



**Conseil Economique  
et Social**

Distr.  
GENERALE

CES/AC.71/1999/4 (SUM)  
25 novembre 1998

FRANÇAIS  
Original : ANGLAIS

**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 iii) : Intégration des activités statistiques aux niveaux national et international, notamment stratégies de modélisation des données et normes nécessaires pour l'intégration des données statistiques

**UNE ARCHITECTURE DES SYSTÈMES D'INFORMATION DESTINÉE AUX ORGANISMES  
NATIONAUX ET INTERNATIONAUX DE STATISTIQUE**

Communication de Statistics Sweden 1/

**COMMUNICATION SOLLICITÉE**

**RÉSUMÉ**

1. Ce document est un rapport méthodologique fondé sur la longue expérience que l'auteur a acquise en ce qui concerne la conception d'architectures de systèmes d'information en statistique (SIS) destinées aux organismes nationaux et internationaux de statistique. Il a pour objet d'aider les services de statistique à concevoir des architectures efficaces tenant compte des exigences croissantes des utilisateurs, du renforcement de la coopération internationale et de l'évolution permanente des technologies de l'information.

2. Ce rapport analyse en profondeur les types de systèmes d'information existant dans les services nationaux et internationaux de statistique, et donne un aperçu des orientations futures. L'auteur examine les aspects techniques de l'architecture proposée des SIS, et formule des recommandations pratiques sur les modalités de mise en oeuvre d'une stratégie réaliste en

---

1/ Elaborée par Bo Sundgren.

matière de technologie de l'information compte tenu de l'évolution constante et rapide de cette technologie.

3. Dans le premier chapitre, l'auteur introduit les concepts fondamentaux utilisés dans le rapport : les différents types d'opérations et de données statistiques, les applications et l'infrastructure statistiques ainsi que le flux de données et de métadonnées qui intervient au cours d'une enquête.

4. Un système d'information en statistique est composé de sous-systèmes (applications) qui permettent de recueillir, de stocker, de récupérer, d'analyser et de diffuser des données statistiques. L'architecture des systèmes d'information (ASI) d'un service de statistique est un cadre commun dans lequel les différents sous-systèmes ont, d'une part, un rôle qui leur est propre et, d'autre part, des relations réciproques. Le document analyse aussi certains rapports existant entre l'ASI et l'architecture organisationnelle d'un service de statistique, et il décrit certaines particularités d'une ASI en statistique.

5. L'ASI doit correspondre aux objectifs et aux tâches du service de statistique. Les décalages entre l'ASI et l'architecture organisationnelle existante des services de statistique peuvent notamment être dus au fait que le mode d'organisation traditionnel de ces services s'articulant autour des enquêtes sectorielles est inadapté aux besoins intersectoriels en information des utilisateurs de statistiques. Puisque les enquêtes dépendent très souvent de la collecte de données, l'organisation des services de statistique est aussi "orientée saisie". Il en résulte que la coordination et le contrôle intersectoriels souhaitables sont difficilement réalisables.

6. Certains services de statistique ont créé des unités spéciales destinées à pourvoir aux besoins de catégories particulières d'utilisateurs. Au vu de l'évolution de la technologie, il pourrait s'avérer plus pratique de mettre sur pied un centre d'échange unique conçu en fonction des utilisateurs et disposant d'une infrastructure souple et évolutive.

7. Le deuxième chapitre analyse de façon détaillée les principaux types suivants de SIS :

- systèmes de traitement de l'enquête,
- systèmes de centre d'échange, "entrepôts de données",
- registres,
- systèmes de traitement analytique.

8. L'auteur examine les tâches, les fonctions et les exigences de chacun de ces principaux systèmes d'information. Il accorde une attention toute particulière aux systèmes de traitement de l'enquête en prenant en compte l'ensemble des phases d'une enquête statistique : son organisation, son exécution et son évaluation.

9. Au cours de la phase d'organisation, les concepteurs de l'enquête décident des objectifs fondamentaux et des principaux utilisateurs de l'enquête, des données d'entrée et résultats qu'il est essentiel d'obtenir, et des opérations à effectuer pour recueillir ces données et les transformer en résultats. Il est utile que les concepteurs d'une enquête statistique puissent accéder à une base de connaissances contenant des informations sur la

conception d'enquêtes analogues ou apparentées. Afin de pouvoir tirer des enseignements de l'expérience acquise, toutes les informations importantes sur la conception des enquêtes statistiques doivent être documentées. Les métadonnées sur la qualité et le contenu des données et les opérations représentent une part très importante de cette documentation.

10. La phase d'exploitation de l'enquête comprend principalement les opérations suivantes : création de la base de sondage, échantillonnage, mesure, préparation des données (saisie, codage et édition des données), création du registre des observations, évaluation, analyse, et présentation et diffusion des résultats. La procédure d'évaluation est souvent combinée avec une analyse orientée vers la production de façon à améliorer la qualité et l'efficacité des futures enquêtes.

11. Les résultats de l'enquête doivent être mis à la disposition des utilisateurs sous une forme convenant à ceux-ci et par des moyens de distribution appropriés. En principe, le cycle de production d'une enquête est terminé lorsque les résultats de l'enquête ont été publiés. On peut observer une tendance à inclure la diffusion électronique des résultats parmi les moyens de publication. Les centres d'échange et Internet pourraient être des éléments importants du nouveau système de publication.

12. L'évaluation de l'enquête consiste à vérifier et à évaluer si les produits finals indiqués ont été fournis, si les résultats ont été publiés et leur publication annoncée comme il convient, et si les métadonnées sont documentées et stockées, ainsi qu'à évaluer les métadonnées orientées vers la production et la réaction des utilisateurs.

13. Le registre a pour fonction de contenir des informations mises à jour sur toutes les entités qui composent une population. Les registres peuvent servir de bases d'échantillonnage pour les enquêtes. Outre qu'il reflète l'état actuel de la population, le registre doit permettre de reconstruire l'ensemble de la population d'entités à tout moment, et de reproduire l'état initial et tous les événements qui ont eu une incidence sur les entités. Les registres contenant des métadonnées, telles que les définitions, les liens avec les enquêtes et les ensembles de données, les formats normalisés, les séries de valeurs, etc., sont des registres d'un type particulier.

14. Le centre d'échange a pour fonction de faciliter l'échange de données et de métadonnées relevant de différents types d'enquêtes, de registres ou d'analyses, y compris avec les utilisateurs externes. Une autre dénomination en est "entrepôt de données". Le centre d'échange reçoit et fournit des données et des métadonnées conformes à des formats normalisés spécifiques, selon des modalités précises de livraison.

15. Outre l'analyse susmentionnée, orientée vers la production, les services de statistique effectuent des analyses qui sont davantage orientées vers l'utilisateur. Lorsqu'une telle analyse nécessite des données issues de plusieurs enquêtes statistiques ou d'autres sources, le système de traitement analytique peut être considéré comme un système distinct possédant des interfaces avec les systèmes de production des enquêtes.

16. Le troisième chapitre décrit une future architecture des systèmes d'information pour un organisme de statistique. Cette architecture est fondée sur les quatre principaux types de systèmes d'information considérés dans le

précédent chapitre. L'entrepôt de données de l'organisme, englobant le centre d'échange et les registres dans leur ensemble, est un composant important de l'architecture proposée.

17. Le futur entrepôt de données d'un organisme de statistique comprend les cinq compartiments suivants :

- données et métadonnées brutes;
- registres des observations finales;
- statistiques multidimensionnelles finales;
- documents électroniques;
- métadonnées globales (registres compris).

18. Les données et les métadonnées figurant dans le compartiment des données brutes ne se présenteront pas toujours sous un format normalisé. Des logiciels généraux permettant de les normaliser doivent être prévus. En outre, il faut prévoir des outils logiciels généraux prenant en charge toutes les opérations et les sous-opérations importantes des systèmes de dépouillement des enquêtes et des systèmes d'analyse.

19. Dans le cas des organismes internationaux, la plupart des pays membres fournissent les données et au moins certaines métadonnées par voie électronique. Mais, même en étant acheminées de cette manière, les données peuvent se présenter sous de nombreux formats différents. Une première étape consistera donc à normaliser les données et les métadonnées reçues. Cette étape peut être évitée si les pays acceptent de fournir les données et les métadonnées conformément à une norme internationale, par exemple la norme EDIFACT pour les messages statistiques génériques (GESMES). Il est important de noter que le format normalisé doit comprendre des normes tant pour les données que pour les métadonnées.

20. Le chapitre 4 analyse les aspects techniques de l'architecture proposée. Comme les systèmes d'information en statistique doivent répondre à la demande d'utilisateurs très différents ayant des besoins divers et parfois contradictoires, la souplesse est un élément particulièrement important pour l'ensemble des composants matériel, logiciels et relatifs aux données.

21. Un service de statistique utilise très souvent un nombre relativement élevé d'applications statistiques différentes. Toutefois, beaucoup parmi elles sont assez semblables en ce sens qu'elles effectuent un nombre limité de tâches qui sont typiques des systèmes d'information utilisés au cours des diverses phases des enquêtes statistiques, à savoir l'organisation, l'exécution et l'évaluation des enquêtes. Lorsqu'on analyse une tâche statistique principale ou secondaire, on observe qu'à un certain stade on atteint un niveau où les composants logiciels ne doivent plus nécessairement correspondre aux besoins spécifiques des applications statistiques. On peut utiliser à leur place des composants logiciels normalisés généraux. Aujourd'hui, ces composants de "niveau général" peuvent occuper une position assez élevée dans l'architecture des systèmes d'une application statistique.

22. Le même principe qui consiste à privilégier des composants normalisés et réutilisables s'applique au matériel. L'ordinateur personnel compatible IBM est depuis longtemps devenu un composant matériel normalisé de fait pour les organismes de statistique ainsi que pour les utilisateurs de leurs données et

leurs clients.

23. Les données incluses dans un système d'information sont stockées soit sur des supports faisant partie intégrante du système logiciel de l'application, soit dans des bases de données ou fichiers distincts. L'indépendance programme/données est une exigence importante, en ce sens que les composants "logiciels" et "données" d'un système d'information peuvent être élaborés et gérés de manière assez autonome les uns par rapport aux autres. Une modification du contenu, de la structure ou du stockage des données ne doit pas nécessiter des modifications des programmes utilisant ces données. Par ailleurs, il doit être possible de modifier ou d'ajouter des composants logiciels sans devoir redéfinir les composants relatifs aux données.

24. L'analyse de différentes architectures de systèmes d'information conduit à proposer une architecture de réseau à plusieurs niveaux qui permette d'équilibrer les besoins en matière de centralisation et de décentralisation dans un organisme de statistique moderne.

25. Le chapitre 5 est axé sur les aspects concernant la mise en oeuvre de l'architecture informatique proposée. La technologie de l'information évoluant rapidement, on a besoin d'un plan réaliste de mise en oeuvre permettant de tenir compte des modifications qui surviendraient au cours de la procédure de mise en oeuvre elle-même. Les quelques principes suivants sont à prendre en compte :

i) puisque le rapport qualité/prix des matériel et logiciels ne fait que s'améliorer, il vaut mieux acheter des composants standard plutôt que développer ses propres solutions, et investir davantage dans la capacité du matériel plutôt que compliquer une solution logicielle simple;

ii) il est plus sûr d'effectuer la normalisation au niveau des interfaces entre les composants plutôt que des composants eux-mêmes;

iii) au lieu d'attendre que les normes, matériel et logiciels se perfectionnent, on achètera les composants matériel et logiciels de pointe disponibles au moment considéré et on remplacera un composant par un meilleur dès que possible, sans devoir changer les autres;

iv) on aura une image nette de l'architecture globale des systèmes d'information de l'organisme et on définira un certain nombre d'interfaces importantes sur le plan stratégique;

v) les projets de développement ne porteront pas sur plus de deux ans, les projets complexes devraient être subdivisés en projets subsidiaires dont les résultats et les échéances seraient clairement définis, et les objectifs trop nombreux et trop ambitieux devraient être évités;

vi) lors du passage à une nouvelle plate-forme technique, on pourra profiter de l'occasion pour améliorer en même temps le contenu et la qualité des statistiques, mais seulement dans la mesure où ces activités ne remettent pas en cause le calendrier du projet; suivant l'importance réelle de ces améliorations, on pourrait admettre certaines insuffisances et y remédier à long terme en procédant à une amélioration durable.

26. Le personnel de direction des organismes de statistique joue un rôle essentiel dans ce contexte. Toutefois, il a besoin du soutien des statisticiens travaillant dans les différents secteurs de la statistique. Le projet devrait être axé sur les tâches statistiques, les technologies de l'information étant un outil important dont on dispose pour le mener à bien. Il faudrait repérer les possibilités d'amélioration de la coordination statistique et les exploiter activement, notamment par le biais du composant relatif aux métadonnées globales de l'entrepôt de données.

-----