



Conseil Economique et Social

Distr.
GENERALE

CES/AC.71/1999/6
11 novembre 1998

ORIGINAL: FRANCAIS
FRANCAIS ET ANGLAIS
SEULEMENT

COMMISSION DE STATISTIQUE et COMMISSION ECONOMIQUE POUR L'EUROPE

CONFERENCE DES STATISTICIENS EUROPEENS

Réunion sur la gestion de la technologie de l'information statistique
(Genève, Suisse, 15-17 février 1999)

Point (i): Les répercussions d'Internet sur le processus de production et de diffusion statistiques

EDITION DE BASES DE DONNEES DANS INTERNET

Rapport envoyé par Statistique Canada¹

I. INTERNET COMME OUTIL DE DIFFUSION

1. Depuis 1995, Internet est devenu un outil de diffusion important pour les bureaux nationaux de la statistique (BNS). Bien que les BNS n'aient pas tous adopté en même temps Internet à des fins de diffusion, cet outil est considéré maintenant comme le principal organe de diffusion pour l'avenir. Comme c'était le cas pour les autres organisations à ce moment-là, les premiers sites Web des BNS visaient à donner un aperçu général aux visiteurs. Par suite des commentaires des clients, l'orientation a changé rapidement, et les sites sont devenus des sites d'information statistique, visant à fournir des statistiques officielles sous diverses formes à une gamme variée de clients.

2. Les avantages d'Internet comme organe de diffusion sont devenus évidents:

- endroit unique (le site Internet du BNS), où toute la gamme de données publiées et diffusées par un BNS est accessible, peu importe le moment

¹ Préparé par Martin Podehl.

ou l'endroit;

- diffusion rapide des données les plus récentes, auxquelles les clients ont accès instantanément;
- possibilité de diffuser de l'information beaucoup plus approfondie que sur papier;
- possibilité de diffuser de l'information avec un contexte beaucoup plus défini, grâce aux hyperliens avec des données connexes, par exemple, détails, notes explicatives, données publiées précédemment, indicateurs de la qualité, méthodologie sous-jacente, etc.;
- réduction des coûts par rapport à la diffusion de publications sur papier, chaque copie additionnelle entraînant des coûts d'impression, de traitement des commandes, d'expédition, de facturation, etc.; dans Internet, les coûts marginaux de l'accès par un client additionnel à des données existantes sont à peu près nuls, tant pour le client que pour le BNS.

3. Aujourd'hui, il est assez évident que les services d'information électronique disponibles par Internet, ou leurs versions futures, se répandront partout dans la société. Il ne faut pas se demander si cela se fera, mais plutôt quand cela se fera. La rapidité de la transition dépend de nombreux facteurs, lesquels semblent évoluer différemment : micro-ordinateurs à la maison; accès à Internet au travail, à l'école ou à la maison; largeur de bande de communication accrue; coût des raccordements Internet; convivialité des fonctions d'accès, de navigation et d'affichage.

4. Même s'il est vrai que par rapport aux publications sur papier, les coûts marginaux d'information d'un client additionnel par Internet sont très faibles, et parfois même à peu près nuls, la mise en place d'un site Internet, de même que l'élaboration et la mise à jour de son contenu, sont coûteuses. De façon plus particulière, au fur et à mesure de l'élargissement du contenu (p. ex., Statistique Canada compte maintenant plus de 60 000 pages dans son site Web), les coûts de la mise à jour et du maintien des pages HTML (Hyper Text Mark-up Language) individuelles prennent de l'importance. Selon les chiffres fournis par l'industrie, il faut actuellement en moyenne de 4 à 5 heures pour créer manuellement et tenir à jour une page HTML. Des méthodes doivent être utilisées pour la création ou la mise à jour de telles pages, de façon dynamique et automatique, à partir d'un ensemble organisé d'information. On appelle cela l'édition de bases de données.

5. Le concept de base de l'édition de bases de données est de faire une distinction entre la mise à jour de l'information sous-jacente et la représentation de son contenu dans les pages HTML. Cela comporte deux avantages:

- au fur et à mesure que de nouveaux renseignements sont ajoutés à la base de données, des pages HTML nouvelles ou mises à jour peuvent être produites automatiquement, sans intervention ni codage manuel;
- en séparant ces deux fonctions, on peut améliorer l'une d'elle sans que

cela ait nécessairement des répercussions sur l'autre.

6. Statistique Canada a adopté le concept de l'édition de bases de données comme principe fondamental de la conception de son service Internet. L'information dans notre site est regroupée en catégories appelées « inventaires de données », chacune correspondant à un ensemble particulier de pages ou de documents de la même nature. Voici des exemples de nos inventaires les plus populaires : *Le Quotidien* : notre bulletin de nouvelles quotidiennes; *Le Canada en statistiques* : un ensemble de tableaux statistiques au sujet du Canada; CANSIM (Système canadien d'information socio-économique) : notre base des séries chronologiques; Commerce : notre base de données détaillées sur les exportations et les importations mensuelles de marchandises; les publications téléchargeables : versions électroniques de nos publications officielles; IPS (Information sur nos produits et services) : notre catalogue de tous les produits et services. Certains de ces inventaires sont de vraies bases de données, dans le sens où l'on entend un SGBD (Système de gestion de base de données). D'autres prennent la forme d'un ensemble organisé de documents ou de pages. Certains de ces inventaires seront décrits de façon plus détaillée ci-après, ainsi que les méthodes d'édition de bases de données utilisées pour les rendre accessibles et les relier entre eux dans notre site Internet.

II. LE QUOTIDIEN

7. L'élément le plus populaire du site Web de Statistique Canada est *Le Quotidien*. Il s'agit d'un véhicule qui sert à communiquer en primeur (et officiellement) les données et publications statistiques de Statistique Canada. Il résume les faits saillants des nouvelles données publiées et indique comment on peut obtenir de l'information plus détaillée. Il comprend le calendrier hebdomadaire et mensuel des principaux communiqués à venir et annonce les nouveaux produits électroniques et services. *Le Quotidien* est publié tous les jours ouvrables à 8 h 30. Il est destiné aux médias, mais revêt aussi un grand intérêt pour les analystes du gouvernement et de l'industrie. *Le Quotidien* est un document très structuré et révisé en profondeur. Les statistiques principales sont résumées à la rubrique « Principaux communiqués », et sont accompagnées de faits saillants, de graphiques statistiques et de tableaux sommaires. D'autres statistiques font l'objet de petits paragraphes de quelques lignes. Par ailleurs, *Le Quotidien* fait référence (au moyen d'hyperliens) à des titres de publications, ainsi qu'à leurs numéros de catalogue et de matrice dans les séries chronologiques CANSIM (voir ci-dessous), lesquelles comprennent plus de détails sur les données émanant de chacun des programmes statistiques.

8. Chaque parution du *Quotidien* est versée dans un répertoire. Cet ensemble de publications individuelles, dont le nombre ne cesse de grandir, fonctionne comme une base de données, c'est-à-dire que l'on peut effectuer des recherches par mot clé dans toutes les anciennes parutions. En outre, des liens à partir d'autres pages de notre site Internet font référence de façon particulière à des numéros du *Quotidien*. D'un point de vue technique, *Le Quotidien* est produit tous les jours comme document SGML (Standard

Generalized Mark-up Language) entièrement codé, lequel est par la suite transposé en une gamme variée de formats de diffusion et de présentation. Il s'agit notamment : de documents HTML sur le Web; de versions imprimées; de messages électroniques aux 1 500 abonnés des listes LISTSERV du *Quotidien*; d'un service de synthétiseur vocal pour les personnes ayant un handicap visuel; ainsi que d'un fichier de texte ASCII que des distributeurs secondaires téléchargent dans leur propre réseau de diffusion. L'édition de bases de données, dans le cas du *Quotidien*, signifie la création d'un document structuré chaque jour (texte, tableaux, graphiques, hyperliens), à partir duquel sont élaborées toutes les versions diffusées, la parution la plus récente étant ajoutée à un répertoire, comme nouvel « enregistrement », pour accès futur.

III. BASE DES SÉRIES CHRONOLOGIQUES CANSIM

9. CANSIM est la base de séries chronologiques en direct de Statistique Canada. Toutes les statistiques socio-économiques principales sont conservées de façon détaillée dans CANSIM, à titre de séries chronologiques comportant des fréquences et des longueurs diverses, certaines remontant à 1914 (p. ex., l'Indice des prix à la consommation, qui paraît sur une base mensuelle). La base de données est mise à jour quotidiennement, et les derniers points de données sont diffusés en même temps que les données sommaires dans *Le Quotidien*. À l'heure actuelle, CANSIM comprend environ 700 000 séries chronologiques. De 1973 à 1996, les données CANSIM n'étaient mises à la disposition du public que par la voie de services commerciaux de bases de données en direct (p. ex., Reuters, Wefa, Datastream, etc.), aux termes de licences accordées par Statistique Canada.

10. En 1996, Statistique Canada a lancé son propre service commercial de diffusion en direct, grâce à une interface entre CANSIM et son site Internet. Cette base de données mise à jour quotidiennement est à la source de deux catégories de services.

- **Accès en direct aux séries chronologiques :** À partir d'une interface programmée en CGI (Système d'interface en infographie) pour les spécifications d'entrée, ainsi que de pages HTML pour la présentation des résultats, les clients consultent le répertoire des métadonnées CANSIM, choisissent la série chronologique qui les intéresse, précisent les paramètres de recherche, paient les droits particuliers d'extraction (prix unitaire fondé sur le nombre de séries demandées) par carte de crédit, grâce à un service de commerce électronique (exploité par un fournisseur de services Internet et une banque), et reçoivent les séries chronologiques dans le format souhaité, à l'écran, ou encore en divers formats, téléchargés dans leur micro-ordinateur. Récemment, on a aussi ajouté une option d'affichage graphique dynamique. Cette interface offre, dans un sens, le même service en direct aux experts de l'analyse. L'innovation dans ce cas vient de la facilité d'utilisation et de la rapidité de réponse par Internet, ainsi que du paiement électronique.

- **Mise à jour des tableaux statistiques dans Internet :** Comme de nombreux autres BNS, Statistique Canada a commencé à publier un aperçu statistique du Canada, des Canadiens et des institutions dans son site Web, à l'intérieur d'un ensemble de tableaux sommaires appelés *Le Canada en statistiques*. Ces tableaux sont regroupés sous quatre grands thèmes : L'économie, Le territoire, La société, L'État. En 1995, on lançait *Le Canada en statistiques* avec environ 100 tableaux. On en compte maintenant 300, et le nombre ne cesse d'augmenter. Chaque tableau présente un sujet particulier et son affichage a été optimisé en fonction de l'écran, c'est-à-dire que l'on évite le défilement dans la mesure du possible. L'ensemble initial de tableaux a été créé manuellement et tenu à jour de la même façon. On s'est vite rendu compte que la mise à jour manuelle ne pourrait se poursuivre compte tenu des ressources limitées consacrées au projet. Étant donné que la plupart des statistiques figurent dans CANSIM, nous avons eu l'idée de mettre à jour les tableaux compris dans *Le Canada en statistiques* de façon automatique à partir de la copie d'interface Internet de la base de données CANSIM. Des modèles de logiciel ont été élaborés pour tous les tableaux dont les données proviennent de CANSIM. Tous les matins, à 8 h 30 précises, un processus utilisant une horloge informatisée recherche les derniers points de données de la base CANSIM, met à jour les tableaux et les affiche dans le site Internet. On utilise le même processus pour le tableau des données économiques et financières, qui est mis à jour quotidiennement et correspond aux données comprises dans le tableau d'affichage des normes de diffusion du Fonds monétaire international.

11. Ce processus de mise à jour des tableaux compris dans *Le Canada en statistiques* constitue un excellent exemple de l'édition de bases de données. Il comporte les avantages suivants:

- aucune intervention humaine n'est nécessaire pour maintenir les tableaux à jour;
- la présentation de tous les tableaux demeure constante;
- l'intégrité des chiffres est assurée, étant donné que ceux-ci sont tirés d'une base de données vérifiée et autorisée;
- les données sont diffusées rapidement et sont toujours à jour.

12. Les tableaux compris dans *Le Canada en statistiques* sont devenus les deuxièmes en popularité parmi les éléments de notre site Web, après *Le Quotidien*. Au moment de la création de ces tableaux, nous avons profité de la caractéristique intrinsèque d'Internet, à savoir les hyperliens entre chaque tableau et des renseignements plus détaillés, par exemple, les séries chronologiques, grâce auxquels un client peut avoir accès à l'ensemble des séries chronologiques dont le tableau est tiré, s'il le souhaite.

IV. CANSIM COMME ENTREPÔT DE DONNÉES POUR LA DIFFUSION

13. Encouragés par le succès obtenu quant à l'utilisation de la base de

données existante CANSIM comme source de données pour l'édition électronique, nous avons entrepris plusieurs projets pour raffermir le rôle de CANSIM à cet égard.

14. **CANSIM II** : Le logiciel sous-jacent de base de données CANSIM fait l'objet de nouveaux travaux de développement (au moyen d'un logiciel SGBDR), afin de pouvoir traiter des tableaux multidimensionnels, et non pas uniquement des séries chronologiques individuelles. Le projet s'appelle CANSIM II. CANSIM II deviendra l'entrepôt pour toutes les données agrégées disponibles dans notre site Internet, et servira de source pour l'accès direct aux données, ainsi que l'édition de bases de données à portée élargie. (Il existe des exceptions, à savoir les données du Recensement de la population et du logement et la base de données existante sur l'exportation/l'importation des marchandises, qui continuent d'avoir leurs propres systèmes de base de données pour le moment.)

15. **Logiciels de navigation multidimensionnelle** : Ces outils logiciels ont récemment vu le jour (p. ex., Beyond 20/20 de la compagnie Ivision Inc.). Ils permettent une navigation souple et pratique dans les tableaux multidimensionnels (cubes), qui sont présentés de façon bidimensionnelle à l'écran. Ils sont des outils puissants d'accès à la base de données structurée de façon souple, et permettent aussi le maintien de la présentation des statistiques en tableaux plats ou en ensembles de séries chronologiques, plus accessibles.

16. **Logiciel de création de tableaux** : La plupart des publications sur papier sont constituées de tableaux. Étant donné que toutes les données qu'ils comprennent sont entreposées dans CANSIM II et que la plupart des publications seront remaniées pour devenir des publications électroniques sur Internet, il est possible de produire des tableaux de façon automatique à partir de CANSIM II (comme c'est déjà le cas avec CANSIM I). Par exemple, la publication mensuelle sur les ventes de véhicules automobiles neufs pourrait être mise à jour de façon automatique, dès que les dernières estimations sont incluses dans CANSIM. Puis, à 8 h 30 le jour de la diffusion des chiffres de ventes de véhicules automobiles neufs dans *Le Quotidien*, les tableaux connexes seraient disponibles dans Internet, avec les détails les plus récents. Le logiciel nécessaire à cette fin est en voie d'élaboration à Statistique Canada, SGML servant de langage structuré pour la conversion des tableaux en HTML, en vue de leur diffusion dans Internet.

17. **Services d'édition personnalisés** : Selon notre expérience récente, les ventes de publication courantes chutent constamment, et on assiste à un accroissement de la demande de services personnalisés. L'édition de bases de données pourrait être utilisée en vue de créer une publication personnalisée, qui comprendrait des tableaux tirés d'une gamme variée de programmes statistiques (enquêtes), les données de ces tableaux provenant toutes de CANSIM II. Du fait que nous disposons déjà d'une interface de commerce électronique dans notre site, les coûts pourraient être facturés aux clients

et remboursés par ces derniers de façon pratique.

V. CATALOGUE ET AUTRES MÉTADONNÉES

18. Statistique Canada tient à jour et publie deux bases de métadonnées, qui sont déjà disponibles dans notre site Internet et qui continueront de l'être.

19. La première base de métadonnées est un catalogue exhaustif de tous les produits et services offerts par Statistique Canada. Un enregistrement dans **IPS (Information sur nos produits et services)** se rapporte à un produit ou à un service particulier et utilise jusqu'à 60 zones pour fournir une description détaillée (p. ex., numéro de catalogue, auteur, résumé, mots clés, prix, personnes-ressources, etc.). Cette base de données, qui comporte environ 6 000 enregistrements dans chaque langue officielle (anglais et français), repose sur SGBD ORACLE, et utilise un serveur de fichier interne. Elle est continuellement mise à jour. Une fois par jour, les changements les plus récents sont chargés dans notre site Internet externe et entreposés dans des pages Web HTML (une page correspondant à un enregistrement). À l'heure actuelle, cet ensemble de pages HTML correspond à la base de données IPS de notre site Internet, que les clients peuvent consulter directement et qui est aussi accessible grâce à des liens à partir d'autres inventaires de données, par exemple, *Le Quotidien*, *Le Canada en statistiques* et CANSIM. Les enregistrements IPS comportent aussi un lien avec la fonction de commande des produits, ces derniers pouvant être livrés sur support électronique (c.-à-d. téléchargés à partir de notre site) ou de façon traditionnelle. À l'avenir, nous avons l'intention d'entreposer les enregistrements IPS dans une base de données SGML de notre site Internet et de produire des pages individuelles HTML, de façon spontanée, toutes les fois qu'un enregistrement IPS est demandé, dans le cadre d'une recherche ou grâce à un hyperlien. Le catalogue IPS est devenu la pierre angulaire de notre site Internet. Les enregistrements qui y figurent comportent des hyperliens avec d'autres inventaires de données. Par exemple, *Le Quotidien* comprend une liste des numéros de catalogue des publications et produits publiés récemment, avec hyperliens IPS. Les enregistrements IPS permettent d'afficher toute l'information pertinente au sujet de chaque publication ou produit et comportent aussi un hyperlien avec des fonctions comme le téléchargement d'une version électronique ou la commande d'une publication sur papier. Tous ces hyperliens sont produits automatiquement, au fur et à mesure que de nouvelles données sont ajoutées à notre site.

20. La deuxième base de métadonnées prend la forme d'une description exhaustive des concepts, définitions, sujets, variables, méthodologies et indicateurs de la qualité de nos programmes statistiques. Cette base de données a été lancée en 1981 sous le nom de **SDDS (Système de documentation des données statistiques)**. Elle a été élargie et améliorée pour devenir la **BMI (Base de métadonnées intégrée)**. Un enregistrement dans cette base se rapporte à un programme source statistique, comme une enquête, un programme d'acquisition de données administratives ou un recensement. Il couvre en outre des programmes statistiques dérivés, par exemple, les divers programmes des Comptes nationaux, qui produisent des statistiques à partir de sources de

données primaires ou secondaires. Chaque enregistrement comporte un numéro d'identification unique (appelé numéro SDDS) et jusqu'à 120 zones dans lesquelles les diverses métadonnées concernant le programme source sont entreposées. À l'automne de 1998, le SDDS/la BMI seront accessibles directement dans notre site Internet (navigation ou recherche par mot clé), et grâce à des hyperliens avec d'autres inventaires de données, au moyen du numéro SDDS. D'ailleurs, les séries chronologiques de CANSIM font déjà référence au programme statistique qui est à la source d'un ensemble de séries chronologiques. Les hyperliens entre le répertoire de données CANSIM et les enregistrements SDDS/BMI permettront aux clients de vérifier la source des séries chronologiques particulières auxquelles ils ont choisi d'avoir accès et qu'ils veulent télécharger.

21. Statistique Canada a toujours eu comme politique de fournir des indications de la qualité des données et de la méthodologie sous-jacente pour toutes les statistiques publiées. Dans le cas des publications sur papier, cela prend généralement la forme d'un chapitre distinct dans chaque publication, ou parfois même d'une publication distincte. Étant donné qu'Internet est devenu l'outil de diffusion d'une proportion de plus en plus grande de nos données, nous avons dû élaborer une approche pratique pour répondre à cette exigence de la politique pour les statistiques diffusées dans Internet. De tels renseignements sur la qualité et la méthodologie sous-jacente doivent être ajoutés à chacun des enregistrements de la BMI. Mais, comme nous l'avons mentionné ci-dessus, ces renseignements existent déjà sous forme textuelle dans de nombreuses publications. Il a donc été décidé de créer, au moins à court terme, des documents distincts dans Internet, à partir des chapitres sur la qualité de toutes les publications, de même que des hyperliens avec les divers inventaires de statistiques de notre site Internet.

Par exemple, le texte diffusé dans *Le Quotidien* pour une enquête particulière peut comporter la mention suivante en conclusion : « Pour plus de renseignements sur la qualité des données, les concepts et la méthodologie, cliquez ici », et un hyperlien permettra de trouver l'enregistrement connexe dans le SDDS/la BMI ou le document comprenant une indication de la qualité. Des liens similaires seront fournis à partir de *Le Canada en statistiques* et de CANSIM, ainsi que des enregistrements du catalogue IPS.

VI. PUBLICATIONS TÉLÉCHARGEABLES

22. Comme c'est le cas pour les autres BNS, Statistique Canada a commencé de convertir ses publications sur papier, en vue de leur diffusion électronique sous forme de documents téléchargeables à partir d'Internet, en formats HTML et PDF (Portable Document Format, particulièrement Adobe/Acrobat). Cette activité ne peut être désignée de façon proprement dite comme de l'édition de bases de données. Toutefois, si l'on perçoit l'ensemble du site Internet comme un genre de « base de données » structurée, chaque publication peut alors être considérée comme un « enregistrement » de l'inventaire des publications, lequel fait partie de la base de données globale d'Internet. Comme c'est le cas pour tous les anciens numéros du *Quotidien*, cet inventaire

des publications peut être consulté par mot clé et grâce à un hyperlien, afin d'établir un lien entre les enregistrements des divers inventaires de notre site Internet.

VII. ENJEUX

23. L'édition de bases de données nécessite l'intervention d'experts pour le développement ponctuel des bases de données, des systèmes et des procédures nécessaires. Dans le cas d'un programme d'édition occasionnelle ou moins fréquente, il peut être plus simple et économique d'utiliser des outils logiciels pour créer manuellement des pages HTML, à partir de textes ou de données figurant dans des tableaux. Les outils de conversion HTML sont maintenant plus faciles à utiliser. Il faut évaluer les avantages et les inconvénients du processus manuel par rapport au processus d'édition de bases de données automatisé dans chaque cas. Par ailleurs, une fois qu'une base de données est établie, de nouvelles possibilités peuvent être exploitées, qui ne pourraient l'être si une telle base n'existait pas.

24. Des procédures strictes d'assurance de la qualité des données doivent être mises en place pour vérifier la précision de l'information avant qu'elle soit entrée dans la base de données. Cela s'applique à la fois aux données et aux métadonnées. Il faut être absolument certain que les données de la base sont « correctes » et que l'édition électronique de bases de données peut se faire sans autre vérification manuelle de la qualité des données. Nous avons eu plusieurs cas de visiteurs Internet qui nous ont mentionné des incohérences réelles ou apparentes dans les tableaux compris dans *Le Canada en statistiques*, qui sont produits automatiquement à partir de CANSIM. Cela a toutefois comme avantage qu'une fois ces erreurs décelées et corrigées dans la base de données, toutes les présentations futures extraites de la base de données sont correctes. (Dans des publications sur papier ayant une large diffusion, de telles erreurs ne pourraient être corrigées.)

25. Étant donné que l'édition sur papier cède de plus en plus de place aux services d'information Internet, le temps exploitable du serveur Internet joue un rôle de premier plan. Si le serveur est en panne, personne n'a accès à l'information. Cela est encore plus grave dans le cas de l'édition de bases de données. En effet, si la base de données n'est pas accessible, rien ne peut être publié.

26. L'extraction des données à partir d'une base et leur présentation finale à l'écran doivent être fondées sur des interfaces robustes et uniformes, afin que les changements touchant la technologie de présentation Internet n'entraînent pas de changement quant à l'interface d'accès aux bases de données. Statistique Canada a eu de bonnes expériences avec SGML à cet égard. Dans la mesure du possible, nous élaborons ces interfaces au moyen de SGML, à titre de format provisoire pour le transfert d'information du niveau base de données au niveau présentation. [1][2]

27. Les changements technologiques se produisent actuellement à une vitesse phénoménale. De nouveaux dispositifs d'accès et de présentation Internet voient constamment le jour, particulièrement sous forme d'unités enfichables.

Évidemment, on doit tirer parti de ces dispositifs facilement accessibles. Par ailleurs, de nombreux clients n'ont pas la plate-forme nécessaire (ordinateurs de 16 bits par rapport à ordinateurs de 32 bits) ou les compétences techniques pour effectuer des chargements complexes, etc. Il faut donc établir un équilibre entre l'innovation dans la conception et les hypothèses conservatrices quant aux compétences et à l'infrastructure dont disposent les clients.

VIII. CONCLUSION

28. Internet a entraîné une modification fondamentale de la façon dont les BNS diffusent les statistiques officielles. Il offre des possibilités d'atteindre davantage de clients, d'offrir davantage d'information plus rapidement, et de réduire en outre le coût de l'ensemble du processus de diffusion à long terme. La seule façon de réaliser des économies c'est d'automatiser le plus grand nombre d'étapes possibles de la chaîne de production statistique, de la collecte des données d'enquête jusqu'à la diffusion aux clients.

29. Dans cette chaîne, un entrepôt de données statistiques publiées ou publiables (données agrégées) jouera un rôle central comme zone de transit : les programmes d'enquête et les autres programmes statistiques intégreront leurs estimations dans l'entrepôt de données; les divers processus de diffusion chercheront les données dans les entrepôts pour les diffuser sous diverses formes et par diverses voies, la principale étant Internet à l'avenir.

30. L'entrepôt de données doit pouvoir comprendre à la fois les estimations réelles exprimées en valeurs numériques, ainsi que des étiquettes, des explications, des indicateurs de la qualité, des notes méthodologiques, etc., liés aux statistiques. Un tel entrepôt de données peut être la source première de l'édition sous forme électronique dans Internet, selon une gamme variée de formats.

REFERENCES

[1] F.E. Hutton, W.M. Podehl: Statistical Systems to Support Analysis in Statistics Canada, Actes de la 49^e séance de l'Institut international de statistique, Florence, Italie, 1993

[2] M. Podehl, M. Parisian: Effective Presentation of Statistics in Electronic Dissemination, Actes de la 51^e séance de l'Institut international de statistique, Istanbul, Turquie, 1997.