

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

Genève, 23-25 juin 2021

Point 2 b) de l'ordre du jour provisoire

**Poursuite de l'innovation dans l'organisation et dans les produits
des organismes nationaux de statistique après la COVID-19 :****Innovation dans l'organisation et les modalités de travail
des organismes nationaux de statistique****Gestion des données opérationnelles relatives à la COVID-19 –
Modèle de production pour la nouvelle génération
de statistiques*****Note établie par la Lituanie***Résumé*

Le présent document passe en revue l'expérience de Statistics Lithuania dans la gestion des données opérationnelles relatives à la pandémie de COVID-19, gestion qui revêt une importance vitale dans la lutte, le suivi et la prise de décisions visant, face à cette pandémie, à préserver la santé et les vies humaines. La situation d'urgence a révélé l'importance de disposer d'un espace au sein de l'État où l'ensemble de l'écosystème des données puisse être régi et qui permette de répondre aux besoins d'information allant de la collecte des données au recours à des services extérieurs, en passant par une gestion souple des droits des utilisateurs, afin de prendre des décisions « ici et maintenant ».

Cette tâche importante a renforcé Statistics Lithuania en tant qu'institution et a augmenté l'efficacité des activités de l'organisation, qui était auparavant insuffisante. Cependant, si elle a eu une forte influence sur la croissance de la réputation et de la notoriété de Statistics Lithuania, cette étape représente surtout un changement stratégique à long terme dans l'activité principale de l'organisation – la production de statistiques officielles. Un modèle efficace de gestion des données, élaboré en peu de temps, est mis en place afin de produire des statistiques officielles et expérimentales. Cela permet d'espérer des changements importants dans le niveau de détail et la pertinence des statistiques, ainsi que de nouveaux services de qualité au profit de l'État, de la science, des entreprises et de la société. Les décisions opportunes de Statistics Lithuania et des actions rapides pour rendre opérationnelle la gestion des statistiques relatives à la COVID-19 ont créé les conditions préalables à la mise en œuvre d'une ambition beaucoup plus grande – changer fondamentalement le rôle de Statistics Lithuania et devenir un gestionnaire des données publiques.

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



Ce document est présenté pour discussion à la session de la Conférence des statisticiens européens consacrée au point de l'ordre du jour intitulé « Poursuite de l'innovation dans l'organisation et dans les produits des organismes nationaux de statistique après la COVID-19 : Innovation dans l'organisation et les modalités de travail des organismes nationaux de statistique.

I. Introduction : aperçu de la situation antérieure à la COVID-19 en ce qui concerne la gestion des données publiques

1. En Lituanie, les données du secteur public sont entièrement décentralisées. Il en résulte une grande diversité de sources administratives sur les plans technologique et fonctionnel. La forte fragmentation des sources de données et ses conséquences négatives sur la capacité à utiliser les données sont devenues évidentes avec le début de la pandémie de COVID-19. Dans le cas de la Lituanie, la collecte et le traitement des données opérationnelles ont constitué un problème majeur dans la gestion de la pandémie. Et c'est la pandémie de COVID-19 qui a inspiré des décisions décisives et rapides dans le domaine de la gestion des données publiques. Statistics Lithuania est devenu un acteur central, démontrant son ambition d'assumer un rôle fondamentalement nouveau en tant que gestionnaire de données.

2. Quelle est la situation actuelle de la gestion des données publiques en Lituanie ? Officiellement, il existe 96 registres et 285 systèmes publics d'information. Ces registres et systèmes d'information sont gérés par 139 responsables différents et exploités par 168 responsables principaux. La plupart des systèmes et des registres répondent principalement à des objectifs opérationnels internes, reposent sur la capacité financière et les compétences des propriétaires ou des gestionnaires et ne tiennent pas nécessairement compte de la nécessité d'une approche systématique des données publiques, du contexte général, des besoins des utilisateurs, etc.

3. Par conséquent, les problèmes à partir desquels nous avons commencé à mettre en œuvre notre rôle d'intendant des données officielles sont clairs :

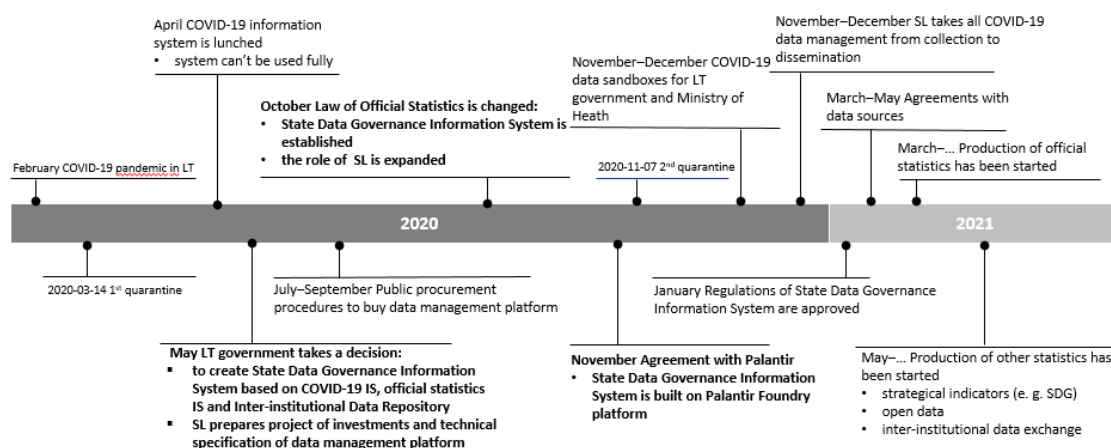
- Les données isolées et fragmentées ne créent pas beaucoup de valeur. Les informations utiles sont obtenues en combinant et en analysant les données provenant de différents domaines. Or, la multiplicité des sources de données fondées sur des solutions technologiques très diverses, déjà mentionnée, rend les données du secteur public inopérantes. Autrement dit, ces données ne sont destinées qu'à satisfaire les besoins internes limités de leurs propriétaires et ne peuvent être utilisées dans la pratique qu'à cette fin. La plupart des registres et des systèmes d'information sont fermés par nature et axés uniquement sur les besoins de leurs propriétaires. Les processus d'ouverture ou de réutilisation des données ne sont pas centralisés ;
- Des statistiques officielles rigides et lentes. Les statistiques officielles sont produites sur la base de la législation européenne et nationale, dans le strict respect des méthodes et des délais établis. Cela permet de garantir la qualité des informations statistiques. Toutefois, de nos jours, les utilisateurs ont davantage besoin de données récentes et granulaires que de données très précises. En moyenne, nous préparons les informations statistiques annuelles en 150 jours, trimestrielles en 58 jours et mensuelles en 24 jours, ce qui n'est pas acceptable dans le contexte d'une pandémie ou d'autres catastrophes éventuelles. Les enquêtes statistiques traditionnelles ne permettent pas d'atteindre un haut niveau d'efficacité et augmentent également la charge de travail liée à l'établissement des statistiques. Pour ces raisons, l'utilisation de sources administratives est l'une des principales solutions. Elle permet de produire des statistiques « plus rapides » qui rendent compte des changements socioéconomiques en temps quasi réel. Par conséquent, les données publiques deviennent de plus en plus importantes dans notre production. Et c'est là que nous nous heurtons à un autre problème de taille : bien que nous ayons le droit de collecter des données administratives à des fins statistiques officielles, nous ne sommes pas en mesure d'exploiter pleinement le potentiel de ces données, car nous n'avons pas le droit de les utiliser à d'autres fins, contrairement aux propriétaires de ces données (conformément au règlement général sur la protection des données (RGPD), bien sûr) ;
- Des initiatives inefficaces d'échange, de réutilisation et d'ouverture des données. Tant le secteur public que les entreprises possèdent d'énormes ressources de données. Toutefois, le partage de ces données est compliqué en raison de la fragmentation des sources de données mentionnée précédemment. Les utilisateurs sont obligés de contacter les différents propriétaires ou gestionnaires de données pour obtenir des

données du secteur public, selon diverses procédures. Les entreprises ne sont pas non plus motivées pour partager leurs mégadonnées à forte valeur ajoutée avec l'État. Les initiatives existantes d'ouverture des données ne résolvent pas le problème de la fragmentation des données du secteur public et n'augmentent pas leur accessibilité. En outre, l'ouverture des données augmente les coûts supportés par les propriétaires pour préparer et fournir ces données ou en rendre la réutilisation possible. En outre, ces coûts augmentent alors que la valeur des données ouvertes reste discutable en raison de défauts de qualité. La fragmentation de la collecte et de l'ouverture des données empêche la science et les entreprises d'utiliser ces sources d'informations publiques pour développer des produits à forte valeur économique et sociale.

4. Les problèmes mentionnés ci-dessus ont conduit Statistics Lithuania à créer un système de gouvernance des données (ou, pour ainsi dire, un lac de données publiques) qui a été lancé en novembre 2020. Statistics Lithuania a acquis une plateforme de gestion des données prête à l'emploi, qui permet de créer un écosystème de données durable et flexible disposant de services d'infrastructure en nuage. Comme nous l'avons déjà mentionné, l'idée de créer un système de gouvernance des données a été (et est toujours) motivée par la pandémie de COVID-19.

Figure 1

Chronologie : de l'idée à la mise en œuvre



5. Repères chronologiques : décision du Gouvernement sur la mise en place du système de gouvernance des données publiques, modification de la loi sur les statistiques officielles et achat d'une plateforme de gestion des données permettant de mettre en œuvre le système d'information. Nous sommes actuellement dans le processus d'assimilation des données primaires à partir des sources de données et de production. Cela signifie que nous signons des accords avec les propriétaires de données, organisons la collecte de données à partir des systèmes de ces derniers et commençons le transfert des enquêtes statistiques et la production d'informations statistiques dans le système de gouvernance des données publiques. En parallèle, nous établissons des modèles pour garantir la confidentialité des données, l'évaluation de la qualité et la prestation de services. Tout cela se fait dans le cadre d'un processus de gestion des données relatives à la COVID-19 déjà existant, ce qui signifie que ce que nous voulons réaliser à grande échelle a déjà été réalisé avec ces données.

II. Base juridique : un timide pas en avant

6. En vertu des nouvelles dispositions de la [Loi sur les statistiques officielles](#), Statistics Lithuania devient le propriétaire et le gestionnaire du système de gouvernance des données publiques, dont le champ d'application couvre un espace de données et des objectifs d'utilisation plus larges que ceux des statistiques officielles. Ainsi, Statistics Lithuania devient un acteur à double rôle : il utilise le nouveau système pour produire des statistiques officielles et, en même temps, il obtient le droit de collecter des données et de produire des informations autres que des statistiques officielles.

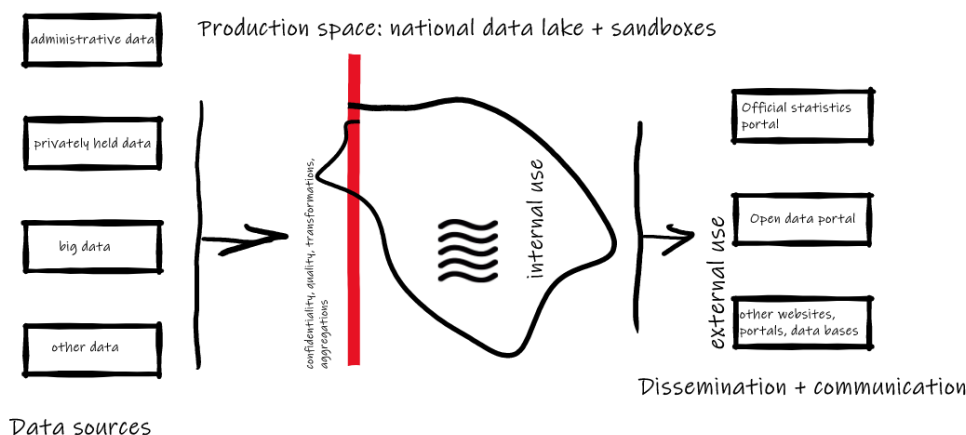
7. Il est à noter que nous constatons que la modification de la loi a légèrement augmenté la capacité de Statistics Lithuania à répondre aux besoins d'informations de l'État et de la société, mais pas suffisamment, car il existe des contradictions et des obstacles importants en matière d'ouverture et de réutilisation des données. Les dispositions de la Loi sur les statistiques officielles ne permettent d'utiliser les ressources du système de gouvernance des données publiques qu'à des fins statistiques. Cela signifie que :

- Les données ne peuvent pas être fournies à des fins administratives, à moins que ne soient engagées de longues procédures bureaucratiques (par exemple, la signature d'accords tripartites entre le propriétaire des données, Statistics Lithuania et l'utilisateur des données), ce qui rend impossible la mise en œuvre de l'idée principale du système de gouvernance des données publiques, à savoir l'échange et la fourniture rapides de données. Exemples concrets : un espace analytique efficace de recherche sur l'économie souterraine, dirigé par les services de l'inspection fiscale, pourrait être créé dans le système de gouvernance des données publiques, mais sans réglementation juridique supplémentaire, le système ne peut être utilisé à cette fin. Une situation analogue est celle de la fourniture de données détaillées aux municipalités afin d'aider celles-ci à lutter contre la pandémie de COVID-19 et à prendre localement des décisions à ce sujet : techniquement, le système de gouvernance des données publiques peut fournir des données opérationnelles et détaillées (dépersonnalisées ou cryptées) utiles à la prise de décisions concernant la gestion des épidémies, les mesures préventives, la logistique de vaccination, etc., mais le cadre juridique actuel ne le permet pas ;
- L'ouverture des données, bien que techniquement très rapide (par exemple, l'ouverture automatisée des données relatives à la COVID-19 sur le portail national de données ouvertes a déjà été mise en œuvre), doit être centralisée. Statistics Lithuania prévoit d'ouvrir, en coordination avec les propriétaires de données, toutes les sources de données du système de gouvernance des données publiques qui pourraient être nécessaires au public. Cela nécessite des données primaires, car c'est le seul moyen de garantir une ouverture maximale des données. Toutefois, il existe un risque que, en l'absence d'obligation légale, les propriétaires de données ne soient pas enclins à partager les données qu'ils détiennent ;
- Du fait de la forte fragmentation des données du secteur public dans différents systèmes et registres, la réception des données dans le système de gouvernance des données publiques, dont la mise en œuvre est techniquement rapide, peut être artificiellement retardée par les gestionnaires de systèmes ou de registres en raison du manque d'intérêt à fournir des données primaires. Cela est déjà évident maintenant, quand les contrats d'acquisition de données font l'objet de négociations relativement longues, que les données sont refusées avec la plus grande facilité, etc. Un autre argument est la législation nationale distincte interdisant la réutilisation des données à des fins autres que celles qui sont spécifiées.

III. La solution et sa mise en œuvre

8. Quelle est la solution pour que Statistics Lithuania devienne le principal acteur de la gestion des données publiques ? C'est très simple : il s'agit de créer un espace de données commun (ou lac de données national) qui permette de consolider et d'exploiter les données publiques importantes. Cela suppose de mettre en place un système de gouvernance des données publiques sur la base d'une plateforme de gestion des données multifonctionnelle et performante, qui permettrait un accès rapide et efficace à différentes sources de données, afin de fournir des services à l'État, aux scientifiques et aux entreprises. Le renforcement du rôle de Statistics Lithuania par la Loi sur les statistiques officielles crée les conditions juridiques préalables nécessaires à la réalisation de cet objectif.

Figure 2
**Espace de production : lac de données national et cadres d'expérimentation
 (bacs à sable)**

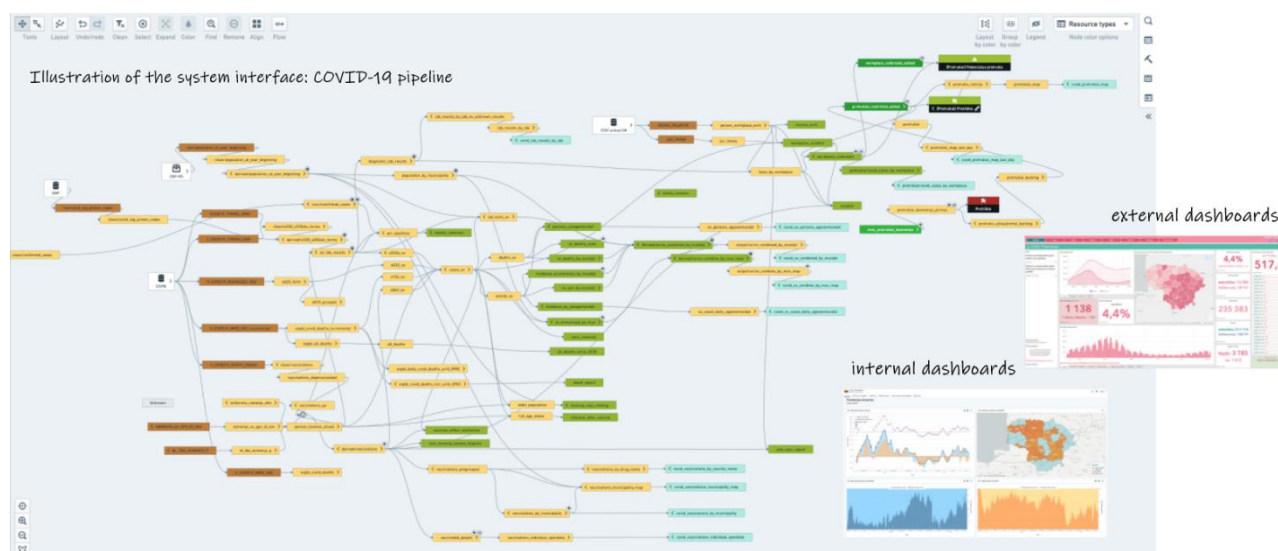


A. Gestion des données relatives à la COVID-19

9. En Lituanie, la principale source de données opérationnelles sur la COVID-19 est le système de santé en ligne géré par le Ministère de la santé et exploité par le Centre des registres, qui est principalement destiné à la collecte et au stockage des données des patients par les établissements de santé et les médecins, ainsi qu'à la fourniture de services électroniques liés aux traitements. Cela signifie que le système n'est pas adapté à l'analyse des données. Ainsi, face à la pandémie de COVID-19, les données collectées dans le système de santé en ligne ne sont pas exploitables dans la pratique. Actuellement, ce système fournit quotidiennement des données primaires relatives, entre autres, au diagnostic, à la morbidité et à la mortalité liés à la COVID-19. On trouvera ci-après un bref aperçu des principaux processus de gestion de ces données : morbidité et mortalité, vaccination et gestion des flambées épidémiques.

- *Processus de gestion des données sur la morbidité et la mortalité.* La figure 3 ci-dessous (et cette courte [vidéo](#)) présentent des exemples de l'interface du système : le parcours complet des données sur les tests, les cas confirmés et les décès liés à la COVID-19, jusqu'au produit final, à savoir les tableaux de bord et rapports internes et externes. Il est possible de répondre aux besoins supplémentaires créés par de nouveaux indicateurs sans détruire la séquence globale des données. Par exemple, si le Ministère de la santé ou un autre organisme public doit soumettre de nouveaux rapports aux organisations internationales, le système de gouvernance des données publiques y pourvoit très rapidement.

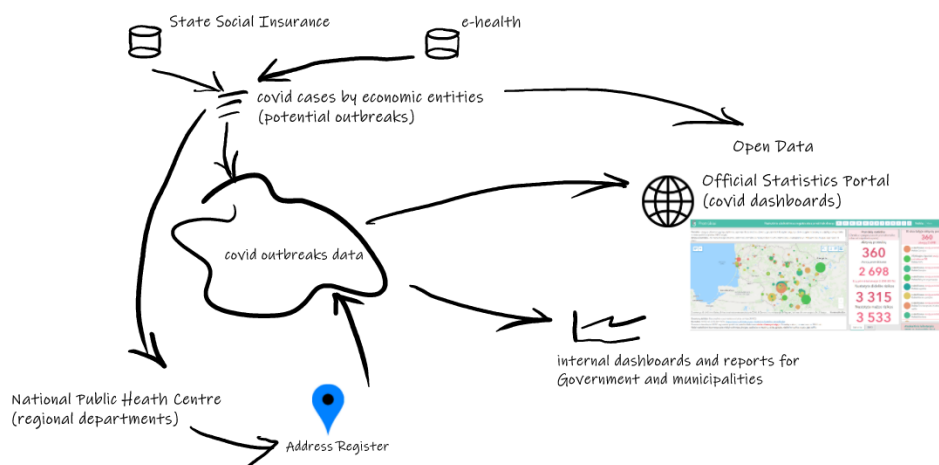
Figure 3

Illustration de l'interface du système : oléoduc, tableau de bord interne et tableau de bord externe de la COVID-19

- *Mise en œuvre de la logistique de vaccination et du processus de gestion des données.* Dans le système de gouvernance des données publiques, le module de données sur la vaccination comprend non seulement la gestion des données mais aussi la logistique des vaccins. C'est un exemple de collaboration réussie avec le Ministère de la Santé – la plateforme gérée par Statistics Lithuania étant, en quelque sorte, « louée » pour des tâches supplémentaires. Le Ministère de la santé saisit les informations sur l'arrivée des vaccins dans le système de gouvernance des données publiques et les vaccins sont automatiquement distribués aux centres de vaccination selon l'algorithme établi. La distribution tient compte de la taille et de la composition de la population par âge et par groupes prioritaires définis par le Gouvernement.
- *Carte des données sur les foyers épidémiques.* La surveillance des flambées de COVID-19 est l'une des tâches les plus difficiles dans la lutte contre la pandémie. Avant d'automatiser la gestion des données sur les foyers épidémiques dans le système de gouvernance des données publiques, les unités régionales du Centre national de santé publique utilisaient leur propre base de données qui, comme le système de santé en ligne, n'a pas de composante analytique ; en outre, toute interface avec d'autres systèmes est difficile à mettre en œuvre en raison d'une technologie obsolète. Par ailleurs, le Gouvernement et le public n'ont pas reçu d'informations claires sur les épidémies et leur gestion. Le processus de gestion des flambées épidémiques diffère des autres processus par le mode de collecte des données : celles-ci sont téléchargées par le personnel des unités régionales et le système de gouvernance des données publiques permet de procéder à une analyse intégrée de la situation et à la diffusion des données.

Figure 4
Données opérationnelles sur les flambées de COVID-19

operational data of covid-19 outbreaks



- *Indicateurs relatifs aux hôpitaux.* La collecte des données hospitalières est organisée de manière similaire à celle des données épidémiologiques. Les établissements de santé régionaux fournissent des données quotidiennes sur les nouveaux patients atteints de la COVID-19, les lits de traitement contre l'infection et la disponibilité des médecins et des infirmières.

10. Toutes les données clés sur la COVID-19 se retrouvent chaque matin dans un [rapport quotidien](#), qui constitue une source de données essentielle pour les médias et le public. Dans le même temps, elles sont mises à jour dans les [tableaux de bord spéciaux de la COVID-19](#), ainsi que sur le [portail de données ouvertes de Statistics Lithuania](#) et le [portail national de données ouvertes](#).

B. Préparation à la production de statistiques officielles

11. Comme le laissent entendre les précédents propos, les objectifs du système de gouvernance des données publiques sont en résumé les suivants : consolidation des données, production d'informations, échange de données et services d'analyse. La portée actuelle (premier semestre 2021) du système de gouvernance des données publiques s'étend à 28 propriétaires de données et à 53 des registres ou systèmes d'information gérés par ces derniers. Cela couvre la quasi-totalité des besoins actuels en matière de statistiques officielles, ainsi qu'une partie importante des indicateurs stratégiques nationaux à construire. Il convient de noter que l'objectif est de consolider les données primaires. Cela permettra la production centralisée d'une nouvelle génération de statistiques, beaucoup plus opérationnelles et détaillées.

12. Tâches prioritaires au cours de cette période : 1) absorption des données ; 2) élaboration et application d'un modèle d'anonymisation/de cryptage garantissant la sécurité des données ; 3) élaboration d'un système de droits des utilisateurs garantissant la sécurité des données ; 4) transfert des algorithmes actuels vers le calcul des indicateurs de la statistique officielle.

IV. Influence sur l'organisation : changements d'image et de réputation

13. Après la prise de contrôle de la gestion et de la diffusion des données relatives à la COVID-19 par Statistics Lithuania, d'importants changements ont été constatés dans la notoriété et la réputation de l'organisation. Celle-ci mesure la valeur de sa production dans la société en calculant deux indices : la facilité d'utilisation des statistiques et la satisfaction des consommateurs. La facilité d'utilisation des statistiques est mesurée par des indicateurs

aits d'évaluation objective : facilité d'utilisation du portail des statistiques officielles, de la base de données sur les indicateurs et des réseaux sociaux, mention de Statistics Lithuania et de sa production dans les médias, demandes individuelles des utilisateurs. L'accessibilité du nom et de la production de Statistics Lithuania dans les principaux médias nationaux est passée de 466,8 millions à 3,8 milliards et, dans le réseau social le plus populaire, Facebook, de 420 000 à 1,0 million, le niveau de satisfaction des utilisateurs étant passé de 72,4 à 75,5.

14. Dans certains cas, Statistics Lithuania a enregistré une croissance multipliée par plusieurs centaines. De toute évidence, cela était dû à la publication des données opérationnelles relatives à la COVID-19. Toutefois, il convient de noter que non seulement les statistiques sur la pandémie mais aussi celles qui portaient sur d'autres domaines ont suscité un intérêt considérable. Cela signifie que Statistics Lithuania et ses activités sont devenus beaucoup plus visibles et connus du public. Il est à noter que la COVID-19 a également inspiré une certaine évolution dans l'utilisation des données ouvertes. Par exemple, les données ouvertes sur la vaccination ont été consultées 500 000 fois depuis la mi-février sur le portail de données ouvertes de Statistics Lithuania et, en moyenne, les fichiers de données ouvertes sont consultés jusqu'à 7 000 fois par jour.

15. En ce qui concerne les changements internes plus subtils au sein de l'organisation, il est nécessaire de mentionner que l'émergence d'une idée commune totalement nouvelle a, en quelque sorte, donné un nouveau souffle à Statistics Lithuania, organisation bureaucratique relativement importante en Lituanie. Avec le projet de système de gouvernance des données publiques, près de 30 nouveaux spécialistes hautement qualifiés ont rejoint Statistics Lithuania et ont apporté une perspective, une culture du travail et des habitudes différentes. En plus des changements structurels apportés dans l'organisation, l'environnement de travail change aussi pendant une pandémie, ce qui inévitablement favorise et encourage le changement : un effet boule de neige semble commencer à se produire.

V. Conclusions

16. La principale conclusion de cette période extrêmement dynamique et de cette expérience extraordinaire est que devenir un responsable des données publiques n'est pas un problème technologique. Il s'agit avant tout d'un problème juridique. Les autorités statistiques ont besoin d'un mandat légal pour élargir leur rôle afin de devenir des gestionnaires de données et pour renforcer leurs propres compétences et données. Actuellement, l'ouverture, le partage et la réutilisation des données sont difficiles à mettre en œuvre en raison des dispositions légales qui restreignent leurs activités. Et, point très important, si les autorités statistiques, qui sont les plus compétentes en matière de gestion des données, n'élargissent pas leur rôle, non seulement en produisant des statistiques officielles, mais aussi en partageant leurs ressources, elles deviendront redondantes car, avec la Loi sur les données, de nouveaux acteurs feront leur entrée sur le marché des données.

VI. Projets à venir : la production centralisée des données ouvertes du secteur public

17. Les objectifs de Statistics Lithuania en 2022-2026 sont liés à la réforme de l'ouverture des données publiques. Le but de cette réforme est d'assurer la disponibilité de données fiables du secteur public et la possibilité de partager ces données entre les secteurs économiques et de les réutiliser, ainsi que de créer les conditions préalables nécessaires pour que les institutions du secteur public prennent des décisions sur la base de données et pour que les entreprises créent des innovations numériques. Il est prévu d'utiliser à cette fin les fonds du mécanisme de relance et de résilience.

18. Dans le contexte de la réforme des données, le système de gouvernance des données publiques est perçu comme un lac de données national permettant de consolider les données publiques importantes qui, à leur tour, grâce à des méthodes normalisées d'anonymisation, de confidentialité et d'ouverture, sont utilisées pour une large gamme de production d'informations : données ouvertes, statistiques officielles, expérimentales et opérationnelles,

suivi du développement stratégique de l'État, tâches gouvernementales spéciales, espaces de données sectorielles, recherche, innovation commerciale.

19. La réforme vise à augmenter la quantité de données publiées dans le système de gouvernance des données publiques en utilisant 376 sources publiques d'informations, à créer les conditions préalables à la réutilisation des données dans tous les systèmes d'information et registres publics et à ouvrir des données de grande valeur au public, aux entreprises et aux scientifiques. Le fait que l'ensemble de l'écosystème des données soit mis en œuvre dans le système de gouvernance des données publiques créé sur la base d'une plateforme de gestion des (méga)données prête à l'emploi montre que ce système est capable de mettre en œuvre d'autres services en plus de son objectif principal : ouverture des données, espaces analytiques (bacs à sable), extraction de mégadonnées à partir de sources commerciales ou autres et divers échanges entre ces sources et avec le monde extérieur.

20. La réutilisation des données détenues par les organismes du secteur public crée une valeur ajoutée au profit des réutilisateurs, des utilisateurs finaux et de la société dans son ensemble. Parallèlement, dans de nombreux cas, c'est l'organisme du secteur public lui-même qui en bénéficie, car en favorisant la transparence et la responsabilité et en recevant un retour d'information de la part des utilisateurs et des utilisateurs finaux, il est encouragé à garantir la qualité des informations collectées et l'exécution de ses missions publiques.
