

Distr.  
GÉNÉRALE

CES/AC.71/2005/16 (Summary)  
2 février 2005

FRANÇAIS  
Original: ANGLAIS

**COMMISSION DE STATISTIQUE et  
COMMISSION ÉCONOMIQUE  
POUR L'EUROPE (CEE)**

**COMMISSION EUROPÉENNE**

**CONFÉRENCE DES STATISTICIENS  
EUROPÉENS**

**OFFICE STATISTIQUE DES  
COMMUNAUTÉS EUROPÉENNES  
(EUROSTAT)**

**ORGANISATION DE COOPÉRATION ET DE  
DÉVELOPPEMENT ÉCONOMIQUES (OCDE)  
DIRECTION DES STATISTIQUES**

**Réunion mixte CEE/Eurostat/OCDE sur la gestion des systèmes d'information statistique  
(Bratislava, Slovaquie, 18-20 avril 2005)**

Thème ii): Stratégies pour le développement des systèmes d'information statistique

**SYSTÈME DE MÉTAINFORMATION STATISTIQUE ET ENTREPÔT DE DONNÉES  
EN TANT QUE PRINCIPAUX SOUS-SYSTÈMES DU SYSTÈME D'INFORMATION  
STATISTIQUE: SITUATION ACTUELLE ET PROJETS D'AVENIR**

**Document d'appui**

Émanant du Bureau central de statistique polonais<sup>1</sup>

1. Le système d'information statistique est un système d'ensemble qui englobera pratiquement tous les systèmes de traitement statistique du Bureau central de statistique polonais. Le système de métainformation statistique et l'entrepôt de données sont deux grands sous-systèmes qui ont été mis en place au cours des trois dernières années.
2. Un des principaux objectifs du système d'information statistique consistait à intégrer les bases de données statistiques réparties existantes pour assurer la continuité des services proposés par la statistique officielle en améliorant l'accès à l'information déjà collectée.

---

<sup>1</sup> Établi par Krzysztof Kurkowski (k.kurkowski@stat.gov.pl) et Stanislaw Sieluzycki (s.sieluzycki@stat.gov.pl).

3. L'entrepôt de données facilite la réalisation d'analyses et les enquêtes secondaires, sur la base de sources de données réparties, en permettant la diffusion des résultats d'origines différentes à partir d'un seul lieu où toutes les données nécessaires sont stockées. Les données, que ce soit les microdonnées ou les macrodonnées, peuvent être consultées en ligne par les statisticiens. Elles sont protégées contre toute détérioration et tout accès non autorisé.
4. L'entrepôt de données s'articule autour des métadonnées, en ce sens que les données stockées sont gérées par les métadonnées et que toute modification de la structure des données doit être précédée par des modifications des métadonnées. La construction du système d'information statistique s'est effectuée à partir de l'intégration des métadonnées («intégration par domaine») et la création de l'entrepôt de données basé sur le système de métainformation («intégration physique»).
5. L'intégration par domaine consiste à adopter une définition unique des termes et des classifications utilisés pour étayer les enquêtes statistiques, alors que l'intégration physique a trait à la cohérence interne des données stockées. L'entrepôt de données est construit de façon à pouvoir obtenir un même résultat de différentes façons, par exemple à partir de données individuelles (microdonnées) ou d'agrégats (macrodonnées).
6. Le système de métainformation est le premier système de cette nature qui comprend une participation directe des statisticiens. Les statisticiens tiendront à jour un glossaire des termes, des classifications et d'autres éléments de métainformation en ligne. La mise à jour de l'entrepôt de données sera confiée à des informaticiens.
7. Le système de métainformation coiffe en quelque sorte tous les autres systèmes: les métadonnées utilisées par les différents systèmes ne peuvent être puisées que dans une seule source, la base de métadonnées. L'entrepôt de données est un système central dans lequel sont stockées les données. Mais il existe d'autres sous-systèmes, qui seront aussi incorporés au système d'information statistique. Ces systèmes peuvent être considérés comme des systèmes en amont ou en aval du point de vue de l'entrepôt de données. Sont inclus dans les systèmes en amont les systèmes de conception et de gestion des enquêtes (y compris le système de documentation); le système de formulaires électroniques pour la saisie, l'édition et la vérification des données; et le système d'exploitation des données de source administrative. Les systèmes en aval sont les suivants: systèmes d'établissement de rapports et d'analyse, systèmes de diffusion des données (portail, système de gestion du contenu, entrepôt de données accessible au public, y compris les systèmes d'information géographique (SIG)).
8. L'élaboration des bases de sondage et des systèmes de documentation s'articulera autour du système de métainformation. Les statisticiens auront la possibilité de créer des bases en utilisant les termes nécessaires ainsi que les informations contenues dans la base de métadonnées. Ils introduiront la documentation des enquêtes et mettront au point dans ce contexte les critères de contrôle qui devront s'appliquer au moment de la saisie des données. D'une façon générale, ces conditions constituent les éléments les plus délicats du système, car elles requièrent l'élaboration d'un langage interne interprété et exécuté automatiquement durant les phases ultérieures du traitement des données.
9. Le système de formulaires électroniques est basé sur l'idée selon laquelle l'Internet constituera, dans un proche avenir, le support de communication utilisé pour la plupart des

enquêtes statistiques. Même si cette tendance ne fait aucun doute, les problèmes de mise en oeuvre sont encore nombreux et, comme on le constate dans d'autres pays, la mise en place de tels systèmes demande du temps. Les informations utilisées dans le système de formulaires électroniques seront tirées du système de bases de sondage et du système de métainformation. Les problèmes de sécurité sont aussi importants, car les données individuelles sont transférées directement par Internet. Le Bureau central de statistique met actuellement en oeuvre le premier système (DG1) qui couvre les données économiques et suivant lequel les entreprises enverront leurs données directement par Internet.

10. L'exploitation des sources administratives requiert tout d'abord la signature de nombreux accords bilatéraux entre le Bureau central de statistique et d'autres organismes. Ces accords couvrent non seulement les aspects juridiques, mais aussi les solutions techniques. XML est souvent la norme retenue en la matière.

11. Pour ce qui est des systèmes en aval, l'utilisation d'un système d'établissement de rapports et d'analyse est particulièrement fréquente. En règle générale, ces systèmes dépendent du système de gestion de base de données relationnelle (RDBMS) utilisé. L'entrepôt de données du Bureau central de statistique est basé sur Oracle, et les statisticiens utilisent donc les instruments proposés par cette société. Oracle Discoverer permet aux statisticiens de créer leurs rapports directement en ligne, sans qu'il leur soit nécessaire de maîtriser le langage SQL. Oracle Reports est utilisé dans les cas plus complexes, qui nécessitent une programmation (confiée à des informaticiens). Les données ainsi communiquées peuvent être exportées vers les formats Excel ou CSV (valeurs séparées par des virgules). Par le terme «instruments d'analyse», nous entendons principalement le progiciel d'analyse statistique SAS. La connexion entre Oracle et SAS n'est pas encore utilisée, mais elle sera techniquement réalisable dans un proche avenir.

12. Enfin et surtout, il y a les systèmes de diffusion de données. Le portail est la formule la plus fréquemment utilisée, car il affiche non seulement les documents gérés par le système de gestion du contenu, mais aussi les données et les métadonnées prises directement dans l'entrepôt de données accessibles au public et dans le système de métainformation. Le Bureau central de statistique prépare actuellement un appel d'offres en vue de la création d'un nouveau portail, qui sera réalisé à l'aide des technologies les plus récentes et sur la base de l'expérience et des meilleures pratiques des autres pays.

-----