Nations Unies ECE/CES/2023/5

Conseil économique et social

Distr. générale
11 avril 2023
Français
Original : anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

Soixante et onzième réunion plénière
Genève, 22-23 juin 2023
Points 4 a) et 8 de l'ordre du jour provisoire
Coordination des activités statistiques internationales dans la région de la Commission économique pour l'Europe : Résultats des examens approfondis récemment réalisés par le Bureau de la Conférence des statisticiens européens

L'éthique des données − un facteur clef de l'acceptabilité sociale

Examen approfondi de l'éthique des données

Document établi par le Canada et le Royaume-Uni avec la contribution d'Eurostat

Résumé

L'examen approfondi fait le point sur l'éthique des données, sur ce que celle-ci englobe et sur ce que devrait être son évolution. Le présent document donne une définition de l'éthique des données, résume l'état des pratiques et l'évolution de l'éthique des données dans un certain nombre de pays et d'organisations professionnelles statistiques, étudie les problèmes et les enjeux connexes, et présente des conclusions et des recommandations. La dernière section résume les débats que le Bureau de la Conférence des statisticiens européens a tenus à sa réunion d'octobre 2022.

La Conférence est invitée à approuver les résultats de l'examen approfondi. La question de « l'éthique des données − un facteur clef de l'acceptabilité sociale » sera examinée plus en détail au point 8 de l'ordre du jour.

GE.23-06638 (F) 030523 040523

Merci de recycler

I. Cadre général

1. Au fur et à mesure que les sociétés évoluent, elles deviennent de plus en plus complexes. Par conséquent, le monde des statistiques officielles doit constamment s'adapter. Les organismes nationaux de statistique (ONS) doivent revoir dans une large mesure les méthodes d'organisation de leur programme statistique pour répondre à la demande de statistiques plus fines, plus diverses, plus actuelles et plus nombreuses. À cette fin, ils recourent désormais, au-delà de l'approche traditionnelle des enquêtes, à de nouvelles sources de données. Leurs méthodes et leurs boîtes à outils font appel à la science des données et aux approches modernes de collecte et d'intégration des données. Des avancées aussi vastes et rapides font ressortir l'importance de veiller à ce que toutes les approches suivies soient éthiques, d'où la nécessité d'une éthique des données.

2. Bien que l'expression « éthique des données » ne soit utilisée que depuis quelques années, le concept et les idées qui la sous-tendent ne sont pas tout à fait nouveaux. Dans leurs processus et leur culture, les ONS ont été guidés par des valeurs et des principes tels que l'importance de produire des statistiques professionnelles dépourvues de toute influence intéressée, de prévenir ou de limiter les préjugés, de protéger la confidentialité des données et de fonder les méthodes suivies sur la science. Cependant, lorsqu'ils pensent à l'éthique, les ONS se concentrent habituellement sur l'éthique des affaires du point de vue du comportement du personnel et des dirigeants, et lorsqu'ils se réfèrent aux données, l'accent est surtout mis sur les aspects liés à la protection de la vie privée et à la confidentialité. Par exemple, les ONS ne s'interrogeaient pas vraiment sur les questions liées à l'acquisition des données (au-delà de la question juridique), mais se préoccupaient plutôt de garantir la sécurité des données une fois qu'ils les possédaient et de limiter l'accès aux microdonnées.

3. La société étant de plus en plus associée et sensibilisée aux données par la présence accrue du monde numérique et virtuel, les particuliers en sont venus à apprécier l'importance des données et à craindre les effets négatifs que leur utilisation pouvait avoir. Dans ce contexte, les ONS ne peuvent pas considérer comme acquise l'acceptation sociale de leurs activités statistiques, que celles-ci soient nouvelles et innovantes ou même traditionnelles. L'acceptation sociale doit être gagnée, entretenue et reconquise lorsqu'elle est amoindrie pour toutes sortes de raisons qui peuvent être directement liées ou non aux actions des ONS. L'éthique des données joue un rôle important en vue de gagner la confiance et l'acceptation de la société, en éclairant les ONS sur ce qui peut être accompli par rapport aux objectifs souhaitables.

4. D'une certaine manière, de nombreux aspects de l'éthique des données sont déjà en place dans les ONS depuis longtemps. Par exemple, la protection des données a toujours fait l'objet d'une grande attention et l'on s'est toujours assuré que seules les personnes qui avaient besoin d'accéder à des informations détaillées y avaient accès. Elle a principalement consisté à garantir la sécurité informatique et la protection de la confidentialité. Cependant, compte tenu des nouvelles possibilités de collecte de données et d'un certain nombre de manquements très médiatisés qui se sont produits dans de grandes organisations internationales, le public considère qu'il incombe encore davantage aux ONS de garantir que l'ensemble des données sont collectées et conservées de manière plus éthique. L'acceptation par le public passe désormais par un renforcement de l'éthique des données et par une transparence accrue tout au long du cycle de vie des données[1].

5. L'éthique des données porte sur tous les aspects des données à toutes les étapes des programmes statistiques. Par exemple, il peut être nécessaire de prendre des décisions éthiques à toutes les étapes du Modèle générique du processus de production statistique (GSBPM)[2]. Dès le début de la conception d'un programme, des questions se posent sur les données à collecter et/ou à acquérir. Même si certains types de données ou ensembles de données seraient parfaitement adaptés pour répondre aux besoins, la société peut ne pas en

[1] Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle (l'approche scientifique en tant que facteur de transparence tout au long du cycle de vie des données). *Statistical journal of the IAOS* 35, 549-558. DOI 10.3233/SJI-190581. IOS Press.

[2] CEE (2022). Modèle générique du processus de production statistique. Generic Statistical Business Process Model − Generic Statistical Business Process Model − UNECE Statswiki.

comprendre l'importance et ne pas vouloir que ces données soient utilisées, ou elle peut être sensible à la partie exacte qui sera utilisée et à la manière dont celle-ci le sera. Lorsque les données étaient collectées dans le cadre d'enquêtes, le processus paraissait plus transparent, car les répondants savaient quelles informations étaient recueillies à leur sujet. Aujourd'hui, le recours aux données administratives ou à d'autres types de données fait que la transparence doit prendre de nouvelles formes et que d'autres valeurs éthiques doivent être prises en compte.

6. Pour illustrer l'éthique des données, il serait considéré comme contraire à l'éthique qu'un ONS qui envisage d'acquérir de nouveaux types de données, en particulier des microdonnées, s'appuie simplement sur les dispositions légales (si elles existent) et n'anticipe pas la fourniture de ces informations aux citoyens. En outre, il faut prendre en compte de manière manifeste les besoins et le bien public. Qui plus est, une approche totalement transparente qui exposerait clairement ce qui doit être acquis et pourquoi cela est nécessaire est une approche éthique.

7. Au sein de la communauté internationale, l'expérience de la conception et de la mise en œuvre d'approches et de méthodes en matière d'éthique des données est différente d'un pays à l'autre. Étant un nouveau sujet, l'éthique des données doit être bien définie et doit être étudiée plus avant. Cet examen fait le point sur l'éthique des données, sur ce que celle-ci englobe et sur ce que devrait être son évolution.

8. L'examen est subdivisé en plusieurs sections. La section II en expose l'objet et la portée. La section III présente la définition et le champ d'application de l'éthique des données, ainsi que ses liens avec d'autres aspects de l'éthique. La section IV décrit l'état des pratiques et l'évolution de l'éthique des données dans un certain nombre de pays ainsi que dans des organisations professionnelles de statistique. La section V examine ensuite les problèmes et les enjeux ayant trait à l'éthique des données et la section VI présente les conclusions et les recommandations.

II. Introduction

9. Le Bureau de la Conférence des statisticiens européens examine régulièrement et en profondeur certains domaines statistiques. L'objectif des examens est d'améliorer la coordination des activités statistiques menées dans la région de la CEE, de recenser les lacunes ou les chevauchements d'activités et d'aborder les nouvelles problématiques. L'examen est axé sur les questions stratégiques et met en évidence les préoccupations des organismes de statistique, tant sur le plan conceptuel que sur celui de la coordination. Le présent document constitue la base de cet examen en résumant les activités statistiques internationales dans le domaine considéré, en répertoriant les questions et les problèmes et en formulant des recommandations sur d'éventuelles activités de suivi.

10. Le Bureau du CES a choisi « l'éthique des données » pour faire l'objet d'un examen approfondi lors de sa réunion d'octobre 2022. Le Royaume-Uni et le Canada, avec l'aide de l'équipe spéciale de la CEE sur le leadership éthique, ont été invités à rédiger le principal document servant de base à l'examen.

11. Un certain nombre de pays (Royaume-Uni, Canada, Albanie, Hongrie) ainsi qu'Eurostat et l'Institut international de statistique ont été consultés. En outre, le contexte de nombreux ONS a été examiné dans le cadre d'une enquête sur l'éthique menée en 2021 par l'équipe spéciale de la CEE sur le leadership éthique, dirigée par l'Italie et le Canada.

III. Portée, définitions et analyse documentaire

A. Qu'est-ce que l'éthique des données ?

12. Les définitions de l'éthique des données reposent sur la transformation du concept d'éthique de l'informatique en concept d'éthique de l'information, l'accent étant désormais mis sur l'information créée par les ordinateurs et autres technologies, plutôt que sur la

technologie elle-même (Floridi et Taddeo 2016). Compte tenu du large éventail de types de données et de méthodes qui existent et du fait que l'utilisation et les applications des données varient en fonction de la culture et du contexte, il est difficile de définir de manière générale l'éthique des données. Les définitions générales se rapportent aux problèmes moraux qui se posent et aux solutions morales qui sont apportées en relation avec les données, les algorithmes et les pratiques (Floridi et Taddeo 2016).

13. Parmi les questions fréquentes en matière d'éthique des données figurent la dépersonnalisation et la réidentification, la confiance et la transparence, la responsabilité, le consentement et l'utilisation secondaire des données. On estime que les données elles-mêmes ne constituent pas un problème éthique, mais que c'est la manière dont celles-ci sont utilisées qui inquiète, en particulier parce que les progrès technologiques constants font que les utilisations futures sont inconnues (Hand 2018). Par exemple, la mise en correspondance d'informations détaillées sur le conjoint ou sur les enfants pourrait conduire à des problèmes d'équité et peut-être à certaines formes de stigmatisation, par exemple lorsque des questions sont posées sur les abus ou les violences à l'égard d'enfants. Dans ce cas, l'approche éthique pourrait consister à restreindre les informations demandées, à atténuer le processus ou simplement à repenser l'approche, voire, dans certains cas, à annuler le projet. Inversement, il serait contraire à l'éthique de poursuivre le projet sans tenir compte des questions éthiques. Il faut trouver un juste milieu entre les risques et les avantages pour profiter au maximum des possibilités offertes par les nouvelles technologies et l'utilisation des données. Floridi et Taddeo (2016) affirment que la réglementation éthique peut amoindrir l'intérêt que revêt la science des données dans la société, mais que cela devrait être compensé par une prise en compte appropriée des préoccupations éthiques afin de réduire les préjudices potentiels.

B. L'intelligence artificielle et la confidentialité des données

14. L'éthique des données s'applique aussi à l'intelligence artificielle (IA) qui, en tant que méthode de traitement et d'analyse des mégadonnées, fait l'objet de considérations éthiques particulières. L'éthique de l'IA s'est développée grâce à la disponibilité croissante de mégadonnées, à la capacité de stockage et aux progrès des algorithmes d'apprentissage automatique. L'évolution rapide de l'apprentissage automatique et des capacités algorithmiques fait que des préoccupations nouvelles et inattendues en matière d'éthique des données peuvent apparaître et entraîner des préjudices.

15. Les lignes directrices éthiques sur l'utilisation de l'IA visent à ce que la technologie soit exploitée de manière sûre. Selon le guide de l'Institut Alan Turing[3] pour une conception et une mise en œuvre responsables des systèmes d'IA, si l'on veut éviter des dommages imprévus, il est important que la responsabilité humaine dans la conception et l'utilisation de l'IA contribue à préserver et à protéger les individus susceptibles d'être touchés par les effets de ces systèmes sur la société. Toutefois, on estime que des lignes directrices et des principes de portée générale peuvent être vagues étant donné les contextes différents et complexes de l'utilisation de l'IA, que des cadres rigides, en particulier dans le domaine du droit, peuvent limiter l'utilisation de l'IA (Hagendorff 2020), et que les effets pourraient dans une certaine mesure être contraires à l'éthique.

C. L'éthique des affaires et l'éthique des données

16. L'éthique des affaires porte sur les valeurs et les obligations morales des organisations qui touchent à la responsabilité des entreprises, aux valeurs organisationnelles et au comportement des salariés. Ici, le contexte est lié à la manière dont les salariés et/ou les dirigeants se comporteraient face à des situations particulières. L'éthique des affaires comporte donc une dimension thématique et une dimension humaine. Elle recoupe l'éthique des données lorsque les pratiques des organisations sont prises en compte, telles que la collecte, le stockage et l'utilisation des données. Il s'agit en grande partie des données que

[3] https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

les entreprises détiennent sur leurs clients. L'Institute of Business Ethics[4] souligne que les entreprises ont la responsabilité de promouvoir la transparence et d'empêcher l'utilisation abusive des données à caractère personnel, car une mauvaise conduite peut influer sur la réputation, les relations avec les clients et, en fin de compte, les revenus.

17. La réutilisation des données des entreprises, souvent des données personnelles des clients, fait l'objet de débats, surtout lorsque les données sont utilisées à des fins différentes de celles pour lesquelles elles ont été collectées à l'origine. Les ONS font face à un problème général de réutilisation accrue des données. Un autre argument en matière d'éthique des données commerciales concerne la propriété des données qui sont collectées et détenues par des organisations privées. L'accès aux données des entreprises peut être contrôlé par les organisations alors que ces données pourraient être utilisées pour aider à mesurer les éléments non pris en compte dans les statistiques nationales, ce qui serait finalement utile à la prise de décisions par les autorités et pourrait bénéficier à la société dans son ensemble (Taylor 2016).

D. La protection de la vie privée et l'éthique des données

18. Parmi les considérations ayant trait à la protection de la vie privée dans le cadre de l'éthique des données figurent la confidentialité, le consentement éclairé, la confiance, la transparence et l'utilisation appropriée des données personnelles. On estime que la protection de la vie privée est une question éthique et que les préoccupations relatives à la protection de la vie privée dans le cadre des travaux de recherche portent sur les conséquences éthiques de l'utilisation des données et les droits des personnes (Kisselburgh 2020). Bien que l'utilisation des données soit réglementée par les lois sur la protection de la vie privée et que les données personnelles soient protégées par l'anonymisation et l'agrégation, on considère que les données agrégées peuvent encore s'avérer sensibles car elles permettent de prédire dans une large mesure des caractéristiques personnelles comme l'âge et le sexe (Taylor 2016).

19. Les initiatives des décideurs en matière d'éthique des données peuvent compléter la législation sur la confidentialité des données, mais l'éthique des données a des incidences complexes qui dépendent de la culture, ce qui signifie qu'il peut être difficile de définir l'éthique des données dans un tel contexte (Hasselbalch 2019). La législation sur la confidentialité des données peut également être difficile à appliquer aux technologies émergentes et aux nouvelles utilisations des données (Mulligan, Koopman et Doty 2016). Si les lois abordent certains éléments de l'éthique des données, d'autres considérations relatives aux problèmes moraux sont moins aisées à définir, d'autant plus que la compréhension et l'attitude du public à l'égard de la vie privée, de l'utilisation des données et de l'éthique évoluent (Hand 2018).

E. Les principes fondamentaux, les valeurs essentielles et les codes de bonnes pratiques

20. L'éthique des données est codifiée dans les principes, les valeurs et les codes de bonnes pratiques des organisations statistiques internationales et nationales. L'adoption et le respect de ces principes garantissent que les statistiques sont produites et utilisées de manière éthique, dans l'intérêt de tous, et qu'elles réduisent les préjudices. Les Principes fondamentaux de la statistique officielle de l'Organisation des Nations Unies (ONU) mettent en évidence des considérations communes liées à l'éthique des données, telles que la confidentialité, l'égalité, l'impartialité, la responsabilité, la transparence et la prévention de l'utilisation abusive des données. L'accent est mis sur la responsabilité des organismes statistiques de s'assurer que leurs pratiques en matière de données sont éthiques en suivant des méthodes et des procédures de collecte, de traitement, de stockage et de présentation des données, et sur le fait que les statistiques officielles − des informations fiables et objectives − jouent un rôle crucial dans la prise de décisions. Ces principes ont été élaborés et adoptés dans la région de la CEE il y a trente ans, puis par la Commission de statistique de l'ONU en 1994 et par l'Assemblée générale des Nations Unies en janvier 2014.

[4] https://www.ibe.org.uk/resource/business-ethics-and-big-data.html.

21. En 2022, les valeurs fondamentales de la statistique officielle ont été approuvées par la Conférence des statisticiens européens[5]. Elles sont défendues par les acteurs de la statistique officielle ; elles comprennent la pertinence (par exemple, par « une approche volontariste, adaptable et axée sur l'utilisateur, fondée sur une communication claire et régulière avec toutes les parties prenantes »), l'impartialité (par exemple, les acteurs « sont objectifs dans tous les aspects de leur travail et agissent avec équité et intégrité pour servir le bien public ») et la transparence (« les statistiques officielles, les méthodes, les processus, les produits et les rapports sur la qualité sont communiqués au public par les voies appropriées et peuvent faire l'objet d'un examen approfondi ») ainsi que le respect de la confidentialité (la communauté statistique « protège la vie privée en veillant à ce que la collecte des données soit limitée aux données nécessaires et proportionnelles au besoin de données ; assure la confidentialité des données sur les personnes, les ménages et les entreprises ; agit de façon éthique, en respectant les normes d'éthique professionnelle »). À titre d'exemple national, l'Office for Statistical Regulation Code of Practice for Statistics (code de bonnes pratiques de l'organisme de réglementation statistique) du Royaume-Uni fixe des normes pour les producteurs de statistiques officielles. Le code repose sur trois piliers : la fiabilité, la qualité et l'intérêt. Un élément du pilier « fiabilité » est que les normes d'éthique des données sont respectées afin d'être transparent sur la façon dont les données sont gérées et utilisées, en reconnaissant les droits des sujets de données.

IV. Pratiques des pays

A. L'éthique des données à Statistique Canada

22. Statistique Canada a toujours adhéré aux normes éthiques les plus élevées. Compte tenu de tous les changements survenus dans la société et de l'essor du numérique, les approches doivent être modernisées non seulement pour que les informations fournies aux citoyens restent pertinentes, mais aussi pour préserver le niveau d'éthique. Afin de mieux servir la société grâce à ses données, Statistique Canada est passé d'une approche axée sur les enquêtes à une approche privilégiant l'utilisation de sources de données non traditionnelles ou autres. Le recours à ces types de données a été au cœur des initiatives de modernisation de Statistique Canada. L'utilisation de sources de données non traditionnelles s'est accompagnée de progrès technologiques. Par exemple, l'apprentissage automatique et le moissonnage du Web sont des techniques qui peuvent servir à obtenir et à comprendre des données. Mais l'innovation soulève de nouvelles questions éthiques. Ce n'est pas parce qu'il est possible de faire quelque chose qu'il faut le faire, même si l'on dispose d'un mandat légal.

23. C'est dans ce contexte que Statistique Canada a adopté un cadre de nécessité et de proportionnalité et a mis en place un secrétariat de l'éthique des données ainsi que des comités de gouvernance, afin de garantir que les données sont utilisées de manière appropriée tout au long de leur cycle de vie.

1. Cadre de nécessité et de proportionnalité

24. En 2019, Statistique Canada a élaboré et mis en place un cadre de nécessité et de proportionnalité. Celui-ci est une adaptation de l'approche scientifique visant à optimiser la protection de la vie privée et la production d'informations lors de la conception d'une approche de collecte de données. La nécessité est le principe qui a trait aux besoins en données − informations à fournir et raisons pour lesquelles celles-ci sont nécessaires. La proportionnalité est l'ampleur de l'effort à faire pour obtenir les informations requises d'une manière qui soit cohérente avec les avantages escomptés d'un projet. En prenant en compte l'intrusion dans la vie privée, d'autres questions éthiques et la qualité des données, le principe de proportionnalité garantit que l'on ne collectera pas plus d'informations que ce qui est nécessaire pour produire les avantages escomptés. Dans ce cadre, toute proposition de nouveau projet ou d'acquisition de données doit non seulement expliquer pourquoi elle est importante, quels sont les avantages pour les Canadiens et qui a besoin de l'information, mais

[5] https://unece.org/sites/default/files/2022-07/ECE_CES_2022_2-2211176E.pdf.

aussi répondre à des considérations éthiques telles que le respect de la vie privée, la transparence et l'équité. Au début, le cadre n'était appliqué qu'aux initiatives de collecte de données provenant de sources administratives et autres, mais il a rapidement été étendu aux enquêtes. Il s'applique également aux initiatives de couplage des données et désormais à tout projet impliquant de nouvelles données ou de nouvelles utilisations des données.

2. Secrétariat de l'éthique des données

25. Le secrétariat de l'éthique des données a été créé pour appliquer le cadre de nécessité et de proportionnalité dans les activités quotidiennes de Statistique Canada. Dans le cadre de ces activités, il soumet les nouveaux projets ou les acquisitions de données à des examens éthiques, en organisant des discussions avec les gestionnaires de programmes et en soulevant des considérations éthiques. Celles-ci reposent sur six principes directeurs :

- Avantages pour les Canadiens ;
- Confidentialité et sécurité ;
- Transparence et responsabilité ;
- Confiance et durabilité ;
- Qualité des données ;
- Équité et absence de préjudice.

26. L'examen aboutit à une recommandation adressée au fonctionnaire principal responsable de l'éthique des données et de l'intégrité scientifique, dont le rôle est de veiller à ce que les scientifiques, les chercheurs et les gestionnaires de Statistique Canada respectent l'éthique des données, se conforment aux normes d'excellence scientifique et se conduisent d'une manière compatible avec les principes de l'intégrité scientifique.

27. Les cas plus sensibles sont soumis à un comité interne d'éthique des données composé de membres issus de différents milieux et ayant des points de vue différents. Les membres proviennent de tous les secteurs de l'organisation : statistiques sociales, statistiques économiques, collecte de données, sécurité informatique, bureau de gestion de la vie privée, etc.

28. Le secrétariat de l'éthique des données collabore à la rédaction de documents sur des sujets précis tels que les considérations éthiques s'appliquant à certains segments de la population, par exemple les mineurs ou les personnes vulnérables et marginalisées. Il organise également des discussions avec des partenaires à l'intérieur et à l'extérieur de l'organisation afin de tenir informés les collègues de Statistique Canada, mais aussi d'affiner et d'améliorer le processus interne en s'enrichissant de l'expérience des partenaires présents au Canada et à l'étranger.

29. L'éthique des données comprend également l'éthique de l'IA/apprentissage automatique. Les projets ayant recours à des techniques de science des données sont également analysés d'un point de vue éthique et soumis à des examens éthiques. En outre, un cadre d'apprentissage automatique responsable, reposant à la fois sur le respect des personnes et sur le respect des données, a été mis au point pour examiner les différentes utilisations des techniques d'IA/apprentissage automatique et s'assurer que celles-ci respectent les valeurs et les principes éthiques.

3. Conseil consultatif sur l'éthique et la modernisation de l'accès aux microdonnées

30. Le Conseil consultatif sur l'éthique et la modernisation de l'accès aux microdonnées a également été créé pour donner à Statistique Canada des orientations appropriées en matière d'accès aux données, de protection de la vie privée et de gouvernance des données, de façon à aider à satisfaire les besoins en données des Canadiens. Les connaissances et l'expérience des membres du Conseil consultatif sont utiles à l'organisme, car Statistique Canada s'efforce de faciliter l'accès des chercheurs aux microdonnées anonymes et d'améliorer la sécurité des données et les protocoles de gestion des risques. Les questions éthiques soumises au Conseil ne sont pas des cas particuliers ou transactionnels, mais concernent plutôt des thèmes

généraux et des orientations pour l'organisme. Le Conseil se réunit deux fois par an et ses rapports sont accessibles au public.

B. L'éthique des données à l'Office central hongrois de statistique

31. L'Office central hongrois de statistique (HCSO) traite les questions éthiques au cas par cas et cherche à intégrer ces procédures dans un cadre éthique plus large couvrant à la fois la production statistique et les opérations institutionnelles. En tant que membre de l'équipe spéciale sur le leadership éthique du Groupe de haut niveau sur la modernisation de la statistique officielle de la CEE, le HCSO suit avec intérêt les meilleures pratiques et les expériences d'autres pays afin de contribuer à l'élaboration de son propre cadre éthique.

32. Les pratiques éthiques actuelles du HCSO concernant le cycle de vie des données peuvent être analysées en examinant les différentes phases du processus statistique : acquisition, traitement et résultat.

33. Le HCSO ne demande que les données strictement nécessaires à l'élaboration, à la production et à la diffusion des statistiques officielles[6]. Dans les rares cas où les fournisseurs de données transmettent un surplus d'informations auxquelles il n'a pas accès, les données sont supprimées ou renvoyées aux fournisseurs en leur demandant de n'envoyer que les données auxquelles le HCSO a un accès justifié. Tout en reconnaissant la charge qui pèse sur les fournisseurs de données, la ligne de conduite est déterminée par des considérations éthiques prises par le chef de l'unité concernée et étayées par le mandat légal du HCSO.

34. Lors de la phase de traitement, le personnel du HCSO est chargé de traiter les données uniquement en vue de la production de statistiques officielles[7]. Outre la signature d'accords de confidentialité pour la protection des données, il est interdit aux salariés du HCSO d'utiliser les données à des fins personnelles et à d'autres fins que la production de statistiques relevant de la responsabilité de la personne et de l'unité concernées. Des contrôles de sécurité et d'accès informatiques sont en place pour empêcher tout accès non autorisé. D'après l'expérience du HCSO, ces mesures de protection sont secondaires par rapport à une culture d'entreprise définie par l'intégrité et par l'exemplarité que donne un leadership éthique.

35. Le résultat du processus de production statistique arrive entre les mains des utilisateurs sous la forme de statistiques officielles. Les discussions menées avec le public et au sein de la communauté statistique[8] au cours des dernières années ont mis en lumière les questions éthiques et les problèmes liés au traitement des informations erronées et de la désinformation, y compris l'étendue du rôle des ONS en la matière. Il s'agit d'un domaine qui exige une analyse conceptuelle plus poussée afin de définir ce qu'est l'utilisation non éthique des statistiques officielles et de formuler des lignes directrices à l'intention des ONS. Le HCSO ne vérifie pas que les statistiques sont utilisées de manière éthique par les utilisateurs. Toutefois, lorsqu'il a reçu des demandes de données ambiguës à des fins de recherche, il a demandé aux utilisateurs des précisions sur la nature et l'objectif des travaux de recherche afin de mieux comprendre leurs besoins et d'y répondre.

---

[6] Voir « Hungarian Generic Statistical Business Process Model, version 2.3 » (modèle générique hongrois de processus d'affaires statistiques, version 2.3) à l'adresse https://www.ksh.hu/docs/bemutatkozas/eng/estfm_eng.pdf.

[7] Voir « Confidentiality policy of the Hungarian Central Statistical Office » (Politique de confidentialité de l'Office central hongrois de statistique) à l'adresse suivante : https://www.ksh.hu/docs/bemutatkozas/eng/avpol_web_eng.pdf.

[8] Voir « Rules of engagement for handling disinformation, 2022 edition » (Règles de participation au traitement de la désinformation, édition 2022) d'Eurostat à l'adresse https://ec.europa.eu/eurostat/documents/3859598/14279253/KS-GQ-22-003-EN-N.pdf/d0017a54-e394-3887-d4ce-875aea0606fe?t=1643963795410.

C. L'éthique des données à l'Institut albanais de statistique

36. L'éthique, en tant qu'élément ayant trait non seulement aux méthodes et procédures choisies, mais aussi aux comportements, fait partie intégrante du maintien de la confiance dans les statistiques officielles.

37. L'éthique est considérée comme un principe important à respecter non seulement sur le plan du comportement organisationnel, mais aussi dans le processus de production et en matière d'accès aux données. L'Institut albanais de statistique suit le Code de bonnes pratiques de la statistique européenne et les Principes fondamentaux de la statistique officielle, qui sont importants en vue de garantir que les statistiques sont appropriées et fiables lorsque l'éthique en fait partie intégrante.

38. Le Conseil statistique est un organe consultatif composé de spécialistes des statistiques, qui aide l'INSTAT à veiller à ce que les statistiques officielles répondent aux besoins des utilisateurs en matière d'information et à promouvoir l'application des principes statistiques et des normes professionnelles. Il examine également, avec la direction des structures statistiques des ministères de tutelle ou d'autres organismes statistiques, les questions qui nécessitent leur intervention pour garantir que les statistiques officielles répondent aux normes professionnelles.

39. Alors que le Conseil statistique est la voix des utilisateurs, le Comité des producteurs de statistiques officielles, créé par un accord interinstitutionnel, rassemble les plus hauts représentants de l'INSTAT et d'autres organismes nationaux (Banque d'Albanie et Ministère des finances) pour coordonner le travail d'élaboration, d'exécution et de suivi du programme de statistiques officielles. Le Comité examine les questions liées à l'expansion des statistiques officielles dans le pays, mais aussi le rapport coût-efficacité résultant de la coopération interinstitutionnelle.

40. Le Comité de la confidentialité et le Groupe technique sur la confidentialité ont été mis en place pour garantir la confidentialité. La création du Comité et du Groupe technique a pour but de garantir le respect du principe de la confidentialité des statistiques grâce à des mesures préventives s'appliquant au processus de collecte, de traitement et de publication des statistiques officielles, ainsi qu'à l'interdiction de l'utilisation des données à des fins non statistiques. Un règlement sur l'accès aux microdonnées est accessible au public et vise à définir des mesures organisationnelles et techniques garantissant la sécurité dans l'administration et l'accès aux microdonnées à des fins de recherche scientifique en faveur de la communauté scientifique.

41. En outre, le Comité de la qualité s'occupe de l'évaluation des normes établies par les lignes directrices en matière de qualité, de la gestion de la qualité et des propositions visant à améliorer ces normes. Les améliorations du processus de production statistique peuvent provenir des recommandations formulées au cours des processus d'audit statistique et d'auto-évaluation. Le Comité est chargé de proposer des normes de qualité et il est responsable de la structure des rapports sur la qualité de tous les produits statistiques.

42. Le Code de déontologie, fondé sur les lois nationales ainsi que sur les normes internationales en matière d'éthique et d'intégrité des salariés, régit le comportement éthique et professionnel du personnel de l'INSTAT, pour ce qui est des décisions prises quotidiennement ou en cohérence avec sa mission, des principes qui le guide et des valeurs fondamentales. Il ne prétend pas apporter toutes les réponses aux problèmes éthiques qui peuvent se poser. Le Code ne remplace pas le bon sens et ne supprime pas les diverses réglementations appliquées par l'INSTAT. Il fait au quotidien le lien entre les besoins et les décisions, et traduit l'engagement en faveur de la mission de l'INSTAT. Il permet également d'encourager les discussions sur l'éthique et d'améliorer les attitudes ou les réactions au travail (face aux dilemmes et aux ambiguïtés éthiques). Les nouveaux membres du personnel reçoivent une formation périodique sur le Code de déontologie.

43. Les organismes de statistique doivent traiter d'éthique dans toutes les typologies utilisées − institutionnelle, échange et disponibilité de données, ressources humaines. Pour la première fois en 2021, un module distinct consacré au comportement éthique a été inclus dans l'enquête de satisfaction du personnel. Sur la base des résultats, un indice a été calculé pour mesurer l'environnement éthique en tenant compte des dimensions suivantes :

- La gestion éthique (ressources humaines, éthique des affaires et éthique des données) ;
- Les pratiques éthiques (par exemple, comportements transparents, égalité de traitement des personnes et style de leadership éthique) ;
- L'éthique des données du point de vue du personnel de l'INSTAT ;
- Les perceptions qu'ont les témoins d'un événement particulier (illégal ou contraire à l'éthique) ;
- Les normes éthiques de l'organisation.

44. Le personnel a donné son avis sur l'importance de l'éthique dans le cycle de production des statistiques officielles, sur les problèmes et les comportements et sur ce qui doit être amélioré en matière de comportement éthique.

45. Dans le processus de production, l'INSTAT utilise le modèle GSBPM pour la normalisation depuis 2017. En 2018, il a élaboré un cadre d'évaluation des résultats et a défini un indicateur de performance pour les mesurer. Ce cadre ressemble au GAMSO. L'année prochaine, l'INSTAT prévoit de remplacer ledit cadre par le GAMSO.

46. Il est temps de définir les instruments appropriés pour intégrer l'éthique des données dans toute la chaîne de production statistique, de diffuser l'éthique des données auprès des producteurs et des utilisateurs de statistiques officielles, d'aller de l'avant et de renforcer la confiance dans les statistiques officielles.

D. L'éthique des données à l'Autorité statistique du Royaume-Uni (UKSA)

1. Comment l'Autorité statistique du Royaume-Uni définit-elle l'éthique des données ?

47. Les travaux de l'UKSA en matière d'éthique des données visent à garantir que la collecte, l'utilisation et la diffusion des données au sein de la communauté des analystes sont conformes à l'éthique. L'éthique des données est définie à l'aide des principes éthiques suivants :

- L'utilisation des données présente des avantages évidents pour les utilisateurs et concourt au bien public ;
- L'identité du sujet de données (qu'il s'agisse d'une personne ou d'une organisation) est protégée, les informations restent confidentielles et sécurisées, et la question du consentement est examinée de manière appropriée ;
- Les risques et les limites des nouvelles technologies sont pris en compte et la surveillance humaine est suffisante pour que les méthodes employées soient conformes aux normes reconnues d'intégrité et de qualité ;
- Les données utilisées et les méthodes employées sont conformes aux exigences légales énoncées notamment dans la législation sur la protection des données, la loi de 1998 sur les droits de l'homme, la loi de 2007 sur les statistiques et le service d'enregistrement et le devoir de confidentialité inscrit dans la *common law* ;
- Les opinions du public sont prises en compte à la lumière des données utilisées et des avantages perçus des travaux de recherche ;
- La disponibilité, l'utilisation et le partage des données sont transparents et font l'objet d'une communication claire et accessible au public.

48. Ces principes éthiques ont été élaborés en collaboration avec le Comité consultatif sur l'éthique des données du Statisticien en chef (NSDEC), qui fournit au Statisticien en chef du Royaume-Uni des conseils éthiques indépendants et transparents sur la collecte et l'utilisation de données à des fins d'analyse. La majorité des membres du NSDEC sont indépendants des autorités et possèdent des compétences dans des domaines tels que le droit des données, l'éthique des données, la recherche sociale, la recherche en matière de santé et la science des données. Il y a également un profane qui donne un point de vue précieux de non-spécialiste.

2. Comment l'Autorité statistique du Royaume-Uni applique-t-elle l'éthique des données ?

49. La stratégie quinquennale de l'UKSA (« Statistics for the Public Good »[9]), publiée en 2020, affiche l'ambition de l'organisation d'être un leader mondial reconnu en matière d'éthique appliquée des données. En février 2021, l'UKSA a créé un centre d'éthique appliquée des données (CADE), constitué de cinq membres du personnel, pour réaliser cette ambition. Jusqu'à présent, le CADE s'est concentré sur la fourniture de ressources (par exemple, des outils, des conseils et des formations) et sur l'assistance aux utilisateurs pour aider les chercheurs à appliquer efficacement les bonnes pratiques en matière d'éthique des données lors de la collecte et de l'utilisation des données. Afin de garantir la visibilité de ces ressources au sein de la communauté des analystes, le CADE a présenté les services qu'il propose à l'occasion de 32 conférences analytiques et manifestations participatives, en diffusant des études de cas qui montraient comment ses services avaient été utilisés avec succès dans le cadre d'une série de projets de recherche de premier plan. Le CADE veille à ce que ses services aient un impact en dehors de l'administration en les rendant accessibles aux chercheurs universitaires et commerciaux qui désirent accéder aux données publiques en toute sécurité, grâce à des conditions de recherche fiables dans l'ensemble du Royaume-Uni.

50. Le CADE a principalement employé son énergie à rendre l'éthique des données accessible à la communauté des chercheurs, afin de leur permettre d'adopter dans les faits une conception éthique de la collecte et de l'utilisation des données à des fins d'analyse. Il s'agissait d'aller au-delà des principes éthiques de haut niveau et de mettre en valeur des ressources pour aider les chercheurs à appliquer efficacement ces principes dans leurs travaux. À cette fin, le CADE a mis au point un outil d'auto-évaluation éthique qui subdivise les principes éthiques en éléments individuels facilement assimilables et permet aux chercheurs d'évaluer rapidement leur projet en fonction des différents aspects de chaque principe éthique. Cette démarche est appuyée par une formation, des conseils et une assistance aux utilisateurs qui permettent de contrôler que l'outil est utilisé de manière appropriée.

51. L'auto-évaluation a été intégrée dans les processus d'accès sécurisé aux données de l'UKSA, de sorte que la plupart des chercheurs qui accèdent aux données à des fins de recherche doivent utiliser cet outil pour obtenir l'autorisation de collecter des données et/ou d'y accéder. Elle permet de repérer rapidement les projets à haut risque sur le plan éthique, projets qui sont ensuite examinés minutieusement par le NSDEC en toute indépendance. Des précisions sur cet examen indépendant sont publiées sur le site Web de l'UKSA.

52. Conformément à la stratégie de l'UKSA, l'un des principaux objectifs de ces travaux est de permettre une utilisation efficace de ces services afin de rendre possible l'accès en temps voulu aux données à des fins d'analyse. Grâce à l'assistance, aux conseils et à la formation dont le CADE fait bénéficier les utilisateurs, le chercheur est soutenu tout au long de la période pendant laquelle il prend en compte l'éthique des données, ce qui contribue à garantir que le processus est relativement solide et efficace. Actuellement, les chercheurs qui soumettent une demande obtiennent une réponse dans les deux jours ouvrables.

53. L'UKSA mesure l'impact de ces travaux sur les différents groupes de parties prenantes au moyen d'évaluations régulières avec les utilisateurs de ses services d'éthique des données et d'indicateurs clefs de la prestation de services. Les résultats de ces études sont régulièrement communiqués à un comité consultatif indépendant composé de parties prenantes issues des ministères, des organismes de financement de la recherche et de la

[9] Statistics for the public good – UK Statistics Authority.

communauté des analystes, qui donnent des avis et des conseils transparents sur l'orientation stratégique, les résultats et l'impact du Centre. Ces évaluations des utilisateurs et ces indicateurs jouent un rôle important, en contribuant à éclairer la conception des services que le Centre offrira à l'avenir.

54. L'année dernière, 324 projets de recherche provenant de l'ensemble de la communauté des analystes ont suivi les principes éthiques de l'UKSA pour examiner de manière efficace la dimension éthique de leur collecte et/ou de leur utilisation des données. Le cadre éthique de l'UKSA est ainsi devenu l'un des cadres éthiques les plus utilisés au Royaume-Uni.

55. Pour diffuser ses travaux auprès du public, l'UKSA a publié une politique d'éthique des données et il est désormais présent sur le site Web de l'Autorité statistique britannique, qui communique de manière transparente les mesures prises par l'organisation pour s'assurer que les données sont utilisées de manière éthique et met gratuitement ses services à la disposition de la communauté des analystes. Il s'engage également à utiliser les données de manière éthique dans le cadre des informations fournies par les participants dans l'ensemble de ses enquêtes et collabore avec les équipes de communication pour faire de l'éthique des données un thème central des travaux de l'organisation sur l'utilisation des données à des fins d'analyse, auxquels participe le public.

56. L'éthique des données est également un principe clef du cadre de partage des données du Royaume-Uni inscrit dans la loi de 2017 sur l'économie numérique. Le Research Accreditation Panel est un comité de contrôle indépendant qui, en sa qualité d'organisme d'accréditation prévu par la loi, conseille l'UKSA sur ce cadre permettant de partager des données des autorités publiques avec la communauté des chercheurs à des fins d'analyse, au profit de la société et de l'économie britanniques. Toutes les parties prenantes à la diffusion, au traitement ou à l'utilisation des données accessibles dans ce cadre sont légalement tenues de respecter des normes éthiques adaptées à la nature des travaux de recherche. En plus de s'assurer que les normes éthiques appropriées sont systématiquement respectées dans toutes les utilisations du cadre de partage des données, le Research Accreditation Panel donne des conseils indépendants à l'UKSA sur la manière dont l'utilisation dudit cadre peut être maximisée dans la communauté des chercheurs, afin d'accroître les avantages publics que les autorités compétentes procurent aux citoyens britanniques. Il s'assure que l'UKSA continue à mettre en œuvre le cadre de partage des données d'une manière éthiquement appropriée et sûre, qui respecte la vie privée et la confidentialité des données.

3. Comment l'Autorité statistique du Royaume-Uni contribue à élargir la réflexion sur l'éthique des données ?

57. Afin de faire de l'UKSA un leader d'opinion dans le domaine de l'éthique des données, le CADE a mené à bien les activités suivantes :

- Une table ronde a été organisée pour examiner les questions ayant trait à la meilleure façon de surmonter les problèmes éthiques qui se font jour dans l'utilisation des données pour la recherche et les statistiques, et à la façon dont nous pouvons aider les chercheurs à y parvenir. Cette manifestation a été suivie par environ 30 participants issus d'organisations diverses provenant de l'ensemble du Royaume-Uni et intéressés par l'éthique des données. Les résultats de ces travaux ont éclairé le plan de travail ;
- Une collaboration a été menée avec une série d'autres acteurs pour produire des conseils pratiques à l'intention des utilisateurs sur une série de sujets relatifs à l'éthique appliquée des données, qui sont énumérés ci-dessous. Ces thèmes ont été choisis après avoir consulté les utilisateurs, les analystes ayant demandé de recevoir des conseils en matière d'éthique sur ces sujets. Des documents d'orientation sont disponibles sur le site Web de l'UKSA et sont mis en avant sur les blogs de hauts responsables qui sont diffusés dans les médias sociaux de l'Autorité :
    - Intégrer des considérations éthiques dans l'utilisation des données géospatiales ;
    - Tenir compte du bien public dans les travaux de recherche et les statistiques ;

- Prendre en compte les opinions et la participation du public concernant l'utilisation des données pour la recherche et les statistiques ;
- Utiliser des données provenant de tiers à des fins de recherche et de statistiques ;
- Prendre en compte l'éthique du travail avec les groupes vulnérables ;

- Participer régulièrement à des réunions avec d'autres services de l'administration s'occupant de l'éthique des données afin de s'assurer que la voix des chercheurs et des statisticiens contribue à une réflexion plus large des autorités. Ces services sont également représentés au sein du comité consultatif indépendant du CADE afin de s'assurer que le Centre travaille de manière coordonnée avec d'autres parties prenantes ;
- Dispenser une formation à l'éthique des données à divers services ministériels et groupes professionnels au sein de l'administration.

58. Au niveau international, l'UKSA a joué un rôle de premier plan dans la contribution du Royaume-Uni à la réflexion ayant trait à l'éthique des données sur la scène internationale. Dans ce domaine, le CADE s'est concentré sur la direction de travaux consacrés à l'étude de l'éthique des données dans l'apprentissage automatique pour produire des statistiques officielles dans le cadre d'un programme de travail de la CEE sur l'apprentissage automatique. De nouvelles orientations en matière d'éthique des données axées sur l'apprentissage automatique pour la recherche et les statistiques officielles ont notamment été publiées. En outre, le CADE a dispensé une formation à l'éthique des données lors de manifestations organisées par la CEE et la CESAP à l'intention des statisticiens travaillant dans les organismes nationaux de statistique à travers l'Europe et dans la région Asie-Pacifique. Le Centre a également organisé et présidé une séance réussie sur l'éthique des données lors d'une conférence de la CEE sur la modernisation de la législation statistique et a contribué à des initiatives internationales, telles que l'équipe spéciale de la CEE sur le leadership éthique et les manifestations que la CESAP a organisées en Asie-Pacifique sur les mégadonnées et la gouvernance.

E. L'éthique des données à Eurostat

59. L'objectif d'Eurostat, à savoir « fournir des statistiques et des données de haute qualité sur l'Europe », et ses valeurs fondamentales sont précisés dans l'énoncé de sa mission[10]. Eurostat produit des statistiques européennes en partenariat avec les organismes nationaux de statistique et d'autres autorités nationales de l'Union européenne. L'élaboration, la production et la diffusion des statistiques européennes sont régis par les principes statistiques suivants : *indépendance professionnelle, impartialité, objectivité, fiabilité, secret statistique et rapport coût-efficacité,* qui sont définis dans le règlement relatif aux statistiques européennes. Les principes statistiques sont précisés dans le Code de bonnes pratiques de la statistique européenne[11], qui fait également partie du règlement relatif à la statistique européenne[12]. Le Code définit des principes supplémentaires ayant trait à la qualité des résultats, aux processus statistiques et à l'environnement institutionnel. L'adhésion du système statistique européen (SSE) à ces principes est évaluée par rapport à une série d'indicateurs relatifs à chacun des principes par un conseil indépendant, le Conseil consultatif européen pour la gouvernance statistique. Celui-ci analyse régulièrement l'application du Code et conseille sur les mesures à prendre pour l'améliorer.

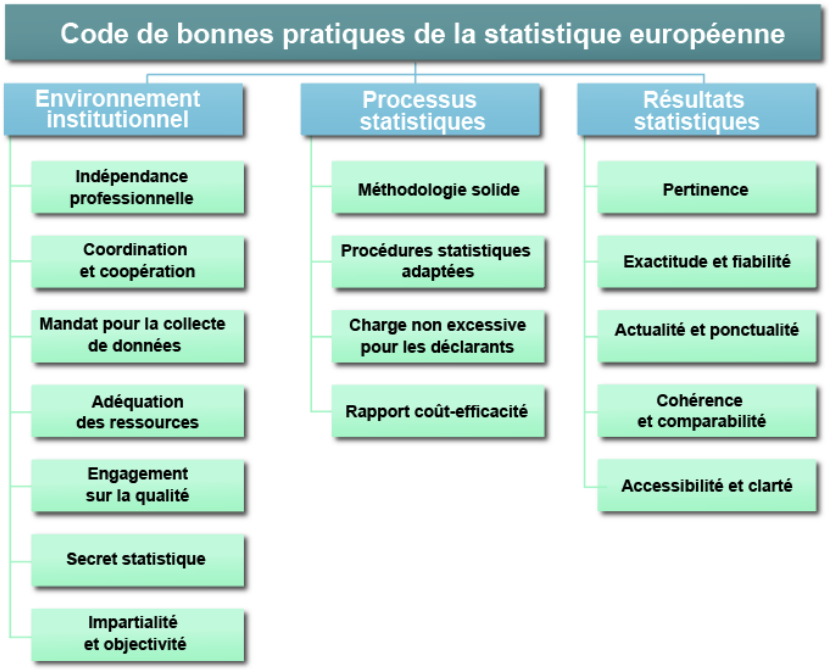
60. Le Code constitue la pierre angulaire du cadre européen de la qualité, qui se compose en outre du cadre d'assurance de la qualité et des principes généraux de gestion de la qualité.

[10] Voir https://ec.europa.eu/eurostat/web/main/about-us/who-we-are.

[11] Voir https://ec.europa.eu/eurostat/documents/4031688/9332274/KS-02-18-142-FR-N.pdf/130905e7-45a7-4475-b37c-8f699b5e33e1?t=1540215834000.

[12] Règlement (CE) nº 223/2009 du Parlement européen et du Conseil de mars 2009, https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=celex%3A32009R0223.

Figure 1
Le Code de bonnes pratiques de la statistique européenne



61. En tant qu'instruments d'autoréglementation non législative, le Code et le cadre de qualité pour les statistiques européennes constituent un cadre éthique régissant l'élaboration, la production et la diffusion des statistiques européennes. En particulier, les principes relatifs à *l'engagement sur la qualité, au secret statistique, à l'impartialité et à l'objectivité, à une méthodologie solide* ainsi que les principes relatifs à la production statistique sont étroitement liés à l'éthique des données car ils garantissent l'utilisation éthique et la qualité des données tout au long du processus de production statistique.

62. Les indicateurs ont été élaborés à l'origine en pensant aux enquêtes et aux recensements comme mode de collecte des données. Ils ont été étendus en tenant compte de l'utilisation de sources administratives, qui ont des caractéristiques communes avec les sources de données que sont les mégadonnées. Dans le cadre des activités visant à inclure de nouvelles sources de données dans la production de statistiques européennes, Eurostat a analysé les principales questions relatives à leur intégration dans la production de statistiques officielles et à leur impact sur les principes du Code et leurs indicateurs[13]. L'analyse a été réalisée en suivant les principales étapes du processus de production statistique (acquisition des données, traitement des données et diffusion des résultats statistiques) et en tenant compte des caractéristiques des sources de mégadonnées choisies. Les principaux problèmes recensés tiennent au fait que ces sources de données ne sont pas conçues pour produire des statistiques officielles, à la sélection des données et des fournisseurs de données, à l'accès aux données, au besoin de métadonnées, à l'application d'algorithmes adéquats pour l'analyse des données, à la transparence du traitement et de l'analyse, ainsi qu'à la collaboration avec les universités et les fournisseurs de données (lien vers les rapports). L'analyse s'est appuyée sur une recherche documentaire, les sources de mégadonnées n'étant guère utilisées dans la production de statistiques officielles à l'époque. Il sera donc nécessaire d'examiner les problèmes et de définir des indicateurs pour compléter le Code. Les travaux sur les problèmes éthiques potentiels ont abouti à un projet de lignes directrices pour l'élaboration, la production et la diffusion de statistiques basées sur des sources de mégadonnées[14]. Les lignes directrices établissent une distinction entre les différents types de sources de données, les problèmes éthiques potentiels variant considérablement de l'une à l'autre. Le projet de lignes directrices relatives au moissonnage du Web a servi de point de

[13] Voir https://ec.europa.eu/eurostat/cros/content/ethical-review-big-data-0_en.
[14] Voir https://ec.europa.eu/eurostat/cros/content/draft-ethical-guidelines-big-data-and-european-statistics_en.

départ à l'élaboration de lignes directrices dans le cadre du pôle de renseignement Web d'Eurostat.

63. Le pôle de renseignement Web du SSE[15], qui est en cours de conception et vise à acquérir des capacités de récupération de contenu Web et de moissonnage du Web pour le système statistique européen, est un domaine d'application. Le SSE a récemment adopté un ensemble de lignes directrices régissant l'extraction et l'utilisation des données collectées sur le Web[16]. Ces lignes directrices seront suivies en définissant un ensemble de règles appliquées, qui sont en cours d'élaboration et d'examen au sein du SSE.

64. En plus d'être la principale organisation du système statistique européen, Eurostat est une direction générale de la Commission européenne. Entre 2018 et 2021, la Commission a élaboré une stratégie globale sur l'intelligence artificielle[17], qui a abouti à proposer un cadre réglementaire visant à favoriser l'adoption de l'IA, tout en tenant compte des risques et en protégeant les droits fondamentaux des citoyens de l'UE.

65. Compte tenu de ce qui précède, la Commission européenne s'est engagée à concevoir un ensemble de lignes directrices internes, une gouvernance interne spéciale et un code de conduite pour la mise au point et l'utilisation des systèmes d'IA. Les lignes directrices accorderont une attention particulière à la formation d'une main-d'œuvre mature en matière de données et d'IA.

F. L'éthique des données et les organisations professionnelles de la statistique

66. En 2020, la Royal Statistical Society (Royaume-Uni) a créé une section sur l'éthique et la gouvernance des données après avoir soutenu pendant trois ans les travaux d'un groupe d'intérêt sur l'éthique des données. L'intérêt est grand et croissant pour des domaines tels que les personnes dans les données, les groupes, les pratiques, les impacts et les structures collectives dignes de confiance.

67. Le Conseil consultatif d'éthique de l'Institut international de statistique investit dans l'élargissement du champ d'application de ses principes pour y inclure de nouvelles disciplines relatives à la science des données et à l'IA et sera donc associé aux travaux sur l'éthique des données.

68. De son côté, l'Association internationale pour les statistiques officielles (IAOS) a créé en 2022 le groupe de Cracovie. Celui-ci vise à contribuer à la résolution des problèmes pressants qui se posent à la production de statistiques officielles dans les sociétés d'aujourd'hui où les données sont de plus en plus nombreuses[18].

69. Enfin, les conférences et les revues statistiques ont enregistré une nette augmentation du nombre de sessions et d'articles consacrés à l'éthique des données et, plus généralement, aux questions éthiques liées au contexte des données dans les ONS.

V. Problèmes et enjeux

70. Cette analyse approfondie a permis de recenser plusieurs problèmes et enjeux qui se posent aux ONS et qui seront abordés dans cette section. En voici quelques-uns :

a) Intégrer avec succès les politiques et les principes d'éthique des données dans les activités de collecte et d'analyse ;

b) Obtenir l'adhésion aux politiques et principes d'éthique des données à tous les niveaux de la communauté statistique ;

[15] Voir https://ec.europa.eu/eurostat/cros/content/project-overview_en.
[16] Voir https://ec.europa.eu/eurostat/cros/content/web-content-retrieval-guidelines-0_en.
[17] Voir https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.
[18] https://www.iaos-isi.org/index.php/latestnews/339-krakow-group.

c) Faire savoir aux différentes parties prenantes que les politiques et les principes d'éthique des données sont efficaces ;

d) Inciter à une réflexion plus large sur l'éthique des données.

A. Intégrer avec succès les politiques et les principes d'éthique des données dans les activités de collecte et d'analyse

71. L'un des enjeux consiste à s'assurer que les politiques et les principes d'éthique des données sont bien appliqués dans l'ensemble de la communauté statistique. Il ne suffit pas que les ONS disent avoir des politiques et des principes en matière d'éthique des données. Pour avoir un sens, les unes et les autres doivent être intégrés et appliqués de manière cohérente dans l'ensemble des processus de collecte de données et de production de statistiques. Ce n'est qu'ainsi que les politiques et les principes d'éthique des données peuvent avoir un impact notable sur toutes les activités du système statistique.

72. Pour réussir à intégrer et à appliquer des principes et des politiques d'éthique des données dans un système statistique, il est essentiel de veiller à ce que les analystes aient accès à des politiques et à des principes de haut niveau. Les analystes ont besoin d'aide pour appliquer les principes et les politiques d'éthique des données. Cette aide peut prendre la forme d'outils leur permettant d'appliquer des politiques et des principes de haut niveau, d'une assistance aux utilisateurs et d'une formation pour sensibiliser à l'intérêt d'une bonne éthique des données au sein de la communauté des analystes. Il est important que ces activités démystifient les principes et les politiques de haut niveau pour les analystes et qu'elles les rendent applicables aux scénarios d'analyse que ceux-ci connaissent réellement. Il est ainsi possible de faire passer l'éthique des données du monde de la théorie de haut niveau à des pratiques appliquées qui peuvent être intégrées avec succès dans les processus de collecte de données et de statistique.

73. Compte tenu des pressions croissantes exercées sur les ONS pour que ceux-ci produisent des analyses rapides et des statistiques en temps voulu, il est également essentiel de veiller à ce que les analystes puissent utiliser ces ressources de manière efficace afin d'intégrer et d'appliquer avec succès des politiques et des principes d'éthique des données. En bref, s'ils ne sont pas faciles à utiliser et s'ils sont inefficaces, ces processus sont moins susceptibles d'être utilisés par les analystes. Des équipes spécialisées dans le soutien à l'éthique des données doivent veiller à ce que ces processus soient faciles à intégrer pour les analystes. Une bonne éthique des données au sein des ONS ne se fait pas toute seule, elle nécessite des ressources pour concevoir de bonnes pratiques en matière d'éthique des données dans un contexte statistique et pour soutenir l'application de ces pratiques de manière efficace au sein de la communauté statistique. En outre, des outils et des pratiques en matière d'éthique des données devraient être conçus de telle manière que les activités liées à l'éthique des données ne soient pas perçues comme un obstacle par des gestionnaires qui considèrent qu'ils doivent en passer par là pour poursuivre leur programme.

B. Obtenir l'adhésion aux politiques et principes d'éthique des données à tous les niveaux de la communauté statistique

74. Un autre enjeu qui a été recensé au cours de cette étude approfondie est la nécessité d'obtenir l'adhésion aux politiques et principes d'éthique des données à tous les niveaux de la communauté statistique. Il est essentiel de parvenir à une compréhension commune de l'importance de l'éthique des données dans l'ensemble du système statistique pour faire en sorte que tous les travaux qu'un ONS consacre à cette question aient un impact notable et durable.

75. Pour obtenir une telle adhésion, il est nécessaire d'établir, avec des parties prenantes à un haut niveau et à un niveau opérationnel, des relations ciblées qui permettent de promouvoir l'importance de l'éthique des données. Dans le cadre de ces relations, il faudrait notamment sensibiliser aux conséquences notables que peuvent avoir une collecte et une utilisation des données qui ne seraient pas éthiques et expliquer pourquoi l'éthique des données diffère d'autres processus mis en place pour protéger l'utilisation des données, tels

que les politiques de sécurité et la législation sur la protection des données. Il s'agit également de faire comprendre pourquoi l'éthique des données est importante aujourd'hui. Les ONS ont actuellement accès à plus de données que jamais auparavant. Les possibilités que l'analyse contribue au bien public et cause du tort sont donc plus nombreuses qu'auparavant. Ces possibilités accrues s'accompagnent d'une responsabilité encore plus grande, celle de se comporter de manière appropriée sur le plan éthique. Ce nouveau contexte dans lequel se trouvent les ONS rend encore plus importante la nécessité de ne pas se contenter de réfléchir à ce que chacun d'eux **peut** faire, mais de s'interroger également sur ce qui **devrait** être fait et **comment** d'un point de vue éthique.

76. Il est également important que les parties prenantes reconnaissent la contribution que l'adhésion généralisée aux principes et politiques d'éthique des données peut apporter à une collecte et une analyse solides des données au sein d'une organisation. Le respect des politiques et des principes d'éthique des données, tels que l'examen des personnes pouvant être exclues d'un échantillon ou l'examen des informations sur les participants fournies à un échantillon, améliore souvent de manière notable la qualité des activités statistiques des ONS. Ces activités ne sont pas seulement judicieuses d'un point de vue éthique, elles le sont aussi souvent d'un point de vue méthodologique car elles enrichissent la qualité des résultats statistiques produits.

C. Faire savoir aux différentes parties prenantes que les politiques et les principes d'éthique des données sont efficaces

77. Il est parfois difficile pour les ONS de surmonter le scepticisme du public et de montrer que les activités d'éthique des données sont efficaces. Il est donc important que les ONS soient transparents dans leurs activités. Cette transparence doit aller au-delà de la diffusion des politiques, des principes et des processus en matière de données et préciser dans quelle mesure ces activités sont menées de manière cohérente dans l'ensemble de la communauté statistique. Les ONS peuvent ainsi prouver qu'ils font ce qu'ils disent et conserver ou accroître la confiance du public dans leurs activités.

78. Les parties prenantes peuvent remettre en question les décisions prises concernant l'éthique de la collecte et de l'utilisation des données par des personnes appartenant à un ONS dont le portefeuille de produits statistiques bénéficiera de ces activités. Il est donc important que les ONS acceptent que de leurs décisions ayant trait à l'éthique de la collecte et de l'utilisation des données fassent l'objet d'un examen indépendant et que ces décisions soient éclairées par un avis d'expert indépendant approprié. Les résultats de cet examen indépendant devraient être transmis aux parties prenantes. En outre, les ONS devraient veiller à documenter leurs activités et décisions en matière d'éthique des données afin d'être en mesure de mettre en évidence et de démontrer le bien-fondé de leur approche. Ces activités aident les ONS à persuader les parties prenantes qu'un processus de prise de décisions éthique est crédible et indépendant de tout intérêt particulier.

D. Inciter à une réflexion plus large sur l'éthique des données

79. Un large éventail d'organisations universitaires, d'organisations gouvernementales, d'organisations commerciales et d'organisations du secteur tertiaire influencent les débats sur l'éthique des données dans les pays du monde entier. Compte tenu de ces voix différentes, il peut parfois être difficile pour les ONS de faire entendre leur voix et de susciter une réflexion plus large sur l'éthique des données. Il est important que les ONS y parviennent car ce sont de grands utilisateurs de données et que, par conséquent, la communauté statistique influence de manière notable les travaux plus larges sur l'éthique des données à la fois dans les pays et au niveau international. Les ONS peuvent le faire en diffusant en toute confiance leur travaux sur l'éthique des données auprès de ces autres secteurs et en collaborant avec eux sur des sujets qui y ont trait. Le risque que les ONS ne soient pas activement impliqués dans l'élaboration de la réflexion sur l'éthique des données s'en trouve ainsi réduit.

VI. Conclusions

80. L'éthique des données va au-delà des thèmes traditionnels de la sécurité et de la protection de la vie privée et englobe des considérations éthiques tout au long du cycle de vie des données, de leur collecte à leur diffusion.

81. Les considérations et principes éthiques ne sont pas nouveaux dans les statistiques officielles. Toutefois, l'évolution du paysage des données et des attitudes sociales à l'égard de la collecte et de l'utilisation des données les place au centre de l'attention en tant que composante nécessaire de la confiance et de l'acceptabilité sociale à acquérir et à entretenir. Pour y parvenir, les ONS commencent à aborder l'éthique des données d'une manière beaucoup plus transparente, pratique et systémique. Ces pratiques ont un rôle important à jouer dans la consolidation de leur réputation en tant que collecteurs, utilisateurs et producteurs de données fiables.

82. Malgré l'importance croissante de l'éthique des données dans les statistiques officielles, il n'existe pas encore de définition internationale commune ni de consensus sur les pratiques recommandées. La compréhension de l'éthique des données et la manière de l'appliquer au sein des ONS peuvent varier en fonction du contexte.

83. Le niveau de développement et d'application des pratiques en matière d'éthique des données est en nette augmentation, mais varie d'un ONS à l'autre. Il ressort clairement du présent document que certains ONS ont accompli des progrès considérables et ont montré qu'il était possible d'utiliser les données de manière ambitieuse et efficace à des fins statistiques tout en veillant à ce que la collecte et l'utilisation de ces données soient conformes à l'éthique. L'expérience de ces pays montre qu'une prise en compte solide et opportune de l'éthique des données est un facteur important en vue de faciliter l'adoption de pratiques innovantes que le public juge acceptables.

84. Les processus d'éthique des données ne font pas obstacle à la collecte et à l'utilisation innovantes de données à des fins statistiques. Des processus transparents qui sont bien intégrés au sein d'un ONS permettent d'établir des relations de confiance avec les parties prenantes (par exemple, les sujets de données, le grand public et les fournisseurs de données) et de garantir l'acceptabilité sociale. Mais avant cela, ils suscitent une culture dans laquelle est davantage pris en compte l'intérêt public à satisfaire les besoins de la société en matière de données. Une éthique des données solide prend également en compte la manière dont la recherche est effectuée et dont celle-ci est communiquée. Par conséquent, elle contribue également à garantir des pratiques méthodologiques solides en matière de données et des niveaux appropriés de transparence au sein d'un ONS.

85. Pour que les bonnes pratiques en matière d'éthique des données soient intégrées avec succès dans une organisation, il faudrait que soient mises en place des ressources spécialisées qui contribuent à développer ces pratiques et à aider les analystes à les appliquer. Il est également important qu'une culture de l'éthique des données soit adoptée par tous les salariés et cadres d'une organisation afin de garantir que ces ressources sont utilisées et prises au sérieux.

86. Les ONS ne sont pas les seuls acteurs dans le domaine de l'éthique des données. Ils risquent de ne jouer qu'un rôle limité dans l'élaboration d'une réflexion plus large. Cela doit changer si les ONS veulent influer sur l'avenir de l'éthique des données plutôt que d'en subir les conséquences.

VII. Recommandations

87. Il est nécessaire de mieux définir et faire savoir ce que l'on entend par éthique des données dans différents contextes et pourquoi celle-ci joue un rôle important dans l'utilisation des données à des fins statistiques. Cela devrait contribuer à une compréhension commune dans différents contextes internationaux. Pour ce faire, il conviendrait d'organiser une séance à l'intention des hauts responsables des ONS, qui examineraient les meilleures pratiques et élaboreraient des normes internationales de haut niveau en matière d'éthique des données pour leurs organisations. Cette séance s'appuierait sur cette analyse approfondie

réussie et permettrait de recueillir davantage d'exemples de pratiques en matière d'éthique des données au sein de la communauté internationale et de promouvoir une plus grande collaboration au sein du système statistique international. Elle contribuerait grandement à éclairer les travaux des ONS en matière d'éthique des données, à renforcer les normes au niveau international et à aider la communauté statistique à parler d'une voix plus unie dans le domaine plus large de l'éthique des données.

88. Pour promouvoir au sein des ONS une culture où l'éthique des données est prise au sérieux, les dirigeants de ces organismes devraient faire part à leurs salariés de leur ferme intention d'adhérer aux principes de l'éthique des données ; le concept de l'éthique des données, les avantages qui en découlent, les valeurs et les questions y relatives devraient être largement diffusés au sein des ONS. L'éthique des données devrait également relever de la responsabilité d'un comité d'évaluation (qui serait soit un comité d'éthique des données, soit un comité d'éthique plus général), ou être ajoutée au mandat d'un comité de gestion/scientifique. Les ONS devraient faire siéger des membres indépendants compétents dans ces comités afin que les décisions relatives à l'éthique des données puissent être prises par des experts indépendants et des membres de la société civile en dehors de tout intérêt particulier.

89. Afin de garantir que les principes d'éthique des données sont bien intégrés dans leurs activités tout au long du cycle de vie des données, allant de la collecte à la diffusion, les ONS devraient désigner une personne/un groupe responsable de la communication et de l'assistance dans ce domaine. Cette personne ou ce groupe devrait disposer de moyens de pression sur les gestionnaires de programmes pour les aider à respecter les principes éthiques. Il lui incombe de fournir des outils faciles à utiliser, une assistance aux utilisateurs et une formation pour permettre aux analystes de mener des analyses conformes à ces principes. Ces outils et activités doivent être simples et aussi peu contraignants que possible pour les analystes qui en bénéficient.

90. Les ONS devraient être transparents sur leurs processus d'éthique des données et diffuser largement leurs travaux en la matière afin de montrer aux parties prenantes qu'ils prennent l'éthique des données au sérieux. Ils ne devraient pas se contenter de dire ce qu'ils font pour garantir que leur collecte et leur utilisation des données sont éthiques, mais ils devraient également prouver qu'ils font ce qu'ils disent et présenter l'impact de leurs travaux en matière d'éthique des données.

91. En tant qu'importants producteurs et utilisateurs de données, les ONS devraient diffuser en toute confiance leurs travaux sur l'éthique des données auprès d'autres parties concernées afin qu'ils puissent être des acteurs de premier plan dans l'élaboration de la réflexion sur l'éthique des données et les pratiques y relatives, tant au niveau national qu'au niveau international.

VIII. Examen par le Bureau de la Conférence des statisticiens européens

92. Le Bureau a procédé à un examen approfondi de l'éthique des données en octobre 2022 sur la base d'un document établi par le Canada et le Royaume-Uni, d'une contribution d'Eurostat et d'une note du secrétariat. Les observations suivantes ont été formulées :

a) Le Bureau a jugé le document très opportun et pertinent. Les ONS participent à de nouvelles activités et opèrent dans un nouveau contexte − la législation sur la protection de la vie privée devient plus stricte, l'acceptabilité du public ne peut être considérée comme acquise, un partenariat avec les propriétaires de données est nécessaire et tout faux pas peut avoir de graves conséquences pour l'ensemble de la communauté ;

b) Le concept d'éthique des données doit être expliqué et diffusé à la fois dans la communauté statistique internationale et au sein de chaque ONS − son essence, les raisons pour lesquelles il est important aujourd'hui et ses liens avec les cadres existants que sont les Principes fondamentaux de la statistique officielle, les valeurs fondamentales, le code de bonnes pratiques de l'UE, les recommandations du Comité des statistiques et de la politique statistique de l'OCDE et la législation statistique, ainsi qu'avec d'autres concepts examinés.

Il pourrait être considéré comme une méta-orientation susceptible de révéler des lacunes dans les cadres existants : par exemple, la vie privée est déjà bien couverte, mais d'autres valeurs sociales peuvent ne pas l'être ;

c) Les pratiques en matière d'éthique des données visent à garantir au public que ses données sont en sécurité, qu'elles sont utilisées de manière appropriée et qu'elles concourent au bien public. L'éthique doit être considérée comme un système dont font partie intégrante la participation du public et la transparence des actions. Dans la communication avec le public, il est plus efficace de se référer à des cas d'utilisation concrets qu'à des données générales. Il conviendrait également de préciser que la question de l'éthique n'est pas examinée en raison de mauvaises pratiques, mais plutôt en raison du nouveau contexte ;

d) D'un point de vue pratique, les décisions en matière d'éthique des données font de plus en plus partie des tâches quotidiennes accomplies tout au long du cycle de vie des données. Il est nécessaire de partager des expériences et des conseils pratiques, y compris des exemples de comportements contraires à l'éthique que l'on peut observer chez certains fournisseurs du secteur privé. Ces décisions ne peuvent pas être basées sur un algorithme, car les situations ne sont jamais claires et le contexte change constamment. Il semble donc préférable de disposer d'un petit nombre de principes de haut niveau, d'un mécanisme permettant de suivre l'évolution de la situation et d'un groupe indépendant de personnes diverses qui peuvent dialoguer en s'appuyant sur le cadre convenu. Toutefois, seuls quelques pays ayant déjà mis en place de telles pratiques, leur impact doit encore être évalué ;

e) La communauté statistique a besoin de participer au débat plus large sur l'éthique des données. Les personnes qui ne travaillent pas dans le domaine des statistiques parlent implicitement de l'éthique des données, mais elles ne considèrent pas que les statistiques sont véritablement concernées. Il est nécessaire de débattre de la manière dont les statistiques peuvent être intégrées dans un dialogue plus large.

93. En conclusion, le Bureau de la Conférence des statisticiens européens est favorable à la poursuite des travaux dans ce domaine et approuve les recommandations présentées dans le document d'examen approfondi, y compris l'organisation d'un séminaire de la Conférence sur l'éthique des données.

IX. Bibliographie

Floridi et Taddeo, 2016, https://doi.org/10.1098/rsta.2016.0360

Hagendorff 2020, https://doi.org/10.1007/s11023-020-09517-8

Hasselbalch 2019, https://doi.org/10.14763/2019.2.1401

Main 2018, https://doi.org/10.1089/big.2018.0083

Kisselburgh 2020, https://doi.org/10.1007/978-3-030-82786-1_17

Mulligan, Koopman et Doty 2016, https://doi.org/10.1098/rsta.2016.0118

Taylor 2016, https://doi.org/10.1098/rsta.2016.0116