

GROUPE DE TRAVAIL DE DEMOGRAPHIE AFRICAINE
(IDP, INED, INSEE, MICOOP, ORSTOM)

SYNTHESE des RECENSEMENTS AFRICAINS

DOCUMENTS D'INFORMATION

35

LES FICHIERS INFORMATISES



Paris

Decembre 1980

PRESENTATION

La Synthèse des Recensements Africains est un projet du Groupe parisien de Démographie Africaine (IDP, INED, INSEE, MICOOP, ORSTOM) réalisé en liaison avec le RIPS (Accra - Ghana). Son programme comporte la publication de Monographies Méthodologiques sur les derniers recensements, sous forme séparée et provisoire d'abord, regroupées dans leur version définitive ensuite. La synthèse proprement dite de ces monographies doit intervenir dans un second temps.

Parallèlement à cet effort, divers Documents d'Information sont également diffusés sur les sujets susceptibles d'intéresser les responsables de recensements.

Les deux séries de documents sont publiées sans périodicité régulière; leur diffusion est restreinte. Pour toute correspondance à leur sujet, s'adresser à :

Robert BLANC
I.N.E.D.
27, rue du Commandeur
75675 PARIS CEDEX 14
Tel. 320-13-45

PUBLICATIONS DEJA PARUES

Monographies Méthodologiques

- 1 - Mauritanie
- 2 - Haute-Volta
- 3 - Tunisie
- 4 - Algérie
- 5 - Somalie
- 6 - Congo
- 7 - Côte d'Ivoire
- 8 - Ghana
- 9 - Sénégal
- 10 - Gambie
- 11 - Libye
- 12 - Soudan
- 13 - Niger
- 14 - Mali
- 15 - Togo
- 16 - Liberia
- 17 - Cameroun
- 18 - Bénin
- 19 - Swaziland (à paraître)
- 20 - Malawi (à paraître)
- 21 - Gabon
- 22 - Guinée-Bissau
- 23 - Maroc
- 24 - Sierra Leone
- 25 - Madagascar
- 26 - Mayotte
- 27 - République Centrafricaine
- 28 - Burundi
- 29 - Egypte
- 30 - Maurice

PUBLICATIONS DEJA PARUES

Documents d'Information

- 1 - Estimations indirectes de fécondité et de mortalité.
- 2 - Estimations indirectes de migrations internationales.
- 3 - Estimations de population de fait et démographie de la population non résidente.
- 4 - Le besoin d'un système d'état-civil.
- 5 - Méthodes indirectes d'estimation des migrations internationales (suite du Document d'Information N° 2).
- 6 - Enquête par sondage et recensement.
- 7 - Les migrations internationales dans le recensement de la Haute-Volta.
- 8 - Le dénombrement des nomades.
- 9 - Introduction à la synthèse des recensements africains.
- 10 - Ce que pensent les recensés d'un recensement.
- 11 - Activité économique et statut de l'emploi.
- 12 - Nombres et noms dans les recensements.
- 13 - Introduction à la synthèse des recensements africains (suite).
- 14 - Sur le lien de parenté avec le chef de ménage.
- 15 - Synthèse de recensements (Pacifique Sud).
- 16 - Etudes longitudinales et taux de réponse.
- 17 - Recensements et migrations internes.
- 18 - Echantillons renouvelés à la place de recensements.
- 19 - Croissance urbaine dans les pays en développement.
- 20 - Bilan du programme.
- 21 - La fécondité diminue-t-elle dans les pays en développement ?
- 22 - Urbanisation dans les pays arabes.
- 23 - Questions sur la fécondité et la mortalité infantiles.
- 24 - Approche du monde nomade afghan.
- 25 - Les recensements de 1980.
- 26 - L'évaluation du recensement du Maroc de 1971.
- 27 - Migrations internationales (suite).
- 28 - Le recensement et la politique.
- 29 - Cartographie.
- 30 - Enquête mondiale sur la fécondité (problèmes de collecte).
- 31 - Evaluation d'un recensement (France et Gambie).
- 32 - Migrations : Réunion de l'Enquête Mondiale Fécondité, Enquête CESAP.
- 33 - Deuxième bilan du programme.
- 34 - Contrôle, correction, ajustement : fondements et limites.
- 35 - Les fichiers informatisés. (présent fascicule)

"POPULATION INDEX" ouvre une rubrique
sur les fichiers publics de données informatiques*

par Judith S. ROWE**

Avec le numéro de Janvier 1980, Population Index sera la première publication bibliographique à comprendre des références de fichiers de données informatiques. Chaque service central de statistique ou autre bureau principal producteur de statistiques dans le monde utilise des ordinateurs pour stocker et tabuler les données de recensement, d'état civil et d'enquête. Nombre de ces bureaux font maintenant des tableaux résumés, disponibles sur bandes, généralement plus étendus que ceux figurant dans les rapports imprimés et quelquefois mettent gratuitement à la disposition du public l'utilisation d'échantillons. Les chercheurs universitaires peuvent, à travers le monde, convertir les réponses d'enquêtes et les données d'autres types sous forme informatique et déposer les fichiers produits dans des archives de données sociales. Même des données obtenues à partir de rapports imprimés, de registres paroissiaux ou d'enregistrements sur micro-films sont convertis, par simple routine, sous

* Traduction de l'article: "Population Index to cite publicly available machine-readable data files", publié dans "Population Index - Octobre 1979" - Office of Population Research - Princeton University.

** L'auteur est membre du Centre Informatique de l'Université de Princeton - 87 Prospect Avenue - Princeton, NJ 08544.

forme informatique et parfois celle-ci est, elle aussi disponible pour le public. A la fois pour l'analyse primaire et secondaire, les fichiers de données informatiques ont, dans une large mesure, remplacé les volumes imprimés, comme source de données pour la recherche démographique. L'analyse secondaire de fichiers de données produits à d'autres fins devient rapidement un mode opératoire commun pour les démographes et même des fichiers considérés comme ayant été entièrement exploités par l'investigateur original, ont, entre les mains - ou par l'ordinateur - d'analystes secondaires, fourni de nouveaux et importants résultats.

Il est coûteux de recueillir un volume substantiel de données et encore plus coûteux de produire, en outre, un fichier de données propre, facilement transportable et bien documenté. Toutefois les utilisations supplémentaires, qui sont rendues possibles par cet effort additionnel, font plus qu'en justifier le coût. Certains types de recherches ne seraient pas possibles si des fichiers de données n'étaient pas déjà disponibles pour l'analyse. Pensons par exemple au chercheur qui n'a besoin que d'une demi-douzaine de rubriques concernant la population de nombreuses villes, au chercheur qui désire retracer la composition ethnique d'une seule zone durant de nombreuses décades ou au chercheur qui aimerait tester une hypothèse sur de nombreuses populations diverses.

Le nombre de fichiers de données informatiques disponibles pour des analyses secondaires s'accroît rapidement mais il est encore difficile pour les chercheurs individuels de repérer les fichiers appropriés à leurs besoins. Il en résulte que des fichiers de données volumineux et coûteux sont sous-utilisés et que des chercheurs abandonnent des projets potentiellement fructueux parce qu'ils croient que les données nécessaires ne sont pas disponibles. Deux facteurs principaux ont empêché un plus large accès aux données informatisées. L'un a été la tendance du chercheur à regarder la donnée non comme un produit du projet mais plutôt comme quelque chose d'analogue à des notes de recherche. En conséquence, de nombreux fichiers de données n'ont circulé que parmi un groupe étroit de collègues et sont restés inconnus ou non disponibles en dehors de ce groupe. Cette procédure a laissé insatisfaits à la fois les chercheurs primaires et secondaires et un nombre croissant de fichiers est transféré vers les archives pour distribution. En même temps les agences gouvernementales remplacent la distribution informelle de données par le développement des données publiques, des codes et de la formation. Comme les fichiers de

données gagnent rapidement le domaine public, le besoin de bibliographies standardisées et accessibles devient plus apparent. Avec la publication de règles pour le classement des fichiers de données informatiques et de standards recommandés pour leur citation, il est devenu possible d'inclure ces matériaux dans des bibliographies standards.

L'inventaire des fichiers de données disponibles pour le public présente déjà beaucoup d'intérêt pour les démographes. Ceux-ci vont des tableaux résumés de données nationales de recensement pour des pays comme les Etats-Unis, le Canada, le Royaume Uni, la Norvège et le Danemark à l'utilisation publique d'échantillons de résultats de recensement provenant de l'enquête mensuelle sur la population des Etats-Unis ou de l'enquête trimestrielle sur les ménages du Royaume-Uni. Encore plus nombreux sont les enquêtes ou les sondages d'opinion traitant de fécondité, avortement, mariage, divorce, migration, mode de vie, retraite et emploi des femmes. Sont également disponibles mais souvent plus difficiles à repérer les fichiers de données créés par des chercheurs individuels généralement des étudiants gradés. De tels fichiers, qui contiennent des données relatives à des petites zones de recensement ou de registres paroissiaux, transférés à partir de rapports imprimés, de microfilms ou de registres ont maintenant leur place parmi les archives de données en Amérique du Nord ou en Europe occidentale. Parmi ces données se trouvent les caractéristiques, le comportement, les croyances, les attitudes et les valeurs des personnes de tous âges et de groupes raciaux ou ethniques divers, de nombreuses régions du monde, aussi bien de personnes vivant aujourd'hui que d'autres décédées il y a des centaines d'années. Le chercheur doué d'imagination peut utiliser des données sur d'autres populations pour créer des séries temporelles, pour faire des analyses internationales ou interculturelles, pour comparer les attitudes et le comportement et pour tester d'innombrables hypothèses sur des sous-populations variées. De même que les matériaux imprimés, les fichiers informatiques vont de la couverture étendue à la focalisation sur des sujets spécifiques. Par nature cependant, les fichiers informatiques sont habituellement assez souples pour être utilisables dans de nombreux domaines. Par exemple, une étude longitudinale, quel que soit son objectif, a besoin de données sur le changement de résidence. Si cette variable est incluse dans le fichier de données, il pourra être utilisé pour une étude de la migration, même si la migration n'intéresse pas le chercheur initial. La même étude peut

également comprendre les données sur le mariage et la fécondité afin de corrélérer ces variables avec la participation à la main d'oeuvre ou les changements de revenu familial. En outre, ces données peuvent être reliées à d'autres variables non économiques. Deux études longitudinales essentielles, dont les données sont disponibles depuis quelque temps, sont l'étude par sondage de James N. Morgan sur les variations du revenu et l'étude longitudinale nationale sur la main d'oeuvre, quelquefois connue sous le nom de "Parnes Study". La diversité des publications secondaires dans lesquelles ces données sont utilisées illustre ce point. Certains fichiers bien qu'encore utiles pour des analyses secondaires, sont plus focalisés sur des sujets précis. Ils sont le produit de recherches concernant des problèmes démographiques spécifiques. Les études nationales sur la fécondité, par exemple, ont des fichiers de données qui sont largement utilisés pour des analyses secondaires et les fichiers de l'Enquête Mondiale sur la Fécondité seront sans doute encore plus massivement utilisés. Ils peuvent être l'équivalent informatisé de volumes imprimés tels que les volumes de données sur les comtés et les villes des Etats-Unis ou bien ils peuvent fournir des tabulations qui complètent les volumes imprimés, comme c'est le cas pour la plupart des fichiers résumés des recensements.

Généralement, les fichiers informatiques contiennent soit des données macro, soit des données micro. Les fichiers de données macro contiennent des tableaux agrégés ou résumés. Ils sont habituellement classés par zones géographiques, qui peuvent être aussi petites que des flots de villes ou aussi grandes que des pays; ou bien par périodes de temps, telles que le jour, le mois ou l'année. Parfois, les fichiers sont transposés et l'unité d'analyse est une variable, comme le total des naissances, avec au premier rang le premier pays, au second le second pays, etc. Les fichiers de données micro comportent des indications individuelles sur les personnes, les familles, les ménages. S'il n'y a qu'un type de données, le fichier est à plat, s'il y en a deux ou plus, il est hiérarchisé. D'autres types de structures complexes existent mais les fichiers correspondants sont habituellement utilisés avec des systèmes de traitement de bases de données qui les rendent moins mobiles. Pour un traitement plus efficace, les fichiers complexes peuvent être rendus rectangulaires par addition de toutes les variables, depuis les données de plus haut niveau jusqu'à celles

du plus bas niveau. Les fichiers de données micro rectangulaires présentent une plus grande souplesse que les fichiers de données macro parce qu'il permettent aux chercheurs de créer leurs propres tableaux résumés et de sélectionner leurs propres sous-groupes de populations. Ainsi, il n'y a qu'avec des données micro que l'on peut produire toute une série de tableaux décrivant les caractéristiques socio-économiques d'un groupe ethnique réduit. Par exemple, les rapports imprimés et les tableaux résumés provenant des recensements des Etats-Unis comportent des tabulations décrivant les Blancs, les Noirs et les Hispaniques mais pas les Norvégiens ou les Hongrois. De telles tabulations ont été produites avec les données micro d'usage public. De même, seuls les fichiers de données micro permettent d'obtenir le nombre moyen d'enfants mis au monde parmi une sous-population donnée, telle que les femmes n'ayant pas plus qu'un certain niveau d'instruction, dont le mari travaille, avec un revenu familial ne dépassant pas 15.000 dollars et qui se sont mariées avant 20 ans. De semblables données ne sont généralement pas publiées et elles ne le sont jamais à un niveau sous-national. En utilisant un fichier de données micro, le chercheur peut sélectionner et tabuler n'importe laquelle des variables présentes. La disponibilité de tels fichiers accroît la capacité de l'analyste secondaire à effectuer aussi bien des recherches comparatives que longitudinales.

Sources de fichiers publics de données informatisées

La plupart des fichiers qui vont être cités peuvent être acquis auprès de l'agence ou de l'organisme qui les a produits ou auprès d'archives de données sociales. Pour toutes les institutions disposant de bibliothèques de données, cette bibliothèque doit d'abord être interrogée car le fichier ou la documentation correspondante peut avoir été obtenue déjà pour un autre chercheur.

Afin d'illustrer le genre de données détenues par les producteurs ou distributeurs, quelques résumés de ces données peuvent s'avérer utiles. Un marché plus important qu'ailleurs pour les fichiers de données s'est créé aux Etats-Unis parce qu'il s'y trouve davantage d'ordinateurs et que l'intérêt pour la recherche quantitative y est plus grand. Il n'est pas surprenant que les principaux producteurs et

distributeurs de données informatisées soient aux Etats-Unis puisque ce marché potentiellement vaste justifie le coût de constitution et d'entretien d'un fichier. C'est un marché qui n'est pas encore développé dans d'autres parties du monde et qui n'existait pas aux Etats-Unis, il y a seulement une dizaine d'années. Néanmoins, il se trouve des producteurs et des distributeurs au Canada, en Europe Occidentale, en Amérique latine et ailleurs. Des informations sur certains d'entre eux figureront ci-après et les références de fichiers spécifiques apparaîtront comme elles ont été annoncées.

Etats-Unis

Le plus gros volume de données publiques informatisées dans le monde est produit par le Bureau du Recensement des Etats-Unis. L'information sur les données produites à partir du recensement décennal de la population de 1970 a été largement diffusée (voir Population Index, Vol. 38, n°4, Octobre 1972, pages 409-424) et une collection semblable de fichiers de données sera constituée à l'issue du recensement de 1980. Dans les années récentes, le Bureau a également fourni les fichiers démographiques annuels, basés sur les enquêtes régulières de la population du mois de Mars et des efforts sont actuellement entrepris pour que des fichiers analogues soient également disponibles pour les autres mois. Ceci vaudra dire une production meilleure, à moindre prix. Les données seront épurées et vérifiées, une documentation complète sera fournie et le prix sera 80 dollars la bobine (soit 160 dollars pour la majorité des mois). Parmi les mois déjà disponibles pour le traitement, ceux d'intérêt prioritaire pour les démographes sont les suppléments au mois de Juin pour la période 1971-1977. L'information de base collectée par le Bureau des Statistiques du Travail est la même chaque mois. En Juin 1972, les questions supplémentaires ont traité de la fécondité des femmes mariées; en 1973, 74 et 76, elles ont traité de la fécondité et des probabilités de naissances et en 1977, ce furent la fécondité, les probabilités de naissances et les soins aux enfants. Les études de 1971 et 1975 comportaient nombre de questions analogues ou plus détaillées sur l'histoire maritale et l'espacement des naissances. D'autres fichiers susceptibles d'intéresser les démographes sont les enquêtes annuelles sur l'habitat et la bande des nouvelles données approfondies

sur les comtés et les villes. L'index citera les nouveaux titres tels qu'ils seront fournis. Les demandes générales concernant les données produites par le Bureau ou les plans futurs doivent être adressées à la Division des Services de données pour l'utilisateur (Data User Services Division, U.S. Bureau of the Census, Washington, DC 20230).

Le Centre National de Statistiques de Santé est l'autre grande agence gouvernementale des Etats-Unis dont les données intéressent les démographes. Ses données sont décrites dans l'ouvrage récemment mis à jour "Standardized Micro-Data Tape Transcripts" (voir Population Index, 44, Title 4873). Mentions spéciales doivent être faites de: l'enquête nationale sur l'accroissement familial, les enquêtes nationales suivies sur la natalité et la mortalité, l'interrogatoire de santé et les enquêtes d'examen de santé. D'autres sources gouvernementales potentielles de données utiles sont la Sécurité Sociale, le Centre National de Statistiques d'Education et la division des archives informatiques des Archives Nationales. Des fichiers sélectionnés en provenance d'autres agences peuvent aussi intéresser certains chercheurs.

Les deux principales archives de données américaines sur les sciences sociales disposent de recueils qui traitent certains aspects de l'étude de la population. Le Consortium pour la Recherche politique et sociale sert d'intermédiaire entre le Centre d'Etudes Politiques de l'Université de Michigan, une division de l'Institut pour la Recherche Sociale et plus de 240 collèges et universités d'Amérique du Nord, d'Europe, du Japon, d'Australie, d'Israël et du Koweït. Le consortium est un distributeur secondaire pour une grande collection de fichiers de données gouvernementales qui sont disponibles gratuitement pour les institutions membres et moyennant paiement pour les autres utilisateurs. Il est aussi le distributeur primaire de nombre de fichiers contenant des matériaux de démographie historique, tels que "Caractéristiques démographiques de la population des Etats-Unis, 1930-1950 (Niveau des Comtés)", de John Gardner et William Cohen, "Les villes du Sud des Etats-Unis en 1870 et 1880: une étude des individus et des familles", de Claudia Goldin, l'ouvrage du Consortium: "Données historiques démographiques et d'instruction, pour la France, 1801-1897", et l'étude du colonialisme interne: "Intégration nationale dans les Iles Britanniques, 1851-1966", de Michael Hechter. L'Administration des Gens âgés du Ministère américain de la Santé, Education et Bien-être a récemment

fondé une archive nationale de données informatiques sur les gens âgés rattachée au Consortium. Le contenu de ces archives concernera la santé et le bien-être social, économique et psychologique des personnes âgées. Le Consortium publie un "Guide annuel des Ressources et Services" qui décrit complètement ses collections actuelles. Des exemplaires du Guide peuvent être obtenus en écrivant au Consortium (ICPSR, The University of Michigan, P.O. Box 1248, Ann Arbor, MI 48106).

La distribution de données par le Centre Roper, nouvellement réorganisé est maintenant assurée par l'Université de Connecticut. Pour des informations détaillées sur les droits, les services et les sujets de recherches, écrire au "Roper Center", Box W164R, Storrs, CT 06268. Le Roper Center est la plus grande archive du monde pour les données d'enquêtes par sondage. Sa collection de données consiste essentiellement en fiches d'opinions et contient plus de mille fichiers avec des données sur la fécondité, le mariage, le divorce, etc. y compris des données provenant de plus de 70 enquêtes CAP (comportements, attitudes, pratiques). Beaucoup de ces fichiers sont mentionnés dans "Population Control: Une bibliographie des données d'enquêtes 1938-1970", de Philip K. Hasting (voir Population Index, 36: Title 4559) et dans: "Données de Population dans les archives de données sur les sciences sociales: Enquêtes détenues par le Roper Center de Recherche sur l'Opinion publique", de Paul R. Voss, dans "Rapport de la Neuvième Conférence Annuelle, de l'Association Internationale des bibliothèques et centres d'information sur la Population et la planification familiale", 1976, pages 161-169. Quelques autres archives de données universitaires aux Etats-Unis sont le Service de bibliothèque pour les programmes et les données, de l'University de Wisconsin, qui possède notamment les études nationales sur la fécondité de 1965 et 1970, la croissance des familles américaines de 1955 et 1960 et l'étude récemment déposée sur les changements d'activité dans une génération; et d'autre part le Centre de données Louis Harris à l'Université de Nord-Caroline qui est le dépositaire des fichiers Louis Harris. Beaucoup de ces fichiers comportent des questions sur des sujets tels que l'avortement et la taille de la famille et l'un d'eux "Harris 1969 Science, Sex and Morality" concerne les attitudes des Américains sur les progrès de la génétique et des techniques de reproduction et leurs effets sur les relations familiales. Il comprend des questions spéciales sur le surpeuplement, le contrôle des naissances,

l'insémination artificielle et les bébés éprouvettes. D'autres centres sans but lucratif, aussi bien d'Universités que d'Instituts de Recherches possèdent de grosses collections de données gouvernementales et fournissent des extraits de fichiers, des copies de bandes et des rapports analytiques à prix coûtant. De semblables services sont aussi fournis par nombre d'organismes commerciaux de données.

Canada

La distribution de données au Canada est très similaire de ce qu'elle est aux Etats-Unis. Les statistiques du Canada sont la principale source de fichiers de données gouvernementales. Elles distribuent à la fois des bandes résumées et des échantillons publics provenant des recensements et fournissent des données pour le CANSIM, la base de données nationales en séries chronologiques qui est régulièrement mise à jour et disponible pour les usagers par l'intermédiaire du traitement par lots ou en temps partagé. Pour plus de détails sur les fichiers de données de recensement, contacter le "Census User Inquiry Services, Statistics Canada, Ottawa". D'autres sources essentielles de données au Canada sont la Division des Archives informatisées des Archives publiques à Ottawa, la bibliothèque de données à l'Université de Colombie Britannique, l'Institut de Recherche sur le comportement à l'Université d'York, le Centre de sondage à l'Université de Montréal, l'Archive de données en sciences sociales à l'Université de Carleton et le laboratoire de recherches sur la population de l'Université d'Alberta. Deux fichiers d'intérêt spécial à Carleton sont les taux de fécondité légitime de période, pour 1926-1969, de Atlan Titus et ses taux de fécondité légitime de cohorte, pour 1882-1954. Une étude essentielle de type CAP disponible en laboratoire de recherche sur la population est l'étude de la croissance des familles d'Alberta, de Krotki et Krishnan. Un inventaire satisfaisant des fichiers de données en sciences sociales au Canada effectué par le bureau Central de données pour les sciences sociales a également identifié de nombreux fichiers tenus dans une grande variété de services gouvernementaux fédéraux et provinciaux. Pour des copies de l'Inventaire des données de sciences sociales de 1977 ou une version informatisée mise à jour de la publication, écrire à "Data Clearinghouse for the Social Sciences, 151 Slater, Ottawa K1P 5 N1"

Royaume Uni

En Europe Occidentale, quelques fichiers sont disponibles directement auprès des agences gouvernementales ou des instituts de recherche, mais un nombre croissant de fichiers sont déposés aux archives nationales de données qui ont déjà été établies dans huit pays. Au Royaume Uni ce service national est implanté à l'Université d'Essex, Colchester, C04 3SQ et est connu sous le nom de "SSRC Survey Archive". En plus des résultats de la recherche universitaire ce service acquiert et distribue les données provenant du Bureau des Recensements et des Enquêtes de Population, y compris les données résumées par petites zones provenant du Recensement du Royaume Uni de 1971 et les données micro de l'enquête générale trimestrielle sur les ménages; l'étude nationale sur le développement de l'enfant, effectuée par le bureau national de l'enfance; et de nombreuses études sur la planification familiale, l'avortement et les soins médicaux conduites pour le Ministère de la Santé et la Sécurité Sociale par l'Institut pour les études sociales en soins médicaux. Deux acquisitions récentes intéressant les démographes historiens sont la migration dans les Highlands du Nord, de Ruth Hildebrandt (1851-1891) et l'histoire démographique de neuf paroisses des Tendring Hundred (Essex, 1538-1838), de R.R. Barker. Le SSRC Survey Archive publie plusieurs fois par an un bulletin qui contient l'information sur les activités actuelles et les nouvelles acquisitions.

Danemark

L'archive danoise de données a récemment déménagé de Copenhague à Odense. Sa nouvelle adresse est: "Dansk Data Arkiv, University of Odense, Campusvej 55, DK-5230 Odense M." Avec cette installation à une adresse permanente, elle deviendra sans doute plus active aussi bien pour l'acquisition que pour la distribution des données. Elle compte déjà dans ses collections des fichiers contenant des données agrégées des recensements de 1920 à 1970. Pour plus d'information, écrire pour demander le catalogue courant et le bulletin.

Suède

Les archives de données suédoises sont implantées à l'Université de Göteborg. Elles sont assez récentes et n'ont pas encore publié de

catalogue de leurs collections. Les demandes doivent être adressées directement à "Institute of Political Science, FACK, University of Göteborg, S-400-25, Göteborg".

Norvège

La plus ancienne des archives scandinaves est le service de données norvégiennes en sciences sociales (NSSDS) implanté à l'Université de Bergen. Quoique son domaine soit large et divers, son activité essentielle s'est récemment déployée en cartographie informatique de données pour petites zones écologiques. C'est également le centre de développement de bases de données chronologiques conjointes pour les cinq pays nordiques. Cette base de données aura des variables sociales, économiques et politiques comportant plusieurs liaisons avec le sexe, l'âge, l'état matrimonial, la fécondité, le divorce, la mortalité et la migration. Le "European Political Data Newsletter" est produit et distribué par NSSDS et l'information le concernant peut être obtenue en écrivant à l'archive, à "Christiegst, 15-19, N-5014 Bergen University".

République Fédérale d'Allemagne

Il y a plusieurs archives de données en République Fédérale d'Allemagne, mais la plus grande et la plus active est l'archive centrale à l'Université de Cologne. Cette archive s'est concentrée sur l'acquisition de données d'enquêtes et gouvernementales et est une bonne source d'information sur des données de toutes sortes relatives à l'Allemagne de l'Ouest. Les demandes de données spéciales et de copies de la revue "Za-Information" doivent être adressées à "Zentralarchiv für Empirische Sozial forschung, Universität zu Köln, 5 Köln, 41 (Lindenthal), Bachemer Strasse, 40".

Belgique

Les archives belges pour les sciences sociales (BASS) sont implantées à "Place Montesquieu 1, Boîte 18, 1348 Louvain-la-Neuve". Parmi d'autres fichiers publics de données informatisées, elles distribuent les données agrégées du recensement belge de 1970 par commune et par secteur statistique. D'après le supplément 1978 de leur dernier catalogue,

le fichier commune sera périodiquement mis à jour avec l'information tirée des sources statistiques disponibles.

Italie

Les Archives et Programmes pour les Sciences Sociales sont implantées "Via Glantoni, 4, 20144 Milan". Elles détiennent des fichiers villes, provinces et communes métropolitaines contenant les données des recensements de 1951, 1961 et 1971, notamment les effectifs de population, l'âge, l'activité, l'instruction, la branche d'activité, la taille de la famille selon l'activité du chef et les membres du ménage, l'habitat et les rubriques économiques et politiques.

Amérique latine

Un service spécial de données pour les données démographiques d'Amérique latine est assuré par la banque de données du CELADE, qui est une partie du programme INFOPAL. Son adresse est "Casilla 91, Santiago, Chile". Cette archive de données est une source exceptionnellement riche de données démographiques. Elle contient des données micro des recensements de 1960 pour l'Argentine, le Brésil, le Chili, la Colombie, le Costa-Rica, la République Dominicaine, l'Equateur, El Salvador, le Guatemala, le Honduras, le Mexique, Panama, le Paraguay et l'Uruguay ainsi que les données micro des recensements de 1970 pour tous ces pays (sauf le Brésil) et en plus la Bolivie, Haïti, le Nicaragua et le Venezuela. En outre, le CELADE possède de nombreux fichiers sur des sujets comme la migration internationale et interne, la fécondité et l'avortement et les enquêtes de ménage de la plupart des pays latino-américains. Bien que l'accès à ces fichiers requière l'autorisation de l'agence gouvernementale détentrice, les demandes pour l'information appropriée sur les données disponibles et les restrictions d'accès peuvent être adressées directement au CELADE.

Asie et Océanie

Des facilités existent aussi en Australie, à "ACSPRI, Survey Research Center, Australian National University, Box 4, Canberra, A.C.T. 2600, Australia" et se développent en Inde, Israël et Japon.

Bibliographie de l'accès aux
fichiers de données informatisées.

Depuis 45 ans, Population Index a été la source primordiale de références bibliographiques pour une gamme étendue d'informations démographiques imprimées. Comme de plus en plus d'information est "publiée" ou produite seulement sous forme informatique et comme les démographes commencent à mettre de plus en plus sous cette forme leurs données, il convient que l'Index prenne l'initiative de fournir l'assistance nécessaire en référant les matériaux informatiques concernés. C'est seulement de cette façon qu'il pourra continuer à fournir la couverture convenable de la littérature sur la population nécessaire pour satisfaire les besoins de ses lecteurs.

Bien que de nombreux fichiers soient encore transférés officieusement entre collègues, l'Index ne citera que ceux qui sont disponibles pour le public. Ceci veut dire que le fichier et son code manuel ou guide d'accompagnement peuvent être obtenus, habituellement au prix de la reproduction, par tout chercheur ou bibliothèque de données. On n'essaiera pas de couvrir rétrospectivement toutes les données informatisées actuellement disponibles, mais quelques fichiers très anciens dont la disponibilité n'est pas très connue seront cités, de même que les nouvelles éditions et les mises à jour de titres anciens. Toutes les citations de fichiers informatisés apparaîtront dans une section séparée de l'Index, Section T: "Machine Readable Data Files".

Dans quelques pays, comme la Suède et le Canada, les données agrégées de population font partie d'une base de données centrale qui est disponible par un programme approprié. D'autres pays pourront vraisemblablement fournir un accès similaire dans l'avenir. En attendant, cependant, les analystes secondaires compteront surtout sur le transfert de fichiers sur des bandes magnétiques par d'autres chercheurs, services gouvernementaux et archives de données. Comme on l'a vu, une grande quantité de données démographiques existent sous forme informatique: données micro de recensements et d'enquêtes, données agrégées publiées ou non publiées, fichiers comportant des données provenant de sources multiples, fichiers nationaux croisés et fichiers chronologiques, dont certains portent sur plusieurs siècles. D'énormes quantités d'argent, de temps et d'effort intellectuel ont été investies dans la

création de chacun de ces fichiers. Peu ont bénéficié de la publicité convenable et ont par conséquent été largement utilisés mais la plupart sont virtuellement inconnus, même par les chercheurs qui en auraient le plus besoin. Il y a plusieurs raisons à cela. Tout d'abord, il n'y a pas de liste unique même des détenteurs de sources essentielles de données; aucun moyen, par conséquent, d'apprendre que les données du recensement de la Finlande sont disponibles à l'Université Carleton du Canada et que des échantillons d'Indonésie, de Hong-Kong et de Corée peuvent être obtenus au Service de bibliothèque programme et données de Wisconsin. Le second problème est le caractère vague des références de fichiers de données dans la littérature des sciences sociales. Une citation typique est par exemple: "des statistiques sur bande à ICPSR", ou "des données provenant du Roper Center", ou encore "l'étude nationale longitudinale". Jusqu'à une époque récente, même le chercheur consciencieux était embarrassé pour fournir une référence convenable. Il n'y avait pas de modèle accepté pour citer un fichier informatisé et même les éléments bibliographiques les plus simples faisaient souvent défaut. Quelquefois, les titres sont également manquants; parfois, il y a plusieurs titres et l'on ne sait lequel choisir. Ces problèmes existent encore mais des solutions commencent à apparaître.

En 1970, l'Association des bibliothèques américaines a constitué un Comité des Règles, pour cataloguer les fichiers de données informatisées. Ce fut la première reconnaissance formelle par la communauté des bibliothèques et centres d'information du besoin de règles pour le contrôle bibliographique et l'accès à ces matériaux. Le rapport final du Comité a été soumis en 1976 et a été incorporé presque complètement dans la 2ème édition récemment publiée des Règlements de catalogue anglo-américains. Entre temps, Sue A. Dodd, coordinateur du Groupe de Classification de IASSIST (Association Internationale pour l'Information et la Technique des Sciences Sociales) a adressé nombre d'indications sur les fichiers informatisés, y compris la création de titres et les pages de titres, la création de références, le développement d'un modèle et la publication (probablement en 1980) d'un manuel pour la mise en catalogue de ces fichiers. Le résultat de tout ceci est qu'il est maintenant possible pour des revues bibliographiques, telles que Population Index, de commencer à inclure des références de fichiers informatisés. Toutefois, afin de fournir une couverture adéquate, l'Index sollicite la

coopération de ses lecteurs pour lui faire parvenir les codes ou l'information bibliographique complète sur les nouveaux fichiers soumis à leur attention.

Pour illustrer ce qui est nécessaire, deux exemples de citations convenables sont proposés ci-après:

- Davis, James A.

Enquêtes sociales générales, 1972-1978: Données cumulatives (fichier de données informatisé). Investigateur principal, James A. Davis, directeur d'étude associé, Tom W. Smith; chercheur adjoint, C. Bruce Stephenson. Chicago, Centre national de recherche d'opinion (producteur), 1978. New Have, Coun., Centre Roger de recherche d'opinion publique, Université de Yale (distributeur). Un fichier de données (10.652 enregistrements logiques) et code (340pp.).

- Etats-Unis, Bureau du Recensement

Population mondiale, 1950-1975 (fichier de données informatisé). Réuni par le Centre International de Programmes Statistiques du Bureau du Recensement des Etats-Unis. Washington, Bureau du Recensement des Etats-Unis (producteur), 1975. Ann Arbor, Mich., Consortium inter-universitaire pour la recherche politique et sociale. Université de Michigan (distributeur). Un fichier de données (201 enregistrements logiques).

Les éléments bibliographiques requis sont consistants avec le règlement ANSI et avec son application pour les fichiers de données en sciences sociales comme l'a énoncé Sue Dodd (1977 et 1979):

(1) Auteur: nom complet de l'auteur (s) ou de l'organisme responsable du contenu intellectuel du fichier (principal investigateur, directeur de projet ou agence responsable).

(2) Titre: titre complet (pas de sigle) contenant les mots descriptifs ou des phrases et dates. Une discussion sur la signification du titre dans le cas de fichiers informatisés figure dans les articles de Dodd.

(3) Sous-titre: titre secondaire pour amplifier ou restreindre le titre principal, comme dans une série de travaux ou des travaux en plusieurs parties.

(4) Désignation générale du matériel: forme générique ou type de matériel cité. "Fichiers de données informatisées" est recommandé dans le chapitre 9 des Règles de catalogue anglo-américaines (1978).

(5) Rôle de l'auteur: indique la relation entre le travail et la personne ou l'organisme nommé en auteur de tête ou d'autres parties significatives telles que: investigateur principal, responsable, ou même agence de financement (par ex.: conduit par Louis Harris et Cie, pour la Sous-Commission sénatoriale sur les relations intergouvernementales).

(6) Edition: fournit aux usagers l'information précieuse sur la source, la date, ou les révisions; les nouvelles éditions indiquent l'addition ou le retrait de données, de variables ou de domaines: le recodage ou la restructuration du fichier; l'addition ou le retrait d'enregistrements logiques (cas, observations, etc.) ou, dans le cas d'un programme, les changements dans le langage du programme.

(7) Impression: désignation du producteur et du distributeur. Le producteur est responsable de la création du fichier de données informatisé et le nom, le lieu et la date doivent en être fournis. Comme des copies ne sont pas, habituellement, stockées, c'est la date d'édition plutôt que celle de distribution qui est mentionnée afin d'indiquer les différentes copies.

(8) Etendue du fichier: pour le fichier l'équivalent du nombre de pages est le nombre d'enregistrements logiques. Les autres caractéristiques physiques telles que taille de la bande, densité d'enregistrement, etc. sont exclues puisque les distributeurs offrent normalement diverses options d'enregistrement. Cette information est sujette à changement et ne doit pas affecter l'identification bibliographique qui est déterminée par le contenu du fichier et non par le support.

(9) Matériel complémentaire: comprend les codes, rapports, manuels et autres matériaux associés sous forme imprimée ou informatique. De tels matériels peuvent ou ne peuvent pas mériter une mention séparée.

(10) Désignation de la série: titre collectif et nombre de rubriques ou autres indications de la série.

Comme il est coutumier pour les autres citations de l'Index, chaque référence sera suivie, le cas échéant, d'une brève description du contenu du fichier. Tout fichier de données informatisé apparaîtra dans une seule section, mais les matériels imprimés concernés seront cités sous les rubriques appropriées. L'information pour les références doit comporter le maximum de documentation et doit être adressée à l'éditeur, Population Index, 21 Prospect Avenue, Princeton, NJ 08544.

*

* *

R E F E R E N C E S

- DODD, Sue, 1977 - Titres: la nouvelle priorité des fichiers de données informatisé en matière de bibliographie en sciences sociales. IASSIST Newsletter (Madison, Wis.), 1 (4): 11-18.
- DODD; Sue A., 1979 - Références bibliographiques pour les fichiers de données sociales numériques: suggestions de directives. Journal of the American Society for Information Science (Washington) 30 (2): 77-82.
- GORMAN, Michael et WINKLER, Paul W., Editeurs, 1978 - Règles de catalogue anglo-américaines. 2ème édition. Chicago: American Library Association; et Ottawa: Canadian Library Association. XVII, 620pp.
-

