

Distr.  
GÉNÉRALE

CES/SEM.47/2  
10 mai 2002

FRANÇAIS  
Original: ANGLAIS

**COMMISSION DE STATISTIQUE et  
COMMISSION ÉCONOMIQUE  
POUR L'EUROPE**

**COMMISSION DES  
COMMUNAUTÉS EUROPÉENNES**

**CONFÉRENCE DES STATISTICIENS  
EUROPÉENS**

**EUROSTAT**

**Séminaire commun CEE-Eurostat sur  
les systèmes intégrés d'information statistique  
et les questions connexes (ISIS 2002)**

(Genève, Suisse, 17-19 avril 2002)

## **RAPPORT DU SÉMINAIRE SUR LES SYSTÈMES INTÉGRÉS D'INFORMATION STATISTIQUE TENU EN AVRIL 2002**

1. Le Séminaire commun CEE-Eurostat sur les systèmes intégrés d'information statistique et les questions connexes s'est tenu à Genève (Suisse) du 17 au 19 avril 2002. Y ont participé des représentants des pays suivants: Allemagne, Arménie, Autriche, Azerbaïdjan, Belgique, Bulgarie, Canada, Chypre, Croatie, Danemark, Estonie, États-Unis d'Amérique, ex-République yougoslave de Macédoine, Fédération de Russie, Finlande, France, Grèce, Hongrie, Irlande, Israël, Italie, Kazakhstan, Lettonie, Lituanie, Norvège, Pays-Bas, Pologne, Roumanie, République tchèque, Slovaquie, Slovénie, Suède et Suisse. La République de Corée y a participé en vertu de l'article 11 du mandat de la Commission économique pour l'Europe. La Commission européenne était représentée par Eurostat. Des représentants des organismes internationaux ci-après étaient aussi présents: Institut de statistique pour l'Asie et le Pacifique de l'ONU (ISAP), Organisation de coopération et de développement économiques (OCDE), Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO), Banque mondiale, Organisation mondiale du commerce (OMC) et Banque centrale européenne (BCE). Des représentants de l'Université d'Athènes ont également assisté au Séminaire à l'invitation d'Eurostat.

### **ORGANISATION DU SÉMINAIRE**

2. Le programme du séminaire a porté sur les questions de fond ci-après:
- i) Utilisation de la technologie du Web pour l'intégration des statistiques;
  - ii) Sécurité des communications et confidentialité des données;

- iii) Technologies orientées objet et architectures à composants; et
- iv) Moyens permettant aux systèmes d'information statistique de mieux répondre aux besoins des utilisateurs.

3. M. Mel TURNER (Canada) et M. Wolfgang KNUEPPEL (Eurostat) ont fait office respectivement de Président et de Coprésident. Les débats sur les questions i) à iv) susmentionnées ont été organisés par: M. Jan BYFUGLIEN (Norvège), M. Richard SWARTZ (États-Unis), M. Georges PONGAS (Eurostat) et M. Marton VUCSAN (Pays-Bas).

4. Des recommandations concernant les travaux futurs figurent ci-après. Les autres conclusions dégagées par les participants, lors du Séminaire, en ce qui concerne les questions susmentionnées sont présentées dans l'annexe du présent rapport.

### **RECOMMANDATIONS CONCERNANT LES TRAVAUX FUTURS**

5. De l'avis des participants, il serait utile qu'à l'avenir les pays échangent aussi des données d'expérience sur les évolutions et les meilleures pratiques dans ce domaine. Ils étaient cependant conscients que deux séries similaires de réunions – les Réunions sur la gestion de la technologie de l'information en statistique (tous les deux ans, les années impaires) et les Séminaires sur les systèmes intégrés d'information statistique et les questions connexes (tous les deux ans, les années paires) – avaient lieu régulièrement dans le cadre du programme de travail de la Conférence. De ce fait, ils ont recommandé de les fusionner en une seule réunion qui serait organisée à la fois par la CEE et Eurostat. Ils ont suggéré en outre de créer un groupe directeur qui définirait pour la nouvelle série un cadre comprenant les éléments suivants: i) objectifs de la série; ii) public visé; iii) partenariats; et iv) résultats escomptés.

6. Les représentants ci-après se sont portés volontaires pour devenir membres du Groupe directeur: M. Mel Turner (Canada); M. Jan Byfuglien (Norvège); M. Marton Vuksan (Pays-Bas); M. Richard Swartz (États-Unis); M. Wolfgang Knüppel, M. Daniel Defays (Eurostat), M. Gérard Salou (OCDE) et M. Karlis Zeila (Lettonie) et sont convenus de faire avancer leurs travaux en communiquant par voie électronique d'ici novembre 2002.

7. Les participants ont recommandé qu'un site Web sur les meilleures pratiques en matière de technologie de l'information dans les services de statistique soit créé, ainsi qu'il était proposé dans le document CES/SEM.47/CRP.3. L'OCDE a proposé d'héberger ce site dont le mandat correspondrait à celui défini dans le document susmentionné. Les auteurs de la proposition ont été invités à la réviser en tenant compte des divers groupes d'intérêts apparaissant dans le cadre de l'activité de programme 2 de la Conférence. MM. Markku Huttunen (Finlande), Richard Swartz (États-Unis), Peter Lubkert et Gérard Salou (OCDE) ont accepté de faire partie du groupe de coordination.

## ANNEXE

### Résumé des principales conclusions dégagées lors du Séminaire sur les systèmes intégrés d'information statistique tenu en avril 2002

**Thème i):** Utilisation de la technologie du Web pour l'intégration des statistiques

**Documentation:** Communications sollicitées du Canada, des États-Unis et de l'OCDE. Documents d'appui d'Israël, du Kazakhstan et de la Suède.

**Animateur:** M. Jan Byfuglien (Norvège)

1. Les échanges de vues consacrés à ce thème ont mis l'accent sur l'intégration des définitions statistiques et des fonctions de collecte, de traitement et de diffusion et sur la façon d'utiliser les technologies du Web pour y parvenir. De l'avis général, le Web pouvait apporter des solutions unifiées dans les divers domaines d'activité statistique. À l'origine, le Web était considéré comme un outil de diffusion, mais il existait un vaste domaine d'applications possibles depuis la saisie jusqu'à la fourniture des résultats.
2. L'application des technologies du Web exigeait une bonne coordination entre informaticiens et statisticiens. Les participants ont débattu du rôle des informaticiens dans les domaines suivants: création et mise en place de logiciel Internet, systèmes de sécurité/barrières de sécurité, administration des bases de données et assistance en la matière et autres tâches informatiques traditionnelles. Sur le plan technique, l'XML et les langages et technologies dérivés ont été jugés très intéressants pour les statistiques. On a relevé que d'autres questions concernant le Web méritaient d'être examinées plus avant: les incohérences du logiciel de navigation et la non-conformité aux normes W3C, l'authentification de l'utilisateur, le cryptage et la sécurité des données. Cette dernière question est examinée en détail au titre du thème ii) ci-après.
3. En ce qui concerne le rôle des statisticiens, la plupart des communications présentées soulignaient l'importance de l'intégration de métadonnées et encourageraient la poursuite de l'examen de cette question. Les participants sont convenus que l'intégration des métadonnées ainsi que l'harmonisation et la normalisation des concepts statistiques faisaient naturellement partie des activités statistiques menées à partir du Web, de préférence au niveau international. Il a été reconnu que le Web incitait davantage à l'harmonisation et à l'intégration plutôt qu'il n'en était un instrument direct. Un exemple d'harmonisation des métadonnées dans le cadre d'OLIS.NET a été présenté au Séminaire, en même temps que la mise en œuvre d'une classification commune des activités statistiques, l'élaboration de catalogues communs de données et un glossaire de termes statistiques. L'intégration exigeait aussi que l'on réfléchisse aux questions touchant les données, les règles d'entreprise, l'identification des utilisateurs et les fonctions liées aux bases de données communes.
4. Les participants ont également examiné les aspects budgétaires découlant de l'adoption des technologies du Web. On a estimé que si celles-ci permettaient de réduire le nombre de formulaires imprimés et d'économiser sur les frais d'affranchissement et d'entrée des données, les sommes ainsi économisées devraient être investies dans le domaine de la sécurité et de l'acquisition de nouvelles compétences techniques. En d'autres termes, il ne s'agissait pas de

réduire les coûts (les dépenses étant simplement déplacées vers d'autres domaines) mais d'améliorer l'actualité, la fiabilité, la cohérence et la comparabilité des données et de fournir des services aux utilisateurs finals dans de meilleurs délais.

5. Les services statistiques fournis par l'Internet devaient être conçus selon les principes fondamentaux suivants: être axés sur l'utilisateur (cohérence, concepts commune), simplicité, actualité (informations actualisées et interface en temps réel) et utilité (ensemble clair de services orientés sur des tâches précises). Les questions à résoudre concernaient davantage l'intégration des applications statistiques existantes que la technologie Web elle-même, celle-ci étant assez simple.

6. S'agissant du cadre architectural possible d'une intégration à partir du Web, les participants ont examiné les possibilités offertes par un «portail Web» contrôlant toutes les transactions lorsque l'on accède à une application existante, de préférence sous mode de lecture seule. Le portail Web représentait une interface intégrée avec les applications existantes; cependant, il pouvait être nécessaire de modifier en partie celles-ci à des fins de cohérence et pour simplifier les solutions techniques. On a souligné qu'il faudrait conserver les principales fonctions des composantes existantes tout en offrant de nouveaux services et formules de traitement.

7. On a présenté un exemple de l'application du Web à la série de recensements de la population et de l'habitation réalisés en 2000 et l'on a insisté sur la nécessité d'une bonne coordination et d'une harmonisation entre les diverses unités du service de statistique. On a estimé que la technologie Web pouvait remplacer la reconnaissance optique de caractères (OCR), mais certains participants ont souligné que celle-ci pouvait encore démontrer son utilité. Un autre exemple de collecte des données au moyen de la technologie Web était fourni par les entretiens individuels assistés par ordinateur (CAPI).

8. En conclusion, l'animateur a souligné qu'il importait d'entreprendre de nouvelles études sur les possibilités de saisie des données par le Web, ce qui nécessitait de résoudre plusieurs problèmes méthodologiques et aussi de réfléchir à la qualité. Il a par ailleurs souligné que le Web contribuait à détecter les défauts de concordance. Il a conclu enfin qu'il existait encore relativement peu d'exemples de l'application du Web à la phase de traitement des données.

## **Thème ii): Sécurité des communications et confidentialité des données**

**Documentation:** Communications sollicitées des États-Unis, de la France et des Pays-Bas. Documents d'appui des États-Unis, de la Norvège, de la Suède, du Centre de calcul international de l'ONU et de la Commission européenne.

**Animateur:** M. Richard Swartz (États-Unis)

9. L'examen de ce point de l'ordre du jour a porté sur les problèmes de confidentialité des données statistiques et de protection des personnes interrogées, compte tenu du recours aux réseaux informatiques et de la demande croissante de données plus détaillées et de microdonnées. Les participants se sont notamment penchés sur l'utilisation des infrastructures à clefs publiques (ICP), les questions de confidentialité découlant de l'accès direct des utilisateurs

aux bases de données, les outils logiciels qui permettent de contrôler la divulgation des données et la sécurité de l'échange des données.

10. On a souligné la nécessité d'adopter une stratégie pour la mise en place des ICP. On a relevé que leur utilisation entraînait d'importantes conséquences organisationnelles, en particulier la nécessité d'aligner le modèle ICP avec les pratiques commerciales. Les participants ont également examiné le problème de l'interopérabilité entre l'ICP de l'organisation et les ICP des partenaires privés. L'expérience pratique d'un institut de statistique concernant l'utilisation de l'infrastructure à clefs publiques a été présentée aux participants. L'ICP devrait être utilisée aussi bien au plan interne qu'avec l'extérieur. Les problèmes soulevés concernaient l'attribution des certificats, l'accréditation de certificats externes et la validité des certificats ICP dans les divers services publics. Un autre exemple avait trait à la mise en place d'une ICP par la Commission européenne (IDA PKI). Certains participants ont souligné que dans leur pays, l'ICP pouvait être assez coûteuse et que par conséquent son utilisation à des fins statistiques n'était pas concluante.

11. Pour éviter la divulgation des statistiques, les participants ont examiné la possibilité de séparer les métadonnées qui identifient les unités déclarantes et les microdonnées statistiques (par exemple en assignant des nombres aléatoires tout en supprimant les numéros d'identité personnelle et les renseignements d'ordre géographique). Quoique utilisée actuellement par plusieurs organismes statistiques, cette méthode risquait d'accroître la complexité du traitement des données. Certains participants ont souligné que les données elles-mêmes pouvaient souvent conduire à la divulgation d'informations confidentielles, même si l'on supprimait les métadonnées d'identification, de sorte que dans les publications statistiques il faudrait aussi supprimer certaines cases (cellules) et remplacer certaines cellules liées par des intervalles.

12. Les participants ont examiné des procédés mathématiques et d'autres méthodes permettant d'empêcher que les microdonnées stockées dans des tableaux statistiques multidimensionnels soient utilisées pour identifier l'unité déclarante. Le risque de divulgation provenait de la possibilité de relier les tableaux multidimensionnels et les séries longitudinales. Pour résoudre ces difficultés, on a proposé des stratégies de publication de modèles statistiques à la place des données ou des tableaux originaux. Il ressortait clairement du simple examen de modèles de régression linéaire et de prédiction spatiale qu'il fallait soigneusement réfléchir aux avantages et aux inconvénients d'une telle démarche.

13. Le projet concernant les aspects informatiques de la confidentialité des statistiques (projet CASC) a été présenté aux participants. Il visait à l'élaboration d'outils pratiques (logiciel ARGUS) permettant le contrôle de la divulgation des statistiques et prévoyait aussi des recherches dans le domaine de la protection des microdonnées et des données contenues dans des tableaux.

14. Au cours de l'échange de vues, certains participants ont souligné que la gestion décentralisée des serveurs de données dans les organismes statistiques risquait de rendre plus difficile la protection de la confidentialité. En pareil cas, il était nécessaire de mettre à jour régulièrement et de manière concertée les programmes de correction. Une autre préoccupation en matière de sécurité concernait la généralisation des logiciels commerciaux de l'Internet et des communications alors que les applications de caractère public pouvaient assurer une meilleure sécurité.

15. Il a également été fait mention des virus, très répandus dans l'environnement des ordinateurs personnels. On s'en est en particulier inquiété à propos des incidences qu'ils peuvent avoir dans le contexte des données et de la réalisation des entretiens, où le virus peut copier des fichiers sensibles ou saisir la frappe et la transmettre à une adresse non autorisée.

16. Certains participants ont insisté sur la sécurité interne, d'autres ont estimé que la sécurité externe était beaucoup plus importante. On a fait valoir que le personnel temporaire pouvait représenter un risque plus élevé pour la sécurité interne.

**Thème iii): Technologies orientées objet et architectures à composants**

**Documentation:** Communications sollicitées de la Lettonie, des Pays-Bas et d'Eurostat. Documents d'appui de l'Allemagne, de l'Azerbaïdjan, de la Suède et de la Banque centrale européenne.

**Animateur:** M. Georges Pongas (Eurostat)

17. Lors de l'examen de ce thème, l'accent a été mis sur les architectures qui favorisent la réutilisation et l'utilisation commune de composants de traitement. Les solutions orientées objet peuvent favoriser une approche à un niveau plus granulaire de la réutilisation, la conception d'applications à partir de petits composants et même la mise en commun de ces composants via le réseau ou par l'intermédiaire de bases de données orientées objet. Dans ce contexte, les participants ont examiné les questions concernant l'interopérabilité, les nouveaux modèles d'architectures informatiques pour les services de statistique et les incidences des systèmes considérés sur le plan organisationnel.

18. Un projet visant à revoir entièrement les systèmes d'information statistique d'un institut national de statistique a été présenté. Ce programme, qui tend à remplacer l'ensemble des logiciels, repose sur un schéma idéal exposé lors de la Réunion sur la gestion de la technologie de l'information en statistique tenue en février 1999 (Bo Sundgren, Statistics Sweden: «Une architecture des systèmes d'information destinée aux organismes nationaux et internationaux de statistique»). Un autre exposé était axé sur la combinaison de composants fonctionnellement riches (Argus, Slice, Bascula, Blaise Cristal). Eurostat a présenté la démarche en matière d'interopérabilité des systèmes logiciels qu'elle applique dans le contexte du projet de recherche STATOBJECT IST, au titre duquel les interactions entre les composants sont basées sur la sémantique de l'entreprise qu'ils appuient. D'autres exposés ont porté sur la mise en place d'un réseau sur la base du système d'exploitation Linux, le système d'information statistique GENESIS, le modèle UML du message GEMMES/CB pour l'échange de séries chronologiques et la mise au point de composants.

19. Au cours du débat, on a insisté sur le fait que l'utilisation des technologies orientées objet nécessite une approche conceptuelle. Alors que certains participants ont souligné que le concepteur moyen éprouvait parfois des difficultés à bien maîtriser ces techniques, d'autres ont estimé que les technologies orientées objet allaient probablement s'imposer comme le choix naturel lorsqu'il s'agirait de mettre au point des applications statistiques. Les participants ont également évoqué les contraintes pesant sur les composants logiciels ainsi que l'utilité limitée des applications constituées de composants sans lien rigoureux entre eux.

20. Les participants au Séminaire ont traité de la gestion des projets complexes, où il arrive fréquemment que des équipes indépendantes travaillent sur différents composants. À ce propos, il a été question de l'externalisation et de la relation qui existe entre l'architecture informatique et la structure administrative. Le débat a mis l'accent sur la nécessité d'une coordination, eu égard au fait que l'importation et la création de nouveaux composants devaient être approuvées avant que l'on puisse les intégrer dans le système. Les diverses sources de composants (par exemple les composants importés et les composants «maison», les applications qui se révèlent efficaces, etc.) ont été examinées.

21. Au cours de l'échange de vues, il est apparu difficile de prévoir quelle plate-forme prévaudrait à l'avenir, le passé ayant réservé certaines surprises. On pouvait donc s'attendre à toute une série de plates-formes de Microsoft à Linux. Le choix de la plate-forme n'était pas sans conséquence sur les autres logiciels en ce qui concerne l'appui XML, la modélisation UML, l'interrogation des bases de données, etc.

22. En conclusion du débat, les participants sont convenus que: i) l'architecture à composants est une technique d'avenir mais doit être utilisée avec soin; ii) la mise au point de nouveaux produits est caractérisée par une tendance à la normalisation (par exemple XML, formats d'échanges de données standard, etc.).

**Thème iv): Moyens permettant aux systèmes d'information statistique de mieux répondre aux besoins des utilisateurs**

**Documentation:** Communications sollicitées du Canada, de la Grèce et de l'OCDE. Documents d'appui des États-Unis, des Pays-Bas et de l'ISAP. Proposition concernant la création d'un site Web sur les meilleures pratiques présentée par l'Arménie, le Canada, les Pays-Bas et l'OCDE.

**Animateur:** M. Marton Vucsan (Pays-Bas)

23. Les participants ont accueilli avec satisfaction la proposition du Canada, des Pays-Bas et de l'OCDE tendant à créer un site Web sur les meilleures pratiques en matière de technologie de l'information dans les services de statistique. Ils ont accepté que le site en question soit hébergé par l'OCDE et construit en fonction des normes applicables aux sites de cette organisation, tandis que les pays contributeurs seraient propriétaires du site et responsables de son contenu. Il a été convenu en outre que chaque service national de statistique participant désignerait un agent de liaison chargé d'assurer la communication avec l'OCDE. Dans un premier temps, les chefs des délégations présentes au Séminaire assumeraient les fonctions d'agent de liaison. MM. Markku Huttunen (Finlande), Richard Swartz (États-Unis), Peter Lubkert et Gérard Salou (OCDE) ont accepté de faire partie du groupe de coordination.

24. Les problèmes et les résultats obtenus en matière de modélisation ont été examinés en relation avec les métadonnées. Il a également été souligné que les métadonnées constituaient un élément indispensable d'un bon site Web statistique. L'entreposage des données, comme indiqué au cours du débat, renforçait le pouvoir dont disposaient les statisticiens. Certaines questions stratégiques (par exemple l'utilisation de l'entrepôt de données comme principal gisement de données) et quelques questions techniques ont été étudiées dans ce contexte. L'ensemble des participants a reconnu que le Web avait démontré les grandes capacités qu'il offrait sur le plan

de la diffusion des données à condition d'appliquer une architecture cohérente en matière de données et métadonnées.

25. Au moins trois catégories de méta-données ont été identifiées: i) sémantiques – descriptions informelles, texte libre (en langage naturel); ii) opérationnelles – informations portant essentiellement sur les processus, mais aussi sur la qualité des données, le nombre d'imputations, le nombre d'estimations effectuées, etc; iii) techniques – noms de domaines, types de données, dimensions des données, etc. (au sens informatique et non pas visuel).

26. Un projet visant à créer une base de métadonnées intégrées dans un service national de statistique a été examiné. Il a été noté que la collecte et la mise à jour des métadonnées représentaient des tâches énormes exigeant un effort et un engagement importants de la part de l'organisme concerné. Dans certains cas, le multilinguisme s'imposait: le projet présenté, par exemple, devait prendre en compte deux langues nationales sur un pied d'égalité. Une autre présentation portait sur un modèle de métadonnées sémantiquement riche qui avait pour objet d'appuyer un système de traitement piloté par des métadonnées.

27. L'utilisation novatrice des technologies de l'information et de la communication permettait d'assurer un mode d'accès collectif aux bases de données d'un organisme. À ce propos, l'OCDE a présenté une amélioration introduite dans l'accès à ses données grâce à la mise en place d'un portail commun. Il a été fait état d'un certain nombre de conditions facilitant l'accès des utilisateurs: accès fondé sur l'Internet (pas besoin d'installer un logiciel quelconque sur l'ordinateur de l'utilisateur); disponibilité d'un instrument de recherche, interaction aisée avec les utilisateurs et manipulation facile des données, disponibilité d'instruments d'analyse ou facilité d'accès à ces derniers, accès dynamique et connectivité ODBC/OLEDB.

-----