



Conseil économique et social

Distr. générale
27 mars 2013
Français
Original: anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

Soixante et unième réunion plénière

Genève, 10-12 juin 2013

Point 4 b) de l'ordre du jour provisoire

**Comment les bureaux de statistique nationaux devraient-ils réagir,
en passant du souci d'éviter les risques à la gestion des risques?**

Microdonnées: un atout fondamental des systèmes statistiques

Note de l'Institut national italien de statistique

Résumé

Le présent document décrit les problèmes stratégiques liés à l'amélioration de l'accès aux microdonnées et la manière dont l'Institut national italien de statistique fait face aux difficultés rencontrées dans ce domaine.

Les statistiques officielles sont destinées à répondre à la demande d'information de différentes catégories d'utilisateurs qui ne cessent d'augmenter et de se diversifier. Les bureaux de statistique nationaux sont censés mettre au point des systèmes intégrés de diffusion permettant de combiner des données extraites de sources multiples et d'améliorer l'exhaustivité de l'information statistique. Le développement des technologies de l'information et, surtout, le besoin accru d'information entraînent une augmentation du nombre d'utilisateurs qui doivent pouvoir accéder aux microdonnées.

I. Introduction

1. Les effets des progrès techniques sur les sociétés, l'économie et les priorités des gouvernements, ainsi que l'augmentation du volume et de la diversification des sources de données disponibles gratuitement en ligne ont créé de nouveaux besoins en matière d'information statistique. Les organismes de statistique, nationaux et autres, sont la principale source de statistiques officielles dans chaque pays. Créés à l'origine pour satisfaire essentiellement les intérêts des gouvernements, les organismes de statistique nationaux sont désormais censés offrir un large éventail de produits et services à un vaste ensemble d'utilisateurs, dont les besoins d'information et la capacité à gérer les données statistiques sont très variables. Dans ce contexte, les organismes de statistique nationaux font face à plusieurs problèmes: un volume croissant de statistiques cohérentes et pertinentes doit être produit; la demande de données plus actuelles sur une plus large gamme de sujets augmente sans cesse; des statistiques spatiales plus détaillées sont plus souvent demandées; l'harmonisation de données provenant de sources multiples est nécessaire pour évaluer des questions transversales; et enfin, de nouvelles stratégies et outils de diffusion doivent être mis au point pour garantir la communication effective de l'information à différents utilisateurs.

2. Du point de vue de la gestion de la clientèle, pour obtenir le plus haut retour possible sur leurs propres investissements (et l'investissement de fonds publics), les organismes de statistique devraient planifier et créer leurs produits et services en les alignant sur les besoins et attentes des principaux groupes d'utilisateurs. Ils doivent en outre adapter le contenu et les caractéristiques des produits à la notion d'amélioration de la maîtrise de la statistique. Dans la pratique, il faudra veiller à ce que les statistiques soient ventilées à des degrés de détail différents, qu'elles s'appuient sur des représentations graphiques, qu'elles soient diffusées par le biais de plusieurs canaux, et qu'elles soient organisées de façon à répondre aux questions et aux préoccupations de différents groupes d'utilisateurs. La même approche que celle adoptée pour les produits statistiques classiques devrait être appliquée à l'accès aux microdonnées.

3. Les étudiants, l'administration publique et les instituts privés et publics ont plus souvent besoin d'accéder aux microdonnées, malgré la complexité des procédures régissant la diffusion de microdonnées par les organismes de statistique nationaux. Si l'accès aux microdonnées devrait faire partie des principaux services offerts par les organismes de statistique, il conviendrait de revoir les compétences organisationnelles, les processus de gestion, les normes culturelles, les cadres législatifs et la propriété des microdonnées (McMillan, 2010).

4. Le présent document décrit les expériences acquises au cours des dernières années par l'Institut national italien de statistique (Istat), présente la démarche adoptée par l'Institut pour surmonter les obstacles à l'accès aux microdonnées, et met en avant les questions stratégiques générales à traiter. La Section II fournit des informations sur le projet qui vise la création d'un réseau d'archives de données en Italie où l'Istat joue un rôle clef, et étudie les défis que doivent relever les organismes de statistique nationaux pour assurer l'accès aux microdonnées. La Section III montre que la modernisation de la production de statistiques a des effets considérables sur les microdonnées lorsque les architectures sont alignées sur de nouvelles normes et des modèles de référence concernant les données et processus statistiques. La Section IV présente le système actuel de services d'accès aux microdonnées mis en place par l'Istat et son intégration récente dans la mise en circulation de fichiers à usage public et de nouvelles questions en la matière. Des conclusions sont formulées dans la Section V.

II. Archives de données statistiques officielles

5. L'établissement d'une stratégie de diffusion contribue à améliorer les différentes étapes des activités de diffusion: la définition de politiques de diffusion, la conception de produits et services, l'élaboration et la présentation de statistiques, la diffusion de l'information sur le site Web, ainsi que la promotion et la commercialisation de produits, services et «publications statistiques» auprès des utilisateurs. Cependant, si l'accès aux microdonnées doit devenir l'un des principaux services offerts par les organismes de statistique, des efforts bien plus importants que l'adoption de stratégies de diffusion seront nécessaires.

6. Récemment, la «phase d'entrée» de la production de statistiques, la saisie de données, a été transformée par de nouvelles méthodes de communication (voir, par exemple, Marske et Stempowski, 2008). La «phase de sortie» devrait subir les mêmes changements en ce sens que l'utilisateur – de microdonnées par exemple – se retrouve au centre de l'attention. Il pourrait s'agir de changements stratégiques: créer une source unique pour toutes les informations pertinentes, mettre au point des produits et services de microdonnées adaptés aux besoins des utilisateurs et offrir des services qui facilitent les relations avec les utilisateurs.

7. Pour concevoir un système de diffusion des microdonnées axé sur l'utilisateur, la gouvernance, l'infrastructure et le cadre juridique doivent être définis et mis en place en parallèle avec des politiques concrètes sur l'accès aux données ainsi que des produits et services souples en matière de microdonnées. Afin d'y parvenir, il est également nécessaire d'établir des relations différentes avec les utilisateurs de microdonnées.

A. Gouvernance

8. En Italie, la conception axée sur l'utilisateur de l'accès aux microdonnées est encouragée par un groupe d'institutions. Celles-ci constituent actuellement des sources de données centralisées liées entre elles, avec l'infrastructure connexe, ce qui permettra d'assurer l'accès le plus large possible à des microdonnées financées par des fonds publics. Ce projet s'inscrit dans le cadre de la statistique officielle: l'Istat sera la plaque tournante d'un projet commun visant à développer un réseau d'archives de données qui traitera des microdonnées produites grâce à des fonds publics. Dans le cadre de ce réseau, la Banque d'Italie et l'Istat assureront la gestion des microdonnées dans leurs domaines d'activité respectifs, une troisième antenne sera chargée des microdonnées des instituts de recherche, et le Ministre de l'éducation encouragera les universités à partager leurs microdonnées. À l'avenir, l'Istat devrait devenir le coordonnateur de l'accès aux microdonnées en ce qui concerne les statistiques provenant des administrations publiques (comme les ministères) et du système de statistique national italien.

B. Infrastructure

9. En Italie, plusieurs institutions publiques offrent un accès aux microdonnées. Bien qu'elles soient disponibles, les microdonnées sont dispersées dans un grand nombre d'institutions publiques ou d'instituts de recherche. En termes d'infrastructure, le réseau italien d'archives de données créera un accès unique et cohérent aux produits et services de microdonnées offerts dans le cadre des statistiques publiques. Un catalogue en ligne de toutes les microdonnées sera disponible et accessible aux utilisateurs, à la fois directement par voie de téléchargement ou sur demande, pour les services plus spécialisés (exécution et accès à distance, etc.).

10. L'infrastructure permettant un accès aux microdonnées à distance, mise en place par l'Istat, et le système d'exécution à distance utilisé au sein de la Banque d'Italie pourraient fournir conjointement un accès aux microdonnées de l'ensemble du système national de statistique et à d'autres données financées par des fonds publics. Ce concept de point d'accès unique en ligne a été adopté par le Royaume-Uni par exemple, et est envisagé au niveau européen par le Council of European Social Science Data Archives (CESSDA). Sous les auspices du septième projet de programme-cadre «Data without Boundaries» (<http://www.dwbproject.org/>), 28 institutions s'emploient à assurer la coordination des infrastructures existantes d'accès aux microdonnées officielles en Europe.

C. Cadre juridique

11. L'Istat, en tant que plaque tournante du réseau des archives italiennes de données, devrait avoir la possibilité de fournir un accès aux microdonnées provenant d'autres institutions du système national de statistique et de gérer l'accès à distance aux microdonnées des instituts de recherche. Un cadre juridique appuyant une telle architecture est indispensable. Cette approche devrait être réalisable en Italie après l'amendement, en cours d'élaboration, de la loi italienne sur les statistiques.

D. Recherche de données et accessibilité

12. Un système axé sur l'utilisateur donne un aperçu rapide et clair du type de microdonnées disponibles et des conditions d'accès à ces données. Des programmes de recherche faciles à utiliser seront créés pour aider les utilisateurs à trouver les microdonnées dont ils ont besoin. De plus, des logiciels de métadonnées normalisés, éléments fondamentaux de tout système de diffusion, doivent être soigneusement sélectionnés pour guider l'interprétation des données. Enfin, le réseau d'archives de données s'emploiera à harmoniser les politiques d'accès aux microdonnées au sein du système national italien de statistique afin d'appuyer la gestion des demandes adressées à différentes institutions.

E. Relation avec les utilisateurs

13. Outre la coopération entre les archives de données, le réseau mettra l'accent sur la collaboration avec les utilisateurs et le développement de leur maîtrise de la statistique. Les utilisateurs de microdonnées ont besoin d'acquérir des connaissances sur les statistiques, la recherche analytique, l'analyse des données et la programmation informatique. L'analyse des données devrait être comprise dans sa forme la plus simple par la plupart des citoyens. Elle devrait par conséquent être enseignée aux étudiants et faire partie des compétences de base du personnel de l'administration publique.

14. Le fait de disposer d'outils très efficaces pour analyser les microdonnées statistiques est insuffisant. Des compétences adéquates doivent être acquises afin de choisir les méthodes appropriées, de les appliquer correctement et de comprendre et interpréter les résultats. Le développement des compétences en matière de statistique devrait être constamment encouragé et appuyé par les organismes de statistique nationaux à la fois à l'intérieur de leurs systèmes nationaux de statistique (entités publiques, administrations publiques) et en dehors de ces systèmes (étudiants de premier cycle et en mastère, doctorants, jeunes chercheurs, etc.).

15. En réponse à la demande croissante émanant de la société et de l'administration publique, l'Istat a créé en 2011 l'École supérieure de statistique et d'analyses socioéconomiques (Scuola Superiore di Statistica e Analisi Economica, SAES). Cette école

encourage l'initiation à la statistique et offre des programmes de formation à des techniques d'enquête et des méthodes statistiques poussées, des formations adaptées et des stages. L'objectif est double: accroître les connaissances statistiques et la recherche analytique au sein de l'administration publique et former les futurs utilisateurs de statistiques officielles.

16. Le réseau d'archives de données et la SAES collaboreront en vue d'assurer des programmes de formation répondant à différents besoins. Mais ce réseau annonce surtout des changements dans les relations entre producteurs et utilisateurs de statistiques officielles. Des partenariats sont actuellement établis pour collaborer avec l'Istat, améliorer la conception des enquêtes, proposer des changements aux questionnaires utilisés pour la collecte de données et accroître l'utilité des données. Les utilisateurs ne seront plus de simples analystes de données, mais seront de plus en plus fréquemment sollicités pour contribuer activement à l'amélioration des statistiques officielles.

III. Microdonnées: un produit en soi

A. Normalisation et qualité

17. Au cours des dernières années, les services chargés des statistiques officielles ont été confrontés à la fois à l'accélération des changements survenus au sein de la société et à une pénurie de ressources qui ont incité des statisticiens en chef à poser un regard critique sur toute la structure du processus de production de statistiques. En Europe, cette situation est exposée dans la *Communication de la Commission au Parlement européen et au Conseil concernant la méthode de production des statistiques de l'Union européenne: une vision de la prochaine décennie (COM 404/2009)*. Au niveau mondial, l'initiative du Groupe de haut niveau sur la modernisation de la production et des services statistiques a renforcé le processus de normalisation et d'industrialisation des statistiques officielles.

18. Le Groupe de haut niveau sur la modernisation de la production et des services statistiques cherche à réutiliser et à partager des méthodes, des composantes, des processus et un référentiel pour les données et à adopter une architecture commune «prête à l'emploi» composée de modules, pour gagner en efficacité. La mise en œuvre de cette architecture peut être réalisée uniquement en adoptant des normes communes, telles qu'elles figurent dans le Modèle générique du processus de production statistique (GSBPM) et le Modèle générique d'informations statistiques (GSIM). Le premier de ces deux modèles est nécessaire pour définir les composantes communes du processus statistique, et le deuxième constitue un cadre de référence d'objets d'information qui permet de faire une description générale de la définition, de la gestion et de l'utilisation des données et des métadonnées tout au long du processus de production de statistiques. Ensemble, ils servent de fondement à l'établissement d'un langage commun et à la mise au point d'un système intégré de production de statistiques.

19. En Europe, une équipe stratégique spéciale, Parrainage de la normalisation, a été constituée pour conseiller le Système statistique européen sur les moyens de mener à bien un processus de normalisation et d'intégration (voir Braaksma *et al.*, 2013). À l'Istat, un projet a été lancé. Il s'agit de Stat2015 (voir Falorsi *et al.*, 2013), dont l'objectif est de normaliser et d'industrialiser les processus de production en s'appuyant sur la réutilisation et l'adoption d'un modèle fondé sur les services communs, dans un cadre d'architectures axées sur les services.

20. Selon la vision mise en avant par le Groupe de haut niveau sur la modernisation de la production et des services statistiques, les microdonnées validées découlent d'un processus normalisé et harmonisé et sont accompagnées de métadonnées appropriées grâce auxquelles elles peuvent être utilisées à différentes fins sans ambiguïté. Dans le GSBPM,

les données validées sont isolées dans un sous-processus 5.8 spécifique (Finalize Data Files) afin de souligner l'importance du résultat: les données validées sont, elles-mêmes, le fruit d'un processus normalisé. En outre, l'adoption de la perspective d'un processus statistique entièrement piloté par les métadonnées est utile du point de vue de la qualité des microdonnées. L'utilisation de logiciels normalisés permet d'établir un lien entre les données et les métadonnées dès l'étape de la saisie des données – par exemple, par le biais de la Data Documentation Initiative (DDI) et/ou de l'Échange de données statistiques (SDMS). Pour que les microdonnées soient considérées comme un produit à part entière, elles doivent être de très bonne qualité.

B. Impartialité, transparence et confiance du public

21. Sur son site Web, l'Istat, comme d'autres organismes de statistique, publie de nombreux produits statistiques: indicateurs, tableaux, plans, graphiques et autres. Pour diffuser et communiquer des produits «à valeur ajoutée», il faut choisir quelles données ou quels résultats il convient de privilégier et sur quels résultats et conséquences mettre l'accent. Les produits fondés sur des microdonnées peuvent contribuer à accroître la transparence des travaux des organismes de statistique et, surtout, à renforcer leur crédibilité et la confiance qui leur est accordée par le public. De fait, seul l'accès aux microdonnées permet aux utilisateurs de reproduire les statistiques publiées par les organismes de statistique, d'effectuer des analyses et des comparaisons, et de contribuer ainsi également à l'adaptation continue du système statistique aux besoins de la société en matière d'information.

22. Les niveaux élevés de qualité, ainsi que les principes éthiques et professionnels stricts qui constituent le fondement de la production de données statistiques devraient encourager les organismes de statistique à donner accès à leurs ressources en microdonnées, en pleine conformité avec les principes de confidentialité. La transparence, l'impartialité et la neutralité peuvent être renforcées en considérant les microdonnées comme des produits en soi.

C. Microdonnées compilées à partir de plusieurs sources

23. Pour améliorer l'accès aux microdonnées, les considérations relatives à la qualité et à l'infrastructure sont insuffisantes. Il faut reconnaître que les utilisateurs ont besoin d'accéder aux microdonnées pour analyser des phénomènes complexes qui ne sont pas traités par les statistiques faciles à obtenir, ce qui signifie que les données recueillies grâce à une seule enquête ne suffisent peut-être pas à bien rendre compte des questions complexes de la société. Par ailleurs, les points forts du système de statistiques officielles sont les suivants: bien souvent, les organismes de statistique sont dotés d'un cadre juridique grâce auquel ils peuvent recueillir des données et microdonnées ou extraire des données des registres de l'administration publique à des fins de production de statistiques, ce qui fait de ces organismes des hôtes naturels pour les processus d'intégration de microdonnées dans les systèmes nationaux de statistique.

24. À l'échelle européenne, plusieurs projets permettent actuellement de créer une plate-forme commune pour partager, traiter et redistribuer l'information au sein du Système statistique européen (projet de registre EuroGroups (EGR) et sources de données centralisées contenant des microdonnées sur les revenus et les conditions de vie des Statistiques de l'Union européenne sur le revenu et les conditions de vie (EU-SILC)). Par conséquent, les technologies susceptibles de changer le traitement des microdonnées et l'accès à celles-ci sont déjà en place. De nouvelles voies s'ouvrent alors en ce qui concerne l'utilisation des microdonnées, leur conférant ainsi une plus grande importance en tant que produit en soi. Ces infrastructures techniques et de traitement élargissent la sphère de la (ré)utilisation des microdonnées tout en soulignant une fois de plus la nécessité profonde de considérer les microdonnées comme un produit en soi.

25. Il est probable qu'à court terme la production de données statistiques fondées sur des microdonnées intégrées fera partie des meilleures pratiques des organismes de statistique. Parallèlement, les microdonnées intégrées, en tant que source principale permettant d'examiner des phénomènes complexes, inciteront davantage d'utilisateurs à accéder aux microdonnées dans le but d'améliorer leurs compétences en matière de statistique. Il est évident que ces analyses ne sont utiles que lorsqu'il s'agit d'objets bien définis, c'est-à-dire seulement si les microdonnées primaires sont considérées comme un produit officiel.

IV. Différents services de microdonnées pour différents besoins

26. Parallèlement aux outils et services, les produits sont au cœur de l'archivage de données. Afin de répondre aux nouveaux besoins hétérogènes des utilisateurs de données, l'Istat a varié ces dernières années les modes d'accès existants et créé de nouveaux produits et services.

27. Actuellement, l'offre va du téléchargement gratuit d'agrégats demandés par le biais de sources de diffusion internes centralisées (Istat: <http://dati.istat.it/?lang=en>) au traitement de données personnalisé (<https://contact.istat.it/Index.php?Lingua=Inglese>).

28. S'agissant des microdonnées, outre la mise à disposition de données anonymisées au niveau de l'unité provenant d'enquêtes sociales et accessibles à tous sur demande, sous l'intitulé «file standard», des Fichiers de microdonnées établis à des fins de recherche (Micro-data Files for Research (MFR)) et contenant des informations bien plus détaillées ont été créés à l'intention des universitaires (http://en.istat.it/dati/microdati/file_microdati.html#file_ricerca).

29. Enfin, le réseau de points d'accès au centre de recherche sur les données (Laboratorio per l'Analisi di Dati Elementari) permet aux chercheurs d'analyser toute microdonnée produite par l'Istat dans chacun des 18 bureaux régionaux de l'Istat répartis sur l'ensemble du pays (<http://www.istat.it/en/information/researchers/analysis-of-individual-data>). Ce réseau, fondé sur un système client-léger (réseau de serveurs), permet aux chercheurs d'avoir accès aux données confidentielles originales à travers un circuit sécurisé.

A. Fichier à usage public et principe de réutilisation

30. Le système d'accès aux microdonnées de l'Istat sera encore perfectionné au cours de 2013 avec l'adoption d'un fichier à usage public nommé mIcro.STAT. Ces fichiers sont des échantillons de données individuelles aux caractéristiques bien définies:

- a) Ils sont disponibles gratuitement en ligne (aucun accord d'accès ne doit être signé);
- b) Ils permettent aux utilisateurs de formuler des conclusions sur les phénomènes liés aux données;
- c) Ils sont généralement publiés sans restrictions ni conditions d'utilisation.

31. Les techniques de sous-échantillonnage sont utilisées, avec d'autres méthodes de protection, pour produire des fichiers à usage public (voir Hundepool *et al.*, 2012). Ces techniques réduisent le risque de divulgation en augmentant l'incertitude quant au nombre d'unités démographiques ayant la même combinaison de variables identifiables. Afin d'accroître l'efficacité de la production de produits diversifiés, l'Istat a appliqué une méthode différente.

32. mIcro.STAT illustre parfaitement la «réutilisation» de microdonnées. De fait, pour la production de fichiers à usage public, deux éléments sont réutilisés: un produit existant au préalable, le MFR, et les compétences en matière d'échantillonnage de l'Istat. Pour accroître l'efficacité de la production de fichiers à usage public, une nouvelle solution méthodologique a été adoptée en se fondant sur le sous-échantillonnage du MFR correspondant (voir Foschi *et al.*, 2012). Le fichier à usage public et le MFR ont donc une structure commune. La structure hiérarchique de ces deux fichiers de données simplifie considérablement l'évaluation du risque de divulgation et réduit la perte d'information associée à la procédure d'anonymisation. Cette méthode a d'autres avantages, notamment la préservation des précisions hiérarchiques des données et la cohérence interne des deux ensembles de données. Elle permet donc de réduire les coûts liés à l'élaboration des fichiers, et de gagner ainsi en efficacité.

33. L'utilisation combinée de différentes techniques d'échantillonnage, par exemple la répartition multivariée multisectionnelle (voir Bethel 1989) et l'échantillonnage équilibré (voir Deville J., Tillé 2004) permet d'assurer de façon simultanée la réduction du risque de divulgation et la préservation d'indicateurs de l'utilité des données définis au préalable. L'objectif consiste à fournir un fichier à usage public satisfaisant au plus grand nombre possible de conditions relatives au risque et à l'utilité.

B. Motifs de l'établissement de fichiers à usage public: démocratie et initiation à la statistique

34. La mise en place d'un fichier à usage public dont les détails, la qualité et la complexité soient identiques à ceux du fichier correspondant établi à des fins de recherche est fondée sur les principes de démocratie en matière d'accès et de droit à la recherche. En outre, ce n'est qu'en permettant aux étudiants d'être formés en matière de données officielles complexes que la maîtrise de la statistique sera renforcée dans le pays.

35. Les fichiers didactiques contiennent habituellement très peu de variables, des informations de base et quelques observations, et présentent une structure de données simplifiée. Pour surmonter ces déficiences, les étudiants doivent apprendre à s'entraîner en utilisant de véritables microdonnées d'enquête, à effectuer des analyses statistiques en se fondant sur des données brutes et, surtout, à acquérir des connaissances en étudiant les données. Ces éléments sont essentiels pour accroître l'initiation de la population en général et des étudiants en particulier à la statistique.

36. Lorsque le fichier à usage public répond à des normes de qualité définies au préalable, il peut contribuer positivement au renforcement de la maîtrise de la statistique. La valeur ajoutée d'un fichier à usage public est une conséquence directe de son niveau de qualité défini par rapport à sa capacité à simuler de véritables applications. La taille du fichier, exprimée en nombre de données et de variables, constitue un premier indicateur de qualité. De plus, le processus de production de données et la qualité des données n'étant pas soumis à un examen approfondi lors des conférences sur les statistiques, tout fichier à usage public pourrait contribuer à la réduction de cet écart. Dans le même temps, un grand nombre de variables favoriserait le développement d'un raisonnement critique sur la signification des variables, leur définition opérationnelle, le phénomène étudié, etc. La précision et la justesse des estimations qui pourraient être établies à partir de fichiers à usage public amélioreraient considérablement l'apprentissage de la méthodologie statistique. Le tableau 1 indique quelles caractéristiques mIcro.STAT sont adaptées au processus d'extraction de connaissances.

37. Enfin, les activités liées à la diffusion de fichiers à usage public comprennent l'établissement d'un permis d'exploitation (qui peut être international) et la définition d'outils et de services à mettre en place pour analyser des produits de microdonnées. D'une part, les microdonnées offrent une souplesse maximale à l'analyse d'un phénomène. D'autre part, par définition, les microdonnées ne sont pas faciles à utiliser et ne permettent pas d'obtenir des résultats immédiats. Seules la conception et la mise au point d'outils et de services appropriés pour analyser des microdonnées amélioreraient les possibilités d'utilisation des données. L'établissement d'un processus clair en matière d'octroi de licences et la mise en place de nouveaux services sont des questions stratégiques lorsqu'il s'agit des moyens d'accès aux microdonnées.

Tableau 1

Caractéristiques du fichier à usage public (mIcro.STAT) et des MFR et leur relation avec le processus correspondant d'extraction de connaissances

mIcro.STAT	Extraction de connaissances	MFR
	Définition de problèmes – formulation d'hypothèses	
✓	Lecture de données (format, documentation, classifications)	✓
✓	Première analyse de données	✓
✓	Utilisation de méthodes statistiques (modélisation, analyse typologique, etc.)	✓
✓	Interprétation des résultats	✓
✓	Vérification des résultats	✓
✓	Comparaisons	✓
✓	Conséquences	✓
✓	Redéfinition des hypothèses	✓
✓	Rapport sur la présentation des résultats	✓
✓	Mise au point d'applications statistiques	✓
	Développement de théories socioéconomiques	✓
	Prise de décisions d'orientation	✓

V. Conclusions

38. Les analyses de microdonnées sont des ressources extrêmement précieuses pour permettre aux planificateurs de l'administration publique, aux analystes des marchés, aux chercheurs universitaires et aux citoyens de prendre de bonnes décisions fondées sur des faits établis. Les organismes de statistique doivent relever les défis liés au renforcement de l'accès à leurs microdonnées de qualité produites à des fins de recherches scientifiques et pour la société en général, tout en respectant les conditions des Principes fondamentaux de la statistique officielle.

39. Les groupes d'utilisateurs étant très variés, il est nécessaire d'adopter une approche différenciée en ce qui concerne la diffusion de données statistiques, c'est-à-dire de recenser la demande d'un côté et l'offre de statistiques de l'autre. Dans le cadre d'une approche dynamique, il faut connaître les besoins des utilisateurs en matière d'information, ajuster l'offre d'information en fonction de ces besoins et s'adapter aux changements de priorités survenus au sein de la société dans le domaine de l'information.

40. La modernisation de la production de statistiques donne aux organismes de statistique les moyens de réaliser de nouveaux objectifs ambitieux. À titre d'exemple, elle facilite le passage de l'innovation de secteur à l'innovation systémique continue, et favorise ainsi la transition stratégique des organismes de statistique de l'état de producteurs de données à celui de fournisseurs de produits et services statistiques. Les organismes de statistique peuvent jouer un rôle de chef de file dans la conception d'archives de données pour les microdonnées nationales.

41. La création d'archives de données ne peut être réduite à une simple infrastructure ou à un catalogue de fichiers disponibles. Une culture et un savoir relatifs aux microdonnées doivent être développés. La maîtrise de la statistique doit être améliorée et les organismes de statistique devraient s'employer activement à assurer des services d'accès aux microdonnées fondés sur le savoir qui répondraient aux demandes des utilisateurs en tenant compte des principes de démocratie (fichiers à usage public) et de méritocratie (MFR).

42. Le problème essentiel qui persiste est celui de la gouvernance de tous ces processus. Des produits et des outils peuvent être créés, des systèmes peuvent être conçus et bâtis, des architectures peuvent être adaptées, mais la direction et la gouvernance solides de ces nouveaux éléments sont indispensables pour garantir la bonne gestion de l'accès aux microdonnées.

VI. Références

Braaksma, B., Colasanti, C., Falorsi, P. D., Kloek, W., Martinez Vidal, M. et Museux, J. M. (2013). *Standardisation in the European Statistical System*, NTTS Conferences on New Techniques and Technologies in Statistics, Bruxelles, 5-7 mars 2013, disponible à l'adresse suivante: www.NTTS2013.eu.

Bethel, J. W. (1989). *Sample Allocation in Multivariate Surveys*. Survey Methodology, Vol. 15, p. 47 à 57.

Deville J., Tillé Y. (2004). *Efficient Balanced Sampling: The Cube Method*. Biometrika. 91(4). 893-912.

Falorsi, P. D., Barcaroli, G., Fasano, A. et Mignolli, N. (2013). *A Business Architecture framework for industrialisation and standardisation in a modern National Statistical Institute*. NTTS Conferences on New Techniques and Technologies in Statistics, Bruxelles, 5-7 mars 2013, disponible à l'adresse suivante: www.NTTS2013.eu.

Foschi, F., Casciano, C., Ichim, D. et Franconi, L. (2012). *Designing Multiple Releases from the Small and Medium Enterprises Survey*, dans J. Domingo Ferrer et I. Tinnirello (eds) Proceeding of PSD2012, Vol. 7556, Lecture Notes in Computer Science, p. 200 à 215. Springer, Berlin/Heidelberg.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K. et De Wolf, P. P. (2012). *Statistical Disclosure Control*. Wiley.

MacMillan, P. (2010). *Unlocking Government: How Data Will Transform Democracy*, Commonwealth Innovation, Vol. 16, 2, p. 13 à 17.

Marske R. et Stempowski D. M. (2008). *Approches de communication centrées sur les entreprises pour la gestion de la réponse aux enquêtes-entreprises*. La série des symposiums internationaux de statistique Canada: recueil (2008). Collecte de données: défis, réalisations et nouvelles orientations.