



**Conseil Économique
et Social**

Distr.
GÉNÉRALE

CES/2000/16
30 mars 2000

FRANÇAIS
Original: ANGLAIS

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE POUR L'EUROPE
CONFÉRENCE DES STATISTICIENS EUROPÉENS

Quarante-huitième réunion plénière
(Paris, 13-15 juin 2000)

Thème 2

**UTILISATION DE SOURCES ADMINISTRATIVES POUR L'ÉLABORATION
DE STATISTIQUES ET COMPARABILITÉ SUR LE PLAN INTERNATIONAL**

Communication sollicitée auprès de Statistics Denmark¹

I. Introduction

1. Il existe dans la plupart des pays de la CEE, sinon dans tous, une tendance à exploiter davantage les sources administratives pour alléger la charge pesant sur les individus et les entreprises, et pour économiser les ressources des instituts nationaux de statistique (INS). Cette utilisation croissante des sources administratives comporte à la fois des avantages et des difficultés qui méritent d'être étudiés. Les avantages ont déjà été examinés dans diverses rencontres internationales, et la perspective d'un système statistique général fondé sur des registres a été exposée par exemple dans un document de Statistics Denmark datant de 1995. Dans la présente communication, nous traiterons surtout de quelques problèmes délicats, en nous efforçant d'appuyer l'analyse sur des exemples concrets.

II. Différents types de sources administratives

2. Nous utiliserons ici indifféremment les termes "sources administratives" et "registres". Mais que faut-il entendre par "statistiques fondées sur des registres" ? Un registre est normalement une collection de données recueillies à des fins administratives par un organisme dont le but n'est pas de produire des

¹ Établie par Lars Thygesen.

statistiques. La production de statistiques constitue donc un usage secondaire des données. Quand on parle de registre, on pense souvent seulement aux données recueillies par des administrations publiques, par exemple les autorités fiscales, mais nous inclurons aussi sous ce terme les données recueillies ou rassemblées à l'intérieur de n'importe quelle entité, une entreprise par exemple, à ses propres fins d'administration comme la comptabilité ou la gestion des salaires.

3. Pour que l'on puisse parler de registre, il faut que les données soient conservées dans un système structuré et normalisé, une sorte de base de données consacrée à certains objets ou certaines entités. S'agissant des sources de statistiques, les entités les plus importantes sont les personnes physiques, les entreprises et les unités immobilières (parcelles cadastrales, quartiers urbains, etc.).

4. Les organisateurs de cette réunion ont proposé une typologie des données d'enregistrement qui pourrait être exploitée en statistique, selon leur origine. Ils distinguent les types de registre suivants :

1. *Les registres internes des entreprises* : créés par la direction de l'entreprise (et liés par exemple au système interne de comptabilité).
2. *Les registres tenus par des sociétés de gestion d'entreprises* : créés par la direction d'une entreprise qui tient des registres pour d'autres entreprises (ici se pose la question des pratiques comptables).
3. *Les registres publics* : créés par une administration ou un service d'enregistrement du gouvernement central, local ou régional.
4. *Les registres quasi publics* : créés par une administration ou un service d'enregistrement d'une institution ayant certaines compétences et fonctions spéciales (par exemple, les registres de la sécurité sociale).

5. Le schéma ci-après illustre les différentes sources de statistiques :

6. Les flèches en gras indiquent les voies par lesquelles les données administratives parviennent à l'INS. On n'a pas fait figurer sur le schéma les registres des sociétés de gestion d'entreprises, qui ne constituent qu'une forme particulière des registres internes des entreprises.
7. On peut faire valoir qu'il n'y a pas de grande différence entre la situation dans laquelle l'entreprise répond directement à l'INS (au moyen d'un questionnaire quelconque) et celle dans laquelle les données sont extraites de son système de gestion. Il est raisonnable de penser que, dans les deux cas, les données devront être extraites du même système de gestion mais, tandis que le processus doit être automatisé dans le second cas, on imagine une forme quelconque d'intervention humaine intelligente – de traduction ou transposition – dans le premier. C'est ce processus intelligent de transposition qu'il nous faudra reproduire pour établir les statistiques d'entreprises sur la base de registres.
8. Alors que l'utilisation des registres administratifs publics et quasi publics, en particulier pour l'élaboration de statistiques sociodémographiques, est déjà bien développée dans plusieurs pays, celle des registres internes des entreprises et des registres des sociétés de gestion d'entreprises l'est beaucoup moins. Au Danemark, ces deux méthodes sont employées depuis de nombreuses années pour établir les statistiques relatives aux salaires et rémunérations. Nous fondons tous de grands espoirs sur l'échange de données informatisées (EDI) et les possibilités qu'il offre d'extraire et de présenter des données statistiques concernant l'activité économique directement à partir des systèmes comptables des entreprises :

9. Des applications prometteuses comme celle-là ont été mises au point grâce au projet de recherche-développement TELER de l'Union européenne, dont est issu le logiciel Edisent. L'expérience montre que l'analyse des concepts du système de gestion et leur traduction en concepts statistiques est une tâche lourde, qui exige des communications intensives avec chaque entreprise. Une fois ce travail fait, les avantages potentiels pour les entreprises comme pour les INS sont évidents.

Quelques exemples illustreront l'éventail des tâches statistiques dont nous parlons ici :

1. *Registres internes des entreprises* : utilisation directe des livres comptables de la société comme base de l'élaboration des statistiques d'entreprises. Au Danemark, nous n'avons pas encore pleinement exploité l'approche du type Edisent mentionnée plus haut, mais nous étudions, dans le cadre d'un projet-pilote, la possibilité d'extraire de la comptabilité des entreprises des données pour l'échange de données informatisées. En attendant, les entreprises sont invitées à communiquer leurs comptes internes, que Statistics Denmark doit ensuite interpréter pour les traduire en termes statistiques.

2. *Registres des sociétés de gestion d'entreprises* : l'exemple danois le plus marquant est celui des statistiques des rémunérations. Un petit nombre de sociétés de gestion d'entreprises gère les salaires d'une grande proportion d'entreprises privées, en utilisant des systèmes normalisés. Si l'entreprise y consent, la société envoie à Statistics Denmark, pour chaque salarié, le formulaire normalisé pour les statistiques des salaires établi par Statistics Denmark. Statistics Denmark reste en liaison avec les sociétés de gestion d'entreprises pour leur permettre d'adapter leurs systèmes aux besoins statistiques.

3. *Registres publics* : au Danemark, la plupart des statistiques sociodémographiques reposent sur des registres publics, gouvernementaux et municipaux, qui utilisent des numéros communs d'identification tels que le *numéro personnel*. Pour les statistiques des entreprises, le Registre central des entreprises (CVR) constitue un élément d'une infrastructure analogue. Les statistiques du cheptel offrent un exemple un peu différent : jusqu'ici, la collecte des données se faisait directement auprès des agriculteurs, qui indiquaient par exemple le nombre de leurs vaches, mais une réforme est en cours, et chaque animal est désormais immatriculé dans un registre public, le Registre central du cheptel. On ne sait pas encore avec certitude quelle est la qualité de ces nouvelles statistiques, ni si les nouveaux chiffres seront compatibles avec les anciens.

4. *Registres quasi publics* : statistiques estimant les tendances à court terme de l'emploi, sur la base des contributions totales des employeurs au Régime de pension complémentaire (ATP).

III. Application dans les pays

10. Comment l'institut national de statistiques d'un pays doit-il s'y prendre pour exploiter ces ressources potentielles ? D'abord, il peut établir une liste des administrations et services d'enregistrement nationaux pertinents. Ensuite, il devra évaluer ces sources du point de vue de leur utilité conceptuelle (y compris l'aspect de la pollution), ce qui signifie qu'il devra comparer les concepts utilisés par les services d'enregistrement et les administrations avec les concepts statistiques. Troisièmement, il lui faudra les

examiner du point de vue de leur accessibilité légale et institutionnelle. Quatrièmement, il devra les examiner du point de vue de la gestion des données et de la technologie de l'information, pour déterminer si l'utilisation est possible et sous quelle forme (utilisation primaire et/ou secondaire). Mais, tout au long de cette procédure, il faudra garder à l'esprit que l'analyse doit prendre pour point de départ les variables-cibles - ce que l'on cherche à mesurer – et que nous ne devons pas être obnubilés par le fait que les données sont déjà disponibles.

IV. Quelques difficultés

11. En même temps que nous reconnaissons les avantages qu'il peut y avoir à se lancer dans l'élaboration de statistiques fondées sur des registres, nous devons être conscients des difficultés de cette entreprise, qui sont considérables et très diverses : elles concernent aussi bien la qualité des données, y compris leur comparabilité, que les aspects pratiques et juridiques ou les ressources nécessaires.

IV.1 Validité et pertinence

12. La question la plus importante est naturellement de savoir s'il est possible d'utiliser les données contenues dans les registres pour analyser les concepts qui sont étudiés en statistique. Ces données sont-elles pertinentes ? Le problème vient de ce que nous ne maîtrisons pas le mécanisme de mesure comme nous le faisons dans une enquête statistique traditionnelle. Les sources administratives ont leurs propres concepts, définis selon leurs propres fins, et ces concepts ne correspondent pas nécessairement aux fins statistiques. Il faut donc, d'une manière ou d'une autre, transposer ou traduire les concepts administratifs en concepts statistiques. Dans quelle mesure la validité, la fiabilité et la comparabilité internationale vont-elles être affectées ?

13. Dans le meilleur des cas, le statisticien peut avoir la chance d'exercer une influence sur le contenu des registres. Dans certains pays, comme le Danemark, la législation donne à l'institut national de statistique le pouvoir d'exercer une influence sur les registres qui pourront être utilisés à des fins statistiques. Cependant, nous ne devons pas nous faire trop d'illusions sur ce point.

14. Une autre solution, qui a donné de bons résultats dans de nombreux cas, consiste à combiner les données. Il est parfois possible de raccorder les données de manière à ce que différentes sources se complètent : on obtient alors l'information souhaitée en comparant un certain nombre de données provenant de différents registres qui renseignent plus ou moins parfaitement sur le sujet. L'utilisation de sources multiples de données sera examinée plus loin.

15. En ce qui concerne la pertinence, il faut aussi tenir compte du fait qu'un système statistique exclusivement, ou dans une très large mesure, fondé sur les registres des administrations publiques peut finir par donner du monde l'image que s'en font ces administrations. Les concepts ou les catégories de personnes qui n'existent pas pour les autorités fiscales ou les services sociaux peuvent également ne pas apparaître dans les statistiques : ce pourrait être le cas par exemple des personnes sans domicile fixe.

16. Pour éliminer ce risque, il importe qu'il existe un certain nombre de registres de base, par exemple le Registre central des entreprises (CVR) du Danemark, où sont enregistrées toutes les unités et où sont

réunis des renseignements provenant de nombreuses sources indépendantes, sans finalité administrative précise. Nous savons aussi qu'il est possible de raccorder des fichiers de sources différentes, de manière à fournir une statistique qui n'est liée à aucun point de vue administratif particulier. En même temps, il importe de ne pas oublier que l'on ne peut pas attendre des registres qu'ils fournissent une image complète de la société, et qu'il est donc nécessaire de recueillir des données statistiques précisément pour révéler des aspects qui sont absents des registres. Ces données complémentaires peuvent être recueillies au moyen d'enquêtes par entretiens ou par questionnaires, et, dans l'idéal, elles devraient comporter des identifiants qui permettent de les comparer et de les compléter avec les données provenant des registres.

IV.2 Fiabilité et précision

17. Autre critère auquel doivent satisfaire les données de base pour l'établissement des statistiques : celui de la fiabilité. Ce critère est d'ailleurs applicable que les données proviennent de questionnaires ou de registres. On doit pouvoir considérer avec un grand degré de certitude que les données reflètent fidèlement la réalité. Si par exemple une personne est déclarée comme artiste-peintre alors qu'elle travaille dans les services statistiques du gouvernement, les conclusions qui seront tirées des statistiques risquent d'être inexactes.

18. On pourrait bien entendu faire valoir que, tant que les erreurs ne sont pas systématiques, elles n'entraîneront pas une distribution asymétrique, parce que les erreurs en sens contraire ont tendance à s'annuler mutuellement. Mais lorsque, comme cela est souvent nécessaire, nous examinons l'incidence statistique de plusieurs variables, de telles erreurs sont source de problèmes, parce qu'elles peuvent entraîner une distorsion dans l'étude des corrélations.

19. Si les erreurs tendent systématiquement dans la même direction, les problèmes sont immédiatement amplifiés mais, étant systématiques, ils sont peut-être plus faciles à corriger.

20. Enfin, il est important que les données soient enregistrées avec un degré de précision approprié aux besoins statistiques, autrement dit que l'échelle utilisée soit suffisamment fine. Par exemple, l'information sur l'activité professionnelle doit être relativement détaillée si l'on veut pouvoir l'utiliser pour des études épidémiologiques, dans lesquelles le risque de maladie doit être examiné en fonction d'effets liés au travail.

21. Le besoin de précision demeure, même si l'information n'est pas destinée à être utilisée dans des rapports statistiques au niveau d'enregistrement considéré. Très souvent, des groupes sont formés sur la base de variables obtenues à partir d'éléments discontinus. La définition et l'application de telles variables dans des contextes statistiques dépendent fondamentalement de la précision du matériau de base.

IV.3 Comparabilité dans le temps

22. De grandes difficultés peuvent se poser sur le plan statistique lorsque des modifications des lois ou des règlements entraînent des modifications dans le contenu des registres administratifs. D'une part, il peut être difficile ou impossible d'estimer la tendance à long terme d'un indicateur donné si des

définitions différentes sont utilisées dans le matériau de base. D'autre part, il peut être difficile de décider quels changements dans les valeurs de données doivent être imputés à des événements extérieurs, et quels changements résultent simplement de nouveaux concepts ou définitions.

23. Les changements des données contenues dans les registres administratifs peuvent résulter de modifications de la loi applicable au domaine, et les conséquences statistiques dépendront du type de statistiques considéré. S'il s'agit de statistiques ayant pour but de contrôler l'effet de la législation, c'est-à-dire de montrer comment l'administration d'une loi touche l'homme de la rue, par exemple, ces statistiques peuvent bien entendu se contenter de reprendre les concepts de la nouvelle loi. Ainsi, les statistiques relatives aux prestations sociales devront, à tout moment, refléter les règles applicables à ce domaine et il pourra être difficile d'apprécier des aspects tels que les changements de comportement.

24. Si, au contraire, il s'agissait de statistiques plus générales, l'impossibilité de comparer les concepts statistiques avant et après le changement législatif serait inacceptable. Les statistiques générales cherchent à élucider certains concepts qui ne sont pas définis dans la législation, par exemple le chômage, concept dont la définition se trouve dans les conventions de l'OIT. Si les statistiques ont trait au paiement des prestations sociales (par exemple, une allocation journalière), et que les règles applicables à celles-ci sont modifiées, il risque de devenir difficile, voire impossible, de compenser le changement dans les statistiques. Il faudra du moins s'efforcer d'estimer la signification des changements, de manière à relier dans une certaine mesure des séries chronologiques.

25. Prenons un exemple emprunté au Danemark : celui des statistiques estimant les tendances du chômage à court terme, sur la base des contributions totales des employeurs au régime de pension complémentaire des travailleurs (ATP)². L'objectif essentiel de ces statistiques étant de décrire des tendances à court terme, il est vital d'éliminer ou de réduire l'effet de changements dans les règles sur la mesure de l'emploi. La loi sur l'ATP a été modifiée à diverses reprises : par exemple elle a été étendue pour couvrir de nouvelles catégories, notamment les agents de la fonction publique ou les jeunes salariés, qui y échappaient jusque-là. Dans ces cas, il a fallu corriger les chiffres administratifs en utilisant comme source auxiliaire l'enquête sur les forces de travail.

26. Autre exemple : les statistiques à court terme sur le chiffre d'affaires, qui sont fondées sur les déclarations des entreprises aux fins de la TVA. Jusqu'en 1999, toutes les entreprises - hormis celles du secteur agricole - faisaient des déclarations trimestrielles mais, depuis le 1er janvier 1999, le système a été modifié; désormais les grandes entreprises doivent faire une déclaration mensuelle, les moyennes entreprises une déclaration trimestrielle et les petites entreprises une déclaration semestrielle. Les statistiques, elles, sont établies sur une base trimestrielle. Pour pouvoir continuer à les établir, il faut procéder par estimation du chiffre d'affaires pour les petites entreprises qui n'ont pas fait leur déclaration au moment où les statistiques doivent être compilées et publiées. De cette manière, une incertitude s'est introduite dans des statistiques jusque-là fondées sur le chiffre d'affaires total.

² ATP = *Arbejdsmarkedets Tilloegspensionsordning* (Régime de pension complémentaire administré conjointement par les syndicats et les organisations d'employeurs).

27. Enfin, il y a aussi des cas où les registres administratifs sont modifiés sans que les règlements aient changé. Généralement, ces modifications sont introduites pour des raisons techniques, ou pour des raisons de rationalisation. Cela peut poser un problème, par exemple, s'il est décidé qu'une certaine catégorie de données n'est plus nécessaire.

28. Les modifications du système peuvent aussi avoir un caractère plus local, par exemple lorsque les limites d'une commune ou le système de numérotation d'une rue ont changé, si bien qu'un certain nombre de personnes ont changé d'adresse sans avoir déménagé. Dans ce cas, il est utile que les changements puissent au moins être signalés comme des rectifications.

29. Toutes ces modifications ont, de toute évidence, des conséquences graves pour nous si on les introduit sans tenir compte suffisamment des besoins statistiques. Il est donc important que les statisticiens soient appelés à participer aux modifications, au moins pour en estimer les conséquences avant qu'elles ne soient introduites.

IV.4 Les statisticiens peuvent-ils influencer les données ?

30. Depuis 1970, début du développement des statistiques fondées sur les registres au Danemark, est apparue comme une difficulté importante l'impossibilité d'exercer sur le contenu des données de base le contrôle auquel nous étions habitués (ou croyions l'être) lorsque nous produisions des statistiques à partir de questionnaires. Nous n'avons pas la certitude que les registres couvrent les unités pertinentes avec le même niveau de précision qu'auparavant, ni que les données sont définies conformément aux besoins des utilisateurs. Des problèmes supplémentaires surgissent lorsque le contenu des registres change avec le temps et c'est paradoxalement le cas même si les registres deviennent plus fiables, puisque, même alors, il en résulte une discontinuité dans les données.

31. Il est bien sûr souhaitable que les statisticiens exercent une certaine influence sur le contenu des données, mais il faut se rappeler en même temps que les registres sont tenus pour des fins administratives bien précises, et que la tâche des services d'enregistrement est de servir au mieux ces fins : ils ne peuvent donc pas trop se préoccuper de ce qui serait nécessaire ou souhaitable d'un point de vue purement statistique.

32. Au Danemark, cependant, l'article premier de la loi sur l'institut de statistique (Statistics Denmark) dispose que toute personne qui a l'intention d'établir un registre doit en discuter avec Statistics Denmark pour que le registre soit constitué d'une manière qui corresponde aussi aux besoins statistiques. En outre, il existe une coopération très bonne et très étroite, entre Statistics Denmark et les services d'enregistrement dans de nombreux domaines.

33. Les statisticiens ne doivent pourtant pas multiplier les demandes de données supplémentaires, de définitions différentes etc., en s'attendant à ce que ces demandes soient satisfaites. Les services d'enregistrement doivent veiller strictement à l'efficacité de leurs propres opérations et de l'utilisation de leurs ressources. Aussi les statisticiens doivent-ils faire preuve d'une grande modestie et ne présenter des demandes d'envergure que lorsqu'ils peuvent penser que la modification du registre entraînera des avantages très importants pour la société. On connaît très peu de cas où des données additionnelles sont recueillies par un service d'enregistrement uniquement pour les besoins statistiques.

IV.5 Actualité

34. L'actualité est un aspect important de la qualité des statistiques, et il importe donc de chercher à voir si elle est affectée par l'utilisation de sources administratives. Nous n'avons pas pu cependant déceler de tendances nettes jusqu'ici. Au Danemark, certaines statistiques fondées sur des registres ont des délais de production extrêmement brefs, par exemple les statistiques démographiques générales, qui peuvent être publiées sous forme détaillée en quelques mois. Le modèle Edisent pour les statistiques d'entreprises semble aussi pouvoir accélérer les choses, puisque les procédures sont automatisées. En revanche, certaines statistiques danoises, par exemple celles sur les revenus et sur la population active, sont fondées sur plusieurs sources de l'administration fiscale, qui ont besoin de temps pour obtenir des données suffisamment complètes, ce qui signifie que la publication des statistiques ne peut avoir lieu qu'un an au plus tôt après la date à laquelle elles se rapportent.

IV.6 Comparabilité internationale

35. Depuis leur apparition vers 1970, on a beaucoup débattu sur le plan international du risque que les statistiques modernes fondées sur des registres ne nuisent à la comparabilité internationale des statistiques. En même temps, l'importance de cette comparabilité internationale a été de plus en plus généralement reconnue, à mesure que la coopération internationale se développait. Il en est résulté une normalisation croissante des concepts et des définitions, ainsi que des accords sur les nomenclatures et les systèmes statistiques. Dans la plupart des cas, la normalisation internationale des statistiques a porté surtout sur l'aspect conceptuel et il n'y a pas eu d'accords sur les instruments de mesure à utiliser, ou seulement dans quelques cas comme celui des recensements de population. Cependant, le développement du système statistique européen de l'Union européenne a peu à peu changé cette situation, si bien qu'il existe aujourd'hui quelques outils de mesure totale comme l'Enquête sur les forces de travail (EFT) pour laquelle les États membres sont tenus par des dispositions légales ou conventionnelles d'utiliser une méthode de mesure normalisée.

36. Depuis de nombreuses années, le Danemark plaide en faveur des statistiques fondées sur des registres, en disant que souvent leur qualité n'est pas inférieure, ou est même supérieure, à celle des statistiques traditionnelles. Cela n'a pas toujours été facile, parce que les statisticiens, accoutumés à leurs propres méthodes, ne se demandent pas ce qui se passe réellement lorsqu'un gestionnaire d'entreprise reçoit un questionnaire statistique conçu par des statisticiens qui ne comprennent peut-être pas pleinement les phénomènes qu'ils s'efforcent de mesurer. Ce que nous mesurons a-t-il toujours un sens ?

37. La question se pose en termes légèrement différents aujourd'hui : pouvons-nous produire des chiffres comparables en utilisant deux méthodes de mesure complètement différentes ? Premièrement, nous devons nous rappeler que nous ne devons pas utiliser les concepts des sources administratives sauf si, par hasard, ils coïncident avec nos concepts statistiques. L'existence de données dans une source accessible n'est pas une raison suffisante de faire des statistiques. Au contraire, nous devons partir des concepts que nous voulons mesurer et qui sont souvent définis par concertation internationale. C'est seulement après que nous devons nous demander quelle est la meilleure méthode pour les mesurer.

38. Comme on l'a déjà vu plus haut, on dispose de certaines techniques pour obtenir une mesure centrale de bonne qualité : la traduction ou transposition des informations de base, l'estimation reposant

sur la combinaison de données issues de sources multiples, etc. Il est essentiel que ces méthodes soient fondées sur l'utilisation de l'éventail complet des données disponibles, au niveau de détail le plus poussé. Dans ce cas, on aura de bonnes chances d'obtenir un bon résultat, et ce résultat ne devrait pas être différent de celui auquel on parviendrait par les méthodes traditionnelles, pourvu que celles-ci fonctionnent raisonnablement aussi.

39. Ce qui peut faire obstacle à la comparabilité, cependant, c'est que les statistiques fondées sur des registres présentent des différences qualitatives par rapport aux enquêtes et recensements traditionnels. Si nous considérons l'Enquête sur les forces de travail qui est effectuée dans tous les États membres de l'Union européenne, nous constatons que les taux de non-réponse varient de 30 à 5 % selon les pays - ce qui, d'ailleurs, risque aussi de nuire à la comparabilité des résultats entre les pays. Nous pourrions facilement envisager de remplacer cette enquête dans certains pays par un dénombrement fondé sur des registres, mais cela entraînerait un défaut de comparabilité, parce que les statistiques fondées sur des registres ont un taux de non-réponse en principe nul.

40. On a beaucoup trop peu étudié jusqu'ici les différences entre statistiques fondées sur les enquêtes et statistiques fondées sur les registres, bien que quelques études aient été faites (Statistics Finland). Une étude suédoise récente (Håkan Lindström 1999) traite des normes de qualité des statistiques fondées sur des sources administratives.

41. S'agissant de la comparabilité internationale, on a proposé comme solution d'utiliser un outil commun de mesure fonctionnant exactement de la même manière dans plusieurs pays (comme l'EFT), pour étalonner en les rapportant à une norme commune les estimations d'autres statistiques fondées sur les registres.

IV.7 Faisabilité : législation et confidentialité

42. Même s'il est sans doute très utile de combiner des sources administratives pour créer des informations nouvelles, cette opération peut être interdite par la loi pour des raisons de confidentialité. Nous savons que cela est effectivement le cas dans de nombreux pays. Quelles sont les limites en la matière et comment peut-on les surmonter ? Quel type de loi ou de règlement concernant la confidentialité facilite ou entrave l'utilisation des sources administratives ?

IV.8 Les statistiques fondées sur des registres sont-elles moins coûteuses ?

43. Il est évidemment plus coûteux de recueillir des données *ex nihilo* que d'utiliser des données qui existent déjà dans un système informatique. Cependant, établir des statistiques à partir de registres est loin d'être simple. Cela exige de grandes connaissances de la part de l'INS, et aussi une grande quantité de travail, ce que nous avons parfois tendance à oublier. Un investissement est indispensable dans la phase initiale.

44. Par exemple, si nous voulons exploiter l'information interne d'une entreprise, nous devons comprendre son organisation beaucoup mieux que si nous nous contentons de lui envoyer un questionnaire. Sauf si les entreprises utilisent des systèmes normalisés, il nous faudra établir et maintenir des relations étroites avec chacune d'elles. Cela sera sans doute extrêmement utile, car nous

comprendrons mieux alors la pertinence ou la faisabilité des concepts statistiques que nous utilisons habituellement, c'est-à-dire des renseignements que nous avons l'habitude de demander : peut-être cela nous montrera-t-il que ce que nous demandons habituellement n'est pas tout à fait pertinent. Mais c'est aussi un processus extrêmement coûteux, parce que nous devons bien comprendre les activités des différentes entreprises, en particulier des plus grandes, pour définir convenablement les mécanismes nécessaires de conversion, parce que l'analyse des concepts suppose beaucoup de ressources et de travail, etc. Il en résulte aussi une charge plus lourde pour l'INS, qui doit faire correspondre les données aux concepts harmonisés utilisés.

45. L'expérience du Danemark montre clairement que, une fois que cet investissement a été fait, l'INS peut réaliser des économies considérables dans ses opérations courantes, économies qui compensent très largement l'investissement initial. C'est ce qu'a prouvé le recensement de la population et du logement de 1981, effectué à partir des registres, et dont le coût pour Statistics Denmark n'a pas dépassé le tiers de celui d'un recensement traditionnel. De même, l'expérience montre que l'on peut économiser des ressources en ce qui concerne les statistiques d'entreprises. En même temps, il ne faut pas oublier que les répondants aussi pourront généralement économiser des ressources, puisque les données seront réutilisées et n'auront pas besoin d'être créées à chaque fois. Au cours du débat public et politique sur la question, l'importance d'alléger la charge de réponse des entreprises en particulier a été soulignée. Au Danemark, on a estimé à 245 années-homme le temps que passeront les entreprises privées à remplir des questionnaires en l'an 2000.

V. Normes internationales possibles

46. La transition vers des statistiques fondées sur les sources administratives s'est faite de manière plus ou moins indépendante dans les différents pays, bien que certains se soient naturellement inspirés parfois des pratiques des autres. Dans d'autres domaines de la statistique, les normes internationales se sont avérées utiles, et il est bon d'examiner si une approche analogue pourrait s'appliquer à certains aspects des statistiques fondées sur les registres, pour surmonter quelques-uns des problèmes évoqués plus haut dans la section 4. Cependant, un grand nombre des problèmes, en raison de leur nature, ne peuvent pas être réglés par des normes : c'est le cas, par exemple, des problèmes politiques. Il n'en demeure pas moins qu'un institut national de statistiques peut très bien tirer profit de l'expérience des autres.

47. Il ne faut pas non plus oublier que la première condition, si l'on veut réaliser la comparabilité, est de s'entendre au niveau international sur des définitions communes des concepts à mesurer. C'est aussi la tâche la plus difficile, et elle occupe un grand nombre de statisticiens pendant des années dans le cadre de groupes de travail internationaux.

48. Le cadre légal : l'obstacle le plus sérieux au développement de l'utilisation des sources administratives dans de nombreux pays est sans doute le manque de base légale. Pourrait-on envisager des dispositions-types dans ce domaine ? Les règles qui fonctionnent bien dans certains pays peuvent-elles constituer un modèle pour les autres ? La question est très controversée et de caractère hautement politique.

49. Les techniques de protection des statistiques : l'une des raisons pour lesquelles l'utilisation des statistiques de sources administratives se heurte à une forte opposition politique est qu'elle placerait entre les mains de l'INS un grand nombre de données confidentielles concernant les personnes physiques et les entreprises, ce qui fait craindre que des informations puissent être divulguées ou des individus identifiés à partir d'informations statistiques. Les chercheurs par exemple souhaiteraient beaucoup pouvoir exploiter pour leurs travaux des données détaillées précieuses qui ont déjà été recueillies à des fins statistiques. Au Danemark, les règles de protection sont très strictes. Par exemple, même les microdonnées non identifiées ne sont communiquées à aucun utilisateur hors de l'institut national de statistiques, alors que d'autres pays rendent publics des fichiers de microdonnées. Des pressions peuvent s'exercer sur les INS ayant des règles restrictives, pour les inciter à plus d'ouverture. L'idéal serait d'inventer et de faire accepter sur le plan international une technique permettant de protéger la confidentialité des microdonnées. Certains projets de recherche-développement entrepris dans ce domaine par l'Union européenne sont prometteurs, mais aucune solution n'a encore été trouvée.

50. Les systèmes comptables : il serait très utile, du point de vue de la comparabilité, de pouvoir s'entendre sur des normes internationales contraignantes applicables aux systèmes comptables, compte tenu des normes statistiques internationales. Toutefois, l'expérience montre que c'est là quelque chose de très difficile, même à l'intérieur d'un pays. Au Danemark, nous avons essayé il y a quelques années d'instaurer une norme qui réponde aux besoins statistiques tout en satisfaisant les autorités fiscales et les autres administrations. Cette norme avait même été acceptée, mais avant même qu'elle puisse être appliquée, des groupes de pression des entreprises ont réussi à persuader le Gouvernement qu'elle aurait pour eux des conséquences très lourdes. Le Ministre des finances a progressivement assoupli la norme, qui est ainsi devenue insuffisante comme source unique pour les statistiques des comptes, et Statistics Denmark s'est vue obligée de compléter ce système en adressant des questionnaires à un échantillon d'entreprises, pour pouvoir établir des statistiques des comptes d'une qualité suffisante.

VI. L'avenir

51. À notre avis, le développement de l'exploitation des sources administratives pour l'établissement des statistiques passera essentiellement par l'amélioration de la traduction des concepts administratifs en concepts statistiques, et en particulier par l'harmonisation des règles applicables au traitement des données dans les différents domaines statistiques de manière à permettre d'utiliser la même définition pour un certain concept, quelles que soient les statistiques concernées. Jusqu'ici, le choix des données et des règles relatives à leur traitement a eu tendance à concerner un horizon assez étroit, centré sur un certain type de statistiques. Il y aura à l'avenir davantage de coopération entre les différents domaines statistiques et une meilleure organisation du traitement des données. Cela devrait avoir pour résultat d'accroître la qualité des données statistiques obtenues par combinaison de données administratives et d'améliorer la comparabilité des concepts.

52. Les travaux effectués dans le cadre d'Eurostat pour préparer l'ECHP, ou l'étude qui le remplacera après l'an 2002, sont très intéressants. Ils sont l'occasion d'examiner les problèmes que pose la comparaison des données recueillies directement par des questionnaires ou des entretiens et des données provenant de sources administratives. On espère qu'il se dégagera de ces travaux une comparabilité suffisante pour qu'on puisse utiliser les deux types de données.

53. L'accès aux données statistiques sur les personnes physiques et les entreprises est une question très urgente. L'équilibre entre protection de la personne et besoins statistiques est délicat, et il suscite dans de nombreux pays des problèmes politiques, qui ne sont probablement pas près de disparaître.

Références

Statistics Denmark 1995 : *Statistics on Persons in Denmark, A Register-Based Approach*. Eurostat, Luxembourg 1995

Statistics Finland 1994 : *Evaluation Study of the Census (1994)*. Statistics Finland, Population Census, Volume 6B. Helsinki

Håkan Lindström 1999 : *Kvalitetssäkring i register för statistikproduktion med administrativt underlag*. Statistics Sweden, Örebro.
