

**Commission économique pour l'Europe**

Conférence des statisticiens européens

Soixante-troisième réunion plénière

Genève, 15-17 juin 2015

Point 4 de l'ordre du jour provisoire

**Modernisation de la production et des services statistiques
et gestion à des fins d'efficacité****Production statistique: passage d'une démarche axée
sur les domaines à une démarche axée sur les processus****Note de l'Office statistique de la République de Slovénie***Résumé*

L'Office statistique de la République de Slovénie a commencé à moderniser la production statistique il y a quelques années en lançant un grand projet d'infrastructure. Ce projet se fondait sur le «Modèle générique du processus de production statistique» et sur quelques concepts fondamentaux qui devaient permettre d'élaborer des solutions génériques mais suffisamment souples pour répondre aux besoins de la majorité des enquêtes statistiques. Plutôt que d'élaborer une solution globale pour l'ensemble du processus statistique, l'Office a choisi d'adopter une approche par étapes. Des solutions génériques pour certaines parties du processus, par exemple l'édition des données, l'agrégation des données ou l'estimation normalisée, seront progressivement mises au point et introduites dans la production statistique. La présente note décrit les grands principes sur lesquels repose le projet, rend compte des résultats obtenus jusqu'ici et donne un aperçu des plans de développement ultérieur.

La présente note est soumise pour examen à la 1^{re} séance du séminaire de la Conférence des statisticiens européens organisé sur le thème «Modernisation de la production et des services statistiques et gestion à des fins d'efficacité».



I. Introduction

1. La réalisation d'enquêtes statistiques est une tâche difficile qui demande beaucoup de temps et qui est par conséquent assez coûteuse. C'est tout particulièrement le cas pour les enquêtes de statistiques officielles, où la fiabilité des résultats est d'une importance particulière. Il est vrai que la plupart des enquêtes statistiques officielles portent sur des phénomènes très complexes qui ne sont pas faciles à observer, ce qui rend la tâche des statisticiens encore plus difficile. D'autre part, en raison de la pression constante exercée en faveur de coupes budgétaires, les statisticiens officiels doivent faire face de plus en plus souvent au défi consistant à produire des statistiques de haute qualité avec des ressources considérablement réduites.

2. Le développement rapide des technologies de l'information (TI) ces dernières années offre la possibilité de reconfigurer le processus d'enquête d'une façon tout à fait nouvelle et plus efficace. On constate très clairement que la plupart des bureaux nationaux de statistique ont récemment déployé de gros efforts pour reconfigurer leurs activités au niveau institutionnel. Des expressions telles que «réorganisation, modernisation, nouvelle architecture», etc., reviennent très fréquemment dans les documents stratégiques des bureaux nationaux de statistique. L'objectif final qui est généralement présenté dans ces documents est d'utiliser la large gamme de nouveaux outils et de nouvelles applications informatiques pour rendre l'ensemble du cycle de production plus léger et moins onéreux.

3. L'Office statistique de la République de Slovénie (ci-après dénommé «OSRS») ne fait pas exception à cette règle. Les efforts visant à mettre en place un nouveau système plus moderne de traitement des données remontent déjà à plusieurs années. Il y a environ dix ans, l'OSRS a commencé à planifier la modernisation de ses processus en vue de mettre en place des systèmes pleinement intégrés de production statistique. La mise en œuvre d'un système de ce type n'a pas fonctionné comme prévu, de sorte que la démarche a peu à peu évolué. L'OSRS s'est tourné vers la mise au point de plusieurs solutions génériques autonomes, qui seraient ensuite appliquées à un processus statistique donné pour chaque enquête.

4. L'OSRS réalise actuellement un grand projet d'infrastructure visant à mettre en pratique les principes de modernisation dans le cadre du processus habituel de production statistique. Des solutions génériques pour certaines parties du processus, par exemple l'édition des données, l'agrégation des données ou l'estimation normalisée, seront progressivement mises au point et introduites dans les enquêtes statistiques. La présente note décrit les grands principes sur lesquels repose le projet, rend compte des résultats obtenus jusqu'ici, et donne un aperçu des plans de développement ultérieur.

II. Du système pleinement intégré à la solution modulaire

5. En 2007, l'OSRS a lancé un vaste projet visant à créer un système complexe de production statistique pleinement intégré applicable à l'ensemble du cycle de production pour la grande majorité des enquêtes statistiques. Ce système devait comprendre, entre autres choses, un système générique piloté par les métadonnées pour l'ensemble du cycle de traitement des données. S'agissant de ce système générique, à la fin du projet, l'OSRS est arrivé à la conclusion que les résultats ne répondaient que partiellement aux attentes.

6. Le projet n'a pas atteint son objectif qui était de mettre en place un système pleinement intégré reliant tous les outils particuliers qui avaient été mis au point au cours du projet. La question qui s'est donc posée après l'analyse des résultats du projet était la suivante: ce type de système pleinement intégré est-il vraiment la meilleure orientation

de développement? Autrement dit, ce type de système intégré permet-il véritablement d'assurer la modernisation vers un système de production statistique souple et d'un bon rapport coûts-efficacité? Toutes les discussions et les réflexions ont finalement abouti à renoncer à l'idée d'un système pleinement intégré au profit de solutions légèrement différentes conservant certaines caractéristiques importantes de «l'ancien système» (également appelé système piloté par les métadonnées). Il s'agissait de recourir à un certain degré de fragmentation de façon à favoriser une plus grande souplesse tout en conservant certains éléments généraux du système

7. Le principal changement d'approche a été la décision de l'OSRS de diviser le processus statistique en un ensemble de sous-processus de moindre ampleur et de mettre au point des solutions modulaires pour chacun de ces sous-processus. Ces solutions modulaires devaient être conçues de façon à pouvoir relier de manière aisée et souple les intrants et les résultats des différentes composantes de l'ensemble du processus statistique. Ces composantes (aussi appelées modules) devaient fournir des solutions logicielles génériques pour certaines parties du processus de production statistique et devaient être conçues de façon à pouvoir fonctionner de manière autonome. Les principales caractéristiques de ces modules peuvent être résumées de la manière suivante:

a) Les modules ont été conçus sur la base de principes méthodologiques harmonisés, transparents et largement acceptés qui ont été définis avant la mise au point effective du module concerné;

b) Les modules doivent être suffisamment ouverts pour pouvoir être intégrés et mis en fonction dans différentes bases de données et dans différents environnements (par exemple, Oracle, SAS (*Statistical Analysis Software*)) à condition que ces bases de données soient conformes à certaines règles fondamentales relatives à l'organisation des données;

c) Les modules ont été conçus comme des systèmes entièrement pilotés par les métadonnées, ce qui signifie que les informations déterminant les paramètres d'exécution d'une enquête particulière et d'une période de référence spécifique doivent pouvoir être communiquées en dehors du programme informatique de base utilisé par l'ordinateur. Les informations se rapportant à l'exécution d'une enquête particulière ne doivent pas être insérées dans le programme informatique général mais doivent être fournies par le personnel spécialisé au moyen des tables spéciales de métadonnées;

d) Les métadonnées de processus peuvent aussi être fournies dans différentes bases de données pour chaque enquête dans différents environnements, mais chacune de ces bases de données (de métadonnées) doit se conformer à des règles strictes relatives à sa structure (tables et variables).

8. Les solutions modulaires génériques ont été élaborées en deux étapes. La première version du nouveau système généralisé a été conçue entre 2007 et 2010. Lors de cette phase initiale, l'objectif principal était d'élaborer des programmes génériques fondés sur le logiciel SAS, à même d'effectuer une partie du processus pour différentes enquêtes sans modifier le logiciel SAS lui-même, mais simplement en adaptant les métadonnées de processus (règles de procédure) introduites dans le programme général. Le plus important ici était le changement majeur dans la manière dont une partie spécifique du processus statistique était réalisée.

9. Avant de mettre en place ce nouveau système, l'OSRS appliquait «l'approche cloisonnée classique de la production», dans laquelle les outils de production et les solutions étaient «fonction de l'enquête». La manière dont les opérations statistiques (par exemple, l'édition et les imputations, l'estimation des erreurs d'échantillonnage, le contrôle de la divulgation de données statistiques et la tabulation) étaient organisées dépendait beaucoup de l'équipe chargée de l'enquête. L'approche tributaire de l'enquête était particulièrement efficace pour la mise au point de solutions logicielles. Ces solutions

étaient élaborées le plus souvent en fonction des besoins d'une enquête particulière. Les informaticiens utilisent les «instructions ouvertes» des statisticiens pour mettre au point une solution logicielle, puis l'adaptent (si nécessaire) en fonction des besoins particuliers de mise en application dans le cadre de l'enquête. Un tel système exigeait beaucoup d'efforts en matière de TI lors de la phase de développement, et rendait très ardue la maintenance des logiciels.

10. Le grand progrès apporté par ces nouvelles solutions était que la solution logicielle pour une partie déterminée du processus était mise au point une seule fois, puis adaptée, en fonction des besoins particuliers de mise en application, par simple modification des métadonnées de processus. Ces métadonnées sont fournies au système par les statisticiens chargés de l'enquête, qui sont désormais les gestionnaires de l'ensemble du système. Les informaticiens et les spécialistes des méthodes générales ont pour seule fonction de fournir l'assistance nécessaire en cas de problème.

11. La première solution générale était très souple et ouverte. Les programmes généraux pouvaient être intégrés et mis en fonction dans différents types de base de données, de microdonnées ou de métadonnées de processus, à condition que ces dernières soient organisées conformément aux règles prédéfinies. Un tel système, très générique et ouvert, est sans aucun doute très souple et constitue un bon outil pour construire un processus statistique. Toutefois, ce genre de système ouvert présente aussi un certain nombre de lacunes évidentes, qui sont liées principalement aux procédures de gestion des métadonnées de processus. Comme indiqué précédemment, les bases de données contenant les métadonnées de processus ont une structure définie de manière stricte, mais pour chaque enquête elles peuvent être incorporées dans différentes bases de données, et même dans des environnements différents (par exemple, Oracle, MS Access, SAS). En fait, pour la plupart des enquêtes, les métadonnées de processus étaient conservées dans des bases MS Access. Cela s'explique essentiellement par le fait que les spécialistes dans ce domaine, qui sont principalement chargés de gérer les métadonnées, préfèrent cet environnement en raison de sa simplicité et de sa facilité d'utilisation.

12. Le problème que pose ce type de système dispersé de métadonnées de processus est qu'il est impossible de créer une application générale efficace pour gérer et contrôler les métadonnées insérées. Comme il a été souligné dans les analyses effectuées après la première phase d'utilisation du système fractionné, le principal problème était le nombre considérable d'erreurs dans les métadonnées de processus. Étant donné que les champs dans lesquels sont insérées les règles sont pour le moment entièrement ouverts, la plupart de ces erreurs étaient des erreurs de syntaxe dans les règles (par exemple, des erreurs de parenthèse) ou de cohérence entre les règles et les variables.

13. Pour permettre la mise au point d'un meilleur système de gestion et de consultation des métadonnées de processus, l'OSRS a décidé d'intégrer à nouveau davantage l'ensemble du système. L'objectif de cette intégration renforcée n'est certes pas de recréer le système pleinement intégré qui avait été initialement conçu, mais seulement de rétablir un niveau d'intégration permettant de créer l'outil de gestion général tout en maintenant le haut degré de souplesse de ce système. Il a été décidé de prendre les mesures ci-après pour permettre à nouveau une plus grande intégration:

a) Créer une base de données unique contenant les métadonnées de processus. Cette base de données serait créée dans l'environnement Oracle et gérée par l'application .NET, ce qui permettrait une gestion aisée des métadonnées de processus;

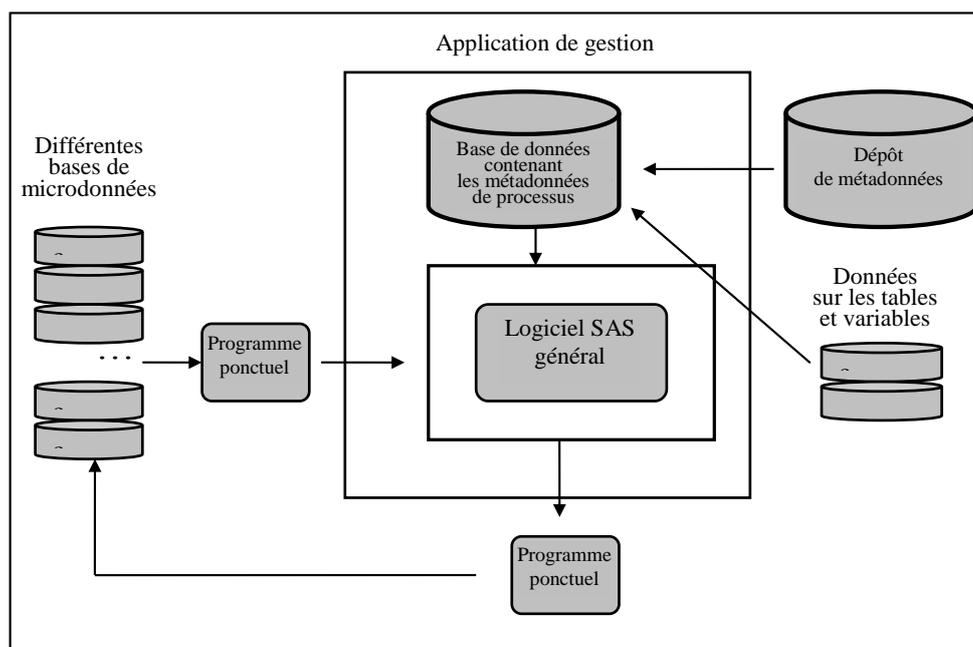
b) Relier le système avec les dépôts de métadonnées, où les données relatives aux enquêtes et des exemples d'enquêtes sont conservés;

c) Permettre à l'application de gestion d'accéder aux données sur les variables dans les tables insérées. Ces données peuvent être insérées dans l'application à partir de deux sources différentes: à partir du registre des variables, si les variables de l'enquête y ont été enregistrées, ou à partir d'autres emplacements où les structures des tables de données sont conservées.

14. On trouvera ci-après un schéma simplifié du nouveau système intégré prévu:

Figure 1

Schéma simplifié du nouveau système intégré



15. La mise en place de ce nouveau système est l'objectif principal du grand projet d'infrastructure mentionné au début de la présente note. L'OSRS a déjà mené à bien les opérations relatives à la validation des données, au système automatisé de correction et d'imputation des données, et nous introduisons progressivement cette partie de l'application dans la production habituelle. La mise au point des modules pour l'agrégation des données, le contrôle de la divulgation des données ou l'estimation des erreurs types et la tabulation nécessitera elle aussi beaucoup de temps encore. L'OSRS prévoit de commencer à introduire ces modules dans la production habituelle à la fin de cette année.

III. Difficultés liées à la mise en place d'un système entièrement nouveau

16. Au début de l'introduction de la nouvelle application (avec la toute nouvelle méthode) dans le processus habituel de production statistique, l'OSRS recueillait régulièrement les informations en retour des statisticiens qui devaient faire face au nouveau mode de traitement des données. Les principaux avantages et inconvénients de la nouvelle méthode, tels que perçus par les statisticiens, peuvent se résumer comme suit.

Principaux avantages:

- Les membres du personnel spécialisé sont beaucoup plus indépendants du service informatique, qui était auparavant responsable de l'exécution technique des opérations;

- Les règles relatives au traitement des données peuvent être modifiées très rapidement par le système centralisé de métadonnées de processus, de sorte que l'ensemble du cycle de traitement des données devient beaucoup plus souple;
- Étant donné que les utilisateurs peuvent exécuter les procédures à plusieurs reprises en peu de temps, il est désormais plus facile de vérifier la faisabilité des différentes méthodes de traitement des données.

Principaux inconvénients:

- Le risque de commettre des erreurs de syntaxe lors de l'insertion des métadonnées est élevé. En cas d'erreur, l'application ne peut pas être exécutée ou est exécutée avec des paramètres inappropriés;
- Les statisticiens spécialisés doivent acquérir de nouvelles compétences, ce qui est parfois un problème à cause de la charge de travail très lourde de la production statistique;
- Si une erreur survient pendant l'exécution de la procédure, les membres du personnel technique doivent être contactés et, s'ils ne sont pas disponibles, l'exécution du processus peut être interrompue pendant un certain temps.

17. La mise au point d'un système entièrement nouveau de production statistique est certainement un grand pas en avant pour l'OSRS. Nous sommes fermement convaincus que les résultats de ce projet nous aideront à mettre en place un nouveau système modernisé de gestion du traitement des données statistiques dans les différents systèmes statistiques. Le passage d'un système cloisonné à des solutions méthodologiques et informatiques centralisées est l'objectif final. L'élément central du système rénové est l'application pilotée par les métadonnées, qui, d'une part, est souple dans la mesure où elle peut être intégrée dans différents environnements de microdonnées et, d'autre part, met en œuvre une gestion centralisée des métadonnées de processus.

18. L'introduction de ce type d'application générique pilotée par les métadonnées pour le traitement des données s'accompagne inévitablement de certains changements au niveau institutionnel général, en ce qui concerne la conception et la réalisation des enquêtes statistiques. En se fondant sur l'expérience acquise jusqu'à présent, on peut résumer les principaux changements comme suit:

a) La répartition du travail entre les spécialistes, les responsables des méthodes générales et les informaticiens est fondamentalement différente. Dans l'ancien système, chaque statisticien avait son propre «programmeur» et ses propres «responsables des méthodes générales», qui s'appuyaient sur les instructions précises du spécialiste dans un domaine pour concevoir et mettre en œuvre les processus spécifiques correspondant à une enquête particulière. À présent, les responsables des méthodes générales et les informaticiens ont pour seule fonction de fournir l'assistance nécessaire s'il survient une erreur dans l'application ou si le processus ne fournit pas les résultats escomptés. Cela signifie que les spécialistes dans un domaine sont maintenant beaucoup plus indépendants du service informatique et du service des méthodes générales;

b) L'évolution du rôle des statisticiens spécialisés dans le processus statistique a également modifié les attentes relatives à leurs compétences et à leurs capacités. On attendait d'eux une connaissance approfondie du domaine concerné et l'aptitude à fournir les instructions écrites (sous forme ouverte) pour l'exécution de certaines parties du processus (par exemple, l'imputation ou l'agrégation). Aujourd'hui, ils doivent être qualifiés et formés de façon à pouvoir écrire eux-mêmes ces règles en langage informatique mathématique;

c) L'ensemble de l'organisation du travail du service informatique et du service des méthodes générales devra être modifié, passant d'une démarche axée sur les domaines à une démarche axée sur les processus. Cette réorganisation implique une conception d'ensemble très différente de l'organisation de l'institution et de la répartition des tâches et constitue par conséquent un véritable défi pour l'organisme responsable des statistiques. L'OSRS en est au premier stade de l'action visant à relever ce défi;

d) La réorientation décrite ci-dessus, prévoyant de passer d'une démarche axée sur les domaines (spécifiques) à une démarche axée sur les processus (généraux) devra être mise en œuvre aussi au niveau des experts en informatique et des spécialistes des méthodes. La mise au point et la maintenance de ce type d'applications génériques nécessitent le travail d'experts à même de fonctionner sur un plan beaucoup plus général et de considérer l'exécution d'une enquête particulière comme une tâche parmi d'autres dans l'ensemble du processus statistique.

Bibliographie

Dolenc, D., Krek, M., Seljak, R. (2011), «Editing Process in the Case of Slovenian Register- based Census», document présenté lors de la réunion de travail de la CEE sur l'édition des données statistiques, tenue à Ljubljana (Slovénie) du 9 au 11 mai 2011;

Seljak, R. (2009), «Integrated statistical systems and their flexibility – How to find the balance?», présenté à la conférence sur le thème des nouvelles techniques et technologies au service des statistiques (New Techniques and Technologies for Statistics (NTTS)), tenue à Bruxelles (Belgique) du 5 au 7 mars 2013;

Seljak, R., Blazic, P. (2011), «Sampling error estimation – SORS practice», présenté au deuxième atelier européen sur l'établissement de statistiques, tenu à Neuchâtel (Suisse) du 12 au 14 septembre 2011;

Seljak, R. (2014), «Metadata driven application for data processing – from local toward global solution», document présenté lors de la réunion de travail de la CEE sur l'édition des données statistiques, tenue à Paris (France) du 28 au 30 avril 2014.