



NATIONS UNIES
CONSEIL ÉCONOMIQUE ET SOCIAL



Distr.
GÉNÉRALE

E/ECA/PSD .5/45
Mars 1988

FRANÇAIS
Original: ANGLAIS

52324

COMMISSION ÉCONOMIQUE POUR L'AFRIQUE

Cinquième session de la Conférence commune
des planificateurs, statisticiens et
démographes africains

Addis-Abeba (Ethiopie), 21-28 mars 1988

QUESTIONS RELATIVES À LA BASE DE DONNÉES
STATISTIQUES DE LA CEA

TABLE DES MATIERES

	<u>Paragrapbes</u>
I. Introduction	1 - 2
II La base de données statistiques de la CEA	3 - 4
2.1 Le troisième niveau de la base de données	5 - 8
2.2 Séries statistiques, définitions et nomenclature	9 - 11
2.3 Sources de données et notes de bas de page	12 - 15
2.4 Système de meta-information	16 - 17
III Séminaires et missions	18 - 20
IV Collecte des données	21 - 23
V Diffusion des données	24 - 25
VI Base de données statistiques sur micro-ordinateurs	26 - 29
VII Conclusion	30 - 31

I. INTRODUCTION

1. La base de données statistiques de la CEA (également dénommée PADIS-STAT) a été créée en 1983. Il s'agit d'un système d'information permettant de compiler, de traiter et de diffuser des données statistiques. Les principaux objectifs de la base de données sont les suivantes :

- compiler des données statistiques sous une forme informatisée;
- automatiser les publications de la CEA relatives à l'information statistique;
- diffuser auprès des utilisateurs des informations statistiques détaillées et à jour, y compris des données obtenues d'autres banques de données internationales;
- aider les Etats membres à créer leurs propres bases de données statistiques;
- donner aux utilisateurs d'autres types de services qu'ils voudraient obtenir du système;
- fournir aux fonctionnaires de la CEA un accès rapide et direct aux données statistiques sur les pays africains.

2. La base de données statistiques de la CEA est installée sur le système informatique matériel/logiciel disponible et des programmes spécialement écrits. Il tourne sur l'ordinateur HP3000/58 avec une mémoire principale de 5 mégabytes, 4 lecteurs de disques ayant une capacité de stockage de 404 mégabytes, 2 dérouleurs de bandes magnétiques d'une densité d'enregistrement de 1600 et de 1600/6250 BPI, des imprimantes et des terminaux. La base de données est accessible à partir de n'importe quel terminal connecté au système dans l'enceinte de la CEA ou à partir du siège de l'OUA par une ligne téléphonique spéciale.

II. LA BASE DE DONNEES STATISTIQUES DE LA CEA

3. La base de données statistiques de la CEA est conçue en trois niveaux intégrés dans un système d'information unique. Le premier niveau donne une image rapide de la situation socio-économique de n'importe quel pays ou sous-région d'Afrique. Le deuxième niveau contient la masse des données statistiques détaillées à son plus bas niveau d'agrégation. Les données étant volumineuses, elles sont gardées sur bandes magnétiques en vue d'un traitement différent et groupé. Le troisième niveau est une donnée chronologique contenant des agrégats de données ainsi que quelques données sociales, démographiques, économiques et de l'environnement fréquemment recherchées par les utilisateurs. La base de données a été améliorée pour contenir des données portant sur 26 ans, de 1965 à 1990.

4. La mise en oeuvre du premier niveau dépend des deuxième et troisième niveaux. A ce jour, ce niveau n'a pas encore été mis en oeuvre. Pour le deuxième niveau, les statistiques du commerce extérieur avec plus de 20 millions d'entrées et les comptabilités nationales avec les déflateurs sectoriels ayant plus de 25 000 d'entrées sont opérationnelles. Certaines de ces données sont utilisées pour mettre à jour le troisième niveau de la base de données. Les composantes restantes du deuxième niveau seront mises en oeuvre dès qu'on disposera des ressources nécessaires.

2.1 Troisième niveau de la base de données

5. Le troisième niveau contient des données sur le commerce, la comptabilité nationale, l'agriculture, l'industrie, les prix, la population, la santé, l'éducation, l'emploi, les finances, les transports et communications ainsi que les statistiques de l'état civil. A ce jour, plus de 136 000 données chronologiques contenant environ 2 millions de valeurs de données statistiques sont entrées dans la base de données.

6. Les principaux produits de la base de données sont les suivants : "Annuaire statistique pour l'Afrique" (4 volumes), "Indicateurs socio-économiques africains", "Statistiques africaines du commerce extérieur : Echanges par pays, Série A", "Statistiques africaines du commerce extérieur : Tableaux récapitulatifs, Série C" et les annexes statistiques de l'Etude des conditions économiques et sociales en Afrique. Divers types de tableaux par pays sont également produits et envoyés aux pays membres et aux organisations internationales.

7. Le troisième niveau de la base de données est construit avec le système de gestion de base de données Image 3000 qui est un produit Hewlett-Packard et avec des programmes spécialement écrits. Image 3000 est utilisé pour définir et créer la base de données et y accéder. Image 3000 étant un système de gestion de base de données général, il n'est pas adapté aux besoins particuliers des données statistiques. Des programmes ont donc été spécialement écrits pour adapter les caractéristiques d'Image 3000 aux besoins de la base de données. Ces programmes sont élaborés pour les traitements de données différenciés et groupés, la création de tableaux pouvant être publiés, les calculs de base, la maintenance de la base de données et la mise à jour régulière des données dérivées.

8. Le temps de réponse pour la saisie et la recherche de données ainsi que pour la production de tableaux avec un grand volume de données s'est avéré très long avec le système de gestion de base de données Image 3000. Afin de résoudre ce problème, on a incorporé au système Image 3000 un fichier séquentiel indexé qui fait office de fichier inversé à accès rapide pour améliorer le rendement sur la base de données.

2.2 Séries statistiques, définitions et nomenclature

9. Les séries statistiques sont catégorisées au moyen de notions, définitions et classifications internationales. Des classifications de séries types utilisant une nomenclature devraient être élaborées pour être utilisées avec la base de données statistiques. Les différents secteurs des statistiques (sociales, démographiques, économiques, de l'environnement) sont apparentés et ne devraient pas être considérés indépendamment les uns des autres. Les classifications des séries statistiques pour les secteurs statistiques pourraient être effectuées ensemble, séparées par chapitres ou par sections. En procédant à la classifications des séries statistiques, il conviendrait d'utiliser autant que possible des définitions et une nomenclature identiques afin de permettre l'établissement d'une relation entre les différentes statistiques figurant dans la base de

données. La réalisation de classifications de séries systématisées facilitera l'agrégation ou la désagrégation des données et l'échange de l'information.

10. Lors de la mise au point de la base de données statistiques, la Classification type pour le commerce international (CTCI), Révision 2 et la Classification par grandes catégories économiques (CGCE) ont été pleinement utilisées pour la classification des statistiques du commerce extérieur. La classification des séries statistiques pour les autres statistiques, à savoir la comptabilité nationale, l'agriculture, l'industrie, les prix, la population, la santé, l'éducation, l'emploi, les finances, les transports et les communications ainsi que les statistiques de l'état civil, a été effectuée sur la base du Répertoire des statistiques internationales, 1981, volume 1 et de la classification internationale type, par industrie, de toutes les branches d'activité économique (CITI). Ces classifications sont effectuées pour quelques séries.

11. Bien que la classification des séries de bases de données puisse avoir des textes descriptifs en français et anglais chacune ayant jusqu'à 40 caractères, les descriptions se réduisent pour le moment à 24 caractères en anglais seulement. Selon le type des séries, il est parfois nécessaire d'allonger la description. En produisant des tableaux statistiques, l'utilisateur fournit toutes les descriptions et notes de bas de page nécessaires pour la série utilisée. Cela rend difficile pour d'autres utilisateurs la détermination et la compréhension des séries statistiques utilisées.

2.3 Sources de données et notes de base de page

12. Chaque donnée de la base de données statistiques de la CEA est marquée par un drapeau sous la forme d'un code à deux caractères pour donner en termes plus larges le type de donnée qui est soit réelle, soit une estimation, soit une projection provenant de sources nationales ou d'organisations internationales.

13. La CEA a commencé à procéder à une classification par sources de données et par notes de bas de page afin de permettre d'identifier la source de la donnée et de mieux préciser sa validité. Ces classifications pourraient être construites comme une donnée séparée au sein de la base de données principale en utilisant la source et la note de bas de page ou les codes descripteurs comme indicateurs. Le fait que cette donnée soit physiquement isolée de la base de données numériques permettra de réduire le stockage et d'éviter le double emploi. Jusqu'à présent plus de 400 différentes sources de données pour tous les pays africains et organisations internationales ont été sélectionnées et classées. Ces sources de données sont les principales publications, les feuilles de sortie d'imprimante, les microfilms et les supports magnétiques utilisés par chaque Etat membre ou organisation internationale pour diffuser ses informations statistiques.

14. Il convient toutefois de noter le temps nécessaire pour déterminer la source de données, la comparer à la source principale et au descripteur que constitue la note de bas de page, la vérifier et l'entrer dans la base de données nécessite des moyens de stockage, du temps machine et des ressources en main-d'oeuvre supplémentaires. Le système de gestion de la base de données devrait également être en mesure de supporter cet ensemble de données descriptives source de données et note de bas de page.

15. Lors de l'élaboration de la base de données statistiques, il y a un compromis à observer entre les ressources disponibles et l'utilité de chaque donnée devant entrer dans la base de données.

2.4 Système de meta-information

16. La donnée statistique est très complexe dans sa structure, son contenu, son application et ses définitions. Il est très difficile d'élaborer des notions, des définitions et des classifications normalisées qui puissent être appliquées à tous les pays et organisations. Il y a de nombreuses divergences d'un pays à l'autre. Les textes descriptifs utilisés avec les classifications de séries statistiques pour la base de données ne sont pas suffisants pour donner une idée claire de la donnée figurant dans la base de données. Celle-ci est construite pour toutes sortes d'utilisateurs qui pourraient ne pas être capables d'interpréter convenablement les données statistiques. Il est donc nécessaire de fournir une description précise de la donnée qui entre dans la base de données, c'est-à-dire une description de toutes les variables ainsi que toutes les valeurs possibles et leur signification. Cela évitera l'incompréhension et la mauvaise interprétation des données et permettra à n'importe quel utilisateur d'utiliser la base de données de façon indépendante. La structure de la base de données, son contenu et la présentation des données devraient également être décrits intégralement. Il est donc nécessaire d'incorporer à la base de données un système d'information pour les données figurant dans la base de données. L'information relative à l'information porte le nom de meta-information ou de meta-donnée. Une base de données statistiques sans meta-données est incomplète.

17. Le système de meta-information donne davantage d'information sur les données, en particulier lorsqu'il y a une divergence par rapport aux notions, définitions ou classifications générales utilisées pour la création de la base de données. Cette meta-donnée peut être interrogée en même temps que les données statistiques. Le système de meta-information de la base de données statistiques de la CEA est gardé à l'esprit par les experts qui interprètent convenablement et règlent ces divergences et donnent les explications nécessaires. La meta-information est également donnée dans les introductions, les notes techniques, les notes de bas de page, les textes descriptifs des tableaux statistiques et autres des produits imprimés à partir de la base de données. Toutes ces informations et autres sur les données statistiques devraient être introduites dans le système de meta-information afin de permettre de répondre à toutes les questions à partir de la base de données et de rendre celle-ci utilisable sans qu'il soit besoin de consulter fréquemment l'administrateur de la base de données.

III. SEMINAIRES ET MISSIONS

18. Depuis la quatrième session de la Conférence commune des planificateurs, statisticiens et démographes africains, des séminaires se sont tenus à la CEA. Les rapports de ces séminaires intitulés "Rapport du Séminaire sur le développement de bases de données statistiques pour les pays africains francophones" (Addis Abeba, 24-28 novembre 1986, E/ECA/PSD.4/46) et "Report of the Training Workshop on the Use of Software Packages in Statistical Computing for English-speaking Government Statisticians and Planners" (Addis-Abeba, 21 septembre - 2 octobre 1987, E/ECA/PSD.4/47) sont présentés à cette conférence. Il ressort de ces séminaires sur la base de données statistiques qu'il faudrait que la CEA organise davantage de séminaires et d'ateliers de formation. Bien que la CEA puisse organiser de tels séminaires, il est nécessaire de mobiliser des ressources financières pour couvrir les frais de voyage et les indemnités journalières de subsistance des participants.

19. Au cours de la période biennale précédente, trois missions ont été organisées respectivement au Maroc (Centre national de documentation du ministère du plan), en Tunisie (Centre de documentation de la Ligue arabe) et au Bénin (Centre national de documentation et d'information). Des séminaires de démonstration et de formation sur les méthodes et l'utilisation de la base de données statistiques en provenance de différentes organisations nationales.

20. La CEA pourrait organiser de telles missions techniques consultatives dans d'autres Etats membres pourvu que ceux-ci en fassent la demande en précisant les objectifs de la mission et en énonçant clairement les problèmes. Ces missions consultatives techniques ne pourraient être utiles que si les Etats membres mettent en oeuvre les recommandations qui leur auront été faites.

IV. COLLECTE DE DONNEES

21. Les données pour la base de données statistiques de la CEA sont compilées et traitées à partir des informations fournies par des organisations nationales, des organisations internationales ainsi que des groupements politiques et économiques. Les organisations nationales des Etats membres devraient être la principale source de données pour la base de données. Cependant, les données en provenance des organisations nationales sont très rares et quelquefois totalement inexistantes. Bien qu'il y ait quelques données dans les fichiers et les registres administratifs, ces données ne sont en général ni compilées, ni traitées, ni diffusées. Même celles qui sont traitées et publiées ne sont pas à jour et leur transmission à la CEA est retardée. La CEA envoie divers imprimés et questionnaires aux Etats membres pour rassembler des données mais le taux de réponse est très faible. La Commission dépend entièrement des organisations nationales pour obtenir des données à jour pour la mise en place de la base de données nationale et sa diffusion sous diverses formes.

22. Les données fournies par les organisations nationales sont sous une forme imprimée dont la compilation nécessite d'énormes ressources pour l'introduction des données dans la base de données. Si des supports magnétiques sont utilisés pour transmettre les données à la CEA, cela permettra de mettre rapidement à jour le système d'information statistique de la CEA.

23. Lorsqu'elle ne reçoit pas de données des organisations nationales, la CEA compile les données à partir d'autres bases de données internationales telles que celles détenues par la Banque mondiale, le Fonds monétaire international (FMI), l'Organisation de coopération et de développement économiques (OCDE), l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO), l'Organisation internationale du Travail (OIT) et autres. Lorsque les données ne sont disponibles de sources nationales ou internationales, la CEA effectue des estimations et des projections chaque fois qu'elle le peut.

V. DIFFUSION DES DONNEES

24. La base de données statistiques de la CEA est un système de production dont le principal objectif est de diffuser l'information auprès des Etats membres. Les données statistiques peuvent être diffusées ou transmises essentiellement par les moyens suivants :

- i) publications techniques;
- ii) feuilles de sortie d'imprimant;

- iii) microfilms;
- iv) supports magnétiques;
- v) réseau local et régional de télétraitement.

25. La CEA diffuse les informations statistiques auprès des Etats membres, et des organisations internationales au moyen de publications techniques et de feuilles de sortie d'imprimante. Les données pourraient également être transmises sur des supports magnétiques, c'est-à-dire bandes magnétiques et disquettes chaque fois que des parties intéressées en font la demande. La base de données statistiques de la CEA peut être transférée sur disquettes en utilisant des micro-ordinateurs en vue de sa transmission ultérieure.

VI. BASE DE DONNEES STATISTIQUES SUR MICRO-ORDINATEURS

26. Les progrès technologiques rapides en matière de matériel et de logiciel informatiques ont fait des micro-ordinateurs, l'un des principaux instruments de traitement des données statistiques. Leur faible coût et leur rendement élevé ont rendu possibles l'acquisition et l'utilisation de micro-ordinateurs dans la plupart des bureaux de statistique. La plupart des micro-ordinateurs sont compatibles IBM et les logiciels utilisés sont la plupart du temps similaires ou identiques. Ces caractéristiques rendent possibles l'échange d'information entre micro-ordinateurs compatibles IBM en utilisant les disquettes.

27. De nombreux logiciels statistiques sur miniordinateurs et ordinateurs centraux ont été adaptés pour qu'ils puissent tourner sur micro-ordinateurs. Certaines applications qui autrefois n'étaient possibles que sur miniordinateurs ou ordinateurs centraux peuvent à présent être effectuées sur micro-ordinateurs. Ces logiciels sont utilisés pour l'analyse statistique avancée, la collecte de données, l'édition des données et la réalisation des tableaux. Les logiciels les plus largement utilisés sont notamment SPSS, SAS et BMDP.

28. Les micro-ordinateurs peuvent être utilisés pour organiser, traiter et présenter les informations statistiques contenues dans les fichiers et questionnaires administratifs. Les micro-ordinateurs offrent une grande capacité de stockage pour traiter les données statistiques soit dans un tableau en utilisant Lotus 1-2-3 ou dans une base de données en utilisant DBASE III. Ils produisent également des tableaux et des graphiques pouvant être imprimés en "qualité lettre" pour la production de publications statistiques. De petites bases de données pourraient être contruites et entretenues en utilisant des micro-ordinateurs. Les mêmes données pourraient également être transférées à de grandes bases de données sur miniordinateurs ou ordinateurs centraux. L'échange de données statistiques sur disquettes en utilisant des micro-ordinateurs est moins coûteux et donc à la portée des bureaux nationaux de statistiques.

29. La base de données statistiques de la CEA mise au point sur le système informatique Hewlett-Packard peut être transférée sur disquettes en vue de la transmission ultérieure des données aux Etats membres. Ces données peuvent être exploitées sur micro-ordinateurs en utilisant le logiciel DBASEIII disponible dans le commerce. La CEA commencera bientôt à mettre au point de petites bases de données pour micro-ordinateurs utilisant DBASE III similaires à la grande base de données. Le micro-ordinateur pourrait être ensuite utilisé pour transférer quelques données statistiques. L'utilisation de micro-ordinateurs permettra de produire des données à jour, de les saisir et de les transmettre très vite.

VII CONCLUSION

30. Dans le processus de mise en place de la base de données de la CEA, on s'est efforcé de collecter et de diffuser autant de données à jour que possible dans les domaines social, démographique et économique. On a essayé dans le présent document d'indiquer les réalisations, les lacunes et les solutions possibles pour la bonne utilisation de la base de données. Un accent particulier a été mis sur :

- l'élaboration de classifications de séries statistiques systématisées et normalisées ainsi que de définitions tant en français qu'en anglais;
- la mise au point d'un système de méta-information;
- la détermination d'une source acceptable de dictionnaire de données;
- l'utilisation de supports magnétiques pour l'échange de l'information statistique.

31. Des efforts sont en cours pour explorer la possibilité de développer la base de données statistique de la CEA en vue de son utilisation sur micro-ordinateurs. Une aide financière et l'assistance d'experts de sources extrabudgétaires seront nécessaires pour ce faire. Au cours de la première phase, ces efforts seront déployés parallèlement à l'examen de la possibilité de réaliser cette activité. Si cette phase réussit, l'expérience qui aura été acquise sera utilisée pour développer des bases de données aux niveaux national et sous-régional.