

**Commission économique pour l'Europe****Conférence des statisticiens européens****Soixante-septième réunion plénière**

Genève, 26-28 juin 2019

Point 6 a) de l'ordre du jour provisoire

Le nouveau rôle des organismes nationaux de statistique dans le domaine de la gestion de données – Première séance : nouvelles possibilités et problématiques relatives au système de données**Le nouveau rôle des organismes nationaux de statistique dans le domaine de la gestion de données¹****Établi par Statistics Denmark***Résumé*

Les organismes de statistique, qui étaient jusqu'ici des « usines » à données sont en train de se transformer en organismes de services modernes qui, outre cette fonction, fournissent des services liés à l'accès aux données et à l'utilisation des microdonnées. Le présent document décrit les possibilités et les difficultés qui se présentent du point de vue d'un petit organisme de statistique doté d'un système de production statistique bien intégré s'appuyant pour l'essentiel sur les données provenant de registres administratifs – et bénéficiant de la relation de confiance forte qui lie les organismes publics à la population.

Statistics Denmark présente l'expérience qu'il a vécue lorsqu'il a dû travailler à la fois dans le domaine statistique et dans celui des données il y a environ trente ans, ouvrant l'accès aux microdonnées aux chercheurs et mettant au point un système de production qui relie les informations résultant de l'enregistrement à l'état civil aux données sur l'éducation, la participation au marché du travail, les revenus, la santé, etc. D'autres composantes du secteur public font de plus en plus appel aux compétences de l'organisme national de statistique pour être sûres aussi de disposer des meilleures données à l'appui des décisions administratives. Cela pose des difficultés mais ouvre aussi des possibilités pour les organismes de statistique d'utiliser plus largement leurs compétences dans l'intérêt général que pour la seule production de statistiques officielles.

Le présent document est présenté pour examen au séminaire organisé en 2019 par la Conférence des statisticiens européens sur le thème « Le nouveau rôle des organismes nationaux de statistique dans le domaine de la gestion de données » dont la première séance porte sur les « nouvelles possibilités et problématiques relatives au système de données ».

¹ Le présent document a été soumis tardivement faute de ressources suffisantes.



I. Introduction

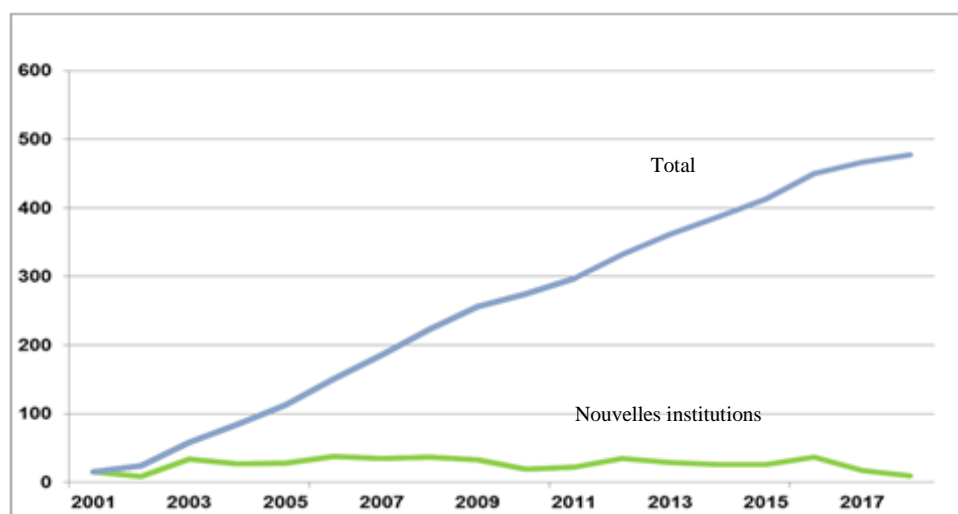
1. Le rôle des organismes de statistique est en train de changer de par l'importance croissante des données et des statistiques qui fondent les évolutions sociétales et les décisions prises dans le domaine politique et dans les entreprises du monde entier. Les organismes de statistique, qui étaient jusqu'ici des « usines » à données, sont déjà en train de se transformer en organismes de services modernes qui, outre cette fonction, fournissent des services liés à l'accès aux données et à l'utilisation des microdonnées pour les travaux de statistique et d'analyse effectués par les utilisateurs extérieurs.
2. Cette évolution offre aux organismes de statistique une excellente occasion d'accroître leur utilité en tant qu'institutions sociales majeures à l'ère des données – mais cela ne va pas sans difficultés.
3. Le présent document décrit à la fois les possibilités et les difficultés qui se présentent du point de vue d'un petit organisme de statistique doté d'un système de production statistique bien intégré s'appuyant pour l'essentiel sur les données provenant de registres administratifs – et bénéficiant de la relation de confiance forte qui lie les organismes publics à la population, comme, par exemple, l'organisme de statistique. D'un côté, ce type d'expérience a toutes les chances d'intéresser les pays et les organismes se trouvant dans une autre situation – mais de l'autre, il va de soi que l'expérience de chaque pays est particulière.

II. Le centre de données de Statistics Denmark

4. Statistics Denmark est devenu un organisme chargé à la fois des statistiques et des données il y a environ trente ans lorsqu'il a ouvert l'accès aux microdonnées aux chercheurs et aux projets de recherche.
5. L'origine de cette évolution a été la mise en place d'un identifiant unique pour chaque personne née au Danemark ou qui y résidait en 1968. C'est à partir de ces éléments que Statistics Denmark a établi un système de production statistique fondé sur des registres tout au long des années 70. Le dernier recensement effectué à partir d'enquêtes a été mené en 1970, et au début des années 80, un système de production reliant les informations résultant de l'enregistrement à l'état civil aux données sur l'éducation, la participation au marché du travail, les revenus, la santé, etc., a été définitivement mis au point.
6. Au début, les chercheurs devaient effectuer leurs analyses dans les locaux de Statistics Denmark (sur site). En 2001, un système en ligne a été créé qui est rapidement devenu le mode d'accès aux données le plus prisé. Les possibilités de travailler sur site ont donc été supprimées en 2008.
7. Le centre de données a eu un immense succès comme le montre la figure 1. Le nombre d'institutions agréées a augmenté : nettement inférieur à 100 en 2001 – lors de la création du service en ligne – il avoisine les 500 aujourd'hui.
8. Un accès est fourni aux institutions agréées, qui, dans la plupart des cas, sont des départements et des unités de recherche universitaires, mais aussi aux services d'analyse des ministères et des municipalités et aux cabinets de conseil privés.

Figure 1

Progression du nombre total d'institutions agréées et du nombre de nouvelles institutions consultant le centre de données de Statistics Denmark, 2001-2017



9. Pour être agréée, une institution doit posséder un environnement stable pour le travail analytique et employer des cadres, des chercheurs et des analystes responsables possédant une expérience du traitement des microdonnées. C'est le Statisticien national qui donne l'agrément.

10. La loi relative à Statistics Denmark constitue la base juridique de l'accès aux données. Elle dispose qu'en plus des obligations qui lui incombent de produire des statistiques, l'organisme peut autoriser l'utilisation de microdonnées du moment que les utilisateurs assument le coût de mise en place de l'opération. La loi de 2018 sur la protection des données du Danemark prescrit que les renseignements recueillis à des fins statistiques peuvent être traités pour être utilisés à d'autres fins statistiques ou scientifiques. La loi sur l'administration publique de 1985 dispose que les individus qui ont accès aux microdonnées à des fins de recherche sont juridiquement tenus à la discrétion et le Code pénal danois sanctionne la violation de la confidentialité par une amende ou une peine d'emprisonnement pouvant atteindre six mois.

11. Les mesures de sécurité des données prises par Statistics Denmark présentent les caractéristiques suivantes :

- Les microdonnées sont désidentifiées – l'utilisateur n'obtient pas l'accès à l'identifiant personnel unique mais à un nombre qui ne contient aucune information au sujet de l'individu concerné ;
- Les microdonnées ne quittent jamais Statistics Denmark – elles sont hébergées par des serveurs dédiés spéciaux qui sont dissociés du système de production statistique de l'organisme ;
- Seuls les résultats qui ne permettent pas d'identifier les personnes et les entreprises doivent être envoyés depuis la machine/le serveur sur lesquels s'effectuent la recherche ;
- Les cellules doivent contenir plus de 3 observations (au minimum) ;
- Chaque résultat est analysé pour détecter les microdonnées avant d'être envoyé ;
- Le chercheur reçoit un avertissement si le système pressent que le résultat contient des microdonnées ;
- Si les utilisateurs emportent des résultats qui enfreignent les règles, leur accès aux microdonnées est fermé pendant un mois et ils doivent produire un plan garantissant que cette infraction ne se reproduira plus.

12. Pour chaque projet de recherche, le centre de données crée un ensemble de données qui lui est dédié. Les ensembles de données sont créés sur la base du contenu du descriptif de projet en concertation avec les chercheurs suivant le principe de ce que l'on a « besoin de savoir ». Les établissements les plus importants qui mènent de nombreux projets de recherche à la fois peuvent obtenir le droit de créer une base de données qui réponde aux besoins d'un plus grand nombre de projets. Le centre de données a pour but de fournir les données convenues un mois au plus tard après la conclusion de l'accord.

13. Ces dernières années, le centre de données destiné aux chercheurs et aux utilisateurs des entreprises privées a été complété par un nouveau dispositif s'adressant aux organismes publics qui offre un accès plus large et plus rapide aux données.

14. Cette évolution répond à la tendance qu'ont les principaux utilisateurs à vouloir un large accès – pratiquement illimité – aux données. Le contexte de cette évolution et les difficultés qu'elle pose font l'objet de la section ci-après du présent document.

15. Le type de données demandées a dicté l'organisation du personnel du centre de données qui se divise en quatre équipes :

- Équipe chargée des clients des ministères ;
- Équipe chargée des clients du secteur de la santé ;
- Équipe chargée des clients du secteur privé ;
- Équipe chargée du grand public.

16. Certaines données sont plus demandées que d'autres. Comme on peut le voir sur la figure 2, celles relatives à la population, à l'éducation et aux revenus se classent aux trois premières places.

Figure 2

Demandes de données adressées au centre de données de Statistics Denmark en 2017, par registre

Registres	Désignation	Classement par nombre d'extractions de données en 2017
BEF	Population (situation au 31 décembre)	1
UDDA	Éducation (situation)	2
IND	Revenus des personnes physiques	3
AKM	Module des nomenclatures de travail	4
LPRDIAG	Patients hospitalisés – diagnostics	5
RAS	Statistiques sur la population active fondées sur des registres	6
LMDB	Base de données sur les médicaments délivrés sur ordonnance	7
IDAP	Base de données intégrée sur la recherche sur le marché du travail	10
IDAN	Base de données intégrée sur la recherche sur le marché du travail – données relatives à l'emploi et aux salaires	11
FAM	Constitution des familles	12

III. Évolutions futures possibles – difficultés et possibilités

17. La demande de données et d'accès rapide aux données semble illimitée. Statistics Denmark compte un nombre considérable d'utilisateurs réguliers du centre de données – et bien que ceux-ci se félicitent généralement de l'accès aux données qui leur est donné – ce sont aussi des clients exigeants qui veulent toujours plus de données et plus vite.

18. Comme indiqué précédemment, l'accès aux données pour les projets de recherche et d'analyse individuels mis en place il y a environ trente ans a récemment été complété par un accès plus large et plus rapide aux données pour les analystes en poste dans des ministères. Ce groupe d'utilisateurs n'établit pas de descriptifs de projets comme le font les chercheurs. Souvent, il n'est pas réaliste d'attendre un mois lorsque les résultats doivent éclairer des négociations et des décisions politiques. En conséquence, l'accès offert est plus

large mais les restrictions imposées à l'utilisation des données mentionnées au paragraphe 11 s'appliquent toujours.

19. À ce jour, les ministères inscrits au nouveau système sont aux alentours de cinq. L'un d'entre eux est allé plus loin encore et a demandé à Statistics Denmark de bâtir un entrepôt de données qui associe les données de Statistics Denmark aux siennes. L'entrepôt de données est pris en charge par le pare-feu de Statistics Denmark. Les règles relatives à l'utilisation des données et aux résultats qui peuvent être importés sont les mêmes pour les utilisateurs de l'entrepôt de données et pour les autres utilisateurs du centre de données.

20. Il s'agit là du premier changement apporté en réponse à une série de demandes d'utilisation de données gérées par des organismes de statistique à des fins administratives – ce qui est interdit comme le savent tous les statisticiens. Mais au lieu de rejeter cette demande, les organismes de statistique pourraient s'enorgueillir du fait que leurs données sont perçues comme l'étalon or des données bien documentées. Bien sûr, les données statistiques ne peuvent pas être utilisées à des fins administratives pour de bonnes et nombreuses raisons. Mais cela ne veut pas dire que les organismes de statistique ne peuvent pas être utiles aux organismes publics qui souhaitent faire appel à leurs compétences en matière de collecte, de traitement, de vérification et de diffusion de données. Nous devons donc trouver des solutions afin que les connaissances acquises grâce à ces compétences puissent aussi être utilisées à des fins autres que statistiques.

21. Lorsque les organismes de statistique cessent progressivement d'être des usines de production de statistiques pour devenir des services de gestion de données pour le secteur public et dans l'intérêt général, cela doit se faire de façon explicite et être communiqué au public. La transparence la plus totale doit régner quant à l'usage qui est fait des données, la manière dont leur exploitation reste sécurisée et les raisons pour lesquelles les organismes de statistique entreprennent une activité nouvelle de cette nature.

IV. Conclusions

22. Dans un monde où les données sont reines, il est naturel que le rôle des organismes de statistique soit remis en question et évolue pour relever les défis.

23. Il est de plus en plus manifeste que les organismes de statistique non seulement détiennent des données dont on a besoin pour de nouveaux usages comme l'accès aux microdonnées – mais possèdent aussi des compétences en matière de collecte, de traitement et de documentation des données. Celles-ci seront de plus en plus demandées par d'autres composantes du secteur public qui souhaitent faire appel à elles pour être absolument sûres d'étayer aussi au mieux leurs décisions administratives. Cela pose des difficultés mais ouvre aussi aux organismes de statistique des possibilités d'utiliser leurs compétences dans l'intérêt général plus largement qu'ils ne le faisaient initialement.