

Distr.
GÉNÉRALE

CES/SEM.47/9 (Summary)
29 janvier 2002

FRANÇAIS
Original: ANGLAIS

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

**COMMISSION DES COMMUNAUTÉS
EUROPÉENNES (EUROSTAT)**

**CONFÉRENCE DES STATISTICIENS
EUROPÉENS**

**Séminaire commun CEE-Eurostat sur les
systèmes intégrés d'information statistique
et les questions connexes (ISIS 2002)**
(Genève, Suisse, 17-19 avril 2002)

Thème II: Sécurité des communications et confidentialité des données

**PROJET CONCERNANT LES ASPECTS INFORMATIQUES
DE LA CONFIDENTIALITÉ DES STATISTIQUES**

Communication sollicitée

de Statistics Netherlands¹

Résumé

Introduction

1. Le contrôle de la divulgation des statistiques (CDS) est un domaine de la statistique qui a suscité beaucoup d'intérêt au cours des dernières années. Les décideurs ont besoin de données statistiques de plus en plus détaillées. Les chercheurs ont les moyens de procéder à une analyse statistique complexe grâce à leurs ordinateurs puissants et demandent des microdonnées détaillées. Par conséquent, tout pousse les services de statistique à publier des informations de plus en plus complètes. Mais ils ne sauraient, pour autant, accorder plus de poids à leur fonction de fournisseur de données qu'à leur obligation de préserver la vie privée des personnes interrogées, qui leur confient des informations personnelles. Sans enquêtes, il n'y aurait pas de données statistiques.

¹ Préparée par Anco Hundepool (ahnl@krypton.vb.cbs.nl).

2. Le projet concernant les aspects informatiques de la confidentialité des statistiques est une initiative visant à coordonner la recherche et le développement en Europe. Il est subventionné en partie par le cinquième Programme-cadre de l'Union européenne. Mis en place dans le prolongement du projet relatif au contrôle de la divulgation des statistiques, il vise à combiner la recherche et l'élaboration d'outils pratiques, tels que le logiciel ARGUS. Dans ce contexte, nous étudions les problèmes de contrôle de la divulgation concernant aussi bien les microdonnées que les données contenues dans des tableaux.

Microdonnées

3. Le logiciel existant μ -ARGUS intègre déjà des fonctions utiles de recodage général et de suppression localisée en vue de protéger les microdonnées sociales. Toutefois, la divulgation sécurisée des microdonnées d'entreprise est bien plus difficile à garantir. Les méthodes traditionnelles de protection ne suffisent plus en l'occurrence, compte tenu des distributions particulières (asymétriques) caractérisant ces ensembles de données. Les pistes de recherche sont les suivantes:

- Ajout de bruit;
- Post-randomisation;
- Micro-agrégation.

4. Ces techniques vont être testées et intégrées aux nouvelles versions de μ -ARGUS. Aussi sera-t-il possible de comparer facilement les résultats des différentes méthodes. On peut ainsi pour la première fois espérer publier des microdonnées d'entreprise sécurisées. Toutefois, pour plus de sécurité, il nous faut envisager des «environnements sûrs», tel un centre d'analyse contrôlé au sein des instituts nationaux de statistique.

5. En outre, nous étudierons les modèles de risque qui peuvent être utilisés pour calculer le degré de sécurité des données personnelles ou d'un ensemble de données.

Données présentées sous forme de tableaux

6. Basée sur des techniques d'optimisation, la version actuelle du logiciel τ -ARGUS a résolu la question de la protection des données diffusées sous la forme de tableaux statistiques non structurés. Toutefois, l'expérience a montré que dans les applications réelles, les tableaux étaient fréquemment structurés de manière plus complexe. La plupart des listes de codes ont une structure hiérarchique, qui rend très difficile la protection de ces tableaux. Un grand nombre de sous-totaux, présentés sous forme de tableaux, pourrait contribuer à annuler la suppression des cellules (cases).

7. Pour résoudre ce problème, on envisage actuellement plusieurs nouvelles approches. Les spécialistes réfléchissent au moyen d'étendre les techniques actuelles d'optimisation aux tableaux hiérarchiques. Les nouvelles versions du logiciel τ -ARGUS intégreront cette fonction, mais d'autres options seront aussi envisagées, car compte tenu de la complexité du problème, on peut s'attendre à ce que le traitement des très grands tableaux en vue de leur protection soit très long. Le logiciel comprendra également des solutions reposant sur un découpage ingénieux du

tableau ainsi qu'une solution s'appuyant sur des hypercubes. Il offrira donc toute une gamme d'options, et l'utilisateur pourra choisir entre une solution lente mais optimale et une solution rapide mais pas tout à fait optimale. L'une des retombées de ces efforts sera la possibilité d'appliquer aux tableaux corrélatifs les solutions mises au point pour les tableaux hiérarchiques.

Expérimentation

8. On accordera une grande attention à l'expérimentation du logiciel et des méthodes. On peut mettre en œuvre de nombreuses techniques prometteuses, mais encore faut-il qu'en fin de compte, les données produites soient utiles. C'est la raison pour laquelle on réfléchit, par exemple, à la «puissance analytique» d'un ensemble de données. L'utilité pratique du logiciel mis au point sera également testée en situation réelle.

Conclusion

9. Nous espérons faire un grand pas en avant tant dans le domaine de la recherche que de l'application pratique des méthodes de contrôle de la divulgation des statistiques. Vous trouverez un complément d'information sur le projet CASC à l'adresse Internet suivante:

<http://neon.vb.cbs.nl/casc/default.htm>.
