

SEMINAIRE

СЕМИНАР

SEMINAR

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE
POUR L'EUROPE



CONFÉRENCE DES STATISTICIENS
EUROPÉENS

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

Thème I : Entreposage des données et mise en place
et utilisation des bases de données statistiques dans
un environnement de réseau

Distr.
GÉNÉRALE

CES/SEM.43/10 (Summary)
3 mars 2000

FRANÇAIS
Original : ANGLAIS

**ÉLABORATION D'UN RÉPERTOIRE DE MÉTADONNÉES SUR
LES SOCIÉTÉS AU BUREAU OF THE CENSUS
DES ÉTATS-UNIS D'AMÉRIQUE**

Communication sollicitée
du Bureau of the Census des États-Unis¹

Résumé

1. Le Bureau of the Census (BOC) des États-Unis d'Amérique, comme la plupart des autres organismes réalisant des enquêtes, achète et élabore depuis de nombreuses années des solutions informatiques pour traiter les résultats des enquêtes. De ce fait, le milieu dans lequel s'effectuent ces opérations est constitué d'un grand nombre de solutions disparates dont très peu se recoupent entre elles. Aussi se trouve-t-on aujourd'hui en présence d'un grand nombre de systèmes qui accèdent à leurs propres séries de données ou qui traitent celles-ci en utilisant pour ce faire une documentation spécifique non partagée. D'où un certain nombre de plaintes connexes très fréquentes selon lesquelles :

- Il faut beaucoup de temps pour adapter un fichier utilisé dans un système donné à la structure requise par un autre système.

¹ Document établi par Samuel N. Highsmith et Daniel W. Gillman.

- Le très faible partage de la documentation ou des procédures entraîne une prolifération naturelle de systèmes différents pour résoudre le même problème.
- Les coûts d'organisation d'un nouveau recensement ou d'une nouvelle enquête sont très élevés, dès lors qu'on ne peut explorer les solutions mises au point dans les systèmes précédents.

2. Nous travaillons depuis de nombreuses années à la recherche d'une solution à ce problème fondée sur un modèle générique. Cette solution est aujourd'hui officiellement appelée Répertoire de métadonnées sur les sociétés (CMR) au BOC. Ce répertoire devrait être à même de décrire une enquête ou un recensement tout au long du cycle économique de telle manière que toute application capable de se connecter au Répertoire pourra accéder à toute donnée d'enquête ou de recensement enregistrée dans celui-ci et l'utiliser sur-le-champ. Imaginez que vous puissiez établir tous vos documents d'enquête en une seule fois et qu'ensuite n'importe quel système capable de comprendre cette documentation puisse accéder facilement aux informations relatives à votre enquête et les utiliser sans rien changer à l'application. Tel est notre objectif.

3. Nous décrirons dans le présent document une manière d'y parvenir et même d'aller beaucoup plus loin.

4. Le Répertoire s'appuie sur un gros travail de recherche et de collaboration. La participation du BOC à des travaux menés avec la Suède, le Canada, l'Australie et à l'atelier sur les métadonnées organisé par la CEE-ONU a fait naître l'idée d'élaborer un modèle de données d'entreprise pour le traitement des enquêtes et des recensements. Il existait plusieurs techniques possibles, notamment des solutions entièrement réparties ou centralisées que l'on pouvait appliquer à l'élaboration d'un répertoire de métadonnées. Le Répertoire créé au BOC est destiné à jouer pour les métadonnées le rôle d'un fichier pour une bibliothèque. Il indique l'emplacement et les caractéristiques de toutes les informations que l'on peut y trouver.

5. En collaboration avec un grand nombre de parties prenantes du Census Bureau et sous la direction d'un consultant rompu aux techniques de répertoires de métadonnées et de stockage de données, un modèle de données d'entreprise (BDM) décrivant le traitement des enquêtes et des recensements a été élaboré. Les résultats de ce travail d'équipe ont révélé que nos organismes d'enquêtes économiques et démographiques et de recensement utilisaient des méthodes très similaires. Ces efforts ont permis d'élaborer avec succès un modèle décrivant le traitement des enquêtes et des recensements au BOC.

6. Nous avons parallèlement confié à un organisme privé, Metadata Management Incorporated, le soin d'élaborer un registre officiel d'éléments de données fondé sur la norme ISO/IEC 11179. En combinant ensuite ce registre d'éléments de données avec le modèle de données d'entreprise, nous avons obtenu un modèle relationnel d'entité unique à l'aide d'un instrument dénommé Erwin. Nous pouvions dès lors générer automatiquement une base de données appliquant ce modèle.

7. C'est là qu'est intervenue la chance dont ce projet avait réellement besoin pour démarrer. Les responsables du recensement décennal, à qui on avait confié la tâche de mettre sur pied un système de diffusion des données pour le recensement 2000, ont décidé d'exploiter le modèle CMR plutôt que d'élaborer le leur. Ils ont décidé toutefois d'élargir le champ de notre

modèle pour y ajouter une fonction unique et obtenir les gains d'efficacité recherchés. L'application actuelle, American FactFinder, accessible à partir du site central www.census.gov sur la Toile, repose entièrement sur des métadonnées extraites d'un répertoire fondé sur le modèle CMR.

8. En décembre 1996, la Division de la recherche statistique a décidé de procéder à une application expérimentale du Répertoire CMR envisagé. Les applications avec lesquelles l'interface devait être assurée étaient l'American FactFinder, l'outil de diffusion des données FERRET et le système de gestion de la documentation de la Direction économique.

9. Nous avons utilisé l'application expérimentale afin d'en démontrer l'utilité pour les différents programmes du BOC. Cette opération a débouché sur une série de mémorandums d'accord dont les signataires s'engageaient à poursuivre ensemble l'élaboration du CMR.

10. La Direction économique a alors adopté le CMR pour le stockage des métadonnées requises par son CSAQ électronique. L'idée d'une application pilote était née. Dans le cas de la Direction économique, l'application pilote porterait sur deux éléments du processus de recensement. Ce serait un outil d'introduction de métadonnées à valeur ajoutée qui servirait à la fois pour la collecte et la diffusion des données. L'application pilote à caractère économique a débuté au cours de l'été 1999 et s'est achevée en décembre 1999. Elle a reçu un excellent accueil et les analystes commencent maintenant à introduire les données en utilisant le Répertoire.

11. L'étape suivante consistait à mettre en place une architecture définissant le CMR, toutes les interfaces et tous les auxiliaires que nous envisagions de créer. Pour cela, il fallait essentiellement :

- Concevoir une architecture ouverte;
- Appliquer des normes ouvertes;
- Respecter les impératifs de sécurité du BOC;
- Favoriser l'utilisation de navigateurs pour les applications du CMR fondées sur la Toile;
- Utiliser une solution logicielle intégrée;
- Assurer la compatibilité avec les solutions relatives aux industries naissantes;
- Utiliser les COTS autant que possible, pour faire baisser les coûts et réduire au minimum les adaptations nécessaires;
- Utiliser le logiciel autorisé pour le site du BOC ou, à défaut, un logiciel ayant un gros potentiel d'utilisateurs du BOC, afin de réduire les coûts;
- Prévoir une interface API ouverte;
- Assurer un échange ouvert de métadonnées fondé sur des normes;

- Appuyer les échanges de métadonnées entre le CMR et les autres systèmes et logiciels du BOC;
- Appuyer les structures de métadonnées d'entrée compatibles avec le CMR : XML et AFF;
- Disposer d'un métamodèle de Répertoire CMR extensible qui soit conforme à la norme ISO/IEC 11179;
- Mettre en place un moyen de partager le métamodèle de Répertoire CMR au sein du BOC;
- Disposer d'un moyen d'incorporer des métadonnées non structurées au Répertoire CMR.

12. Le Répertoire CMR, les auxiliaires et les interfaces devraient être achevés et opérationnels au début de 2001. En fait, la Direction économique du BOC les utilise déjà dans l'enquête annuelle sur les industries manufacturières. Le Répertoire offrira une infrastructure capable d'étayer de nombreuses applications concernant la recherche, l'exploitation et la diffusion des données. Une fois enregistrés les ensembles de données et les métadonnées qui les accompagnent, toute application capable de communiquer avec le CMR pourra utiliser pleinement cette série de données et la totalité des métadonnées du Répertoire sans qu'il soit nécessaire de la modifier en quoi que soit. En bref, il suffira de définir l'application une seule fois pour pouvoir l'utiliser indéfiniment.
