



**Conseil Économique
et Social**

Distr.
GÉNÉRALE

CES/2001/6
21 mars 2001

FRANÇAIS
Original : ANGLAIS

**COMMISSION DE STATISTIQUE et COMMISSION ÉCONOMIQUE POUR L'EUROPE
CONFÉRENCE DES STATISTICIENS EUROPÉENS**

Quarante-neuvième réunion plénière
(Genève, 11-13 juin 2001)

Le développement d'Internet : un moyen d'élargir la diffusion des statistiques¹

Document établi par le secrétariat

I. INTRODUCTION

1. L'élaboration de politiques et de stratégies efficaces de diffusion de l'information statistique dans le contexte de l'essor des technologies de l'information et de la communication, et plus particulièrement du développement des services en ligne sur Internet, constitue une tâche complexe qui continue de poser des problèmes aux bureaux de statistique. Un des principaux objectifs consiste à faire en sorte que les utilisateurs, toujours plus nombreux, puissent avoir accès à l'information statistique dont ils ont besoin et l'utiliser. Pour permettre une meilleure adéquation entre les statistiques officielles et les besoins des clients des bureaux de statistique, qu'ils soient répondants ou simples utilisateurs, il est nécessaire de réorganiser de nombreuses phases de la production de statistiques.

2. Comme n'importe quelle organisation aujourd'hui, les bureaux nationaux et internationaux de statistique doivent décider de la façon dont ils vont s'adapter au développement prodigieux d'Internet et choisir le moment auquel ils comptent en tirer parti. Si les premières activités concernées par l'arrivée d'Internet sont la collecte et la diffusion des données, ce phénomène aura aussi des répercussions sur la gestion de la production de statistiques et sur toute son organisation.

¹ Les délégués qui souhaitent consulter la liste des références sont invités à se mettre en rapport avec le secrétariat.

3. L'objectif du présent document est de mettre en évidence les principaux effets du développement d'Internet sur la production de statistiques. Dans la section II, l'accent est mis sur les principales difficultés que doivent résoudre les bureaux de statistique; dans la section III sont décrites les méthodes et techniques de collecte de données en ligne. La section IV est consacrée à la diffusion des statistiques sur Internet considérée dans la perspective du client, au rôle des entrepôts de données et à l'utilité des métadonnées dans l'optique de l'intégration. Enfin, la section V présente des conclusions.

4. Les avis exprimés dans le présent document tiennent compte de la plupart des points examinés par les États membres de la CEE/ONU et les organisations internationales lors des réunions sur les technologies de l'information appliquées à la statistique, organisées ces deux dernières années par le secrétariat sous l'égide du programme de travail conjoint de la Conférence des statisticiens européens.

II. PRINCIPALES DIFFICULTÉS

5. Internet : un défi pour les bureaux de statistique. À l'heure où Internet est en train de devenir le principal vecteur d'interaction entre le bureau de statistique et ses clients, la demande est vouée à de profonds bouleversements, car les attentes du public face à Internet dépassent largement le cadre traditionnel des sources d'informations gouvernementales :

- **Accessibilité** : le public s'attendra à avoir un accès beaucoup plus large aux statistiques et à pouvoir consulter des données plus détaillées. Toutes les analyses devront pouvoir être consultées moyennant un coût modique, voire nul. Il y aura en outre une forte demande de fichiers informatiques faciles d'accès et susceptibles d'être intégrés à d'autres analyses.
- **Ponctualité** : les utilisateurs s'attendent à pouvoir obtenir et utiliser immédiatement les résultats. Les délais traditionnels d'établissement et de traitement de l'information statistiques seront de moins en moins adaptés à la demande.
- **Cohérence** : les utilisateurs non spécialisés prendront l'habitude de confronter des informations provenant de sources multiples. Il faut donc harmoniser les concepts et les classifications les plus fréquemment utilisés. De nombreux utilisateurs découvriront rapidement les lacunes du système.
- **Pertinence** : les informations seront évaluées en fonction de leur pertinence par rapport aux domaines traités par les clients (groupe de population visé et domaine étudié, accès aux métadonnées). Les données seront intégrées de diverses manières; elles seront notamment intégrées à des informations non statistiques. La publication de statistiques par les bureaux de statistique constituera un label d'authenticité, au regard de l'apparition de nombreuses sources jusqu'ici inconnues.

6. **Participation directe des utilisateurs à la mise en œuvre des stratégies de production de statistiques.** L'instauration d'une communication de plus en plus interactive, via Internet, entre les clients utilisateurs de statistiques officielles et les bureaux de statistique conduit les utilisateurs potentiels à participer de plus en plus directement à l'élaboration de stratégies de collecte et de diffusion des données. De plus, la participation accrue des utilisateurs permet aux bureaux

de statistique d'analyser les réactions de ces derniers, de suivre l'évolution de leur comportement et d'adapter en conséquence leur stratégie de diffusion. Il est donc essentiel de mieux comprendre qui sont les utilisateurs recourant à Internet, car leurs besoins d'information peuvent varier considérablement. La stratégie de diffusion doit être perçue dans un contexte plus large, et les questions de communication et de formation doivent être considérées non seulement du point de vue des producteurs de données mais aussi dans l'optique de l'utilisateur.

7. **Gestion des ressources.** L'utilisation d'Internet a une incidence directe sur le plan des effectifs des bureaux de statistique. Cela est particulièrement vrai pour la collecte des données, car à ce stade, l'utilisation d'Internet entraîne incontestablement un transfert du travail d'édition et de codage du bureau de statistique au répondant. En réaffectant les ressources jusqu'ici mobilisées pour le dépouillement du courrier, la saisie, l'édition et le codage à des activités telles que l'assurance-qualité, la recherche méthodologique, l'analyse, la diffusion ou la conception et la maintenance de systèmes, le bureau de statistique, qui dispose d'effectifs limités, peut affecter une grande partie de son personnel à des activités à forte valeur ajoutée, pour une production peu coûteuse.

8. **Établissement de budgets et planification.** La prise en compte croissante de l'utilisateur par le bureau de statistique peut avoir des effets importants sur l'élaboration des budgets et la planification. Les responsables des bureaux de statistique pourraient en effet disposer d'un instrument leur permettant de mettre en place un système de tarification de la diffusion des données, système qui pourrait être utilisé pour évaluer le degré de pertinence des données produites.

9. **Intégration des entreprises et des systèmes d'information statistique.** L'utilisation d'Internet comme moyen de communication entre les répondants et les bureaux de statistique, notamment pour la collecte des données auprès des entreprises, risque d'entraîner un bouleversement de l'architecture des systèmes d'information statistique. Il faut s'attendre à ce que l'approche fondée sur l'enquête elle-même soit peu à peu délaissée au profit d'une démarche basée sur les concepts thématiques.

10. **Sécurité et confidentialité.** La nécessité de recourir à des méthodes efficaces de protection de la vie privée et du caractère confidentiel des données prendra une importance croissante. Cette question constituera un sujet de préoccupation à la fois pour les particuliers et pour les entreprises. Il faudra accorder une plus grande attention à la communication avec les répondants au moment de la collecte des données, et le législateur devra limiter plus rigoureusement l'établissement de liens entre les différentes données. Les utilisateurs souhaiteront être associés aux négociations publiques consacrées à la protection de la vie privée et de la confidentialité des données. Les bureaux de statistique seront tenus d'informer les répondants au sujet de la nature des données à caractère personnel qu'ils détiennent, ainsi que de l'utilisation à laquelle ces données sont destinées.

11. **Concepts de données et de métadonnées.** Le développement de réseaux nationaux et internationaux nécessitera un effort soutenu de normalisation et d'intégration. Des concepts normalisés régissant les statistiques et les métadonnées, mais aussi des classifications, seront nécessaires pour répondre à la demande de cohérence et de compatibilité des données qui sera exprimée par la clientèle. L'intégration s'opérera par l'adoption de désignations administratives

et législatives harmonisées (numéros d'entreprise communs, notions géographiques normalisées, etc.).

12. **Normalisation du processus de production.** Au sein des bureaux de statistique, l'utilisation d'Internet favorise incontestablement le processus d'unification et de regroupement de nombreuses activités liées à la collecte et à la diffusion de données. Un tel regroupement pourrait bien s'étendre à des domaines thématiques, grâce à l'harmonisation des concepts, à la création d'instruments communs et à l'utilisation de procédés statistiques identiques. Les clients devront avoir accès, par le biais d'interfaces graphiques conviviales, à des informations cohérentes et comparables jusque dans l'utilisation de cadres et de systèmes de classification communs.

13. **Mesure de la qualité des données.** La définition et la mesure de la qualité des statistiques tendront à devenir plus explicites. Le développement de réseaux multidimensionnels intégrant les clients des bureaux de statistique souligne la nécessité d'élaborer un cadre généralement accepté régissant la qualité, afin de permettre une meilleure évaluation de la comparabilité des données diffusées, aux niveaux national et international.

14. Certains pays ont déjà mis au point des systèmes dans lesquels la cohérence, l'interprétabilité, la ponctualité, l'accessibilité, l'exactitude et la pertinence représentent autant de critères mesurables. Il est toutefois possible de retenir d'autres critères de qualité, tels que les mesures de variabilité d'échantillonnage, la couverture, des indicateurs de qualité d'édition ou l'impact de la confidentialité. De plus, d'autres caractéristiques essentielles de la qualité des données sont définies par les utilisateurs eux-mêmes et peuvent donc varier considérablement. Des métadonnées décrivant la qualité des statistiques doivent donc être aisément disponibles. Il serait particulièrement judicieux d'élaborer des modèles et des recommandations internationalement acceptés concernant l'évaluation de la qualité des données.

15. **Les diverses façons d'utiliser Internet.** Internet est appelé à devenir un moyen de communication essentiel, mais les méthodes traditionnelles de collecte et de diffusion des données, et d'interaction avec les bureaux de statistique, ne disparaîtront pas pour autant. Les bureaux de statistique doivent donc se préparer à recourir à divers procédés de transmission, qu'ils devront veiller à rendre compatibles. À court terme, il se peut que les coûts augmentent et que la production de statistiques devienne plus complexe.

16. **Impact sur les systèmes d'information statistique au niveau national (intégration verticale, intégration horizontale).** D'une façon générale, Internet favorisera l'intégration des services gouvernementaux nationaux, qui seront centrés sur le client. En d'autres termes, du point de vue de l'utilisateur, les différences entre bureaux et départements nationaux seront moins importantes. Bien qu'induite par une demande de services toujours plus performants, cette évolution posera aussi le problème du respect de la confidentialité des données. Il serait possible de remédier à ce problème en opérant une séparation entre la branche statistique et l'administration proprement dite. Une telle mesure permettrait d'unir les bureaux de statistique au niveau national et elle s'accompagnerait vraisemblablement d'une législation stricte régissant la protection de la vie privée et de la confidentialité. En d'autres termes, le bureau de statistique deviendrait la seule autorité légalement habilitée à intégrer des données aux fins de statistiques, tandis que les autres départements ne disposeraient que du minimum de données nécessaire à la fourniture de services.

17. **Impact sur l'intégration lors de la diffusion internationale des statistiques.** Internet accélère considérablement le processus d'intégration de la circulation des statistiques au niveau international. Les particuliers, les entreprises et les organismes publics (nationaux et internationaux) exigeront une plus grande cohérence dans leurs rapports avec les bureaux de statistique. Les organisations internationales qui communiquent avec les bureaux nationaux ont un grand rôle à jouer en ce domaine.

18. **Instruments d'intégration, nouvelle définition et réorganisation de la production de statistiques.** Les outils Internet auront un impact considérable sur l'ensemble du processus de production de statistiques. Des technologies telles que le langage XML seront essentielles dans l'élaboration des interfaces et l'automatisation du transfert de l'information entre les nœuds de traitement. L'intégration, dans son acception traditionnelle d'unification monolithique, sera abandonnée, au profit d'unités de traitement réduites susceptibles de communiquer et de coopérer les unes avec les autres en utilisant une syntaxe de transfert commune et des définitions d'interfaces compatibles. Les outils Internet les plus avancés ont fréquemment démontré qu'ils pouvaient être à l'origine de changements dans l'organisation des services de statistique, notamment en obligeant les responsables à restructurer le fonctionnement des bureaux de statistique et, plus précisément, à renoncer au traitement séparé de chaque thème, au profit d'une production intégrée.

III. INTERNET ET LA COLLECTE DE DONNÉES

19. La collecte de données par Internet ou via le Web peut entraîner de profonds bouleversements dans les autres modes de collecte de données assistée par ordinateur (entretien téléphonique assisté par ordinateur, entretien direct assisté par ordinateur, saisie de données par téléphone à numérotation par tonalité, reconnaissance vocale et autres). Une interface Web conviviale permet la saisie et l'édition de données, mais aussi, entre autres, une mise à jour aisée de ces données, y compris par le répondant lui-même. Elle permet également de réduire les coûts, par exemple en diminuant le nombre d'enquêteurs sur le terrain.

20. Internet ne va pas remplacer les autres méthodes de collecte de données. C'est là un point important à souligner, car il ne serait ni exact ni réaliste de penser que la collecte électronique de données via Internet remplacera les autres moyens à brève échéance. L'objectif des bureaux de statistique doit donc être d'élaborer le meilleur dispositif possible, et non de remplacer tous les procédés par la collecte en ligne. Plus généralement, les bureaux de statistique devraient être responsables de l'ensemble du processus d'élaboration et de mise en œuvre des différentes méthodes et techniques de collecte de données.

21. La collecte en ligne s'effectue par le biais d'un système informatique automatisé qui ne nécessite pas la présence d'un enquêteur : un questionnaire électronique au format HTML est affiché dans la fenêtre d'un logiciel de navigation standard, et les réponses sont envoyées au serveur par Internet. Il faudra développer de nouvelles compétences dans les domaines liés à Internet, la présentation de l'information, l'élaboration de questionnaires et la sécurité.

22. Les enquêtes en ligne sont fréquemment vantées comme étant le moyen le moins coûteux de collecter des données, mais on déplore aussi souvent la baisse de la qualité des données ainsi obtenues, un phénomène qui donne lieu à des résultats non valides. Le rapport entre baisse des

coûts et augmentation du risque d'erreurs dans les enquêtes risque donc de poser davantage de problèmes que dans le cas des autres méthodes de collecte. Il convient de noter que certaines caractéristiques de l'enquête en ligne contribuent à améliorer la qualité des données. Il n'y a pas d'erreur d'acheminement (ordre des questions, omissions, erreurs dans l'arborescence); les données peuvent être vérifiées instantanément (contrôle d'étendue et de cohérence); l'ordre des questions et/ou des réponses est aléatoire; il n'y a pas de phase de saisie séparée; les informations concernant le déroulement de l'"entretien" sont disponibles (par exemple le moment auquel il a eu lieu, sa durée) etc.

23. Par ailleurs, la modicité des coûts des enquêtes en ligne permettra d'utiliser des échantillons plus importants avec, a priori, une variance moindre. Toutefois, dans la plupart des cas, cela ne suffira pas à garantir la qualité des données. La faiblesse la plus fréquente des enquêtes en ligne réside dans une base insuffisante, car toutes les unités cibles n'ont pas accès à Internet. Par exemple, dans le cas des enquêtes auprès des ménages, une telle déficience risque d'avoir un impact considérable sur le contexte général de l'enquête (catégories et strates, niveau d'études, répartition entre les sexes et certaines variables liées à des dispositions psychologiques). Ce phénomène semble moins sérieux dans le cas des enquêtes auprès des entreprises.

24. **Taux de non-réponse.** L'augmentation potentielle du taux de non-réponse est un autre problème. Plusieurs facteurs pourraient en effet influencer les taux de réponse dans les enquêtes en ligne : caractéristiques de l'initiateur de l'enquête, obligation ou non de fournir les données, taille des entreprises, sélection des répondants, activités de suivi, importance du thème traité, disponibilité des données demandées, conception de l'enquête, durée de la collecte des données, et bien d'autres encore. Une constante doit également être soulignée : les enquêtes en ligne sont auto-administrées. L'absence d'enquêteurs et les contacts moins soutenus avec les répondants pourraient bien entraîner une baisse des taux de réponse. Néanmoins, certaines études consacrées aux répondants interrogés via Internet ont montré qu'il était possible, moyennant certaines pondérations, de faire en sorte que le profil de ces derniers soit comparable à celui des personnes qui répondent à des enquêtes par téléphone, du moins en ce qui concerne les variables démographiques; il n'en va pas de même dans le cas de questions sur les dispositions d'esprit. En tout état de cause, cette question mérite d'être étudiée plus en profondeur.

25. **Coopération du public.** Dans de nombreux pays, l'expérience a démontré qu'un des moyens les plus efficaces de collecter des données en ligne avec succès était d'encourager la coopération volontaire du public. Même pour les recensements et les enquêtes considérées comme obligatoires, les refus sont toujours possibles et peuvent occasionner la perte de données. Internet fournit l'occasion d'inciter le public à coopérer davantage à la collecte de données, en mettant à sa disposition un moyen supplémentaire, que certains répondants trouveront plus pratique. Les moyens disponibles sont l'enquête par téléphone, assistée par ordinateur, la télécopie et l'échange électronique de données.

26. **Questionnaires auto-administrés informatisés.** La technologie du questionnaire auto-administré informatisé (CSAQ) est également prometteuse. Les CSAQ accessibles par le Web n'en sont qu'au stade des premiers balbutiements, très souvent en phase expérimentale. Le logiciel idéal reste encore à inventer. Certains spécialistes préconisent l'utilisation de logiciels préexistants pour élaborer des CSAQ et d'autres outils d'enquête assistée par ordinateur, mais il est souvent difficile de trouver le moindre logiciel répondant aux exigences propres de la statistique. L'utilisation du HTML et du Java Script semble adaptée. Ces formats offrent le

double avantage de permettre l'édition en temps réel et d'être pleinement compatibles avec une visualisation sur écran. Il se peut néanmoins que ce type de logiciel, assorti des mesures de sécurité nécessaires, entraîne l'élimination de nombreux répondants potentiels.

27. **Mise en œuvre des CSAQ sur le Web.** Les premières expériences tendent à montrer que la méthode du CSAQ convient particulièrement pour les questionnaires brefs (une page) soumis à intervalles réguliers. En revanche, lors des enquêtes complexes et ponctuelles nécessitant des explications de concepts et un travail de saisie important, le taux de réponse décroît considérablement. L'emploi des technologies existantes (logiciels de navigation courants) a limité les possibilités en matière d'édition, de contrôle multissession et de gestion de la sécurité des interactions. Nous devons veiller à ne pas aller au-delà de ces possibilités lorsque nous commencerons à recourir à la collecte de données en ligne. La confidentialité doit être intégrée dès l'élaboration du questionnaire. Pour renforcer la confiance du répondant dans le fait que ses données seront tenues confidentielles, il faut lui indiquer clairement comment elles seront exploitées, dans quel type d'agrégat (données thématiques, géographiques ou longitudinales) elles seront incorporées, et avec quelles informations elles seront reliées.

28. **Quel est le meilleur domaine pour l'application des techniques de collecte électronique de données ?** Récemment, les enquêtes économiques ont été considérées comme le meilleur domaine de mise en œuvre des nouvelles technologies de collecte de données en ligne, car les entreprises ont plus de chances que les ménages d'être équipées de micro-ordinateurs et autre matériel requis et d'être connectées à Internet. L'utilisation et la présentation des questionnaires en ligne peuvent être très différentes selon qu'il s'agit, par exemple, d'une enquête économique ou d'une enquête démographique. L'enquête économique se caractérise à la fois par la nécessité de communiquer des données historiques au répondant, par la longueur du délai nécessaire à sa réalisation et par la faiblesse relative du risque de voir le taux de réponse altéré par le fait que certains répondants potentiels n'ont pas accès à un ordinateur. Pour sa part, l'enquête démographique nécessite rarement l'envoi de données historiques, mais on peut redouter que la couverture soit incomplète, car de nombreux répondants ne disposent pas d'un ordinateur équipé d'un logiciel de navigation récent. Pour éliminer ce risque, il faut laisser au répondant la possibilité de choisir entre le questionnaire en ligne et le questionnaire sur papier.

29. Les résultats de diverses expériences ont conduit les bureaux de statistique à étudier plusieurs moyens que les entreprises et organisations pourraient être invitées à utiliser pour remplir les questionnaires en ligne. Comme il s'agit là, pour l'essentiel, de travaux de recherche, l'étude des résultats nécessite une coopération étroite entre les bureaux de statistique et les instituts de recherche. Les chercheurs souhaitent souvent quantifier les préoccupations des entreprises à propos de l'utilisation d'Internet comme moyen de transmission des données, mais aussi faire en sorte que les bureaux de statistique ne négligent aucune des exigences particulières. En définitive, l'utilisation et la présentation des questionnaires en ligne peuvent être très différentes selon qu'il s'agit d'une enquête économique ou démographique. Dans le cas des enquêtes avec collecte de données en ligne, un système spécial de sécurisation protégeant la confidentialité des données devra être mis au point.

30. Internet renforcera l'intégration des entreprises lors des opérations de collecte de données au sein des bureaux de statistique. La multiplicité des moyens proposés aux répondants pour communiquer leurs données compliquera la phase de la collecte. Potentiellement, Internet peut

réduire la charge de travail des répondants, à condition de développer une interactivité directe avec les applications informatiques utilisées par les entreprises.

31. **Vie privée, confidentialité et sécurité** vont devenir des enjeux stratégiques qui nécessiteront un débat public, de nouvelles législations et des approches novatrices. Il faudra acquérir de nouvelles compétences dans les applications liées à Internet, à l'élaboration de questionnaires et à la sécurisation. Ces compétences seront également très demandées par le secteur privé et deviendront l'enjeu d'une forte compétition, ce qui pourrait bien entraîner un développement de la sous-traitance.

32. Il importe également que le répondant sache qui a accès à ses données, quels sont les mesures de protection (législation) en place et à quelles réparations il peut prétendre en cas de violation de la confidentialité. Les répondants doivent également être informés du niveau de protection garanti par les dispositifs de sécurisation. Les questions qui se posent sont, notamment : comment les données sont-elles cryptées et qui a accès aux clefs ? À quel endroit les données sont-elles stockées en format lisible et cet endroit est-il sûr ?

33. **La place des métadonnées dans la collecte de données via Internet.** Les métadonnées et la gestion de celles-ci constituent la question centrale. Depuis quelque temps, les objectifs de la R-D sur les métadonnées statistiques sont largement influencés par la multiplication accélérée des groupes de clients et de répondants. En tant qu'instruments d'information, les métadonnées sont au cœur des initiatives prises par plusieurs pays en matière de services administratifs en ligne. La tendance à la normalisation des métadonnées et le renforcement de l'interactivité entre les systèmes nationaux d'information contraignent les statisticiens à adapter leurs solutions à celles retenues par les autres organes administratifs en matière de métadonnées.

34. Pour faire face à cette tendance, un rapprochement des producteurs, des utilisateurs et des archives de métadonnées est indispensable. Mais ce processus est à la fois très exigeant et très coûteux. Souvent, les bureaux nationaux de statistique ne disposent ni des moyens nécessaires ni des spécialistes qualifiés dont ils auraient besoin pour mener leurs propres travaux de recherche. C'est pourquoi la R-D a un rôle essentiel à jouer en la matière. Les tâches les plus importantes consistent à mettre au point une terminologie commune des métadonnées statistiques et à constituer des catalogues et des thésaurus de métadonnées permettant la création de banques de métadonnées sur les entreprises. Les métadonnées, les registres et les répertoires ne doivent pas être confidentiels et doivent pouvoir être partagés le plus largement possible. Ils ne doivent renfermer de données sensibles ou ces dernières doivent être séparées des autres données et consultables uniquement en mode sécurisé.

35. **Qualité des données.** Les effets de la collecte en ligne sur la qualité des données sont mal connus. Il faut donc s'employer plus activement à réaliser des études cognitives et à établir des comparaisons entre les différents modes de collecte. Nous devons élaborer des lignes directrices applicables à la réalisation des CSAQ et, plus généralement, à l'élaboration des interfaces. Par exemple, des réponses devront être apportées aux questions suivantes : comment informer les répondants de l'utilisation qui sera faite des données et des dispositions prises pour assurer la protection des informations confidentielles ? Les utilisateurs auront-ils la possibilité de faire des commentaires, de recevoir des conseils et de consulter des métadonnées au moment où ils

remplissent le questionnaire ? Les répondants pourront-ils interagir dans le cadre de multisections ? Auront-ils accès aux réponses aux questionnaires antérieurs et pourront-ils les modifier ? Il faut évaluer l'effet de telles possibilités de choix sur la qualité des données.

IV. INTERNET ET LA DIFFUSION DE DONNÉES

36. Aucun bouleversement des infrastructures n'a, semble-t-il, eu autant de répercussions que le développement d'Internet et l'augmentation spectaculaire du nombre d'utilisateurs de données qui attendent et exigent aujourd'hui d'avoir pleinement accès à un choix de statistiques pratiquement illimité. Les bureaux de statistique ne sont évidemment pas restés inactifs devant les possibilités et les défis découlant de ce nouveau mode de diffusion de données. Si, au début des années 90, seules quelques statistiques économiques étaient disponibles en ligne sur des panneaux d'affichage électroniques accessibles via une connexion téléphonique, aujourd'hui, tout bureau de statistique digne de ce nom possède son site Web, rempli des chiffres les plus récents. Le développement du World Wide Web a mis en évidence certaines difficultés préexistantes. Le besoin d'exactitude, d'actualité et de fiabilité des chiffres existe depuis longtemps, mais il fait aujourd'hui l'objet d'une attention plus grande que jamais.

37. **Les internautes.** Le public tend non seulement à augmenter, mais aussi à se diversifier fortement. Auparavant, les moyens de diffusion des données s'adressaient le plus souvent à un public doté de certaines compétences en économie et en statistique. Il était fréquent que l'utilisateur connaisse aussi bien le sujet traité que le producteur de statistiques lui-même. De toute évidence, tel n'est plus le cas aujourd'hui. Par exemple, les statistiques économiques intéressent aussi bien les économistes professionnels et les décideurs que les profanes qui s'intéressent à ce sujet ou les élèves qui doivent faire des recherches dans le cadre de leurs études. L'érudition économique et statistique de l'utilisateur ne peut donc plus être tenue pour un fait acquis.

38. De plus, Internet offre la possibilité de boucler la boucle entre les fournisseurs de statistiques et les utilisateurs. En effet, certains répondants seront plus motivés pour remplir les questionnaires dans la mesure où ils auront la possibilité de consulter les statistiques produites à partir de leurs données. Dans le domaine institutionnel (santé, éducation et justice, notamment), il est nécessaire de mettre en place un accès partagé aux microdonnées. Dans ces cas, les répondants sont à la fois fournisseurs et utilisateurs de données.

39. Les bureaux de statistique doivent faire un effort particulier pour comprendre qui sont les internautes, car les besoins de ces derniers peuvent varier considérablement. D'une façon générale, selon les catégories ciblées, les utilisateurs de données en ligne se divisent en deux groupes : les utilisateurs internes (statisticiens responsables de la production de données) et les utilisateurs externes.

40. Les chercheurs spécialisés dans le domaine en question, les responsables politiques, les fonctionnaires, les cadres, les enseignants, les étudiants, les bibliothécaires, les journalistes, etc. sont autant d'utilisateurs externes. Le bureau de statistique doit donc impérativement tenir compte de cette diversité. On pourrait également classer les utilisateurs externes sur la base de leur maîtrise de la statistique et de leur capacité d'interprétation des données. Les groupes suivants pourraient ainsi être définis : les utilisateurs peu au fait de l'analyse statistique (le grand public); les utilisateurs avertis, mais peu enclins à lire les métadonnées et préférant disposer de

compilations prêtes à l'emploi; et les utilisateurs expérimentés capables de rechercher et récupérer l'information, d'évaluer la qualité des données ou de les interpréter, voire de produire eux-mêmes des statistiques.

41. Du fait de la diversité de leurs centres d'intérêt et de leurs niveaux de compétence, mais aussi de l'abondance des données existantes, les utilisateurs devront impérativement avoir la possibilité de gérer leurs consultations en personnalisant l'interface. Quel que soit le degré de pertinence de l'information vis-à-vis d'un groupe donné, tout utilisateur de statistiques en ligne a besoin de disposer de métadonnées pour réaliser les tâches suivantes : prendre connaissance des types de données disponibles, contribuer à la recherche d'informations, interpréter l'information et, le cas échéant, apporter son concours au post-traitement de l'information (téléchargement et réutilisation).

42. **Relations avec les différents acteurs.** Parallèlement à la stratégie de diffusion, le bureau de statistique doit mettre en place une structure d'appui destinée aux utilisateurs. Cette structure doit couvrir les questions techniques et fonctionnelles. L'utilisateur doit être clairement informé du niveau de service auquel il peut s'attendre dans chacun de ces domaines. Les bureaux de statistique doivent impérativement élaborer et gérer un mécanisme permettant l'interaction avec les groupes d'utilisateurs susmentionnés. Autre point important : les appréciations des utilisateurs constituent non seulement une source d'information utile en vue de mieux satisfaire ces mêmes utilisateurs, mais aussi un moyen d'améliorer l'ensemble du processus de production de statistiques. Ces mécanismes pourraient prendre la forme de comités ou conseils consultatifs, de comités fédéraux, provinciaux ou territoriaux, de partenariats bilatéraux avec les principaux ministères fédéraux, et d'instances consultatives regroupant des utilisateurs et des distributeurs. Les besoins des utilisateurs peuvent être évalués dans le cadre d'enquêtes et d'études spécifiques. Alors que les programmes en ligne se développent, il faut trouver des moyens pour mesurer le degré de satisfaction des utilisateurs à l'égard de l'offre en ligne.

43. La façon dont les bureaux de statistique pourraient mettre en œuvre de tels mécanismes interactifs dans un environnement en ligne est illustrée par les exemples ci-après :

- **Activités consultatives.** La création d'un site Web pour tous les groupes consultatifs faciliterait le dialogue et permettrait d'héberger toutes les communications et tous les documents.
- **Programmes et activités de recherche conjoints.** Les programmes et initiatives conjoints de recherche ont un rôle important à jouer dans des domaines tels que la justice, la santé et l'éducation, où plusieurs niveaux d'autorité sont concernés et où il est nécessaire d'échanger des données administratives et d'autres données d'enquêtes. Il faudra créer des extranets ou des réseaux virtuels privés pour permettre l'échange de données en toute sécurité. Il n'existe le plus souvent aucune solution toute faite en matière d'extranets, et ces réseaux doivent être créés sur mesure. Une des solutions techniques les plus prometteuses consiste à utiliser le format XML pour acheminer des messages en temps réel aux administrations partenaires. Cette solution permettrait aux différentes administrations publiques de partager leurs sources et de consulter les principaux registres nationaux (registre de la population, registre des entreprises, etc.).

- Consultation des utilisateurs. Il est nécessaire de consulter les utilisateurs à propos du contenu des enquêtes, mais aussi pour aider le bureau de statistique à mieux développer ses produits et ses services. Des produits et services toujours plus nombreux sont proposés en ligne, ce qui a des répercussions sur les bibliothèques, les universités et les chercheurs indépendants. Via Internet, il est plus que jamais possible d'associer ces utilisateurs à l'élaboration de programmes.
- Formation en ligne. Internet permet de dispenser une formation en ligne destinée à faciliter la diffusion de l'information et à sensibiliser le public.

44. **Réalisation de sites Web.** Le World Wide Web offre un vaste potentiel d'amélioration de la diffusion de statistiques. Il y a fort à parier que, dans le futur, la qualité des services proposés par un bureau de statistique sera souvent jugée d'après la qualité de son site Web. La réalisation d'un site Web a beaucoup de points communs avec l'élaboration d'une enquête. C'est pourquoi les bureaux de statistique doivent s'acquitter de cette tâche avec le plus grand soin. Les directeurs des services informatiques doivent garder à l'esprit le fait que pour réaliser un site Web performant et convivial, il faut des ressources. Ils doivent donc faire appel à un personnel compétent et bien formé pour mettre en œuvre une interaction constructive entre l'homme et l'ordinateur, notamment au travers de tests de convivialité des sites. Les services de spécialistes du génie logiciel sont indispensables. Les bureaux de statistique qui peuvent d'ores et déjà compter sur les services d'équipes d'évaluation sont en mesure de confirmer que, pour se familiariser avec les techniques de préparation de questionnaires d'évaluation ou d'expérimentation sur le terrain, il faut d'abord disposer des ressources et du savoir-faire nécessaires à la mise en œuvre de tests de convivialité des systèmes. Toutefois, l'expérience montre aussi que ces tests peuvent être réalisés progressivement, d'abord auprès d'un petit groupe de fonctionnaires intéressés, puis de façon plus large au fur et à mesure qu'ils prouvent leur utilité. Il existe suffisamment d'ouvrages de référence, de cours de formation spécialisés et de programmes universitaires pour pouvoir commencer.

45. **Entrepôts de données.** La gestion des statistiques pourrait subir des transformations importantes du fait de l'abandon progressif de l'approche thématique de la collecte, de la production et de la diffusion des statistiques au profit d'une approche intégrée basée sur l'utilisation d'Internet. Le développement des bases et entrepôts de données pourrait modifier le caractère traditionnellement thématique du processus de production de statistiques. Contrairement à ce qui était le cas avec l'approche thématique, le développement et la maintenance d'entrepôts de données et de bases de données centralisées ne pourront relever de la responsabilité des services de statistique spécialisés dans un thème précis. Le rôle principal d'un entrepôt de données consiste à réunir différentes données et à les rendre accessibles à des utilisateurs aussi nombreux que possible. S'il est bien conçu, un système d'administration de l'entrepôt de données reposant sur une structure hiérarchisée avec, à son sommet, la direction du bureau de statistique, contribuera très largement au succès de ce processus.

46. Un entrepôt de données offre la possibilité d'accumuler des données au fil du temps et de les présenter à l'utilisateur non plus à intervalles réguliers, mais en continu. Il permet en outre d'effectuer toujours le même arbitrage entre actualité et précision des données. L'utilisateur pourra choisir de consulter rapidement des statistiques préliminaires ou d'attendre que des données en quantité suffisante aient été accumulées pour réduire la marge d'erreur.

47. Le principal problème qui se pose lors de la création d'un entrepôt de données statistiques reste celui de la cohérence des données stockées. Pour remédier à ce problème, il faut mettre au point un système perfectionné de métadonnées et des classifications harmonisées. Bien que des métadonnées statistiques soient de plus en plus utilisées dans la production, l'unification fait encore bien souvent défaut, que ce soit au sein du bureau de statistique ou à l'extérieur. Le souci d'obtenir des métadonnées pertinentes doit être présent durant tout le processus de production, à commencer par la collecte des données. La responsabilité de la création d'un système centralisé de métadonnées doit être assurée par la direction du bureau de statistique.

48. **Les métadonnées et leur diffusion.** Le problème de la cohérence des données doit être considéré au niveau national et au niveau international. Il faudrait que les organisations internationales coordonnent la collecte et les échanges de métadonnées et de données, ainsi que les critères auxquels elles doivent répondre, afin de faciliter ces échanges sans surcharger les pays de demandes identiques émanant d'institutions internationales différentes. Par ailleurs, les organisations internationales doivent fournir aux bureaux nationaux de statistique les instruments leur permettant de diffuser des métadonnées aux utilisateurs. À cette fin, une meilleure intégration des métadonnées et des données entre les organisations internationales est nécessaire.

49. Nous pouvons d'ores et déjà identifier les métadonnées susceptibles d'être mises en commun. De nombreuses métadonnées sont disponibles sur les sites Web des organisations internationales et des bureaux nationaux de statistique. Des liens pourraient être créés à partir des métadonnées présentes sur les sites Web des organisations internationales vers les métadonnées plus détaillées qui se trouvent sur les sites nationaux. L'accès pourrait être coordonné via la création d'un portail unique permettant la consultation des données et des métadonnées. Ce portail pourrait être conçu sur la base du tableau d'affichage des normes de diffusion du Fonds monétaire international, susceptible d'être pris comme point de référence. Tel qu'il se présente actuellement, le tableau d'affichage des normes de diffusion autorise la mise à jour constante des métadonnées et il existe une correspondance exacte entre les métadonnées et les données effectivement diffusées par les abonnés sur leurs sites Web respectifs. Un tel accord a déjà été conclu entre le FMI, l'OCDE, Eurostat et d'autres organisations internationales lors de la Réunion de travail sur les métadonnées statistiques de la CEE/ONU qui s'est tenue à Washington au mois de novembre 2000.

V. OBSERVATIONS FINALES

50. Au vu de toutes les considérations évoquées plus haut, de nombreuses questions restent encore en suspens :

- Internet doit-il être considéré comme le principal moyen de diffusion et de collecte de données des bureaux de statistique ?
- Internet est-il en train de révolutionner toute notre conception de la production de statistiques, ou n'est-il qu'un procédé de collecte et de diffusion parmi d'autres ?
- Internet a-t-il un effet sur la commercialisation des statistiques ?

- Les bureaux de statistique doivent-ils s'efforcer d'instaurer des pratiques en matière de protection des données confidentielles sur Internet ?
- Quel doit être notre rôle dans la promotion de normes internationales de présentation et de description des statistiques ?
- Quelle doit être la place de la coopération internationale dans l'élaboration de lignes directrices, de recommandations et de meilleures pratiques ?
- Internet contribue-t-il à élargir ou, au contraire, à combler, le fossé technologique qui sépare les pays en développement des pays développés ?
- Faut-il renforcer la coopération avec les chercheurs et les universitaires pour mener à bien des tâches nouvelles découlant du rôle croissant d'Internet dans la production de statistiques ?
