

SEMINAIRE

СЕМИНАР

SEMINAR

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE
POUR L'EUROPE



Distr.
GÉNÉRALE

CES/SEM.43/32 (Summary)
25 janvier 2000

CONFÉRENCE DES STATISTICIENS
EUROPÉENS

FRANÇAIS
Original : ANGLAIS

Séminaire sur les systèmes intégrés
d'information statistique et les questions
connexes (ISIS 2000)
(Riga, Lettonie, 29-31 mai 2000)

Thème IV : Amélioration des stratégies de diffusion des données

**FEDSTATS : MISE EN PLACE DE L'INFRASTRUCTURE NATIONALE
D'INFORMATION STATISTIQUE POUR LE XXI^e SIÈCLE
AUX ÉTATS-UNIS**

Communication sollicitée

du Bureau of Justice et du Bureau of the Census des États-Unis¹

RÉSUMÉ

1. En mai 1997, le Conseil interinstitutions chargé de la politique statistique (ICSP) des États-Unis a inauguré un site Web servant de guichet unique pour la consultation de données statistiques fédérales et intitulé "FedStats" (<http://www.fedstats.gov/>). FedStats a largement dépassé son objectif de départ qui était de permettre au public d'avoir aisément accès sur Internet, en passant par un point d'entrée initial, à la large gamme de statistiques fédérales disponibles. Ce premier succès n'est que l'amorce de la conception plus vaste d'une infrastructure nationale d'information statistique à l'intérieur des États-Unis.

2. Contrairement à la plupart des autres pays où une seule entité réunit, analyse et diffuse des informations statistiques, les États-Unis ont mis au point un système statistique décentralisé. Par exemple, différents organismes sont chargés de rassembler et diffuser les informations

¹ Document établi par Marshall DeBerry, Service statistique du Bureau of Justice; Valerie J. Gregg et Rachael LaPorte Taylor, Bureau of the Census.

statistiques sur la santé, le chômage, la démographie et la criminalité. Avec le développement d'Internet, toutefois, l'occasion s'est présentée de donner au public une image plus uniforme du système statistique des États-Unis. FedStats a été créé pour fournir aux utilisateurs une passerelle commune qui leur permette d'accéder rapidement aux informations statistiques qu'ils recherchent.

3. Pour renforcer l'utilité de FedStats, des moyens nouveaux de diffusion intégrée des données ont récemment été mis en œuvre. Avec l'adjonction dans le site d'une nouvelle fonction appelée "MapStats", les utilisateurs peuvent maintenant trouver plus aisément des renseignements statistiques relatifs aux unités géographiques des états ou des comtés composant le territoire des États-Unis. Afin de faciliter la création de cet instrument, on a utilisé un logiciel à source ouverte, comme MySQL et Perl. L'emploi de ces instruments solides et aisément disponibles a considérablement facilité le développement rapide des bases de données et scénarios nécessaires pour fournir aux utilisateurs l'information statistique voulue pour une zone géographique donnée.

4. FedStats continue d'être bien accueilli et exploité à la fois par les médias et par le public. Au cours de la première année, il a enregistré plus de 800 000 visites d'utilisateurs et plus de 1,3 million au cours de la deuxième année. En juillet 1999, Yahoo Internet Life a désigné FedStats comme étant l'un des 50 sites "les plus extraordinairement utiles" pour la deuxième année consécutive. Les 14 organismes membres de l'ICSP fournissent chaque année des ressources pour la conception, la mise au point, la tenue à jour et la gestion du site.

5. Au cours du processus de conception et de mise au point de FedStats, l'équipe spéciale compétente a été confrontée à diverses modalités de diffusion des statistiques fédérales sur les sites Web des différents organismes. L'initiative FedStats présente notamment l'avantage de permettre aux organismes de tirer mutuellement parti des connaissances qu'ils ont acquises, de trouver des solutions communes et de rechercher de manière concertée des réponses aux problèmes que chacun d'eux ne parvient pas nécessairement à résoudre par lui-même.

Information géographique

6. L'une des questions les plus fréquemment posées par les utilisateurs est "Où puis-je trouver des renseignements au sujet de mon état, mon comté ou ma collectivité locale de résidence?". La rubrique "Statistiques régionales" actuellement installée sur FedStats est une passerelle qui permet d'accéder à des renseignements par zone géographique en procédant à une recherche par organisme. Cependant, les utilisateurs veulent avoir une vision exhaustive des renseignements disponibles sur un endroit précis, et pas uniquement auprès d'un seul organisme. Pour cela, il faut procéder à des recherches de grande ampleur par le biais des sites Web de différents organismes. Afin de mieux répondre aux besoins des utilisateurs qui recherchent des statistiques établies sur la base de critères géographiques, une nouvelle fonction appelée "MapStats" a été ajoutée au site. Les utilisateurs peuvent sélectionner un "état" ou un "comté" à partir soit de listes déroulantes, soit d'images au format GIF destinées à faciliter l'"extraction" de la composante géographique recherchée. Une fois qu'ils se trouvent au niveau géographique voulu, les utilisateurs peuvent choisir l'information statistique qui les intéresse, allant de données démographiques relatives à la population, à la criminalité et à l'immigration, à des renseignements d'ordre économique et environnemental. Pour un certain nombre de tableaux, les utilisateurs ont également la possibilité de télécharger l'information sous un format qui se prête à l'analyse ultérieure à l'aide d'un logiciel

statistique ou d'un progiciel tableur. Les améliorations qui seront apportées ultérieurement à ce système permettront i) d'y insérer des couches supplémentaires d'informations géographiques ainsi que d'envisager l'utilisation de technologies nouvelles comme le XML pour faciliter l'interrogation de bases de données statistiques hétérogènes, et ii) d'intégrer des séries de données en vue d'un affichage thématique.

Caractéristiques technologiques

7. Le site Web de FedStats utilise des logiciels à source ouverte pour rechercher l'information et la communiquer aux visiteurs du site. Au nombre des logiciels à source ouverte employés par FedStats figurent le serveur Web Apache Group, la base de données relationnelle MySQL, le langage de programmation Perl et Sendmail, un logiciel de courrier électronique. Le serveur Web Apache est utilisé comme serveur de site et a fonctionné avec un minimum d'interruptions depuis son installation. Pour satisfaire certaines demandes de données, agricoles ou démographiques par exemple, on utilise la base de données MySQL qui sert à la recherche et au stockage de l'information. Le langage Perl est utilisé pour transmettre les demandes d'information à la base de données MySQL et adapter l'information recueillie au format HTML approprié. Il est prévu d'établir un serveur de base de données utilisant Linux, un système d'exploitation à source ouverte. L'équipe spéciale chargée de FedStats continue d'envisager des utilisations nouvelles d'un logiciel à source ouverte pour veiller à ce que le site continue de fonctionner de manière fiable et efficace.

L'infrastructure nationale d'information statistique (NSII) pour le XXI^e siècle

8. La conception envisagée pour la NSII consiste à mettre en place une bibliothèque statistique numérique de portée nationale et à structure répartie dotée d'outils de recherche, d'extraction et de réutilisation des données ainsi que de visualisation de l'information, qui permette la transformation des connaissances en intelligence tout en protégeant la vie privée des répondants et la confidentialité des données. Pour donner une forme concrète à cette conception, la NSII aura besoin d'interfaces utilisateur communes et d'instruments d'accès aux données et de recherche utilisables par des personnes ayant des connaissances informatiques et statistiques plus ou moins poussées afin que les données puissent être utilisées de manière appropriée aux fins d'analyses, à l'intérieur des bases de données et d'une base de données à l'autre.

9. Les sources actuelles décentralisées et autonomes d'informations statistiques ont peu de points communs s'agissant des aspects suivants : concepts et définitions; architectures de systèmes, logiciels et matériels; méthodes de mesure; interfaces; ou modalités de diffusion et de présentation. L'interopérabilité constitue un obstacle important dans divers domaines. Les problèmes d'harmonisation des données sont légion. Des difficultés considérables touchant l'informatique et le calcul de haut niveau ainsi que la mise en réseau à grande échelle empêchent de concrétiser la conception de la NSII.

10. L'équipe spéciale est convaincue que ces questions pourront être traitées dans un proche avenir et que les difficultés que posent l'élaboration et la mise à la disposition du public d'une infrastructure nationale cohérente d'information statistique seront pleinement résolues.

Collaboration de FedStats dans le domaine de la recherche sur l'intégration des données avec le Programme d'adaptation de l'administration à l'ère du numérique de la National Science Foundation

11. L'administration est un utilisateur important de technologies de l'information; elle rassemble et tient à jour de très larges séries de données et fournit des services d'information essentiels et exceptionnels aux particuliers, aux états, aux entreprises et à d'autres clients. L'objectif du Programme d'adaptation de l'administration à l'ère du numérique de la National Science Foundation des États-Unis consiste à financer des études qui puissent à la fois intéresser la communauté des chercheurs en informatique et sciences de l'information et répondre aux besoins des services d'information de l'administration. L'Internet, qui est le résultat d'un partenariat fructueux entre des organismes administratifs et les milieux de la recherche sur les technologies de l'information, constitue un élément et un cadre très stimulants pour ce programme.

12. Dans ce contexte, le Programme d'adaptation de l'administration à l'ère du numérique a pour objet d'appuyer des projets novateurs qui étudient efficacement et dans une vaste perspective les possibilités d'améliorer les activités au sein des organismes, entre les organismes et au niveau intergouvernemental et/ou l'interaction entre les administrations et les citoyens. Ce genre d'études doit permettre d'élaborer et d'utiliser un flux continu de technologies informatiques de pointe en vue de les adopter et de les intégrer dès que possible dans les systèmes d'information administratifs.

13. L'atelier intitulé "Le Gouvernement à l'ère du numérique" tenu en mai 1997 (voir : <http://www.isi.edu/nsf/final.html>) a conduit l'ICSP à autoriser la création dans le cadre de FedStats d'un groupe de travail interinstitutions sur la recherche-développement. Dans un premier temps, le groupe de travail a répertorié les problèmes communs auxquels étaient confrontés de nombreux organismes statistiques et qui pourraient éventuellement être surmontés en appliquant des technologies de l'information de pointe. Le groupe de travail s'est adressé à des spécialistes de la recherche universitaire et industrielle choisis parmi des sociologues et des mathématiciens renommés qui connaissent très bien les difficultés auxquelles est confrontée la communauté des statisticiens. Le groupe de travail a encouragé ces chercheurs à envisager les possibilités de collaboration avec des informaticiens au sein de leurs institutions afin qu'en partenariat avec les organismes statistiques fédéraux, ils puissent élaborer et soumettre des projets à plusieurs programmes de recherche de la National Science Foundation. Dans d'autres cas, ce sont des informaticiens qui ont contacté le groupe de travail de FedStats à propos d'activités potentielles de recherche concertée sur les technologies de l'information. Les personnels des organismes statistiques ont consacré beaucoup de temps à familiariser les chercheurs avec les architectures, systèmes et méthodes de collecte, de traitement et de diffusion de l'information statistique utilisés à l'heure actuelle. Ils ont également fourni des explications sur les problèmes qui font obstacle à la concrétisation de la conception de l'infrastructure nationale d'information statistique. Les chercheurs passent beaucoup de temps à analyser ces difficultés et à identifier les sujets qui pourraient faire l'objet de recherches scientifiques intéressantes. Ensemble, les participants à cette collaboration mettent au point des projets de recherche faisant état d'objectifs, d'échéances, et de n'importe quel service en nature fourni par les partenaires du secteur de l'administration et/ou du secteur industriel.