Comité des statistiques publiques

18 octobre 2021 10h00-12h00	Compte-rendu	Version validée

Présences

Nom	Organisation	Fonction
Ferdy ADAM	STATEC	
Serge ALLEGREZZA	STATEC	Président
Ludwig BALMER	Centre de gestion du personnel et de l'organisation de l'État	
Roger BARTHELMY	Service d'économie rurale	
Françoise BERTHET	Ministère de la Santé	
Laurence BRASSEUR	Ministère de la Culture	
Simone CASALI	STATEC	
Eric DE BRABANTER	Ministère de l'Environnement, du Climat et du Développement durable	
Filipe FERREIRA	Association d'assurance accident	
Marc FERRING	STATEC	
Clémentine FRY	Observatoire de la formation	
Lucia GARGANO	STATEC	
Daniel GENGLER	Ministère de l'Énergie et de l'Aménagement du territoire, Département de l'Énergie	
Pierre GRAMME	Agence pour le développement de l'emploi (ADEM)	
Petra GROEN	Ministère de l'Énergie et de l'Aménagement du territoire	
John HAAS	STATEC	Vice-président
Michel HAAS	Ministère des Finances	
Henri HANSEN	Administration de la gestion de l'eau	
Martine HILDGEN	Ministère de la Digitalisation	
Danielle HORPER	Administration du Cadastre et de la Topographie	
Max JENTGEN	Ministère de l'Économie, Observatoire de la Compétitivité	
Xavier KAMPHAUS	Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse	
Marc KOHLL	Institut luxembourgeois de régulation	
René KRIPPES	STATEC	

Thierry LALLEMANG	Commission nationale pour la protection des données	
Sonia LEITE	Ministère de la Santé	
Julien LICHERON	Observatoire de l'habitat	
Anne-Charlotte LORCY	Ministère de la Santé	
Thierry MAZOYER	Inspection générale de la sécurité sociale	
Elisa MAZZUCATO	Ministère de l'Éducation nationale, de l'Enfance et de la Jeunesse	
Yves MELAN	Administration de l'enregistrement et des domaines	
Christian MERSCH	Ministère de la Mobilité et des Travaux publics	
Carole MOLLING	Ministère de la Famille, de l'Intégration et à la Grande Région	
Charles MULLER	Service d'économie rurale	
Marc PAULY	STATEC	Secrétaire
François PELTIER	STATEC	
Virginie RAYMOND	Réseau d'étude sur le marché du travail et de l'emploi (RETEL) et Ministère du Travail de l'Emploi et de l'Économie sociale et solidaire	
Roland RICHARDY	Administration de l'enregistrement et des domaines	
Bruno RODRIGUES	Ministère de l'Enseignement supérieur et de la Recherche	
Luc ROETTGERS	STATEC	
Robert SCHEUEREN	Administration des douanes et accises	
Christiane SCHMIT	Commissariat aux assurances	
Marc SCHNEIDER	LISER	
Philippe SCHRAM	Ministère de l'Intérieur	
Marc SCHUMAN	Ministère de l'Environnement, du Climat et du Développement durable	
David SOPPELSA	STATEC	
Giulia SPALLETTI	Ministère de l'Économie, Observatoire de la compétitivité et Observatoire de la formation des prix	
Astrid SPREITZER	Ministère de la Fonction publique	
Olivier THUNUS	STATEC	
Eric WEBER	Ministère de la Digitalisation	
Guy WEBER	Ministère de la Santé	

Ordre du jour

1. Mot de bienvenue et introduction du Président
2. Présentation de projets
   - Infrastructure nationale de pseudonymisation et d'anonymisation de données pour le gouvernement du Luxembourg (Éric Weber, Ministère de la Digitalisation)
   - Recensement de la population de 2021 (François Peltier, STATEC)
   - Peer Review du Système statistique européen (Simone Casali, STATEC)
   - Initiative Data Science (Marc Pauly, STATEC)
3. Rapports des groupes de travail du Comité des statistiques publiques
4. Discussion libre
5. Conclusion du Président

1. Mot de bienvenue et introduction du Président

Serge Allegrezza souhaite la bienvenue à tous les membres du Comité des statistiques publiques, dont la grande majorité assiste à la réunion par visioconférence. Il salue en particulier les collègues du Ministère de la Santé, qui avaient contribué une présentation en 2020 ; il se réjouit également de la création récente de l'observatoire national de la santé. Sa directrice, Dr Françoise Berthet, assiste à la réunion du Comité.

Serge Allegrezza rappelle aux membres du Comité des statistiques publiques qu'ils constituent le Système statistique luxembourgeois, et qu'ils contribuent à ce que notre pays se dote d'un appareil statistique. Tous sont des producteurs de statistiques d'intérêt général. Le STATEC a le rôle du coordinateur, de l'animateur.

Certains membres ont pu suivre le cours sur le Code de bonnes pratiques assuré par Simone Casali. Ce Code est la charte de qualité à appliquer en matière de statistique partout en Europe et donc au Luxembourg. Il est important de connaître ce Code et d'en respecter les principes.

2. Présentation de projets

2.1. Infrastructure nationale de pseudonymisation et d'anonymisation de données pour le gouvernement du Luxembourg

Le projet est présenté par Éric Weber du Ministère de la Digitalisation. Aujourd'hui le traitement et l'échange de données au niveau national soulèvent des risques techniques et juridiques. L'infrastructure Nationale de Pseudonymisation / Anonymisation (INPA) souhaite apporter une réponse à ces défis, en proposant une solution technologique pour l'anonymisation ou la pseudonymisation de données, en identifiant et anticipant les risques liés à la mise en place d'une telle infrastructure, et enfin en mettant en place un pilote pour tester la validité et la viabilité d'une solution. Les définitions de la pseudonymisation et de l'anonymisation sont évoquées ; les deux méthodes visent, à des degrés divers, à garantir l'anonymat des personnes concernées. Six cas d'usage sont envisagés, du développement logiciel au RGPD, en passant par les data warehouses et le open data.

Le projet, en collaboration avec le cabinet de conseil PwC et le cabinet juridique Nautadulith, se subdivise en 4 lots : analyse des besoins, conception de la solution, cahier des charges et analyse juridique. Le kickoff meeting a déjà eu lieu. Parmi les partenaires on trouve, outre le Ministère de la Digitalisation, les Ministères de la Santé, de l'Enseignement supérieur, de l'Économie, de la Sécurité sociale et de la Justice, ainsi que l'IGSS, l'ADEM, l'Agence e-Santé et le STATEC.

Serge Allegrezza remercie l'orateur. Il juge l'initiative très utile pour des producteurs de statistiques qui ont besoin d'accès à des données. Il convient de faire l'arbitrage entre les besoins respectifs en matière d'analyse, d'un côté, et le besoin en pseudonymisation de l'autre côté. Le programme AI4GOV du Ministère de la Digitalisation est également impacté par cette question. Le STATEC a pu y participer en 2021.

Thierry Lallemang salue l'annonce. La CNPD demande depuis longtemps qu'une telle infrastructure soit mise en place, certes les travaux sont encore à un stade précoce, mais on peut se réjouir du démarrage. Éric Weber précise dans ce contexte que le marché de l'implémentation reste encore à attribuer. Thierry Lallemang demande ensuite quand l'infrastructure serait opérationnelle. Éric Weber pense que la réponse à cette question ne pourra être formulée qu'à partir des besoins collectés. Actuellement on utilise le mot « infrastructure », qui garde un caractère vague, mais c'est à travers la collecte des besoins qu'on pourra décider comment mettre en place concrètement le projet. Il faut comprendre ces besoins et on attend ensuite de la part du consultant qu'une proposition soit formulée.

Thierry Mazoyer intervient pour signaler que le Ministère de la Santé dispose aussi d'un service de pseudonymisation, et que l'IGSS a fourni du travail dans cette direction. Éric Weber confirme que l'IGSS a collecté des expériences intéressantes à ce niveau. Serge Allegrezza illustre le propos : le STATEC a participé au projet « Santé pour tous », en utilisant un bureau virtuel permettant d'accéder à des données de l'IGSS. Il était compliqué de débloquer l'accès, mais une fois cette étape franchie, les agents du STATEC en charge de cette étude ont pu travailler de manière libre et sécurisée.

Françoise Berthet aimerait savoir si le terme « infrastructure » inclut aussi la notion de sa gouvernance sur le plan technique. Éric Weber confirme. Il faut définir les rôles et responsabilités, et la gouvernance en fait partie.

Serge Allegrezza exprime le souhait que le Comité des statistiques publiques soit tenu au courant de l'évolution de ce projet prometteur.

2.2. Recensement de la population de 2021

François Peltier de l'unité « Population et logement » résume pour le Conseil les principales caractéristiques de l'opération du recensement de la population, dont la collecte aura lieu en novembre 2021. Le premier recensement luxembourgeois remonte à 1821, celui de 2021 sera la 37ème édition. Dans les temps anciens, le recensement servait simplement à compter la population ; de nos jours, les recensements ont gagné en complexité. Le STATEC aura des données socio-économiques et démographies, mais aussi des données sur le handicap pour satisfaire une demande formulée par le Ministère de la Famille. Par opposition aux autres enquêtes du STATEC, on travaille ici avec la population exhaustive et non plus avec un échantillon. On pourra dès lors descendre à des niveaux géographiques fins (jusqu'à des cellules carrées 1km x 1km). Le recensement sera tourné vers le digital et utilisera une combinaison de sources (questionnaires et 3 registres administratifs). Idéalement le STATEC voudrait faire ses recensements à 100% par recours à des registres administratifs, ce qui est

déjà la pratique dans certains pays. Outre les réponses aux questionnaires, le STATEC utilisera le Registre National des Personnes Physiques (source administrative principale), des données de l'Inspection générale de la sécurité sociale et enfin des données du Registre des Bâtiments et des Logements (dont le STATEC est le gestionnaire). La collecte par Internet se déroule du 8 au 28 novembre. Les ménages n'ayant pas répondu par voie électronique seront ensuite contactés par les agents recenseurs. Généralement, la distribution du questionnaire papier (à partir du 22 novembre) relance la participation par voie digitale. Le STATEC espère un maximum de participation par la voie digitale, qui inclut aussi des appareils mobiles comme la tablette et le smartphone. Le système fonctionne via MyGuichet.lu, avec ou sans carte LuxTrust, toutefois il n'y a pas de persistance des réponses en cas d'authentification faible. Pour inciter la population à participer massivement, le STATEC organisera une loterie. Le questionnaire digital s'adapte au fur et à mesure que l'utilisateur fournit ses réponses. Le recensement propose 5 langues : les trois langues officielles, ainsi que le portugais et l'anglais. Plus le taux de participation digitale est élevé, plus le traitement des données sera rapide. 2000 agents recenseurs accompagnent l'opération du recensement 2021. La remise du questionnaire papier peut se faire à l'agent recenseur ou à la commune ; il existe aussi la possibilité de l'envoyer par courrier au STATEC. Une campagne de publicité en 5 langues sur divers médias (presse écrite, affiches dans parkings et abribus, cinémas, radio, sous-verres etc.) annoncera l'événement au grand public, à partir du 18 octobre. Une aide aux citoyens sera mise en place pendant la phase de la collecte (Hotline téléphonique dédiée, adresse mail dédiée, site web dédié). Le 20 septembre, le STATEC a tenu une conférence de presse avec le ministre de l'Économie Franz Fayot, et le même jour il a fait une présentation à la Chambre des députés.

Serge Allegrezza souhaite commenter le contexte. Le recensement est l'opération statistique la plus importante dans le pays, et aussi la plus chère. Eurostat voudrait augmenter la fréquence des recensements. La stratégie du STATEC consiste à aller vers des recensements s'appuyant sur des registres de données administratives. Mais il faut reconnaître que nos registres ne contiennent pas encore beaucoup de données, c'est pourquoi le recours aux questionnaires reste nécessaire. En 2021, le STATEC mise fortement sur la participation digitale du public. Avant la pandémie, on pouvait encore observer une certaine résistance à l'outil informatique. Mais entretemps, le Luxembourg est devenu le pays européen qui pratique le plus le télétravail, et les compétences digitales de la population se sont développées avec la pandémie. On passe plus souvent par MyGuichet, notamment pour le Large Scale Testing. Le recensement de la population est ainsi un grand test, dont on espère pouvoir tirer des leçons.

Marc Kohll demande si on pourra dire des choses fines en termes de densité de population, notamment dans les zones rurales. François Peltier estime que le STATEC sera en mesure de le faire.

Serge Allegrezza déplore que le STATEC ait dû retarder à deux reprises la collecte des données pour le recensement. Maintenant la situation s'est améliorée à ce niveau, et l'opération est bien sur les rails.

Guy Weber pose la question si un citoyen a la possibilité de valider les données issues des trois registres employés. François Peltier dit qu'il n'a pas été possible de préremplir les questionnaires électroniques avec les données provenant des registres. Bien évidemment, les citoyens sont informés que le STATEC se procure un accès aux données administratives les concernant.

2.3. Peer Review du Système statistique européen

Simone Casali présente au Comité des statistiques publiques le déroulement du prochain round – le 3ème – de la Peer Review du Système statistique européen. L'objectif est d'évaluer la conformité au Code de bonnes pratiques, d'améliorer la qualité des statistiques et la confiance dans celles-ci, et de formuler des recommandations pour développer les systèmes statistiques nationaux.

Le calendrier se subdivise en 4 phases :

1	Self-assessment questionnaire (SAQ), à remplir par le STATEC et les trois ONAs (other national authorities)	Juillet-octobre 2021
2	Visite des experts au Luxembourg	**24-28 janvier 2022**
3	Recommandations et rapport d'audit	Février-avril 2022
4	Plans d'actions, à rédiger par STATEC et les ONAs audités	Avril-juillet 2022

Le SAQ complété et ses pièces justificatives doivent être remis pour le 22 octobre 2021, en langue anglaise. Pour les ONAs (IGSS, Santé, Service d'économie rurale), le SAQ est allégé.

De très nombreuses fonctions du STATEC et structures du système statistique luxembourgeois seront impactées par la Peer Review. Après la visite des experts sur site, ceux-ci rédigent un rapport avec des recommandations d'amélioration. Il incombe alors au STATEC et aux ONAs de formuler des actions d'amélioration en réponse à ce rapport. À un stade ultérieur de la procédure, il y aura un suivi de la part d'Eurostat.

Serge Allegrezza insiste sur l'importance de faire cet exercice avec sérieux. Dès qu'il y aura les résultats de la Peer Review pour le Luxembourg, le STATEC les partagera avec le Comité des statistiques publiques. Ce sera l'occasion de découvrir quelles forces et quelles faiblesses auront été constatées. Le gouvernement sera lui aussi confronté avec les conclusions de la Peer Review, dont le but premier est d'améliorer le Système statistique européen.

2.4. Initiative Data Science

Marc Pauly propose au Comité un aperçu du projet « Initiative Data Science ». Alors qu'une collecte traditionnelle de données s'adresse directement aux répondants détenteurs de l'information statistique, une approche innovante serait de recourir à une collecte indirecte de cette information. L'essor de nouvelles technologies, notamment dans le domaine informatique, permet de collecter aujourd'hui de grands volumes de données à grande fréquence. On peut penser à la téléphonie mobile, aux réseaux sociaux, aux données de transactions financières, ou encore aux compteurs d'électricité intelligents. Afin de pouvoir traiter efficacement ces quantités importantes de données et d'en extraire le contenu statistique en minimisant la marge d'erreur, il est nécessaire de développer un corpus théorique qui s'appuie à la fois sur la statistique mathématique, l'algorithmique et l'ingénierie hardware. Ce corpus est souvent appelé « data science », dans la mesure où l'objet du traitement est la donnée, vue comme matière première à transformer pour aboutir à un produit fini, dans ce cas un produit statistique. L'objectif de l'initiative Data Science au STATEC est d'installer des compétences en data science de manière durable et organique. Le moyen en serait le traitement de cas d'usage ayant une application immédiate en statistique officielle et nécessitant un apport en données tirées de nouvelles sources, généralement abondantes mais dont l'adéquation avec les critères de qualité de la

statistique officielle ne peut pas être garantie. L'un des volets de ce projet est l'étude, par le STATEC, en collaboration avec POST et une firme de consultance en data science et avec le soutien du Ministère de la Digitalisation, d'un jeu de données issu de la téléphonie mobile. Cette étude a permis de mettre en évidence l'importance du bon choix les traitements à effectuer (temps de calcul longs) et d'une communication soutenue avec le producteur initial des données. L'équipe projet a pu construire un prototype de classificateur des détenteurs de téléphone en 4 profils : résident, frontalier, proche frontière, indéterminé. Par ailleurs, certains effets de la pandémie COVID-19 peuvent être lus sur les données, sans que cette problématique ait pour le moment été approfondie. Les perspectives de ce volet téléphonie mobile restent nombreuses, et incluent notamment une étude « à lieu et temps fixé », par opposition à l'approche « à appareil fixé » poursuivie jusque-là. Il sera également primordial de confronter les résultats obtenus à des chiffres supposés mesurer le même phénomène, mais avec une méthode différente.

Serge Allegrezza est convaincu que la science des données est le futur de la statistique. Tout le monde ou presque en prend aujourd'hui la direction. Il demande s'il existe des expériences dans ce domaine chez les membres du Comité. Bruno Rodrigues fait remarquer que le CTIE apporte son soutien à la technologie DevOps. Il serait intéressant que les data scientists utilisent cette technologie et aient recours aux outils associés.

3. Rapports des groupes de travail du Comité des statistiques publiques

3.1. Logement

Marc Ferring annonce que le groupe de travail s'appelle désormais « Statistiques dans le domaine de l'immobilier », dans la mesure où ce nom est plus cohérent par rapport à la demande de développer des indicateurs dans l'immobilier commercial. Les membres du groupe sont la BCL, la CSSF, l'AED, le LISER et le STATEC. La prochaine réunion se déroulera le 20 octobre 2021.

Serge Allegrezza voit ce groupe de travail comme une réussite, car on travaille maintenant de façon coordonnée, dans un domaine qui suscite de l'intérêt. Ici comme ailleurs, il est important de disposer de données d'une bonne qualité. L'Administration de l'enregistrement et des domaines est louée pour des livraisons de données hautement utiles.

3.2. Finances publiques

David Soppelsa situe la fondation de ce groupe de travail en 2013. Eurostat voulait voir un cadre plus formel pour les interactions entre le STATEC et ses fournisseurs de données, ceci dans le contexte de la notification *Early Deficit Procedure* (EDP). Le groupe est chargé d'assurer l'échange de données statistiques et d'informations ayant une incidence sur les comptes des administrations publiques. La prochaine réunion de ce groupe de travail sera pour le mois de décembre.

Serge Allegrezza déplore que les visites de dialogue effectuées par Eurostat tiennent parfois davantage de la surveillance que du dialogue.

3.3. Environnement

Olivier Thunus rapporte que le groupe de travail se réunit habituellement deux fois par an. Parmi ses sujets de travail, on peut noter la mise à jour de trois tableaux de bord figurant sur le portail des statistiques, consacrés respectivement à l'énergie, à l'économie verte, et aux indicateurs relatifs au changement climatique.

Le STATEC a décidé de créer un site dynamique réunissant les 17 objectifs de développement durable sur sa page d'accueil, depuis laquelle on peut ensuite se focaliser sur l'objectif de son choix. Ce site sera mis en ligne dans les prochains jours.

Serge Allegrezza déclare que l'environnement compte parmi les thèmes les plus importants traités par le STATEC, et qu'il est par conséquent classé en priorité élevée. Il existe des programmes de décarbonation aux horizons 2030 et 2050, de tels sujets ont une grande pertinence dans le débat public. Il s'agit d'avancer sur ces questions de manière concertée, or nos ressources sont limitées et il serait opportun de convaincre le gouvernement de donner un coup de pouce à ce niveau au système statistique national.

Christian Mersch salue le site des 17 objectifs de développement durable comme un effort significatif de rendre les objectifs tangibles.

3.4. Application et interprétation du Code de bonnes pratiques

Simone Casali signale que la réunion inaugurale de ce groupe a conduit à trois décisions : Mettre à profit les plénières du Comité des statistiques publiques pour relancer la participation de ses membres aux formations sur le Code de bonnes pratiques, relancer l'envoi des rapports de qualité par les ONAs (autres autorités statistiques du système statistique national produisant des statistiques européennes) et enfin réaliser une petite enquête auprès des membres du Comité, permettant d'évaluer dans quelle mesure les 84 indicateurs du Code sont implémentés ou non.

Serge Allegrezza voit le Code comme une charte que les statisticiens se doivent de respecter. Le Code est adapté régulièrement, suite aux expériences des membres du Système statistique européen. Il est donc important que les membres du Système statistique luxembourgeois donnent du feedback au STATEC au sujet de ce texte.

4. Discussion libre

Marc Pauly demande la parole pour brièvement informer le Comité des statistiques publiques de l'existence du programme académique *European Master in Official Statistics* (EMOS), en collaboration avec l'Université du Luxembourg. Au sein de la promotion actuelle, 16 candidats seraient a priori intéressés par des stages dans le milieu de la statistique officielle. Le STATEC pourrait en absorber une petite dizaine. Marc Pauly annonce que les membres du Comité des statistiques publiques seraient contactés le cas échéant pour évaluer leur intérêt à accueillir un ou plusieurs étudiants inscrits à ce programme.

Serge Allegrezza invite le Comité à la discussion libre.

Philippe Schram trouve intéressant de proposer des incitations positives aux citoyens pour qu'ils participent au recensement. Il suggère un échange sur le seuil qui déclenche l'envie de participer chez les répondants. Serge Allegrezza pense que la question se relie à l'analyse des taux de réponse enregistrés par le STATEC, qui baissent fortement. Bien sûr, la pandémie joue un rôle, mais l'individualisme dans la société également. Le STATEC a de véritables difficultés à continuer à collecter des données. C'est un thème dont il faut parler, car il est primordial pour des producteurs de statistiques. Le recensement fera l'objet d'une communication importante. Le STATEC pourra voir de quelle façon les gens répondent, et jusqu'où il convient d'insister pour obtenir les réponses aux questions posées. On reviendra certainement sur le sujet à la prochaine réunion du Comité des statistiques publiques. Les données administratives sont l'avenir, il faut que l'enjeu soit bien compris de tous. Si les taux de réponse deviennent trop bas, la qualité des statistiques produites dans de telles conditions deviendra aussi un problème pour le gouvernement.

Thierry Lallemang relate que la Commission Nationale pour la Protection des Données reçoit beaucoup de commentaires de citoyens qui se plaignent de l'aspect répressif associé aux enquêtes, c'est-à-dire de la menace de se voir infliger une amende. La peur d'une sanction éventuelle pousse certains à répondre, mais d'autres se braquent contre une telle approche. Serge Allegrezza voit lui aussi plusieurs cas de figure. Jusqu'à présent le STATEC n'a jamais envoyé la police à un particulier, mais parfois les personnes sont trop réticentes par rapport aux questionnaires, estimant que leur sphère privée est sacrée. On observe un comportement paradoxal de certaines personnes vis-à-vis de la mise à disposition de leurs données, selon l'entité qui les demande. Il peut parfois être utile d'infliger des sanctions, particulièrement chez les entreprises, où le STATEC en arrive parfois à envoyer la police pour un rappel à l'ordre.

5. Conclusion du Président

Les membres du Comité des statistiques publiques doivent se sentir appartenir à la communauté du Système statistique luxembourgeois, qui produit les statistiques publiques pour ce pays. Serge Allegrezza voit un besoin d'échanger beaucoup plus et de discuter des questions fondamentales.

Il y a un an, le STATEC a décidé d'engager un Project Management Officer pour donner à l'institut une vue d'ensemble sur les 20% de son travail constitué d'activités non récurrentes. Les projets ont des objectifs clairement définis, bénéficient de ressources humaines et budgétaires identifiés, et enfin le STATEC fait le suivi de ces projets très régulièrement. Voilà un autre sujet sur lequel le Comité des statistiques publiques pourrait échanger.

Serge Allegrezza adresse des remerciements au Ministère de la Digitalisation, qui a considérablement aidé le STATEC, et qui a des projets intéressants en route. Il clôt la réunion en remerciant tous les participants.