

N° de catalogue 95-16

**MODÉLISATION DES FICHIERS DE DONNÉES
POUR LES ENQUÊTES LONGITUDINALES**

Numéro d'enregistrement du produit 75F0002M

Août 1995

Philip Giles, Division des enquêtes-ménages

Pierre Lafrance, Division des enquêtes-ménages

La série de documents de recherche de l'EDTR est conçue en vue de communiquer les résultats des études ainsi que les décisions importantes ayant trait à l'Enquête sur la dynamique du travail et du revenu. Ils sont offerts en français et en anglais. Pour obtenir une description sommaire des documents disponibles ou un exemplaire de ces documents, communiquez avec Philip Giles, EDTR, par la poste à Édifice Jean-Talon, 11^e étage, section D8, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6; par INTERNET: GILES@STATCAN.CA; par téléphone au (613) 951-2891; ou par télécopieur au (613) 951-3253.

SOMMAIRE

Le présent document a été présenté à la rencontre annuelle de 1995 de l'American Statistical Association à Orlando (Floride).

Les fichiers de données classiques provenant des enquêtes dérivent des réponses aux questions de l'enquête, posées dans l'ordre où elles apparaissent sur le questionnaire. Une série de variables dérivées (c'est-à-dire dérivées de deux ou de plusieurs réponses) s'ajoute à la fin de chaque enregistrement. Les progrès récents réalisés dans le domaine des enquêtes, en l'occurrence l'introduction généralisée des enquêtes longitudinales et des interviews assistées par ordinateur ont considérablement compliqué cette approche, sans pour autant la rendre irréalisable.

Le présent article porte sur le développement de fichiers de données, ou la modélisation des données, pour les enquêtes longitudinales.

TABLE DES MATIÈRES

	Page
1. Introduction	1
2. Principes fondamentaux d'une enquête longitudinale	1
3. Approche à la modélisation des données	3
3.1 Bases de données relationnelles	4
3.2 Description du modèle	6
4. Aperçu de l'Enquête sur la dynamique du travail et du revenu (EDTR)	9
5. Le modèle de données de l'EDTR: principes directeurs	11
6. Élaboration du modèle de l'EDTR	15
7. Problèmes particuliers à l'élaboration du modèle de l'EDTR	16
8. Le modèle de l'EDTR: particularités fondamentales et futures orientations	25
9. Les fichiers de microdonnées de l'EDTR et l'interface avec l'utilisateur	28
ANNEXE 1: Diagrammes entités-relations (DER) et dictionnaire des données (DD) de l'Enquête sur la dynamique du travail et du revenu	
ANNEXE 2: Organisation de la base de données de l'EDTR et liste partielle des variables	

1. Introduction

Les fichiers de données classiques provenant des enquêtes dérivent des réponses aux questions de l'enquête, posées dans l'ordre où elles apparaissent sur le questionnaire. Une série de variables dérivées (c'est-à-dire dérivées de deux ou de plusieurs réponses) s'ajoute à la fin de chaque enregistrement. Les progrès récents réalisés dans le domaine des enquêtes, en l'occurrence l'introduction généralisée des enquêtes longitudinales et des interviews assistées par ordinateur, ont considérablement compliqué cette approche, sans pour autant la rendre irréalisable.

Nous examinerons ici la création de fichiers de données, ou la modélisation des données, dans le contexte des enquêtes longitudinales. Après une description générale, le document analyse une nouvelle enquête longitudinale, soit l'Enquête sur la dynamique du travail et du revenu de Statistique Canada.

2. Principes fondamentaux d'une enquête longitudinale

Lorsqu'on reprend périodiquement une enquête, l'interview d'un nouvel échantillon de personnes, comme on le fait dans la plupart des cas, ou l'interview du même groupe à plusieurs reprises, comme cela se produit lors d'une enquête longitudinale, ne donneront pas les mêmes résultats. L'avantage réputé des échantillons transversaux (c'est-à-dire d'un échantillon au moins en partie différent à chaque enquête) est que ces derniers sont habituellement plus représentatifs et révèlent mieux que les autres les niveaux et les tendances de la population générale et des sous-groupes qui la composent.

Avec l'enquête longitudinale, au lieu de s'arrêter aux mesures statiques et au changement net, on s'intéresse plutôt à la série complète de transitions, de durées et de répétitions propres aux paramètres qu'on souhaite mesurer. Bon nombre d'enquêtes longitudinales prennent comme période de rappel le laps de temps qui

sépare deux interviews. Par exemple, lors d'une enquête longitudinale annuelle, on priera le répondant de fournir des renseignements sur les douze mois antérieurs. De cette manière, les données temporelles seront continues pour chaque répondant, du moins sur le plan théorique¹.

Un autre aspect majeur des enquêtes longitudinales concerne l'élaboration de *règles de suivi*, soit de la procédure qui déterminera quelles personnes seront interrogées à chaque collecte de données. Ces problèmes, simples en apparence, gagnent souvent vite en complexité sur le plan pratique. Ainsi, dans la plupart des enquêtes longitudinales on souhaite suivre les personnes lorsqu'elles déménagent. Malheureusement, il existe plusieurs sortes de déménagements. Pour certaines personnes, le déménagement peut entraîner un changement de situation, si bien que la teneur de l'enquête n'a plus la pertinence qu'elle avait auparavant. D'autres peuvent quitter le pays ou être hospitalisées. Sans doute est-il souhaitable de vérifier le changement de lieu de résidence, mais interroger ces personnes dans le cadre de l'enquête devient futile.

Si l'avènement des enquêtes longitudinales soulève de nouvelles difficultés au niveau de la collecte des données, il en va autant pour la création des fichiers. Nous nous efforcerons d'identifier ces difficultés et verrons comment on a tenté de les surmonter dans le cadre d'une enquête particulière.

¹ Les problèmes habituels que suscitent les enquêtes, notamment les non-réponses et les erreurs de remémoration, rendent ce concept de *continuité temporelle* partiellement véridique. Par ailleurs, les enquêtes longitudinales où l'on ne recueille pas *continuellement* les données ne permettent pas toujours de déterminer la durée des périodes (par exemple, les périodes de tel ou tel état civil ou les périodes de chômage), ni la séquence des événements (ce qui survient en premier : la perte de l'emploi ou la séparation). Néanmoins, ces enquêtes ne poursuivent pas toujours de tels objectifs, si bien que la perte de ces renseignements ne présente pas d'importance.

3. Approche à la modélisation des données

Certaines définitions de base s'imposent. (Bien que quelques définitions et principes qui suivent aient une application générale, la discussion porte essentiellement sur les fichiers de données des enquêtes.) En termes simples, l'expression *modélisation de données* désigne la création d'une base de données structurée dans laquelle seront emmagasinés les résultats de l'enquête. On trouvera plus loin une explication plus détaillée. Une base de données est un ensemble de données organisé de manière à faciliter la saisie, l'extraction et la mise à jour de l'information. Un système de gestion de base de données (SGBD) est un ensemble de logiciels utilisé pour gérer une base de données. Le SGBD traite les demandes relatives à l'ajout, à la manipulation ou à l'extraction des données, ou s'occupe des programmes correspondants.

Les bases de données des enquêtes sont habituellement constituées de fichiers *plats*, c'est-à-dire comptant une entrée par unité d'enquête (personne ou entreprise, par exemple). Les variables associées à l'unité d'enquête font partie du même enregistrement. Les enquêtes gagnent cependant de plus en plus en complexité, ainsi qu'on peut le constater avec les enquêtes longitudinales. Les fichiers classiques présentant des inconvénients, on envisage maintenant d'autres sortes de bases de données. Voici trois possibilités :

- C l'approche relationnelle;
- C l'approche hiérarchique;
- C l'approche du réseau.

Décrire et comparer ces trois approches dépasserait du cadre du présent document. Il suffira de dire que la première est nettement sur le point de devenir la norme au sein de la profession. Oracle, Paradox, FoxPro et DB2 sont des

systèmes de gestion de bases de données relationnelles actuellement en utilisation. Notre discussion se limitera donc à ces dernières.

3.1 Bases de données relationnelles

Par base de données relationnelle, on entend une base de données ayant la forme d'une série de *relations* ou de *tableaux* bidimensionnels. En voici un exemple, que nous appellerons PERSONNE:

IDPERSONNE	NOM	ÂGE	SEXE
011100000001	Luc	22	M
011200000002	Paul	37	M
021100000003	Diane	27	F

La terminologie dérive des mathématiques. Chaque tableau représente un cas particulier du concept mathématique appelé *relation*. Une rangée correspond à un *tuple*, alors qu'une colonne est un *attribut*. Pour les profanes, un tableau relationnel ressemble à un fichier bidimensionnel ou *plat+ ordinaire. Bien que sa définition soit beaucoup plus restreinte au sens strict, en ce qui nous concerne, on peut considérer ces deux expressions équivalentes.

La base de données relationnelle peut comprendre plus d'un tableau. Aux fins qui nous intéressent, c'est cette possibilité même qui distingue la base de données relationnelle du fichier bidimensionnel. Voyons un autre tableau énumérant les emplois d'une personne. Nous l'appellerons EMPLOI.

IDPERSONNE	EMPLOI N°	SALAIRE HORAIRE	ANNÉES DE TRAVAIL
011100000001	1	7,50 \$	1,50
011100000001	2	5,35 \$	1,00
011200000002	1	14,75 \$	12,75
021100000003	1	10,00 \$	7,00
021100000003	2	9,25 \$	1,50
021100000003	3	12,30 \$	0,25

En examinant un seul tableau EMPLOI, on constate (d'après les entrées correspondant au même numéro IDPERSONNE) qu'il y a en fait trois personnes, et que la première a eu deux emplois, la seconde un seul et la dernière trois. Les valeurs de IDPERSONNE sont cependant les mêmes qu'au tableau PERSONNE. En reliant les deux tableaux et faisant correspondre les entrées au moyen de la valeur IDPERSONNE, on obtient d'autres renseignements, notamment l'âge et le sexe de la personne qui détient l'emploi.

Pour manipuler les bases de données relationnelles, on recourt habituellement aux commandes du langage relationnel SQL, une norme de l'ANSI (American National Standards Institute). En choisissant des logiciels qui épousent cette norme pour créer la base de données, l'utilisation d'autres gestionnaires de base de données pour accéder à l'information ne soulève aucune difficulté.

L'usage d'une base de données relationnelle présente plusieurs avantages.

- C Les tableaux sont simples mais bien structurés.
- C Les données qui présentent des liens logiques peuvent se retrouver dans le même tableau (p.ex. renseignements relatifs au certificat ou au diplôme

d'une personne). L'utilisateur peut alors se concentrer sur les aspects qui l'intéressent le plus.

- C On peut réduire le nombre de redondances. (Dans l'exemple qui précède, on ne répéterait pas l'âge et le sexe de la personne pour chaque nouvel emploi.)
- C La structure relationnelle permet d'économiser de l'espace à la saisie des données.
- C Les données peuvent être partagées.
- C Les incohérences sont plus faciles à éviter et on peut préserver l'intégrité des données.
- C Une certaine normalisation est réalisable (à savoir, les dates ont toutes le même format).
- C La source d'information commune est régie par un personnel centralisé.
- C Il est possible d'imposer certaines mesures de sécurité et de présenter les données de diverses manières (p. ex. autoriser l'accès à quelques tableaux ou champs d'un tableau).

3.2 Description du modèle

En termes d'informatique, on appelle *modèle de données+ la représentation graphique d'objets ou d'entités concrètes et des relations qu'ils partagent entre eux. Dans notre exemple, les objets en question sont une personne et un emploi. La modélisation consiste à déterminer exactement les renseignements qui apparaîtront dans la base de données, bref à identifier les entités présentant de l'intérêt et l'information sur ces entités à saisir. La modélisation sert aussi à identifier les liens entre différentes entités.

La modélisation s'effectue au moyen de diagrammes entité-relation (DER) et du dictionnaire des données (DD). On trouvera à l'annexe 1 des exemples de chacun,

provenant de l'Enquête sur la dynamique du travail et du revenu. La version française de ces documents n'existant pas encore, l'annexe est présentée en anglais.

Le DER expose les paramètres et leurs liens. Il intègre les éléments fondamentaux que voici:

C Entités

Il s'agit des tableaux, en l'occurrence dans notre exemple, les tableaux PERSONNE et EMPLOI.

C Relations

Les entités partagent des liens entre elles. En d'autres termes, elles ont des relations dont on peut préciser la cardinalité. Ainsi, chaque Personne *peut avoir* un ou plusieurs Emplois.

C Attributs

Des attributs décrivent les entités. L'*âge*, le *sexe* et le *nom* sont des attributs de l'entité Personne. Ces attributs peuvent être repris dans le DER ou, s'ils sont trop nombreux, être décrits dans le dictionnaire des données.

Le dictionnaire des données (DD) accompagne les DER puisqu'il en définit les composantes. À chaque entité correspond une entrée dans le dictionnaire et celle-ci comprend:

C la définition du but ou du rôle de l'entité dans le modèle;

C les paramètres ou les attributs qui décrivent l'entité;

C les identificateurs permettant de distinguer une entité des autres.

L'analyse théorique qui aboutit à la définition du modèle de données réclame beaucoup d'efforts et de temps, mais la mise en oeuvre s'en trouve accélérée. Dès

qu'on crée un fichier selon le modèle, les éléments du modèle ont leur équivalent dans le fichier:

<u>Modèle</u>	<u>Fichier</u>
Entités	Tableaux
Attributs	Champs
Identificateurs	Zones de contrôle

Voici les principaux avantages d'un modèle de données :

- C Les diagrammes ER graphiques (voir l'exemple de l'annexe 1) sont faciles à comprendre, non seulement pour les analystes de système mais aussi pour les gestionnaires, les spécialistes des domaines, les méthodologistes et les utilisateurs.
- C Les DER et le DD constituent une excellente source d'information.
- C Ils facilitent la communication entre les analystes et les spécialistes des domaines quand vient le moment de préciser les exigences de l'utilisateur en matière d'information.
- C La création du modèle exige qu'on détermine dès le départ en quoi consisteront les données et quelle en sera la précision. De cette façon on clarifie le cadre théorique et évite l'omission de certaines possibilités d'analyse.
- C Le modèle crée une structure facilitant l'exploitation des données.
- C Si la représentation des entités tout comme les relations entre ces dernières sont valables dans le modèle (en théorie), elles devraient l'être également dans les fichiers (à savoir, base de données relationnelle).
- C Un modèle de données simplifie le développement et l'implantation d'un système de traitement et de production.
- C Lors de l'élaboration d'un système de traitement, l'existence d'un modèle de données facilitera l'identification des procédés auxquels on peut ou non avoir recours. On aura une vue d'ensemble des attributs utiles qui peuvent

de l'échantillon. On possédera six années de données longitudinales sur environ 31 000 personnes et trois années de données sur le double de cette population, composée de répondants des deux panels. La figure 1 illustre la rotation de l'échantillon.

Treize interviews auront lieu au cours des six années de l'enquête. Lors du lancement d'un panel, une interview préliminaire servira à recueillir les données de base, soit les renseignements démographiques et des précisions sur l'éducation et l'expérience au travail. Le cycle d'interviews rétrospectives sur le travail et le revenu débute un an plus tard. Chaque année, en janvier, on interrogera la personne au sujet de ses activités sur le marché du travail durant l'année précédente; en mai, on en fera autant mais pour le revenu et sa provenance. Toutes les données (sauf celles de l'interview préliminaire qui a eu lieu en janvier 1993 pour les répondants du premier panel) sont recueillies par interview assistée par ordinateur (IAO).

Les membres du ménage des habitations sélectionnées en janvier 1993 sont considérés comme des répondants longitudinaux. Ils seront donc suivis pendant six ans - jusqu'à la fin de 1998. Les personnes qui emménagent avec eux au cours de ces six années seront intégrées à l'enquête et répondront aux mêmes questions. Avec le temps, on assistera à une hausse du nombre de ménages dans le panel, en raison des ruptures. Cette hausse sera néanmoins compensée par la réduction naturelle de l'échantillon attribuable aux abandons.

On trouvera à l'annexe 2 la liste sommaire des variables de l'enquête et un tableau décrivant les principaux types de données. L'examen des variables de l'EDTR révélera les liens complexes qui existent entre elles, signe de la complexité des structures qui serviront à les stocker.

5. Le modèle de données de l'EDTR: principes directeurs

Plusieurs décisions importantes ont dû être prises d'emblée lors du développement du modèle de données de l'EDTR. Ces décisions en sont devenues les *principes directeurs+.

Il est préférable d'envisager un développement à long terme plutôt que de répondre aux besoins immédiats et de modifier le modèle quand le besoin s'en fait sentir.

Il s'agit de la première grande décision en ce qui concerne les fichiers de sortie. Le développement des premiers fichiers a exigé plus de temps, mais on y a plus que regagné par la suite. Tout d'abord, les utilisateurs ne sont pas contraints de s'adapter à un produit qui changera d'année en année. En second lieu, on estimait qu'une telle approche accélérerait la diffusion ultérieure des fichiers (la troisième, quatrième ou cinquième année). Troisièmement, après une période de familiarisation initiale, l'équipe chargée du projet, les personnes responsables du traitement surtout, aurait une vision plus nette de l'approche retenue. On éliminerait donc une *cible mouvante+.

Concevoir la structure des fichiers de façon à accueillir un maximum de données (à savoir six ans) au lieu de rebâtir les fichiers d'année en année.

Cette application découle directement du principe précédent. En un sens, on pourrait dire qu'elle est indispensable. Elle signifie qu'il fallait réfléchir à l'aspect que le fichier aurait six ans plus tard, exercice délicat mais utile.

Les variables de sortie devraient aider l'utilisateur et pas seulement refléter les questions posées (à savoir, les fichiers devraient s'articuler sur les variables, pas sur les questions).

Avec ce principe, on reconnaît de façon explicite qu'il n'existe pas de lien direct entre les éléments utiles pour l'analyse (variables) et les questions posées lors de l'enquête. La vaste majorité des utilisateurs s'intéressent aux variables, pas nécessairement à la manière dont on les a obtenues.

Cette approche a ralenti l'élaboration du modèle de données et prolongé la mise au point des systèmes de traitement. Néanmoins, nous croyons qu'elle contribuera à simplifier un fichier complexe en réduisant le nombre de manipulations auxquelles doit se plier l'utilisateur. Une de nos craintes était que les personnes techniquement moins à l'aise ne se servent pas des fichiers de l'EDTR à cause de leur complexité. La mesure qui précède devrait régler le problème.

Les données de sortie devraient paraître n'avoir été obtenues qu'une seule fois au bout de six ans.

Les données sont recueillies annuellement, si bien qu'à la fin de l'étude d'un panel, on disposera de six séries de données par sujet, chaque série couvrant une période d'un an différente. L'approche la plus simple consisterait à fournir à l'utilisateur les données sous le même format. Cependant, comme on l'a fait pour les variables, les données des différentes années seront fusionnées afin que l'utilisateur ne s'enlise pas dans les manipulations.

L'exemple illustré à la figure 2 montre comment le fichier élimine les chevauchements. Il représente six emplois qui existaient au cours de la première année de référence.

Les fichiers de microdonnées destinés au public ont une structure identique à celle du fichier principal d'usage interne.

Ce principe n'a pas autant d'incidence que les autres sur le développement du modèle, mais il a d'importantes conséquences pour ce qui est de simplifier l'accès de l'utilisateur aux données de l'EDTR. Ainsi, pour protéger la confidentialité des répondants, le fichier interne exige habituellement deux manipulations : on annule les valeurs de certaines variables (à savoir, on supprime la variable; la date de naissance par exemple) et on agrège les valeurs de certaines variables (c'est-à-dire en donnant le groupe d'âge au lieu de l'âge des sujets).

Les variables supprimées continueront de faire partie de la structure des fichiers d'usage public de l'EDTR. Toutefois, aucune valeur n'y sera assortie. Les variables dont les valeurs ont été agrégées seront considérées comme si elles avaient été supprimées; en d'autres termes, la variable se retrouvera dans la structure du fichier mais aucune donnée ne s'y rapportera. La version agrégée de la variable s'ajoutera au modèle, et les données apparaîtront dans le fichier de microdonnées destiné à la population.

Pourquoi fournir des variables sans données ? Actuellement, les utilisateurs qui aimeraient se servir des données que renferment les fichiers d'usage interne de Statistique Canada doivent obtenir l'information du personnel du Bureau. Dans la mesure où l'information réclamée n'enfreint pas les règles de confidentialité, le personnel accède à la requête de l'utilisateur. Ce dernier paie les frais engagés dans l'exercice, notamment les efforts déployés par les employés pour satisfaire à sa demande et le temps d'utilisation de l'ordinateur. Selon la complexité de la demande, il se peut qu'un certain laps de temps s'écoule avant l'obtention des renseignements souhaités.

En donnant une structure identique aux fichiers de microdonnées d'usage public, on estime que l'utilisateur pourra voir la gamme complète de variables existantes et concevoir un logiciel qui l'aidera à extraire l'information dont il a besoin. Le programme pourra être transmis à Statistique Canada sur support électronique. Un membre de l'équipe l'exécutera et verra s'il respecte les critères de confidentialité, puis renverra les résultats à l'utilisateur de la même manière. Cette approche économisera du temps au personnel de Statistique Canada, ce qui réduira le coût des demandes. On devrait aussi assister à un raccourcissement des délais. Ce «télé-accès», comme on l'appelle, n'est pas encore réalisable à Statistique Canada, mais les travaux vont bon train, et le personnel de l'EDTR voulait s'assurer de pouvoir y recourir le moment voulu.

6. Élaboration du modèle de l'EDTR

Le modèle de données de l'EDTR a vu le jour de la façon décrite ci-après. On s'est servi d'un programme d'ingénierie logicielle assistée par ordinateur pour enregistrer automatiquement toutes les étapes du développement. L'information a été stockée dans une base de données relationnelle, ce qui en a permis l'utilisation pour la rédaction du recueil de codes du fichier de données.

- C On a pris comme point de départ la définition des thèmes de l'enquête dont on trouvera une description à l'annexe 2. Chaque case du diagramme représente un thème.
- C Les questions de l'enquête ont été associées aux thèmes.
- C On a défini le facteur «temps» des questions. Bref, certaines questions ne sont posées qu'une fois et la valeur des réponses ne devrait pas changer (par exemple, la date de naissance). D'autres sont posées chaque année, mais on ne note pas la date à laquelle le changement survient. D'autres

encore reviennent tous les ans, et on inscrit la date du changement, ce qui permet de suivre chronologiquement les résultats.

- C Les différentes unités des données ont été identifiées. Ces unités comprennent notamment la personne, l'emploi, les périodes sans emploi et le diplôme universitaire.
- C On a dressé une liste d'entités. Le thème, les dimensions temporelles et les unités des entités diffèrent.
- C Les questions ont été associées aux entités pour lesquelles on avait défini des variables.
- C Les variables obtenues (celles qu'on peut définir à partir d'une ou de plusieurs autres variables) ont aussi été définies. Il peut s'agir de variables dont on a agrégé les aspects ou du sommaire de plusieurs variables. Puisque les données de l'EDTR sont recueillies séparément pour chaque sujet, toutes les variables familiales sont dérivées.

7. Problèmes particuliers à l'élaboration du modèle de l'EDTR

Utilisation de données longitudinales plutôt que transversales

Bien que le principal objectif de l'EDTR consiste à permettre les analyses longitudinales, la base de données reste vaste et unique, même au niveau transversal (à savoir, pour une année particulière). Le modèle de données de l'EDTR devrait permettre une analyse transversale.

Unités d'analyse

Les données de l'EDTR sont recueillies pour chaque personne de l'échantillon. Toutefois, beaucoup de variables se rapportent à une unité d'analyse qui ne correspond pas à la personne. Ainsi, les paramètres de l'emploi s'appliquent à un emploi particulier détenu par le sujet. Il arrive néanmoins qu'une personne soit sans travail, n'ait qu'un emploi ou en cumule plusieurs. Parmi les autres unités

d'analyse de l'EDTR, on retrouve : les arrêts de travail, le ménage et la famille (examinés plus loin) et une longue série de périodes variées (état civil, sans emploi, chômage).

Représentation du temps: périodes ou vecteurs

Un des principaux aspects d'une enquête longitudinale a trait à la manière dont les fichiers de données représentent le temps. Il existe deux approches fondamentales, à cela : celle de la période et celle du vecteur. Chacune présente des avantages, selon le genre d'analyse souhaitée. La première approche consiste à relier chaque valeur d'une variable à une date de départ et à une date d'arrivée, alors qu'avec la seconde les périodes de temps continues sont représentées par des valeurs discrètes. La périodicité des composantes vectorielles dépend de la fréquence à laquelle change la valeur de la variable.

Voici un exemple illustrant les deux approches. Examinons la situation vis-à-vis de l'activité d'une personne au cours d'une année donnée. Supposons que cette dernière soit occupée du 1^{er} janvier au 11 mars, chô me du 12 mars au 7 mai, soit occupée du 8 mai au 26 mai, soit inactive du 27 mai au 10 septembre, puis fasse de nouveau partie des chô meurs du 11 septembre au 31 décembre.

Avec l'approche *périodique+, cette personne traverse cinq périodes durant l'année, c'est-à-dire :

- C occupée du 1^{er} janvier au 11 mars ;
- C en chô mage du 12 mars au 7 mai;
- C occupée du 8 mai au 26 mai;
- C inactive du 27 mai au 10 septembre;
- C en chô mage du 11 septembre au 31 décembre.

En combinant ces différentes périodes on obtient un tableau complet de la situation vis-à-vis de l'activité de la personne concernée pendant l'année. L'information est complète, mais le nombre de périodes peut varier d'une personne à l'autre, ce qui complique l'exploitation des données. On pourrait aussi créer deux vecteurs mensuels, le premier représentant la situation à un moment précis du mois et le second regroupant les valeurs pour un mois donné. En supposant qu'on s'intéresse à la valeur à la fin de chaque mois, le tableau que voici présente les deux vecteurs:

Mois	Valeur à la fin du mois	Toutes les valeurs possibles
Janvier	Occupée	Occupée seulement
Février	Occupée	Occupée seulement
Mars	En chômage	Occupée et en chômage
Avril	En chômage	En chômage seulement
Mai	Inactive	Occupée, en chômage et inactive
Juin	Inactive	Inactive seulement
Juillet	Inactive	Inactive seulement
Août	Inactive	Inactive seulement
Septembre	En chômage	En chômage et inactive
Octobre	En chômage	En chômage seulement
Novembre	En chômage	En chômage seulement
Décembre	En chômage	En chômage seulement

Les avantages et les inconvénients d'une telle méthode sont évidents. Le vecteur qui identifie une valeur à un point précis dans le mois est beaucoup plus simple, mais ne donne pas autant de renseignements que l'approche périodique. De fait, il omet la brève période de travail du mois de mai. Le deuxième vecteur constitue un

compromis. Il est plus complexe que le premier, puisque chaque composante vectorielle peut avoir un plus grand nombre de valeurs (7 dans ce cas, comparativement à 3). Le nombre de valeurs est constant pour chaque sujet, ce qui n'est pas le cas avec l'approche périodique, mais on perd la date à laquelle le changement survient. Manifestement, il est possible de remédier à certains inconvénients de l'approche vectorielle en raccourcissant la période à une semaine au lieu d'un mois par exemple. De cette façon, on augmente le nombre de composantes vectorielles.

En règle générale, le choix d'une approche ou d'une autre dépend de l'unité d'analyse. Avec une personne, l'approche vectorielle est préférable. Il reste alors à déterminer quelle période représentera chaque composante. Des compromis s'imposent. Si on s'intéresse à la durée du chômage davantage, l'unité d'analyse sera la période sans emploi et il est préférable de recourir à l'approche périodique. Il est toujours possible de passer de cette dernière à la représentation vectorielle. La conversion dans l'autre sens est également réalisable, mais les dates de départ et d'arrivée seront assujetties au cadre temporel utilisé pour les vecteurs.

Le modèle de données de l'EDTR recourt à l'approche périodique et à l'approche vectorielle selon l'usage principal de la variable. Il s'ensuit que l'utilisateur devra manipuler les données s'il désire une représentation différente de celle qui apparaît dans le modèle.

Données sur les familles et les ménages

Lorsqu'on lance une analyse longitudinale, la première réflexion que l'on se fait est : *J'aimerais savoir ce que devient une famille au fil des ans+. Un tel souhait peut paraître raisonnable et relativement simple sur le plan théorique, mais on se rend vite compte que ce n'est pas le cas quand vient le moment de décrire les données requises. Qu'est-ce qu'une *famille longitudinale+, quand la composition

de cette dernière n'est pas stable ? Le même problème se pose pour les ménages. L'EDTR ne définira pas de famille longitudinale, mais donnera les caractéristiques relatives à la famille et au ménage de chaque sujet. Par exemple, on peut étudier la manière dont le revenu familial d'une personne évolue avec le temps en fonction de son activité sur le marché du travail. La principale différence est qu'on prend la personne comme unité d'analyse plutôt que la famille. Pour plus de renseignements à ce sujet, on lira le document de recherche 94-06 de l'EDTR intitulé *Variables de l'EDTR relatives aux ménages et aux familles*.

Lieu de résidence

Les variables géographiques suscitent des difficultés analogues aux variables concernant la famille, mais leur orientation diffère légèrement. C'est pourquoi on les examinera séparément. Comme pour les autres variables qui changent dans le temps, le lieu de résidence doit intégrer une valeur temporelle quelconque.

La décision que cela implique est néanmoins un peu plus complexe que celle relative au choix de l'approche périodique ou vectorielle. Bon nombre d'analystes, notamment ceux des gouvernements provinciaux, s'intéressent aux données infranationales. La demande peut paraître anodine, mais il convient de répondre d'abord à la question suivante:

C Vous intéressez-vous à un point ou à une période donnés dans le temps ?

Dans le premier cas, définir la population souhaitée est aisé. Malheureusement, les utilisateurs s'intéressent habituellement à une période d'un an, auquel cas il faut répondre à d'autres questions:

C Vous intéressez-vous à un moment particulier de la période (la fin, le début ou un autre point du cadre temporel)? Ou aux personnes qui

vivaient dans la région, à un moment quelconque du cadre temporel ?
(L'EDTR comprendra des renseignements sur les déménagements effectués chaque année.)

Les réponses à ces questions peuvent dépendre de l'analyse. Quoi qu'il en soit, les résultats varieront (peut-être pas beaucoup) avec la façon dont la population à laquelle on s'intéresse est définie.

Règles de suivi

La partie 2 décrivait de façon succincte les règles applicables à l'EDTR. Toutefois, on ne sera pas surpris d'apprendre que les décisions relatives au choix des personnes qui seront interrogées à chaque période de collecte des données exerceront une influence sur le modèle.

Modification des variables principales: date de naissance/sexe

Les valeurs de certaines variables jouent un rôle capital dans la collecte des données, car elles déterminent quelles questions on peut ou non poser à la personne interviewée. L'âge (déterminé d'après la date de naissance) et le sexe font partie de ces variables dans la plupart des enquêtes sur les ménages (l'EDTR n'y fait pas exception). L'analyse longitudinale permet néanmoins la correction des erreurs lors d'une interview subséquente. Il s'agit d'un avantage à de nombreux égards, mais certaines difficultés peuvent se présenter. Prenons l'EDTR comme exemple. Toutes les personnes ayant 16 ans et plus à la fin de l'année de référence passeront l'interview annuelle sur le travail. Supposons que selon sa date de naissance, une personne ait 20 ans. Elle peut donc être interviewée, et les données pertinentes sont recueillies. L'interview suivante révèle cependant que la date de naissance était incorrecte et qu'en réalité, la personne interrogée n'avait que 12 ans à l'interview précédente. Faut-il faire marche arrière et supprimer les données recueillies ? On serait tenté de répondre *oui+, mais supposons que la

date de naissance initiale était bel et bien la bonne et qu'on doive procéder à une troisième correction par la suite. Des données utiles auront été supprimées. L'autre possibilité serait de n'autoriser aucune modification des valeurs recueillies au départ, la raison étant que les enquêtes transversales comprennent des réponses erronées pour les diverses variables, mais que ces erreurs passent inaperçues car l'information n'est recueillie qu'une fois. Dans certains cas néanmoins, il peut arriver que l'intervieweur se trouve dans une situation embarrassante s'il doit expliquer au répondant pourquoi sa date de naissance ou son sexe n'ont pas été corrigés. Nous espérons ne pas avoir à modifier souvent la valeur des variables principales, si bien que la décision retenue, quelle qu'elle soit aura peu d'impact sur l'utilisation des données. Une décision doit pourtant être prise. Pour l'EDTR, il a été convenu qu'on ne supprimerait aucune des données recueillies et qu'on mettrait à jour les valeurs relatives à la date de naissance et au sexe, c'est-à-dire qu'on validerait toutes les corrections. L'utilisateur devra donc filtrer les données pour établir l'admissibilité des sujets quant à l'âge et au sexe.

Report de l'interview du mois de mai

Bien qu'on s'entretienne avec les répondants deux fois par année, on considère qu'il s'agit de la même interview, scindée en deux. Des renseignements différents sont recueillis dans chaque cas, et les deux volets prennent l'année civile antérieure comme période de référence. Les renseignements sur le revenu sont obtenus en mai, ce qui améliore la qualité des données, car les déclarations de revenu ont été envoyées peu de temps auparavant et la plupart des personnes se rappelleront leurs sources de revenu. Toutefois, cette interview signifie que les changements de composition du ménage seront notés deux fois par année.

Théoriquement, l'EDTR recueille des données de tous les sujets de l'étude longitudinale et des personnes qui vivaient avec eux à la fin de décembre de l'année de référence. Quelques changements peuvent survenir lorsqu'on recueille

les renseignements sur le travail à la fin de janvier. Néanmoins, beaucoup d'autres modifications se produiront entre cette date et le milieu du mois de mai où on recueillera les données sur le revenu. Voici les possibilités qu'il faut envisager:

- C Changements dans la composition du ménage survenus entre l'interview de mai et celle de janvier - la plupart surviendront avant le 31 décembre, mais certains suivront.
- C Changements dans la composition du ménage survenus entre l'interview de janvier et l'interview de mai.
- C Si on ne recueille pas de données en mai (impossibilité de rejoindre la personne ou non-réponse), changements dans la composition du ménage survenus entre l'interview de janvier d'une année donnée et l'interview de janvier de l'année suivante; certaines modifications se seront produites avant le 31 décembre, d'autres après.
- C Lorsqu'on ne peut recueillir de données en janvier (impossibilité de rejoindre la personne ou non-réponse), changements dans la composition du ménage survenus entre l'interview de mai d'une année donnée et l'interview de mai suivant : quelques modifications se seront produites avant le 31 décembre, d'autres par la suite.
- C Changements non identifiés au moment voulu; par exemple, personne qui a emménagé avant le 31 décembre sans que la chose soit rapportée avant le mois de mai. Pareille situation pourrait résulter de la confusion dans l'esprit du répondant ou de l'intervieweur, l'un ou l'autre croyant que la personne en question ne devrait pas être identifiée. Elle pourrait également résulter d'une véritable ambiguïté au sujet de l'emménagement ou non de la personne concernée au moment de l'interview de janvier.
- C Retrouver les sujets de l'étude longitudinale pourrait déboucher sur des dates contradictoires, ce qui empêcherait l'EDTR d'établir la situation courante. Supposons par exemple qu'un homme se sépare de sa femme et emménage avec son frère, les ex-conjoints faisant partie de l'étude

longitudinale. Dans l'interview suivant la séparation, l'épouse signale la date à laquelle son mari l'a quittée. Après avoir retrouvé celui-ci, on demande au frère à quel moment a emménagé le mari. Si la date mentionnée par l'épouse précède celle indiquée par le frère d'un mois, qui dit la vérité ? Le mari a-t-il vécu ailleurs pendant un mois, les personnes interrogées se sont-elles trompées ou le départ a-t-il été graduel, au lieu de survenir à un moment précis dans le temps ?

Dans la mesure où les données le permettent, on s'efforcera d'établir les dates correctes lors du traitement, en s'assurant que les informations relatives à une période de référence sont cohérentes avec le but théorique de la collecte des données.

Interview avec rétro-information - que faire en cas de démenti ?

Par interview avec rétro-information, on entend l'utilisation des données recueillies lors d'une interview antérieure pour aider une personne à répondre aux questions de la nouvelle interview. Ainsi, dans l'EDTR, on *rappelle* au répondant le nom des employeurs pour lesquels il avait travaillé à la fin de la dernière période de référence. Par exemple, supposons qu'une personne signale qu'elle travaille pour l'entreprise ABC et pour DEF Inc. à la fin de l'année 1. Lors de l'interview sur le travail de l'année suivante, on lui demandera si elle travaillait encore pour ces deux employeurs au début de l'année 2. Sans rétro-information, il se peut que le sujet oublie une des deux entreprises, surtout s'il a cessé de travailler pour elle en début d'année. L'inconvénient de cette méthode est qu'on pourrait obtenir un démenti. Que faire dans pareil cas ? Pour maintenir de bonnes relations avec le répondant, l'intervieweur acceptera l'information, c'est-à-dire le démenti. Une décision s'imposera néanmoins lors du traitement : quelle est la bonne réponse ? Il ne semble pas y avoir de règle générale à suivre. La décision dépendra de l'élément d'information concerné.

Aspects relatifs au traitement

Il s'agit essentiellement d'une question de bon sens, valable pour toutes les enquêtes. Lorsqu'on met au point un modèle de données, il importe de savoir comment seront traitées ces dernières. Ainsi, les responsables de la modélisation veilleront à ce que l'information soit traitée comme on le désire et à ce que le modèle n'introduise pas de restrictions. Garder ou supprimer des données au moment des corrections en est un exemple simple. Si les données brutes ne représentent qu'une étape dans la création du fichier de sortie, on pourra les éliminer (après avoir fait une copie de sauvegarde). Si l'utilisateur souhaite comparer une variable particulière avant et après la correction par contre, le modèle devrait permettre le stockage des deux versions de la variable dans la même base.

8. Le modèle de l'EDTR: particularités fondamentales et futures orientations

Les données de l'EDTR seront emmagasinées dans une base de données relationnelle. Divers facteurs sont à l'origine de cette décision :

- C Nombre des unités d'analyse différentes. Une base de données relationnelle réduira la taille du fichier et les redondances.

- C Le modèle de l'EDTR devrait inclure toutes les données de l'enquête, pas seulement celles requises pour l'analyse (les fichiers relatifs à la collecte, notamment les renseignements préliminaires sur les répondants, font partie du modèle même si les utilisateurs ne pourront y avoir accès pour des raisons évidentes).

- C Les répondants ne sont pas constamment admissibles. On ne pourra accéder aux données des années antérieures pour les nouveaux membres

du ménage. On recueillera également très peu de renseignements sur les sujets de moins de 15 ans, mais certaines personnes qui n'avaient pas 15 ans au début de l'enquête de 6 ans dépasseront cet âge par la suite, de sorte qu'on pourra les interviewer.

À l'heure actuelle, le modèle de données de l'EDTR compte quarante-quatre entités. Il se peut que ce nombre augmente ou baisse légèrement, mais pas beaucoup tant qu'on n'apportera pas de modifications majeures au contenu de l'enquête. Un ajout important envisagé pour 1998 a trait à la collecte de renseignements sur l'actif et le passif du répondant. Il ne fait aucun doute qu'on devra pour cela ajouter des entités au modèle de données.

On continuera d'apporter certaines modifications aux variables avec le temps, principalement à la suite de l'addition de variables dérivées. Toutes les variables dérivées proposées sont ajoutées au modèle. Il sera vraisemblablement impossible d'en calculer certaines, qui seront abandonnées. À l'heure actuelle, le modèle compte environ sept cents variables.

Le nombre de variables peut être trompeur avec une base de données relationnelle. Par exemple, une variable déterminera si l'intéressé est un cadre dans le contexte d'un emploi donné. On recueillera ce renseignement pour tous les emplois, chaque année où le sujet occupe l'emploi concerné. On enregistrera donc plusieurs valeurs pour chaque personne, mais on ne comptera la variable qu'une fois.

Le modèle de données de l'EDTR comprend 4 types fondamentaux d'entités:

- C Les entités fixes sont celles dont la valeur ne change pas. L'entité précisant la langue maternelle, le pays d'origine et le groupe ethnique en est un exemple.

- C Les entités annuelles couvrent une année civile ou incluent les données relatives à un point dans le temps recueillies une seule fois par année. L'entité qui indique le revenu provenant de diverses sources lors d'une année précise, illustre le premier type d'entité annuelle. Un exemple du second est l'entité donnant des renseignements sur les personnes qui signalent une limitation dans leurs activités et les conséquences de cette limitation sur leur participation au marché du travail. Ces renseignements se rapportent à un point précis dans le temps (théoriquement, le 31 décembre de l'année de référence). À la fin de l'enquête de six ans, on possédera six *instantanés+ annuels des données pertinentes.
- C Les entités cumulatives sont des entités annuelles dont on dérive la valeur à partir de l'année antérieure, par majoration des données recueillies durant l'année de référence. L'expérience en est un exemple. L'expérience totale d'une personne à la fin de l'année 2 correspond à l'expérience acquise à la fin de l'année 1, plus la période d'emploi durant l'année 2.
- C Les entités périodiques correspondent aux données s'appliquant à une période. L'entité indiquant l'état civil d'une personne en est un exemple. Il existe une donnée pour chaque état civil et chaque sujet. Toutes les périodes d'état civil d'une personne doivent être regroupées si on veut obtenir des renseignements d'ordre personnel, c'est-à-dire une représentation vectorielle.

9. Les fichiers de microdonnées de l'EDTR et l'interface avec l'utilisateur

Bien que le présent document porte essentiellement sur le développement du modèle de données, il est naturel d'examiner l'incidence de ce dernier sur l'utilisateur. C'est ce que nous aborderons brièvement ici.

Tel qu'indiqué précédemment, au moment d'élaborer le modèle de données de l'EDTR, on s'est interrogé sur les aptitudes techniques des utilisateurs. Certaines personnes seront relativement à l'aise avec une base de données relationnelle, mais il n'en ira pas autant de la majorité. Dans une telle optique, le terme *convivialité+ désigne la capacité d'apprentissage de l'utilisateur en ce qui concerne la manipulation des données, et ne s'applique pas aux contraintes que posent matériel et logiciel. (Ces contraintes susciteront des difficultés, mais elles devraient être minimales étant donné la rapidité avec laquelle progresse la technologie.) Il est donc capital de fournir des outils qui simplifieront l'exploitation des données.

Logiciel de récupération

Les fichiers de microdonnées de l'EDTR seront vraisemblablement distribués sur CD-ROM avec un *logiciel de récupération+ servant d'interface avec la base de données relationnelle. Une série d'écrans aidera l'utilisateur à choisir l'unité d'analyse, la période qui l'intéresse, la population souhaitée et les principales variables. Le logiciel bâtira alors un fichier bidimensionnel sous format ASCII, bref un fichier sur mesure. L'utilisateur n'aura pas à apprendre comment manipuler les données de la base relationnelle. Le fichier ASCII pourra ensuite être exploité avec le logiciel d'analyse préféré de l'utilisateur.

Logiciel pour l'analyse fondamentale

On songe ajouter un autre logiciel au fichier de microdonnées de l'EDTR mais on n'a pas encore déterminé à quoi il ressemblerait. Le principe consiste néanmoins à fournir des instruments analytiques de base, par exemple des statistiques sommaires, des tabulations par recoupement, des graphiques et des diagrammes. L'utilisateur disposerait ainsi d'un point de départ, d'un *aperçu+ des données. Le recours à un autre logiciel s'avérera peut-être superflu pour ceux qui n'ont besoin que d'une tabulation par recoupement.

Ateliers pour l'utilisateur

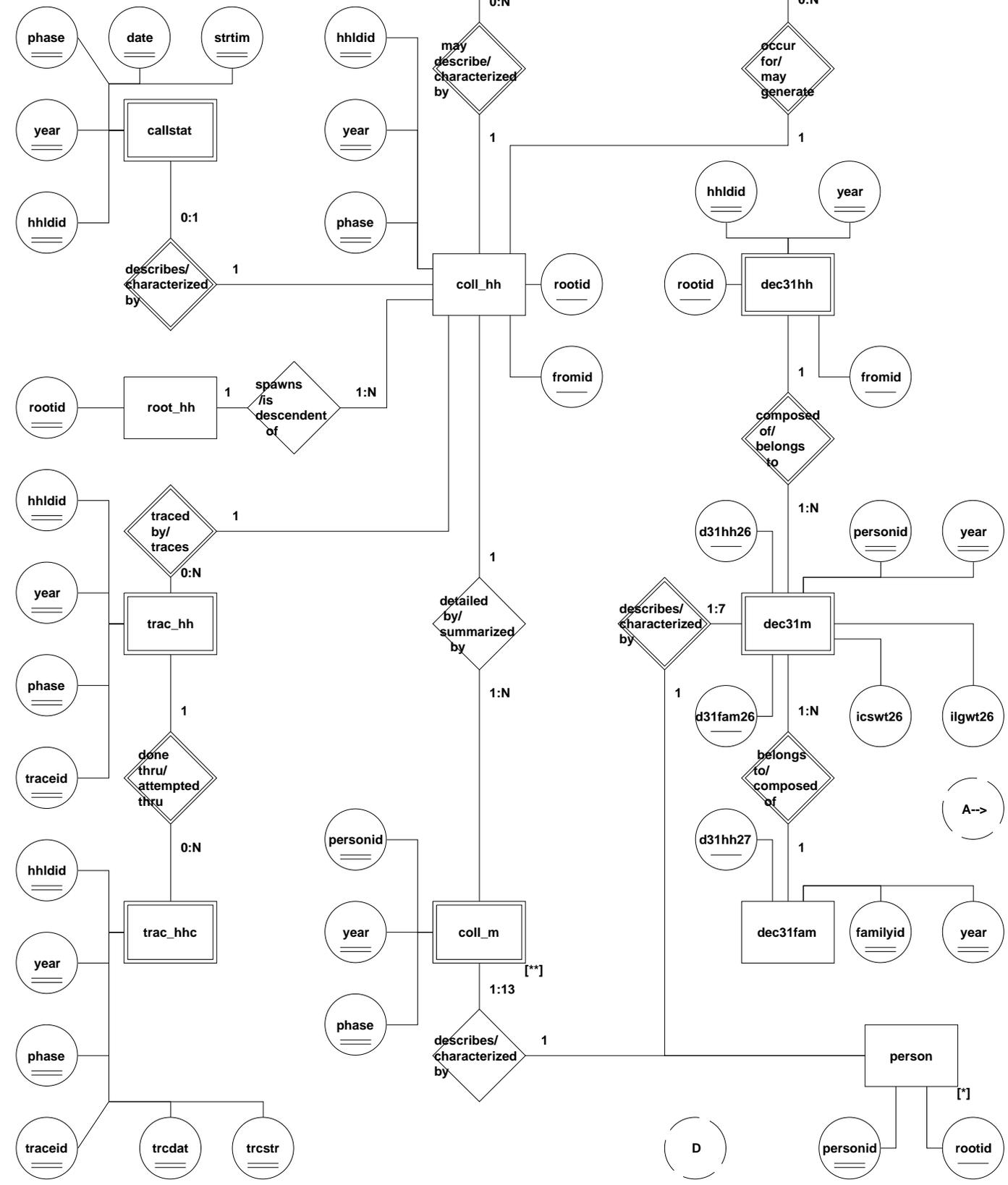
Des questions surgissent souvent au moment où on commence à utiliser les données. Des ateliers illustrant quelques méthodes simples de récupération et de tabulation des données devraient aider l'utilisateur à partir du bon pied.

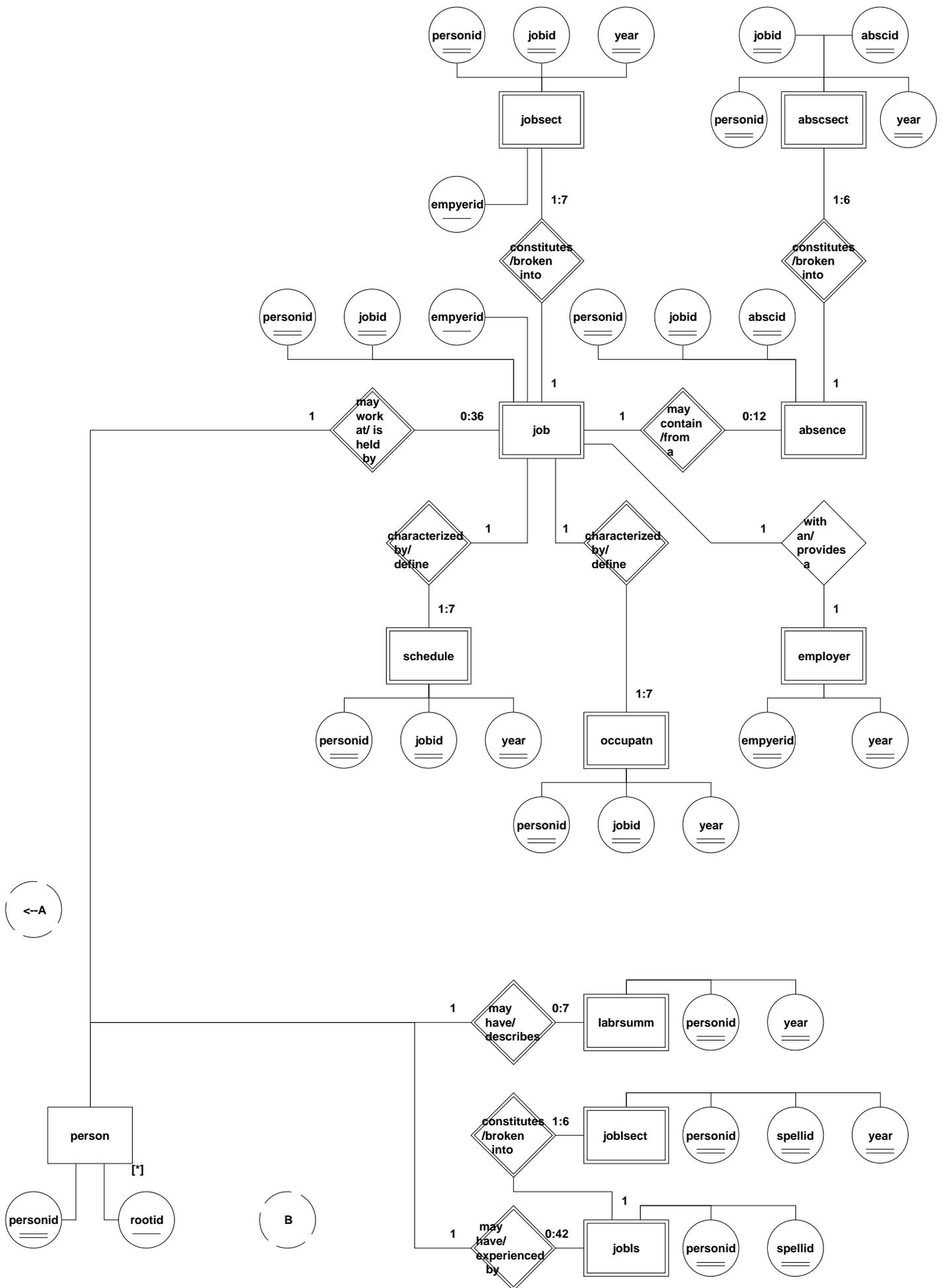
ANNEXE 1:

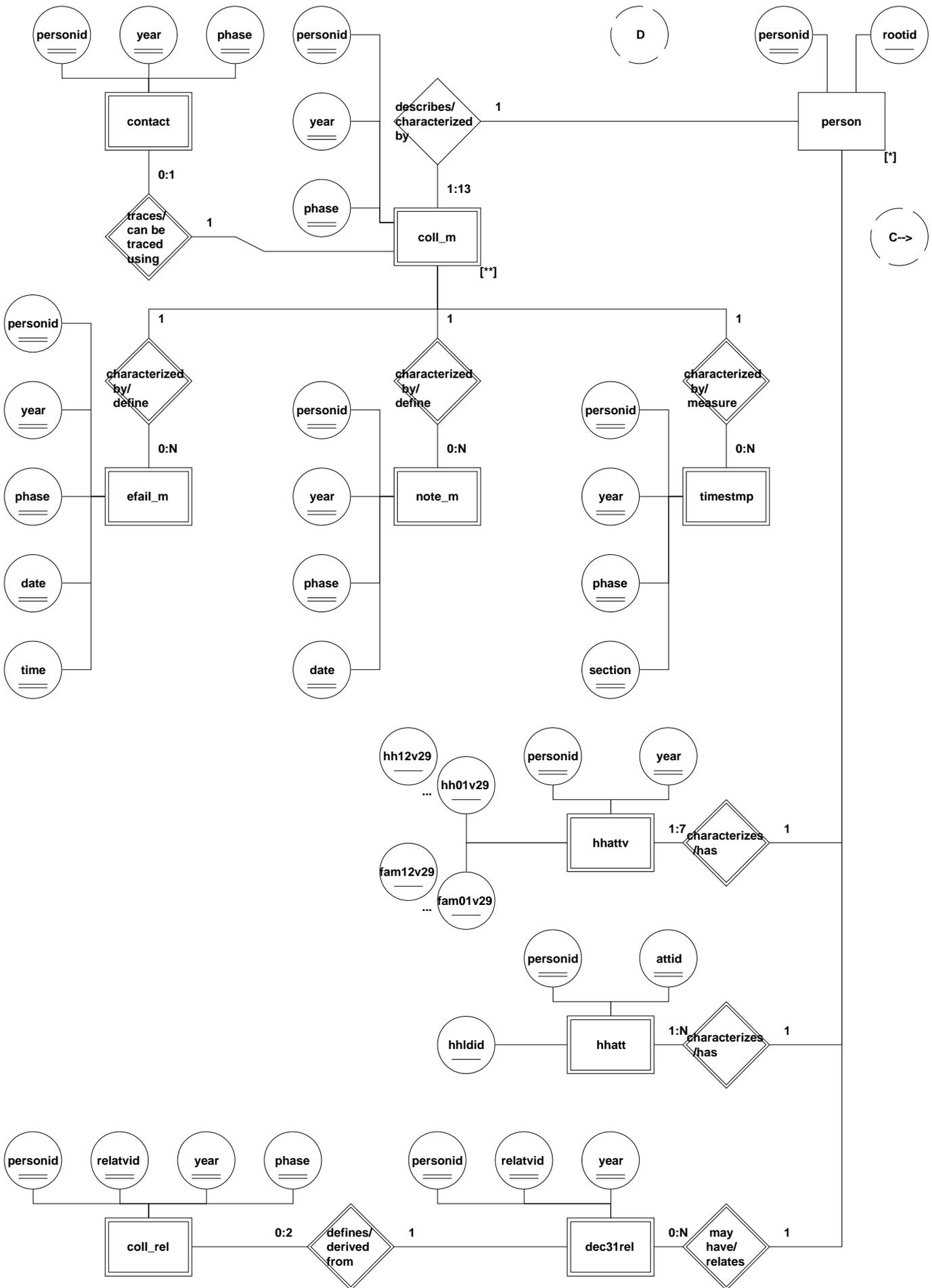
**DIAGRAMMES ENTITÉS-RELATIONS (DER) ET DICTIONNAIRE DES
DONNÉES (DD) DE L'ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL
ET DU REVENU**

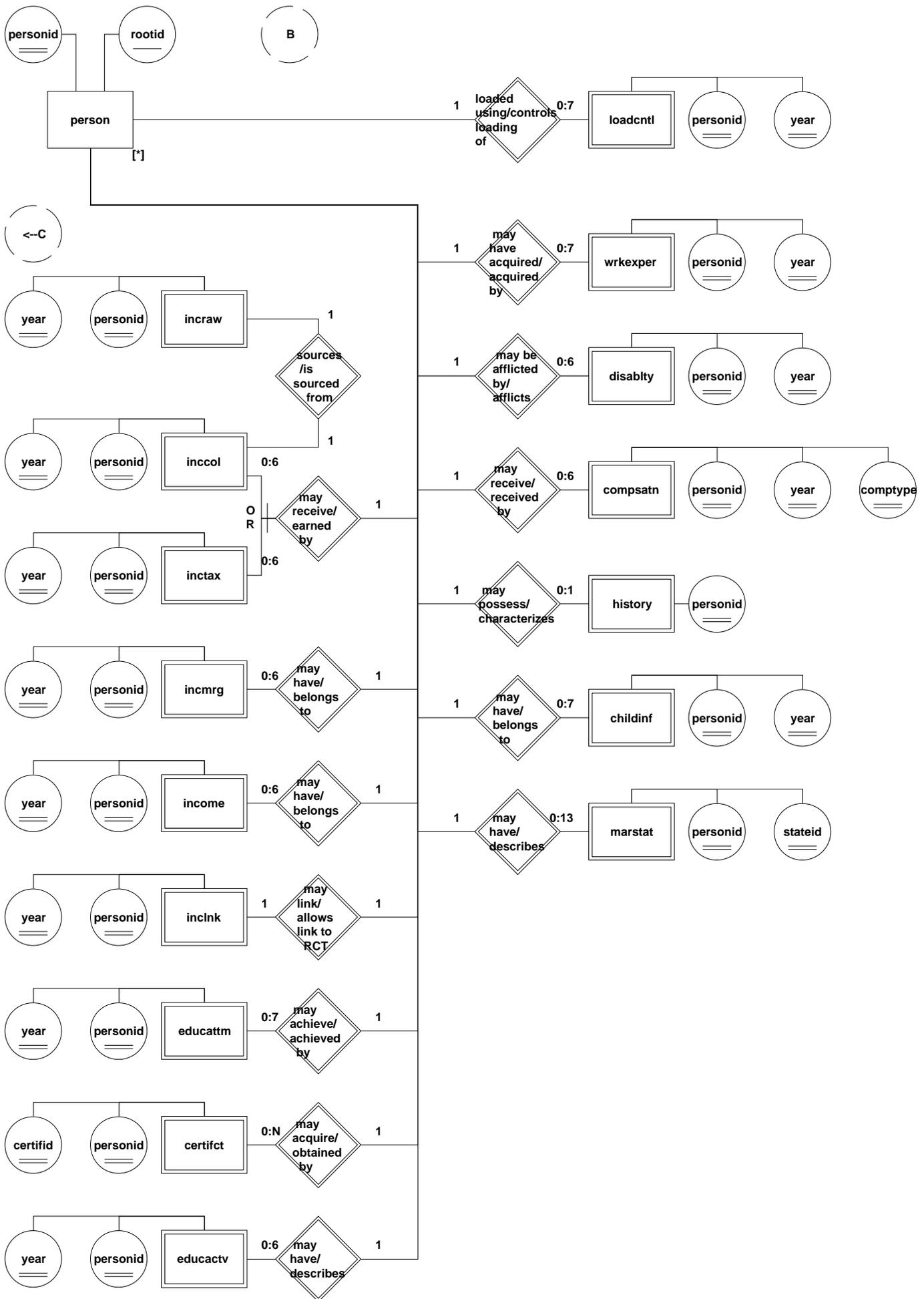
(Remarque : La version française de ces documents, n'étant pas disponible,
l'annexe est présentée en anglais.)

Project Name: SLID Data Model
 Project Path: r:\ecwin\slid\dm\
 Chart File: erd00109.erd
 Chart Name: SLID_DM
 Created On: Jun-10-1994
 Created By: SLID Data Model Team
 Modified On: Aug-26-1997
 Modified By: Pierre Lafrance









Entité: **JOB****Alias:** Entité 9**Unité****d'analyse:** Personne-Emploi**Champs clés:****PERSONID** Caractère (12) Format: PPWFMMMMMMMM*Long nom:* Id de personne*Description:* Identificateur unique pour une personne.

(Note. Valeur supprimée du fichier à grande diffusion.)

Le format de ce champ est le suivant : PPWFMMMMMMMM

P - identificateur de panel

W - vague au cours de laquelle la personne est entrée dans l'échantillon

F - phase où la personne est entrée dans l'échantillon

M - identificateur de la personne

Population: Toutes les personnes*Codes Réservés:* 999999999996 Hors Échantillon

999999999997 Ne sait pas

999999999998 Refus

999999999999 Non Applicable

JOBID Numérique (2)*Long nom:* Id d'épisode d'emploi*Description:* Identificateur unique pour un épisode d'emploi avec un employeur. Le même épisode d'emploi qui couvre plus d'une année de référence aura le même "jobid". Deux épisodes d'emploi distincts avec le même employeur durant la même année de référence auront des "jobid" distincts.*Population:* Personnes âgées de 16 à 69 ans ET ayant un emploi pendant la période de réf. de l'enquête*Gamme:* 00:95*Codes Réservés:* 96 Hors Échantillon

97 Ne sait pas

98 Refus

99 Non Applicable

Attributs:

EMPYER9 Caractère (14) Format: PPWFMMMMMMMMJJ

Long nom: Id d'employeur

Description: Identificateur unique d'employeur.

(Note. Valeur supprimée du fichier à grande diffusion.)

Le format de ce champ est le suivant : P P P P P P P P P P P P J J

P - identificateur de personne

J - identificateur du premier emploi pour cette personne avec cet employeur

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Codes Réservés:

99999999999996	Hors Échantillon
99999999999997	Ne sait pas
99999999999998	Refus
99999999999999	Non Applicable

STRDAT9 Date (8) Format: AAAAMMJJ

Long nom: Date début d'emploi

Description: Date de début d'emploi.

Toutes les variables de type "date" sont constitués en concaténant les trois sous-composantes année, mois et jour en une seule variable. Les codes réservés standards de l'EDTR sont assignés, lorsque nécessaire, aux sous-composantes, et non à la variable de type date en entier. Il est donc possible d'avoir un mélange de codes réservés et de valeurs valides dans une variable de type "date".

(Note. Jour supprimé dans le fichier de microdonnées à grande diffusion.)

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: [DATES-Q2T1]
[DATES-Q2T2]
[DATES-Q3], [DATES-Q5], [DATES-Q6]

Gamme: 18000101:99951231

Codes Réservés:

96	Hors Échantillon
97	Ne sait pas
98	Refus
99	Non Applicable
9996	Hors Échantillon
9997	Ne sait pas
9998	Refus
9999	Non Applicable

STRDA_9 Caractère (3) Format: vecteur 3 octets de format AMJ

Long nom: Drapeaux d'imp de date

Description: Drapeaux d'imputations de dates.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

**MODÈLE DE DONNÉES DE L'EDTR:
STRUCTURE D'ENTITÉS**

Codes: 0 composante de date recueillie
1 composante de date imputée

ENDDAT9 Date (8) Format: AAAAMMJJ

Long nom: Date fin d'emploi

Description: Date de la fin d'emploi.

Toutes les variables de type "date" sont constitués en concaténant les trois sous-composantes année, mois et jour en une seule variable. Les codes réservés standards de l'EDTR sont assignés, lorsque nécessaire, aux sous-composantes, et non à la variable de type date en entier. Il est donc possible d'avoir un mélange de codes réservés et de valeurs valides dans une variable de type "date".

(Note. Jour supprimé dans le fichier de microdonnées à grande diffusion.)

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: [DATES-Q3], [DATES-Q8]
[DATES-Q10], [DATES-Q11], [DATES-Q12]

Gamme: 18000101:99951231

Codes Réservés: 96 Hors Échantillon
97 Ne sait pas
98 Refus
99 Non Applicable
9996 Hors Échantillon
9997 Ne sait pas
9998 Refus
9999 Non Applicable

ENDDA_9 Caractère (3) Format: vecteur 3 octets de format AMJ

Long nom: Drapeaux d'imp de date

Description: Drapeaux d'imputations de dates.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Codes: 0 composante de date recueillie
1 composante de date imputée

FSTDAT9 Date (8) Format: AAAAMMJJ

Long nom: Date 1er emp avec employeur

Description: Date de début du premier emploi avec cet employeur.

Toutes les variables de type "date" sont constituées en concaténant les trois sous-composantes année, mois et jour en une seule variable. Les codes réservés standards de l'EDTR sont assignés, lorsque nécessaire, aux sous-composantes, et non à la variable de type date en entier. Il est donc possible d'avoir un mélange de codes réservés et de valeurs valides dans une variable de type "date".

(Note. Jour supprimé dans le fichier de microdonnées à grande diffusion.)

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: PRELIM-Q21A,Q21B
PRELIM-Q25A,Q25B
[DATES-Q7], [DATES-Q7A]

Gamme: 18000101:99951231

Codes Réservés:

96	Hors Échantillon
97	Ne sait pas
98	Refus
99	Non Applicable
9996	Hors Échantillon
9997	Ne sait pas
9998	Refus
9999	Non Applicable

FSTDA_9 Caractère (3) Format: vecteur 3 octets de format AMJ

Long nom: Drapeaux d'imp de date

Description: Drapeaux d'imputations de dates.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Codes:

0	composante de date recueillie
1	composante de date imputée

ENDED9 Caractère (1)

Long nom: Emploi terminé

Description: Indicateur signalant qu'un emploi s'est terminé avant la fin de la période de référence la plus récente de l'enquête comprise dans le fichier de données.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: [DATES-Q3], [DATES-Q8], [DATES-Q10]
[DATES-Q11], [DATES-Q12]
JOB.ENDDAT9

Codes:

1	Oui
2	Non

Codes Réservés:

6	Hors Échantillon
7	Ne sait pas
8	Refus
9	Non Applicable

ENDTYP9 Caractère (1)

Long nom: Type de fin d'emploi

Description: Raison pour laquelle l'emploi s'est terminé durant le traitement. Cette variable est pour usage interne lors du traitement des données.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Codes:

1	Emploi terminé normalement
---	----------------------------

**MODÈLE DE DONNÉES DE L'EDTR:
STRUCTURE D'ENTITÉS**

<i>Codes:</i>	2	Emploi terminé parce que le répondant a nié l'existence de l'emploi.
	3	Emploi terminé par manque d'information lors de périodes de collectes subséquentes - soit par manque de rétro information, pour cause de non-réponse ou parce que le répondant n'est plus éligible à l'interview du travail
<i>Codes Réservés:</i>	6	Hors Échantillon
	7	Ne sait pas
	8	Refus
	9	Non Applicable
REAEND9	Caractère (2)	
<i>Long nom:</i>	Raison de fin d'emploi	
<i>Description:</i>	Raison de la fin d'emploi.	
<i>Population:</i>	Personnes âgées de 16 à 69 ans ET ayant eu une cessation d'emploi pendant la période de réf. de l'enquête ET pour lesquelles l'emploi s'est terminé	
<i>Source:</i>	[DATES-Q13], [DATES-Q13A] [DATES-Q13A1], [DATES-Q13A2], [DATES-Q13B]	
<i>Codes:</i>	01	Maladie ou incapacité - reliée au travail
	02	Maladie ou incapacité - non reliée au travail
	03	S'occuper de ses enfants
	04	S'occuper des personnes âgées de sa famille
	05	Autres obligations personnelles ou familiales
	06	Études
	07	Nouvel emploi
	08	Déménagement
	09	Rémunération insuffisante
	10	Pas assez d'heures de travail
	11	Trop d'heures de travail
	12	Mauvaises conditions de travail (mauvaise ventilation, bruyant)
	13	Harcèlement sexuel
	14	Incompatibilité avec son employeur ou ses collègues
	15	Trop de pression au travail
	16	La compagnie a déménagé
	17	La compagnie a fermé ses portes
	18	Le travail était de nature saisonnière
	19	Mise à pied/un manque de travail (non saisonnière)
	20	Conflit de travail
	21	A été congédié par l'employeur
	22	Emploi temporaire/fin de contrat
	23	Avait un deuxième emploi
	24	Retraite
	25	Autre
<i>Codes Réservés:</i>	96	Hors Échantillon
	97	Ne sait pas
	98	Refus
	99	Non Applicable

TYPJS9	Caractère (1)	
<i>Long nom:</i>	Type de fin d'emploi	
<i>Description:</i>	Type de fin d'emploi.	
<i>Population:</i>	Personnes âgées de 16 à 69 ans ET ayant eu une cessation d'emploi pendant la période de réf. de l'enquête ET pour lesquelles l'emploi s'est terminé	
<i>Source:</i>	JOB.REAEND9	
<i>Codes:</i>	1	Volontaire
	2	Involontaire
<i>Codes Réservés:</i>	6	Hors Échantillon
	7	Ne sait pas
	8	Refus
	9	Non Applicable
CLWKR9	Caractère (2)	
<i>Long nom:</i>	Cat de travailleur	
<i>Description:</i>	Catégorie de travailleur la plus récente pour cet emploi.	
<i>Population:</i>	Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête	
<i>Source:</i>	PRELIM-F05Q76 PRELIM-Q29,COW2_COD [CHAR-Q3], [CHAR-Q3A], [CHAR-Q3B], [CHAR-Q3C] JOBSECT.CLWKR1	
<i>Codes:</i>	01	Travailleur rémunéré
	02	Travailleur familial non rémunéré
	03	Travailleur autonome avec aide rémunérée (Entreprise constituée en société)
	04	Travailleur autonome sans aide rémunérée (Entreprise constituée en société)
	05	Travailleur autonome avec aide rémunérée (Entreprise non constituée en société)
	06	Travailleur autonome sans aide rémunérée (Entreprise non constituée en société)
<i>Codes Réservés:</i>	96	Hors Échantillon
	97	Ne sait pas
	98	Refus
	99	Non Applicable
RCLWKR9	Caractère (2)	
<i>Long nom:</i>	Cat de travailleur recodé	
<i>Description:</i>	Catégorie de travailleur recodée la plus à jour pour cet emploi. Les propriétaires d'entreprises incorporées sont inclus dans la catégorie "Travailleur rémunéré". De plus, une catégorie distincte a été créée pour les employés du gouvernement.	
<i>Population:</i>	Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête	

**MODÈLE DE DONNÉES DE L'EDTR:
STRUCTURE D'ENTITÉS**

Source: PRELIM-F05Q76 (job 1)
PRELIM-Q29, COW2_COD (job 2)
[CHAR-Q3], [CHAR-Q3A], [CHAR-Q3B], [CHAR-Q3C]
JOBSECT.RCLWKR1

Codes:

01	Travailleur rémunéré
02	Travailleur rémunéré (secteur public)
04	Travailleur autonome avec aide rémunérée (entreprise non constituée en société)
05	Travailleur autonome sans aide rémunérée (entreprise non constituée en société)
06	Travailleur familial non rémunéré

Codes Réservés:

96	Hors Échantillon
97	Ne sait pas
98	Refus
99	Non Applicable

HOWOBT9 Caractère (2)

Long nom: Méthode empl obtenu

Description: Façon dont l'emploi a été obtenu.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête ET pour lesquelles l'emploi a commencé après le 1er janvier 1993 ET étant des travailleurs rémunérés

Source: [CHAR-Q4]

Codes:

01	A contacté l'employeur directement
02	Par l'entremise d'un ami ou d'un parent
03	A fait paraître une annonce ou a répondu à une annonce
04	Agence de placement
05	A été recommandé(e) par un autre employeur
06	A été contacté par l'employeur directement
07	Syndicat
08	Autre

Codes Réservés:

96	Hors Échantillon
97	Ne sait pas
98	Refus
99	Non Applicable

OFRDAT9 Date (8) Format: AAAAMMJJ

Long nom: Date offre reçue

Description: Date de réception de l'offre d'emploi.

Toutes les variables de type "date" sont constitués en concaténant les trois sous-composantes année, mois et jour en une seule variable. Les codes réservés standards de l'EDTR sont assignés, lorsque nécessaire, aux sous-composantes, et non à la variable de type date en entier. Il est donc possible d'avoir un mélange de codes réservés et de valeurs valides dans une variable de type "date".

(Note. Jour supprimé dans le fichier de microdonnées à grande diffusion.)

**MODÈLE DE DONNÉES DE L'EDTR:
STRUCTURE D'ENTITÉS**

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête ET pour lesquelles l'emploi a commencé après le 1er janvier 1993 ET étant des travailleurs rémunérés

Source: [CHAR-Q5]

Gamme: 18000101:99951231

<i>Codes Réservés:</i>	96	Hors Échantillon
	97	Ne sait pas
	98	Refus
	99	Non Applicable
	9996	Hors Échantillon
	9997	Ne sait pas
	9998	Refus
	9999	Non Applicable

OFRDA_9 Caractère (3) Format: vecteur 3 octets de format AMJ

Long nom: Drapeaux d'imp de date

Description: Drapeaux d'imputations de dates.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

<i>Codes:</i>	0	composante de date recueillie
	1	composante de date imputée

NBOCCU9 Numérique (2)

Long nom: Activités changées-cet emp

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: [CHAR-Q11]

Gamme: 01:95

<i>Codes Réservés:</i>	96	Hors Échantillon
	97	Ne sait pas
	98	Refus
	99	Non Applicable

OCCHG9 Caractère (1)

Long nom: Activités changées-cet emp

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

<i>Codes:</i>	1	Oui
	2	Non
<i>Codes Réservés:</i>	6	Hors Échantillon
	7	Ne sait pas
	8	Refus
	9	Non Applicable

NBSCHD9 Numérique (2)

Long nom: Nbre d'hor trav-cet emploi

**MODÈLE DE DONNÉES DE L'EDTR:
STRUCTURE D'ENTITÉS**

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: [CHAR-Q27], [CHAR-Q32]

Gamme: 00:95

Codes Réservés:

96	Hors Échantillon
97	Ne sait pas
98	Refus
99	Non Applicable

NBABS9 Numérique (2)

Long nom: Nbre d'absences-cet emploi

Description: Excluant les vacances rémunérées, nombre de fois que le répondant a été absent de cet emploi pour une période de plus d'une semaine.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: [CHAR-Q46] (for t1 or t3)
[CHAR-Q48], [CHAR-Q48A], [CHAR-Q48B]
[CHAR-Q49], [CHAR-Q49A]

Gamme: 00:95

Codes Réservés:

96	Hors Échantillon
97	Ne sait pas
98	Refus
99	Non Applicable

NBMTWK9 Numérique (2)

Long nom: Nbre mois travaillé-emp

Description: Nombre total de mois durant lesquels du travail a été effectué à cet emploi excluant les mois où la personne était absente de l'emploi pour le mois en entier.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

Source: [DATES-Q2T1], [DATES-Q2T2], [DATES-Q3]
[DATES-Q5], [DATES-Q6], [DATES-Q10], [DATES-Q12]
JOB.STRDAT9
JOB.ENDDAT9
JOB.ENDED9

Gamme: 00:95

Codes Réservés:

96	Hors Échantillon
97	Ne sait pas
98	Refus
99	Non Applicable

JOB DUR9 Numérique (3)

Long nom: Durée de l'emploi en mois

Description: Durée de l'épisode d'emploi (exprimée en mois).

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête

**MODÈLE DE DONNÉES DE L'EDTR:
STRUCTURE D'ENTITÉS**

Source: JOB.STRDAT9
JOB.ENDDAT9

Gamme: 001:995

Codes Réservés: 996 Hors Échantillon
997 Ne sait pas
998 Refus
999 Non Applicable

CONATT9 Caractère (1)

Long nom: Employeur continu

Description: Indicateur signalant si l'emploi auprès de cet employeur a été continu depuis la première fois que le répondant a travaillé pour cet employeur.

Population: Personnes âgées de 16 à 69 ans ET ayant eu un emploi pendant la période de réf. de l'enquête ET emploi continu depuis la date de commencement

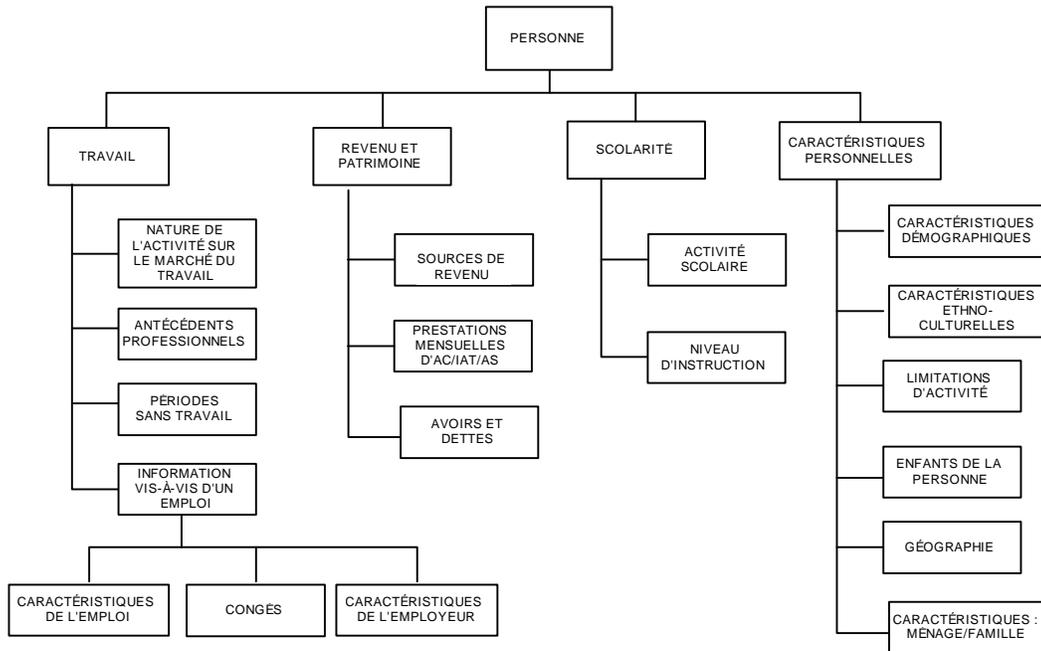
Source: JOB.STRDAT9
JOB.FSTDAT9

Codes: 1 Oui
2 Non

Codes Réservés: 6 Hors Échantillon
7 Ne sait pas
8 Refus
9 Non Applicable

ANNEXE 2: ORGANISATION DE LA BASE DE DONNÉES DE L'EDTR ET LISTE PARTIELLE DES VARIABLES

ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL ET DU REVENU : ORGANISATION DU CONTENU



LISTE PARTIELLE DES VARIABLES

I. Travail

Nature et structure de l'activité sur le marché du travail :

- périodes d'emploi et périodes de chômage (dates de début et de fin, durées)
- situation hebdomadaire vis-à-vis de l'activité
- nombre total de semaines pendant lesquelles la personne a été occupée, en chômage et inactive chaque année
- périodes de cumul d'emplois
- congés

Antécédents professionnels :

- nombre d'années de travail à temps plein et nombre d'années de travail à temps partiel
- Nombre total d'années de travail complètes (équivalents d'années complètes de travail à temps plein)

Caractéristiques des périodes sans travail :

- recherche d'emploi pendant la période
- date des périodes de recherche d'emploi
- désir de travailler
- Raison pour laquelle la personne ne cherchait pas d'emploi

Caractéristiques des emplois: les renseignements sont recueillis pour six emplois au plus par année et mis à jour chaque année; les dates des changements sont enregistrées

- dates de début et de fin, date de début avec cet employeur pour la première fois

- salaire
- horaire de travail (heures et type)
- avantages sociaux
- affiliation syndicale
- profession
- tâches de supervision/gestion
- catégorie de travailleurs
- durée d'emploi
- façon dont la personne a obtenu l'emploi
- raison pour laquelle l'emploi a pris fin

Caractéristiques des congés d'une semaine ou plus : des renseignements détaillés sont recueillis au sujet du premier et du dernier congé chaque année, pour chaque employeur

- dates
- motif
- congé rémunéré ou non rémunéré

Attributs de l'employeur :

- secteur d'activité
- taille de l'entreprise
- secteur public ou privé

II. Revenu et patrimoine

Sources de revenu :

- renseignements annuels concernant environ 25 sources de revenu
- revenu total
- impôt payé
- revenu après impôt

Compensation financière : s'il s'agit de prestations reçues au cours de l'année, et, si oui, mois au cours desquels la personne a touché ces prestations

- Prestations d'assurance-chômage
- Assistance sociale
- Indemnités pour accidents de travail

Avoirs et dettes : renseignements concernant environ 20 catégories d'avoirs et de dettes, pouvant être recueillis une ou deux fois pendant la durée du panel.

III. Études

Activité scolaire :

- inscription à un programme d'études créditées, durée en mois
- genre d'établissement d'enseignement
- fréquentation à temps plein ou à temps partiel
- certificats reçus

Niveau de scolarité : (variables mises à jour chaque année)

- nombre d'années d'études
- grades et diplômes obtenus
- principal domaine d'études

IV. Caractéristiques personnelles

Caractéristiques démographiques :

- date de naissance / âge
- sexe
- durée de l'état matrimonial actuel
- date/ âge du premier mariage

Caractéristiques ethno-culturelles :

- antécédents ethniques
- appartenance à un groupe désigné conformément au programme d'équité en matière d'emploi
- langue maternelle
- date d'immigration
- pays de naissance
- niveau de scolarité du père et de la mère

Limitations d'activité :

- renseignements annuels sur les limitations d'activité et leurs répercussions sur le travail
- la satisfaction à l'égard de l'emploi

Renseignements sur les enfants d'une personne :

- nombre d'enfants nés, élevés
- année où le premier enfant est né et âge de la personne à ce moment

Géographie et mobilité géographique :

- région économique ou région métropolitaine de recensement du domicile actuel
- taille de la collectivité

- a déménagé dans l'année
- date du déménagement
- motif du déménagement
- nature du déménagement (déménagement du ménage complet/fractionnement du ménage)

Renseignements sur le ménage et sur la famille (économique) : renseignements sommaires annuels au niveau (p.ex. taille, genre)

- caractéristiques principales des autres membres du ménage / de la famille (p.ex. âge, sexe, lien avec les autres membres, revenu, nombre annuel d'heures travaillées)
- seuil de faible revenu pertinent
- événements relatifs à la famille (séparation, décès, naissance)
- type de logement et mode d'occupation