 18 mois à compter de la notification de la présente convention.

Les entretiens seront effectués par des enquêteurs ou chercheurs mis à disposition de l'Insee sous la responsabilité du chef de la division « Enquêtes et études démographiques », après signature d'un engagement au secret statistique. Une fois les fiches adresse mises à leur disposition, celles-ci devront être utilisées pour contacter les personnes dans un délai de 3 mois.

Une fois les entretiens réalisés ou en cas d'impossibilité de réalisation, les enquêteurs ou chercheurs s'engagent à détruire les informations nominatives fournies par l'Insee ou recueillies lors des entretiens.

Les enquêteurs ou chercheurs ayant réalisé des entretiens qualitatifs retranscriront ces entretiens de façon anonyme par suppression des noms et prénoms. Le lien avec le fichier quantitatif de l'enquête (identifiant mentionné au point 2) pourra être conservé.

Les publications réalisées à partir de ces entretiens porteront la mention « Entretiens réalisés en [année de réalisation] à la suite de l'enquête TeO (Trajectoires et origines), Ined-Insee, 2008 ».

Le détail des modalités pratiques de réalisation des enquêtes qualitatives fera l'objet d'un avenant.

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 3 :
Projet de convention Insee-Ined

Enquête auprès des enfants des personnes enquêtées

Une enquête postale ou par dépôt retrait sera réalisée auprès des enfants des personnes enquêtées, ayant entre 15 et 24 ans et résidant avec elles. Cette enquête a reçu un avis d'opportunité favorable du CNIS le 24 mai 2007. Elle a également été examinée par la comité du Label le 17 décembre 2007 et elle est entérinée par l'avis de conformité établi le xx décembre 2007. Le fichier issu de la saisie de ces questionnaires sera joint au fichier brut de collecte et après anonymisation au fichier de production et de recherche le plus rapidement possible.

Tribunal compétent

En cas de litige, les parties s'engagent à tenter de le résoudre à l'amiable ; en cas d'échec, le litige sera porté devant le tribunal administratif de Paris.

Pièces annexes

Les annexes suivantes font partie intégrante de la présente convention :

- Annexe 1 : Objectifs de l'enquête
- Annexe 2 : Questionnaire du test de novembre-décembre 2007
- Annexe 3 : Calendrier
- Annexe 4 : Composition des différents groupes
- Annexe 5 : Schéma d'accès aux données
- Annexe 6 : Sécurité des données
- Annexe 7 : Liste des partenaires financiers

Fait à Paris, le
En 2 exemplaires originaux

Pour l'Ined,
Le Directeur

Pour l'Insee,
Le Directeur général

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 4 :
Sommaire du questionnaire principal

Le questionnaire est joint en document autonome ; sa table des matières est présentée ci-après.

TABLE DES MATIERES

1. TABLEAU DES HABITANTS DU LOGEMENTS	page 4
2. REVENUS	page 15
3. NATIONALITÉ ET ORIGINE DES PARENTS	page 17
4. LANGUES	page 23
5. TRAJECTOIRES MIGRATOIRES, RAPPORT AU PAYS D'ORIGINE	page 27
6. RELATIONS FAMILIALES ET TRANSMISSIONS	page 34
7. VOS ENFANTS	page 44
8. IMAGE DE SOI ET REGARD DES AUTRES	page 46
9. EDUCATION	page 50
10. VIE PROFESSIONNELLE	page 65
11. RELIGION	page 94
12. VIE MATRIMONIALE	page 97
13. LOGEMENT ET CADRE DE VIE	page 128
14. VIE CITOYENNE	page 138
15. SANTÉ	page 141
16. DISCRIMINATIONS	page 146
17. RELATIONS SOCIALES	page 150
18. FIN DE L'ENTRETIEN	page 154
19. QUESTIONNAIRE QUALITÉ	page 155

Dénomination du traitement : **TeO**
Enquête « trajectoires et origines » sur la diversité des
populations en France

Documentation complémentaire 5 :
Note sur les questionnaires 15-24 ans

Les questionnaires destinés aux jeunes sont joints en documents autonomes.

Pour des raisons techniques, la version des questionnaires destinés aux jeunes qui figure ci-après est celle que le Comité du label réuni le 17 décembre a examiné ; conformément aux remarques formulées par le Comité du label, les questionnaires seront aménagés de la façon suivante :

1. les questionnaires porteront en première page le cartouche d'informations légales qui figure également sur la lettre-avis (cf. annexe 13)
2. la question A9 qui figure à l'identique sur les deux questionnaires sera corrigée (elle comporte - par erreur - deux fois la modalité « chanson et variété française »)
3. la modalité « Ne sait pas » de la question A20 qui figure à l'identique sur les deux questionnaires (« diriez-vous que vos relations avec vos parents sont ... ») est inappropriée et sera remplacée par la modalité « ne se prononce pas »
4. la question E3 du questionnaire destiné aux étudiants, collégiens et lycéens, qui concerne les discriminations ressenties seront reformulées conformément aux dispositions prévues dans le cadre du questionnaire principal, afin de permettre des comparaisons. Comme cela a été fait dans le questionnaire principal, les principaux motifs de discrimination seront explicitement mentionnés et différenciés (sexe, couleur de peau, origine ou nationalité, état de santé ou handicap).
5. les questions C4 et E3 du questionnaire destiné aux jeunes en emploi ou en recherche d'emploi, qui concernent également les discriminations seront modifiée de la même façon. Pour ces deux questions, l'item « Rien de tout cela » sera remplacé par « Non ».
6. dans la même optique, la question G7 du questionnaire destiné aux jeunes en emploi ou en recherche d'emploi, qui concerne les conditions importantes pour réussir professionnellement, sera modifiée, pour intégrer plus d'items notamment le sexe, l'apparence physique, l'âge et pour distinguer l'origine sociale de l'origine nationale ou du quartier d'origine. Le nombre de modalités à choisir pourraient être en conséquence modifié.
7. dans la rubrique G du questionnaire destiné aux étudiants, collégiens et lycéens, qui concerne l'aide apportée par les parents, les modalités « votre mère » et « votre belle-mère » seront regroupées en une seule « votre mère ou votre belle-mère » ; de même, les modalités « votre père » et « votre beau-père » seront regroupées en une seule « votre père ou votre beau-père »
8. dans la rubrique C du questionnaire destiné aux jeunes en emploi ou en recherche d'emploi, qui concerne la recherche d'emploi, la modalité « vous avez déjà un emploi » sera ajoutée aux modalités de réponse à la question C5 (« pourquoi ne cherchez-vous pas [d'emploi] ? »)