

**91<sup>e</sup> CONFERENCE DES DGINS**

**COPENHAGUE, 26 & 27 MAI 2005**

**Systèmes statistiques nationaux et chercheurs**

*Svante Öberg*

*Directeur Général de l'institut national suédois de la statistique*

# Systèmes statistiques nationaux et chercheurs

Svante ÖBERG

*Directeur Général de l'institut suédois de la statistique*

Dans ce document je présente les relations entre les systèmes statistiques nationaux et les chercheurs du point de vue du producteur de statistiques. J'expose les avantages de l'accès gratuit aux publications et bases de données via Internet et de la mise au point de systèmes d'accès en ligne aux microdonnées à l'attention des chercheurs et je plaide en faveur d'une amélioration de la coopération avec la communauté des chercheurs. J'utilise la situation de l'institut suédois de la statistique comme exemple. Je propose ces réflexions à la discussion mais ne prétends nullement savoir si elles conviennent bien à chacun des instituts nationaux de la statistique.

## **1. Introduction**

La communauté des chercheurs est un important groupe d'utilisateurs de statistiques. L'une des principales missions des différents instituts nationaux de la statistique (INS) consiste à fournir à cette communauté des statistiques et des microdonnées afin d'alimenter ses travaux de recherche. Ce document présente, d'une part, l'accès gratuit aux bases de données via Internet et l'accès facilité aux microdonnées comme moyens de soutenir le travail des chercheurs et comme conditions indispensables à une diffusion encore plus flexible des statistiques. Ce document dresse, d'autre part, les perspectives d'amélioration de la coopération avec les chercheurs au niveau des INS et au niveau européen.

## **2. Accès gratuit aux publications et bases de données via Internet**

Il y a un quart de siècle, les données statistiques étaient généralement diffusées à l'aide de tableaux imprimés insérés dans des publications papier. Cette méthode était relativement peu efficace en termes de flexibilité. Les statisticiens étaient contraints de choisir les tableaux à publier et de concevoir l'étude en conséquence. Lorsque les chercheurs souhaitaient utiliser les statistiques pour effectuer des calculs différents, ils devaient alors transférer les chiffres manuellement ou les faire transférer dans leurs propres programmes.

Au milieu des années 70, l'institut suédois de la statistique a entrepris de construire des bases de données. Elles étaient accessibles contre paiement via des terminaux utilisant le réseau téléphonique. Les souscriptions des utilisateurs permettaient de financer les opérations liées aux bases de données. En janvier 1997, ces bases de données sont devenues accessibles via Internet.

En janvier 2000, nous avons adopté une nouvelle politique en matière de diffusion et avons permis aux utilisateurs d'accéder gratuitement à nos bases de données. Auparavant, chaque utilisateur devait verser environ 650 € par an. Les utilisateurs payants étaient au nombre de 350 et les non payants au nombre de 100 (consultation de bibliothèques...). Les pouvoirs publics ont compensé cette perte de revenus (près de 200 000 €).

En l'espace de trois mois, nous comptons déjà plus de 10 000 utilisateurs. Ce chiffre s'élevait à 30 000 lors de la dernière estimation. Depuis, nous avons arrêté de compter, car, pour accéder aux bases de données, les utilisateurs n'ont plus besoin de s'inscrire. A la suite de ce changement, le nombre d'extractions de bases de données a doublé.

En janvier 2000, nous avons également décidé de diffuser sur notre site Web l'ensemble de nos publications au format électronique et de permettre aux utilisateurs de les télécharger gratuitement. L'une des principales raisons pour lesquelles nous avons décidé de ne pas faire payer les utilisateurs est que cela n'aurait pas un impact majeur sur notre trésorerie. Nous savions que si nos revenus provenant des publications papier s'en trouvaient réduits, il en allait de même de nos coûts. Ces revenus couvraient uniquement le coût du papier, d'impression, de distribution et de gestion.

Notre site Web est désormais notre principal support de diffusion des données statistiques. L'année dernière notre site a totalisé 2,7 millions de visites, sans compter les accès depuis les moteurs de recherche.

Cette expérience souligne les avantages de l'accès gratuit aux publications et bases de données via Internet. Il existe également de nombreux arguments théoriques qui tendent à présenter les données statistiques comme un bien public, qui devrait être financé par l'impôt et mis gratuitement à la disposition du public. En effet, le coût marginal de la diffusion d'un exemplaire supplémentaire d'une publication ou d'une série de données est très faible, beaucoup plus faible que pour une copie papier de la même publication et généralement inférieur au bénéfice qu'en retire l'utilisateur.

Parallèlement, les INS doivent trouver des moyens de financer les coûts de développement. S'il est impossible d'obtenir un financement des pouvoirs publics, il peut s'avérer nécessaire de faire payer les utilisateurs. Une autre solution pour réduire les coûts associés à l'accès gratuit aux publications et bases de données consiste à rationaliser le système de production en se concentrant sur la diffusion Internet comme seul canal de diffusion. Lorsque les publications sont préparées, elles devraient l'être de manière à pouvoir être imprimées et téléchargées depuis le site Web sans engendrer de coûts supplémentaires.

Je suis un fervent défenseur de l'accès gratuit aux publications et bases de données via Internet. En outre, je suis particulièrement satisfait qu'Eurostat ait modifié sa politique de diffusion en ce sens. Les données statistiques relatives aux évolutions de l'Union européenne et de la zone euro prennent une importance de plus en plus grande à mesure que l'intégration des États membres s'intensifie et que la politique européenne se développe.

J'ai deux suggestions à faire. Premièrement, le secrétariat de l'OCDE devrait suivre le mouvement. Ses publications de très grande qualité et son large éventail de précieuses bases de données seraient mieux mis à profit si les utilisateurs pouvaient y accéder gratuitement. Au cours des 10 à 15 dernières années, l'Europe s'est longuement penchée sur l'harmonisation de ses statistiques. Il est temps à présent de s'intéresser à la comparabilité entre l'Europe et les autres régions du monde, notamment les autres pays de l'OCDE. A cet effet, il serait souhaitable que les vastes archives de l'OCDE soient accessibles gratuitement.

Deuxièmement, je recommande aux INS qui n'offrent pas encore un accès gratuit à leurs publications et bases de données de revoir leur politique, de peser le pour et le contre et de réfléchir aux différents moyens de proposer un accès gratuit.

### **3. Systèmes d'accès en ligne aux microdonnées pour chercheurs**

Les fichiers de microdonnées anonymisés, c'est-à-dire des fichiers de données contenant des informations sur des individus ou des entreprises, mais aucun numéro d'identification ou autre, constituent une nouvelle avancée qui permettra aux INS de proposer à leurs utilisateurs une offre plus flexible. Grâce aux bases de données accessibles via Internet, les utilisateurs ont la possibilité de compiler une grande diversité de tableaux. Les fichiers de microdonnées anonymisés offrent encore plus de flexibilité. En outre, les progrès technologiques ont permis aux chercheurs d'intégrer plus facilement les microdonnées à leurs travaux. La demande de microdonnées est en constante augmentation.

L'institut suédois de la statistique a pour tradition de collecter des données administratives et de les convertir en registres compatibles avec les pratiques statistiques. La grande majorité des données statistiques suédoises sont tirées de registres administratifs – de 85 % à 95 % selon le mode de mesure. Le système de registres comprend également des registres basés sur des enquêtes, telles que les enquêtes sur les forces de travail. Depuis des années, l'institut suédois de la statistique diffuse des microdonnées anonymes à un grand nombre de centres de recherche sur bandes magnétiques, CD-Rom, DVD ou autre support. Le volume de publications a progressé parallèlement à

l'augmentation du nombre de communications. Environ 170 publications ont été dénombrées l'année dernière.

La confidentialité des données relatives à des individus ou des entreprises est l'un des principes clés des opérations statistiques officielles. Cet aspect doit être abordé lorsqu'il est question de microdonnées. Les individus doivent être protégés contre toute intrusion dans leur vie privée. A cet effet, l'utilisation des informations statistiques est normalement réglementée par une législation ou un code de bonne conduite.

En Suède, la législation relative à l'utilisation des informations statistiques prévoit que toutes les données collectées à des fins statistiques, y compris les données anonymes, sont confidentielles, quelle qu'en soit la source. Les données collectées à des fins statistiques peuvent uniquement servir à la production de statistiques ou à des travaux de recherche. Seules les personnes autorisées peuvent consulter les données confidentielles. Elles ne sont pas pour autant en mesure d'identifier directement ou indirectement les individus ou entreprises concernés. En pratique, les microdonnées auxquelles l'institut suédois de la statistique permet d'accéder ne contiennent ni nom, ni adresse, ni numéro d'identification. La législation en vigueur en Suède et dans les autres pays nordiques ne contient aucune règle limitative quant au mode de diffusion des microdonnées. Dès lors que les impératifs juridiques généraux sont respectés, l'émetteur peut choisir la méthode la plus appropriée.

L'institut suédois de la statistique a entrepris de mettre au point des procédures qui rendent l'utilisation des microdonnées dans les travaux de recherche plus simple, moins coûteuse et plus sécurisée. Nous pensons qu'il est préférable de conserver les microdonnées à l'institut statistique sous leur forme physique plutôt que de les distribuer aux centres de recherche à l'aide de CD-Roms ou autres types de supports. Nous devons déployer des techniques permettant aux chercheurs de manipuler, via Internet, les microdonnées stockées à l'institut statistique et rechercher des financements pour soutenir le développement et la gestion du système afin de réduire au minimum le coût marginal pour chaque chercheur.

L'année dernière, une nouvelle unité a été mise en place en vue de coordonner cette initiative. Cette unité, baptisée «Register Coordination and Microdata Access» (coordination des registres et accès aux microdonnées), fait partie du département Recherche & Développement. L'un de nos objectifs est de mettre en place un vaste système de registres, baptisé «Statistics Sweden Data Warehouse» (centre de données de l'institut suédois de la statistique), au sein duquel les données des différents registres seront standardisées et interconnectées et les responsabilités clairement définies pour chaque registre et chaque étape du travail de coordination.

L'institut suédois de la statistique a également développé un système d'accès en ligne aux microdonnées. Ce système, inspiré du système danois, est surnommé MONA (Microdata Online Access at Statistics Sweden – accès en ligne aux microdonnées). Les chercheurs peuvent se connecter directement au serveur de l'institut suédois de la statistique. Ils accèdent aux données à partir d'une plate-forme spécifique (SAS ou SPSS, par ex.) et peuvent ensuite les manipuler à leur gré, établir leurs propres jeux de données, effectuer des calculs, etc. Toutes les opérations sont effectuées directement sur le serveur de l'institut suédois de la statistique et aucun téléchargement n'est autorisé. La plupart du temps, les résultats sont transmis aux chercheurs par e-mail sous forme de tableaux.

La prochaine étape consistera à élaborer de nouveaux registres thématiques répondant aux besoins de la communauté des chercheurs (comme le registre LISA qui est spécifiquement dédié aux arrêts maladie et à l'emploi). C'est un travail titanesque qui demande l'intervention à la fois de méthodologistes et d'experts des sujets traités. Nous songeons également à développer des techniques qui permettront de relier des données provenant de différentes sources, internes ou externes.

L'amélioration de l'accès aux microdonnées implique cependant des coûts relativement élevés. En l'absence d'un financement public, comme c'est notre cas, ce sont les chercheurs qui doivent en supporter les coûts. C'est pourquoi l'institut suédois de la statistique a demandé au Conseil suédois pour la recherche de financer un système d'accès aux microdonnées. Ce serait une manière de réduire les coûts marginaux pris en charge par les chercheurs et d'encourager l'utilisation des microdonnées dans les travaux de recherche. Ce mode de fonctionnement – prise en charge du financement par une entité tierce et réduction des coûts marginaux – a été appliqué avec succès à d'autres domaines. Un tel système d'accès aux microdonnées demande en outre une prise en charge informatique complète, un service d'assistance aux chercheurs et la création de bases de données thématiques.

L'institut suédois de la statistique reçoit de nombreuses demandes de microdonnées de la part d'universités et de centres de recherche américains, désireux d'intégrer à leurs propres bases de données des résultats de recensement ou des données de registres. Pour différentes raisons, notamment des restrictions juridiques, nous ne pouvons pas répondre aux demandes impliquant un transfert physique des données. Cependant, il semblerait que la transposition à l'échelle mondiale du modèle danois/suédois décrit précédemment soit techniquement envisageable. Toutes les données pourraient ainsi être conservées à leur emplacement d'origine tout en étant à la disposition des chercheurs du monde entier. L'Australie a déjà entrepris de déployer un réseau similaire au

niveau national, le National Data Network (réseau national de données). Le bureau australien de la statistique est l'un des points névralgiques du réseau, également composé d'agences statistiques et de centres de recherche.

J'ai trois suggestions à faire. Premièrement, je suggère que les INS prêts à développer des systèmes d'accès en ligne aux microdonnées se regroupent au sein d'un réseau afin d'échanger leurs idées et expériences. Je suis convaincu que la communauté des chercheurs en retirerait d'énormes bénéfices. En outre, l'institut suédois de la statistique ayant déjà pris des initiatives en ce sens, il ne fait aucun doute que nous prendrions une part active à ce réseau.

Deuxièmement, dans le contexte international, je suggère que nous procédions de manière progressive, en développant dans un premier temps une politique commune en matière de confidentialité et d'accès aux microdonnées. Un groupe de travail mené par Dennis Trewin, directeur du bureau australien de la statistique, placé sous l'égide de la conférence des statisticiens européens, a été créé à cet effet. L'année dernière, les tentatives d'accélération du processus de la part du Comité du programme statistique (Eurostat) et du Comité des statistiques (OCDE) n'ont pas été particulièrement bien accueillies.

Troisièmement, je suggère que nous nous intéressions aux perspectives de création de réseaux de microdonnées internationaux, qui permettraient aux chercheurs d'autres pays d'accéder à nos données de manière sécurisée.

#### **4. Organisation de la coopération avec les chercheurs**

Presque tous les instituts nationaux de la statistique, sinon tous, coopèrent avec les chercheurs. Cette coopération peut prendre diverses formes selon la taille du pays, les coutumes, etc.

A l'institut suédois de la statistique, nous avons défini douze groupes d'utilisateurs dédiés à différents sujets : économie, protection sociale, démographie, territorialité, etc. Ces groupes d'utilisateurs sont composés d'experts externes, notamment de chercheurs. Ils nous conseillent sur le développement des statistiques relatives à leurs domaines de compétence. Nous disposons également d'un conseil scientifique, d'un comité de l'indice des prix à la consommation et d'un comité du coût de la construction réunissant des représentants de la communauté des chercheurs. Cette année nous avons créé un comité dédié aux comptes de la nation, qui nous conseillera sur les méthodes utilisées dans ce domaine et ce, conformément au Système européen de comptes nationaux.

Par ailleurs, nous coopérons avec des universités, notamment les universités de Stockholm et Örebro, villes dans lesquelles sont implantés nos bureaux. Nous avons signé un accord de coopération avec l'université d'Örebro en 2001. Il prévoyait la création d'une chaire partagée dédiée aux statistiques, le financement conjoint des doctorats et la mise en place de cours réguliers, de séminaires et de stages d'été. Je suis membre du conseil de l'université d'Örebro et plusieurs professeurs de cette université sont membres des différents groupes mentionnés précédemment. A l'université de Stockholm, une chaire est spécialement dédiée aux statistiques officielles. Le professeur concerné consacre une partie de son temps à l'institut suédois de la statistique.

Outre ces formes de coopération assez formelles et étendues, il existe également de nombreuses formes de coopération entre l'institut suédois de la statistique et des chercheurs issus de divers domaines. Certains de nos agents sont des professeurs d'université et il arrive également que des chercheurs universitaires travaillent avec nous à temps partiel. Enfin, nous conduisons régulièrement des enquêtes de satisfaction auprès des utilisateurs et notamment auprès des chercheurs.

En comparant l'institut suédois de la statistique à d'autres INS, j'ai remarqué que les Etats-Unis étaient sensiblement en avance sur nous et savaient pleinement tirer profit de la coopération avec les universités. A cet égard, les INS d'autres pays européens, notamment du Royaume Uni et de la France, sont également mieux positionnés que nous. Par ailleurs, nous ne sommes pas très actifs dans les domaines de la recherche et de l'analyse en interne, contrairement aux INS canadien, australien et norvégien. Et contrairement à l'INSEE en France, nous ne proposons pas de formation supérieure interne en statistiques.

J'ai deux suggestions à faire. Premièrement, je suggère que la plupart des INS renforce la coopération avec les chercheurs. Nous avons tout à gagner à nous inspirer des expériences positives des autres pays. Il se trouve que la plupart des changements qui s'opèrent actuellement au sein de l'institut suédois de la statistique sont le reflet des expériences des autres pays. Je ne suis pas en train de dire que nous devons adopter exactement les mêmes formes de coopération, mais que nous pouvons reprendre les idées compatibles avec nos systèmes nationaux.

Par ailleurs, j'ai été étonné de voir à quel point l'approche des INS et d'Eurostat était différente en matière de relations avec les utilisateurs. Dans la plupart des INS, nous disposons de systèmes très élaborés pour dialoguer avec les utilisateurs. Bien sûr, ces systèmes sont différents d'un pays à un autre. Certains INS ont des conseils nationaux de la statistique couvrant l'intégralité des statistiques officielles. D'autres INS, dont l'institut suédois de la statistique, disposent de groupes d'utilisateurs

dédiés à différents sujets. La plupart des INS entretiennent également des contacts réguliers avec les organisations sectorielles, entre autres.

Ce type de coopération fait encore défaut au niveau européen. Nous avons bien le CEIES (Comité Consultatif européen de l'Information statistique dans les domaines Economique et Social), qui fait actuellement l'objet d'un audit. Je ne pense pas être injuste en disant que le CEIES n'a pas joué un rôle majeur en tant que groupe d'utilisateurs. Les points de vue des utilisateurs ont transité par les diverses directions générales de la Commission et par les groupes de travail et comités d'Eurostat. A la décharge du CEIES, je tiens quand même à rappeler que ce comité a organisé d'excellents séminaires particulièrement appréciés.

Deuxièmement, je suggère qu'Eurostat entreprenne de revoir l'organisation de ses relations avec les utilisateurs, notamment les chercheurs, en tenant compte des expériences des pays membres. Le système statistique européen ne tirerait-il pas profit de conseils composés d'utilisateurs et dédiés à l'ensemble des statistiques européennes ou à des sujets spécifiques? Ne serait-il pas plus efficace s'il disposait d'un conseil scientifique ou de conseils dédiés aux comptes nationaux, à l'indice des prix à la consommation ou à d'autres sujets? N'a-t-il rien à apprendre d'enquêtes de satisfaction portant sur différents domaines statistiques?

## **5. Conclusion**

Ce document présente les relations entre les systèmes statistiques nationaux et les chercheurs du point de vue du producteur de statistiques. Les perspectives d'amélioration et de valorisation de notre production sont nombreuses. Ce document expose notamment les avantages de l'accès gratuit aux publications et bases de données via Internet et du déploiement de systèmes d'accès en ligne aux microdonnées à l'attention des chercheurs. Nous avons également de grands progrès à faire en matière de coopération avec la communauté des chercheurs. Les réflexions présentées ici le sont à titre de discussion et je suis parfaitement conscient que d'autres solutions sont peut-être mieux adaptées à certains instituts nationaux de la statistique.