

SEMINAIRE

E/CN.E/WP.3/2000/1

SEMINAR

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE
POUR L'EUROPE



Distr.
GÉNÉRALE

CONFÉRENCE DES STATISTICIENS
EUROPÉENS

CES/SEM.43/7 (Summary)
10 février 2000

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

FRANÇAIS
Original : ANGLAIS

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

EXPÉRIENCE DE STATISTICS FINLAND EN MATIÈRE D'ARCHITECTURE DE DONNÉES ET DE DIFFUSION DE DONNÉES SUR LA CYBERTOILE

Communication sollicitée

Présentée par Statistics Finland¹

Résumé

I. Architecture des données

1. Pendant la période 1994-1997, le Service finlandais de statistique Statistics Finland s'est attaché à concevoir et à élaborer le modèle de mise en oeuvre - destiné à répondre aux besoins de sa clientèle - de son projet d'architecture des données dans un environnement PC/Unix. Ce projet reposait sur le système de fichiers unifiés et sur la base de données de classification que Statistics Finland avait intégrés dans son processeur central, ainsi que sur une base de données multidimensionnelle adaptée. Un rang de priorité élevé avait été accordé à la mise au point du traitement des métadonnées. Le modèle de mise en oeuvre comprenait trois volets : les fichiers, les produits statistiques et les métadonnées. Il définissait les modalités selon lesquelles les données brutes d'un système statistique, c'est-à-dire l'un des trois niveaux hiérarchiques de fichiers (fichiers distincts, fichiers unifiés et bases de données multidimensionnelles), étaient transformées en produits statistiques.

¹ Établie par Sven Björkqvist et Pirjo Toivonen.

2. Le projet a permis de construire un système de fichiers unifiés avec sa base de données descriptive et la base de données de classification, et de les raccorder. On s'est également efforcé de mettre au point une nouvelle version de base de données de sortie multidimensionnelle devant remplacer la base de données en ligne utilisée depuis les années 80 dans le processeur central. Ces efforts n'ayant pas eu les résultats escomptés faute de ressources, il a été décidé, au début de 1998, de mettre un terme au projet.

3. En mars 1998, Statistics Finland a commencé à rechercher un nouvel entrepôt de données permettant de stocker des données agrégées à usage interne et de diffuser des données sur la cybertaine et sur CD-ROM. Un projet a été lancé pour évaluer les options et proposer un modèle de production. La solution proposée était un entrepôt central où les données agrégées étaient stockées sous forme de tableaux. À l'automne 1998, le système de données de sortie StatLine de Statistics Netherlands a finalement été choisi pour mettre en oeuvre ce modèle de production.

4. En 1998, Statistics Finland a aussi commencé à mettre au point la base de données conceptuelle, qui, à l'image du système StatLine, s'intègre bien à l'architecture de données que le Service avait précédemment définie. Cette architecture décrit la manière dont les données devraient migrer des bases de données de production vers les bases de données de sortie. À l'heure actuelle, le modèle de mise en oeuvre comprend le système des fichiers unifiés avec sa base descriptive, la base de classification, la base conceptuelle, les bases de données de sortie StatLine ainsi que les liens entre toutes ces bases. À Statistics Finland, il est désormais possible et recommandé de produire directement à partir des fichiers unifiés les tableaux StatLine avec les métadonnées nécessaires.

II. Diffusion de données sur la cybertaine

5. Au début de 1999, Statistics Finland a lancé un projet visant à mettre en oeuvre le système de données de sortie StatLine. Grâce aux activités préparatoires qui avaient été entreprises, le projet a pu démarrer facilement et rapidement. Il devait être achevé le 31 décembre 1999. Plusieurs étapes étaient prévues : la mise en place et l'exploitation d'un service interne fournissant des données utilisables dès janvier 1999, puis l'ouverture au public d'un service gratuit diffusant une grande quantité de données dès mai 1999 et d'un service payant dès le début de l'an 2000.

6. Comme le système StatLine a été réalisé aux Pays-Bas, il a fallu l'adapter aux conditions de production de Statistics Finland. Les principaux objectifs du projet étaient notamment : de déterminer la séquence d'instructions que Statistics Finland devait utiliser (définition des concepts, traduction des programmes et des manuels); de prévoir la structure de l'entrepôt interne de données et du service proposé sur la cybertaine; d'aider et de conseiller les départements de statistique quant au contenu des données à fournir au système. Il fallait également adapter les autres outils de production de Statistics Finland pour les intégrer sans problème au système StatLine, définir les meilleures modalités d'utilisation de l'entrepôt de données et du service d'information dans des situations différentes, ainsi qu'élaborer les instructions et les manuels nécessaires. À ces fins, les activités à exécuter ont été réparties en trois grands domaines, à savoir la création de l'infrastructure technique, la définition des concepts et l'entretien du service, et la communication.

7. Le projet a bénéficié du soutien sans réserve de la direction de Statistics Finland et a reçu pendant sa durée, à savoir un an, un budget de 1,5 million de markkaa (environ 254 000 euros). Comme il a eu des incidences sur l'ensemble des activités du Service, il a également occasionné de très nombreux frais dans les autres départements, si bien que les dépenses du groupe de travail institué n'ont représenté qu'une partie du coût total. Pendant les quatre premiers mois, le groupe de travail était constitué de six personnes (dont certaines travaillaient à temps partiel), puis deux autres personnes ont été embauchées.

8. Le passage d'un service en ligne de données de sortie centralisées à un service en ligne sur la cybertoile de données réparties, mais encore administrées de manière centralisée ne s'est pas fait facilement - d'autant que cela ne devait prendre qu'un an. Il était donc essentiel que la direction soutienne le projet. Le groupe d'intérêt spécial, qui était composé de quarante personnes choisies dans les départements de statistique, ont joué un rôle essentiel pour promouvoir le système au sein du Service - sans leur contribution, le projet aurait très probablement échoué. En outre, la formation d'un nombre suffisant de fournisseurs de données a permis de mieux alimenter l'entrepôt. Il était également essentiel d'adapter le nouveau système de diffusion aux instruments existants, afin de réduire le volume de travail que représentait l'entrée des données dans le système.

9. La diffusion en ligne auprès du public exige que les données soient expliquées et décrites dans le détail afin d'éviter les malentendus et de rendre les données utilisables par les professionnels. Ces explications et descriptions devraient émaner de systèmes centralisés de métadonnées comme le système de fichiers unifiés de Statistics Finland. Sinon, une nouvelle description des données à publier représenterait un fardeau énorme pour les départements de statistique.

10. Le projet a été une réussite : Statistics Finland est désormais doté d'un entrepôt interne de données de référence, d'un service statistique accessible au public sur la cybertoile et d'une infrastructure technique permettant de mettre au point des services payants grâce au système StatLine. Statistics Finland a également formé plus d'une centaine de personnes à l'utilisation du système, afin qu'elles produisent des données avec des métadonnées pour les bases. Le public peut accéder à une centaine de tableaux totalisant jusqu'à 30 millions de cases (au 15 novembre 1999), mais ce nombre augmente rapidement. Les réactions des clients sont positives et d'autres producteurs officiels de statistiques en Finlande manifestent un vif intérêt pour le système. La diffusion gratuite d'une grande quantité de données a fortement contribué à produire l'image dynamique d'un fournisseur de données dans la société de l'information.

11. Le système StatLine est actuellement utilisé pour produire quotidiennement des données, mais il est encore nécessaire de développer, d'améliorer et de normaliser les interfaces et les flux de données et de métadonnées, afin de l'intégrer sans problème aux processus de production de Statistics Finland. Cela représente un défi important pour les années qui viennent, de même que la tâche de répondre au flux constant d'informations en retour provenant des utilisateurs. La répartition des activités de coordination n'est pas toujours une bonne chose - il serait nécessaire de disposer au moins d'un comité de rédaction pour coordonner le contenu et la structure du service proposé. La centralisation totale des activités de coordination n'est pas non plus la solution car elle implique une moindre participation et semble confier l'alimentation et la mise à jour de la base de données à l'organe central de coordination. La coordination du contenu des bases est donc un domaine dans lequel beaucoup reste à faire.
