



# Conseil économique et social

Distr. générale  
9 avril 2024  
Français  
Original : anglais

## Commission économique pour l'Europe

### Conférence des statisticiens européens

Soixante-douzième réunion plénière

Genève, 20 et 21 juin 2024

Point 9 de l'ordre du jour provisoire

**Travaux du Groupe de haut niveau sur  
la modernisation de la statistique officielle**

## **Programme de travail du Groupe de haut niveau sur la modernisation de la statistique officielle pour 2024**

**Document élaboré par le secrétariat et la Commission de contrôle du  
Groupe de haut niveau sur la modernisation de la statistique officielle**

### *Résumé*

Le présent document donne un aperçu des travaux que le Groupe de haut niveau sur la modernisation de la statistique officielle (ci-après, le Groupe de haut niveau) prévoit de mener en 2024.

À la réunion de février 2024, le Bureau de la Conférence des statisticiens européens a examiné et approuvé le programme de travail du Groupe de haut niveau. La Conférence est invitée à approuver le programme de travail proposé.



## I. Introduction

1. Le Groupe de haut niveau sur la modernisation de la statistique officielle (ci-après, le Groupe de haut niveau)<sup>1</sup> constitue une plateforme de collaboration qui permet aux experts des organismes statistiques d'élaborer des stratégies et de mettre au point des solutions modernes de manière souple et agile. Il met régulièrement à jour sa vision stratégique afin de l'adapter à un environnement changeant et définit de nouvelles priorités en conséquence. Le programme de travail annuel vise à traduire l'émergence de nouveaux besoins et, dans le même temps, à confirmer ou améliorer les résultats obtenus des années précédentes.
2. Le Groupe de haut niveau est une plateforme sans équivalent qui a apporté plusieurs contributions importantes à la modernisation de la statistique officielle, comme le Modèle générique du processus de production statistique (GSBPM) et le Modèle générique d'informations statistiques (GSIM). Les projets qu'il a menés dans le domaine des mégadonnées, de l'intégration des données, de la communication stratégique et de l'apprentissage automatique ont été à l'origine de la mise en œuvre de nouvelles technologies, de nouvelles méthodes et d'autres moyens au sein des organismes statistiques. En raison de son succès, le mode de travail du Groupe de haut niveau inspire aujourd'hui d'autres initiatives de modernisation.
3. La participation aux travaux du Groupe de haut niveau est ouverte à tous ceux qui souhaitent contribuer à faire avancer et à moderniser la statistique officielle. De nombreux collègues ont contribué aux travaux en ce sens et, chaque année, environ 250 collègues participent concrètement aux activités du programme, tandis que des milliers de collègues en bénéficient indirectement en collaborant avec les membres ou en participant aux réunions d'experts, aux ateliers et aux séminaires en ligne organisés dans le cadre du programme de travail du Groupe de haut niveau. Comme les données d'expérience et tous les résultats obtenus dans ce cadre sont mis à la disposition du public, l'incidence réelle est encore plus importante.
4. Chaque année, tous les membres de la Conférence des statisticiens européens (ci-après, la Conférence) sont invités à soumettre des propositions de projets et d'activités au Groupe de haut niveau pour examen. Ces propositions doivent être conformes à la mission, à la vision à long terme ainsi qu'aux priorités à court et à moyen terme définies par le Groupe de haut niveau. En outre, le Réseau de recherche fondamentale du Groupe de haut niveau est parfois sollicité pour évaluer et renforcer les idées et les propositions émises par la Commission de contrôle du Groupe de haut niveau. La Commission de contrôle fait part de ses observations et choisit les propositions qui donneront lieu à des projets du Groupe de haut niveau ou nécessiteront un suivi. Au cours de l'atelier annuel sur la modernisation de la statistique officielle, qui se tient à la fin du mois de novembre, ces propositions sont présentées et discutées par les experts qui travaillent sur la modernisation. Des débats en petits groupes permettent d'évaluer les propositions de projets et de formuler des suggestions aux fins de leur suivi. La Commission de contrôle examine ensuite les propositions, les affine et évalue les capacités disponibles. Les propositions de projets et d'activités retenues sont finalement soumises au Groupe de haut niveau pour approbation.
5. Le présent document donne une vue d'ensemble du programme de travail pour 2024, qui est inspiré des débats menés au cours de l'atelier de 2023 sur la modernisation de la statistique officielle (21 et 22 novembre, Genève (Suisse))<sup>2</sup> et a été affiné par la Commission de contrôle afin que les membres de la communauté statistique puissent participer aux activités qui leur sont les plus utiles.

## II. Projets

6. Les projets du Groupe de haut niveau sont axés sur les technologies et les idées nouvelles et ont normalement une durée fixe d'un à deux ans.

---

<sup>1</sup> Voir l'annexe I pour la structure du Groupe de haut niveau.

<sup>2</sup> <https://unece.org/statistics/events/HLG2023>.

7. Pour 2024, deux projets sont proposés à l'examen du Groupe de haut niveau: i) les logiciels statistiques libres ; et ii) l'IA générative au service de la statistique officielle. Ces projets portent sur deux changements technologiques récents importants qui présentent un grand potentiel pour la statistique officielle.

8. Les propositions de projets, qui visent à répondre aux besoins de nombreux statisticiens, ont été présentées lors de l'atelier sur la modernisation qui s'est tenu en novembre 2023. Les commentaires des participants et des membres de la Commission de contrôle sont inclus dans la description des projets ci-dessous<sup>3</sup>.

## A. Projet relatif aux logiciels statistiques libres

9. Compte tenu de la nécessité croissante de gagner en ouverture, en transparence et en efficacité, de nombreux organismes statistiques passent des logiciels traditionnels protégés aux logiciels libres. Cette transition pose toutefois des problèmes sur le plan de l'assistance, de la maintenance, de la formation, des conditions de partage et des aspects juridiques. La question des logiciels libres a été examinée lors de la soixante et onzième réunion plénière de la Conférence des statisticiens européens<sup>4</sup> et **le Bureau de la Conférence a demandé au Groupe de haut niveau de se pencher sur le sujet.**

10. L'objectif du projet relatif aux logiciels statistiques libres est de parvenir à une meilleure compréhension commune des avantages et des inconvénients du passage à une utilisation plus large des logiciels libres pour la production de statistiques officielles, ainsi que des choses à faire ou à ne pas faire dans ce domaine, afin que cette utilisation devienne une des clés de la production en question.

11. Après une première phase de cadrage, le projet se concentrera sur les points suivants :

a) **Aspects génériques de l'utilisation systématique de méthodes fondées sur les logiciels libres** pour établir des statistiques officielles, notamment en ce qui concerne l'organisation de la maintenance, du soutien et de la formation ; les normes et les principes ; les aspects juridiques et les responsabilités ; les modèles de licence et la répartition équitable des coûts ; la création d'une communauté, la communication et la participation externe (de la communauté scientifique et du secteur privé, par exemple) ; le processus d'incubation (de l'idéation à la production). Il sera peut-être nécessaire d'accorder une attention différente à ces aspects selon que le point de vue sera celui de l'utilisateur (qui utilise les solutions existantes fondées sur des sources ouvertes) ou celui du producteur (qui produit de nouvelles solutions fondées sur des sources ouvertes) ;

b) **Analyse de cas concrets d'utilisation des logiciels libres** dans les domaines de la collecte, de l'analyse, du traitement et de la diffusion des données. Le choix des cas d'utilisation à analyser pourra être fait indépendamment des travaux susmentionnés ou conjointement avec ces travaux. Les résultats de l'analyse pourront servir, en suggérant des technologies et des méthodes concrètes fondées sur les logiciels libres, à soutenir les travaux menés du sommet à la base sur les technologies de ce type.

## B. Projet relatif à l'IA générative au service de la statistique officielle

12. Les capacités de l'intelligence artificielle (IA) ont fait un bond en avant au cours des dernières années grâce aux progrès des grands modèles de langage (GML) capables de traiter le langage naturel et de générer des textes et la communauté des statisticiens est de plus en plus consciente du potentiel de transformation que recèlent les GML.

13. En réponse à cet intérêt croissant, le Réseau de recherche fondamentale et le Groupe de l'application de la science des données et des méthodes modernes (groupes de modernisation du Groupe de haut niveau) ont entrepris de rédiger un livre blanc sur les grands modèles de langage dans le contexte de la statistique officielle, qui a été achevé dans un délai

<sup>3</sup> La version complète des propositions de projets est disponible à l'annexe II ; une liste des commentaires des participants à l'atelier sur la modernisation est disponible à l'annexe III.

<sup>4</sup> <https://unece.org/statistics/events/CES2023>.

relativement court de quatre mois. Ce document<sup>5</sup> explore les perspectives et les conséquences des grands modèles de langage dans le domaine de la statistique officielle, ainsi que les risques associés, formule des recommandations et présente des considérations d'ordre stratégique.

14. S'appuyant sur le livre blanc consacré aux grands modèles de langage, le projet vise à approfondir l'étude du potentiel de l'IA générative, grande catégorie de système avancé d'intelligence artificielle qui englobe les GML (par exemple, la création d'images) et les considérations stratégiques à prendre en compte lorsque les organismes statistiques veulent utiliser l'IA générative de manière efficace et responsable (par exemple, la gouvernance, les modèles ouverts), ainsi qu'à mettre en évidence les possibilités d'élaborer en commun des solutions concrètes.

15. Le projet débutera par une phase initiale de cadrage, à l'issue de laquelle les trois activités principales suivantes sont prévues :

a) **Mettre en commun les cas d'utilisation, les données d'expérience et les enseignements tirés.** Les exemples d'utilisation ne se limitent pas au domaine de la production, mais s'étendent à d'autres domaines de l'entreprise, tels que les ressources humaines et les finances. Cette activité aidera les organismes statistiques à établir des priorités dans les domaines les plus prometteurs et les plus réalistes pour la statistique officielle ;

b) **Coconstruire des solutions** dans des domaines d'intérêt commun pour de nombreux organismes statistiques (par exemple, la conception d'invites, le copilotage, les dialogueurs, les recherches avancées sur le Web) ;

c) **Compiler les pratiques et les recommandations concrètes** fondées sur les deux premières activités et sur le livre blanc consacré aux GML. Il est essentiel de se concentrer sur quelques thèmes clés qui sont particulièrement pertinents et importants pour les organismes statistiques (par exemple, la confidentialité, la sécurité, l'assurance qualité). Cette activité pourrait également consister à établir des protocoles communs assortis de prescriptions du point de vue de la statistique officielle, qui pourraient ensuite être utilisés pour collaborer avec les entreprises technologiques.

16. Pour les deux projets, il sera important de procéder avec agilité, compte tenu de l'évolution rapide des technologies liées aux sources ouvertes et à l'intelligence artificielle, ainsi que du grand intérêt suscité par les sujets abordés. Les projets nécessiteront en outre un mélange d'expertise technique, commerciale et managériale au sein des équipes de projet, qui seront étroitement guidées par la Commission de contrôle.

17. L'extension du projet ModernStats Carpentries au-delà de 2023 suscite également de l'intérêt. Cette prolongation permettrait de tirer toutes les leçons de la mise en œuvre du module 1 de Python et d'exploiter d'autres possibilités.

### C. Procédures opérationnelles applicables aux projets

18. Les travaux menés dans le cadre des projets du Groupe de haut niveau sont normalement dirigés par un chef de projet qui est soit un membre du Groupe de haut niveau (contribution en nature), soit un contractuel mandaté par la Commission économique pour l'Europe (CEE) et rémunéré au titre du Fonds d'affectation spéciale du Groupe de haut niveau. Les projets peuvent également être dirigés conjointement par un chef de projet et un responsable technique. Les modules de travail ou les groupes thématiques sont normalement dirigés par des présidents choisis parmi les membres du projet. Le wiki et les conférences en ligne de la CEE sont utilisés pour suivre l'avancement des projets. Les chefs de projet font régulièrement rapport à la Commission de contrôle, qui supervise les progrès et oriente les travaux si nécessaire. La CEE apporte son soutien à la gestion des projets et fournit des services de secrétariat supplémentaires en cas de besoin. Des ateliers en mode « sprint » sont souvent utilisés pour élargir le champ d'application et accélérer les travaux. Les progrès et

<sup>5</sup> [https://unece.org/sites/default/files/2023-12/HLGMOS%20LLM%20Paper\\_Preprint\\_1.pdf](https://unece.org/sites/default/files/2023-12/HLGMOS%20LLM%20Paper_Preprint_1.pdf).

les résultats sont présentés lors de réunions d'experts afin d'échanger des informations sur les travaux et de recueillir les commentaires d'un public plus large.

### III. Groupes de modernisation

19. Les groupes de modernisation du Groupe de haut niveau ont été créés pour apporter un soutien continu au sujet des thèmes transversaux importants pour la modernisation des organismes statistiques, comme les normes, les cadres, la communication et les ressources humaines. À la différence des projets du Groupe de haut niveau, les travaux des groupes de modernisation s'inscrivent dans le long terme, mais les activités de chaque groupe (organisé en équipes de travail ou en sous-groupes) changent chaque année pour répondre aux besoins les plus urgents dans les domaines de travail respectifs, ce qui permet de poursuivre la modernisation et de rester à jour sur le plan de l'innovation.

20. Les groupes choisissent un président général et d'autres responsables sont sélectionnés pour les différentes équipes de travail et les différents sous-groupes. Les groupes tiennent des réunions plénières mensuelles et les équipes de travail (ou les sous-groupes) se réunissent virtuellement au moins une fois par mois. Les présidents des groupes informent régulièrement la Commission de contrôle au sujet de l'avancement des travaux.

21. Les groupes peuvent organiser des ateliers en mode sprint afin d'accélérer les travaux. Le wiki de la CEE et d'autres plateformes telles que GitHub sont utilisés à des fins de collaboration et de coordination des travaux. Les services de secrétariat sont toujours assurés par la CEE. Les principales difficultés rencontrées sont la disponibilité des membres du groupe et les ressources dont dispose la CEE pour soutenir et coordonner les travaux. Comme pour toutes les activités du Groupe de haut niveau, la participation est volontaire et tout membre de la communauté de la statistique officielle intéressé par une partie ou une autre des activités est encouragé à participer. La participation aux groupes de modernisation offre des possibilités uniques de renforcement des compétences et de constitution de réseaux en dehors d'un cadre national.

22. Le Groupe de haut niveau a chargé la Commission de contrôle d'examiner régulièrement la structure et les domaines de travail des groupes de modernisation, d'évaluer si ces derniers continuent à respecter les priorités essentielles définies par le Groupe de haut niveau et de fournir des orientations.

23. Les activités que les groupes de modernisation prévoient de mener en 2024 sont brièvement décrites ci-dessous. Des descriptions plus détaillées de toutes ces activités sont disponibles sur le site web du Groupe de haut niveau : <https://statswiki.unece.org/x/lwF-EQ>.

#### A. Groupe de l'application de la science des données et des méthodes modernes

24. Le Groupe de l'application de la science des données et des méthodes modernes, présidé par Statistics New Zealand, vise à dépasser le cadre conceptuel de la science des données et des méthodes modernes et à trouver des moyens concrets de moderniser les processus opérationnels des organismes de statistique. Lors du cycle de recensement des sujets émergents de 2023, trois suggestions ont été faites pour le cycle de 2024 :

a) **Promouvoir l'IA responsable dans les organismes statistiques : faire le lien entre la connaissance et la pratique.** Cette activité vise à créer un répertoire des ressources et des outils afin d'aider les organismes statistiques à adapter leur mode de fonctionnement à une application responsable des principes de l'IA. Trois modules de travail sont prévus sur les thèmes suivants : renforcement des connaissances par la formation, mise au point d'outils pratiques, en particulier d'applications éthiques de l'IA, sur la base du cadre pour une IA responsable élaboré en 2023 par le Groupe de l'application de la science des données et des méthodes modernes et mise en œuvre de programmes pilotes à partir d'études de cas. L'objectif est de favoriser une culture de l'IA responsable, en garantissant le respect des normes et en promouvant la transparence et la confiance ;

b) **Au-delà des prédictions ponctuelles : garantir la fiabilité des statistiques officielles en quantifiant l'incertitude.** Cette activité consistera à effectuer une analyse documentaire en mettant l'accent sur la prédiction conforme, méthode d'apprentissage automatique permettant de quantifier l'incertitude, afin de comprendre l'état actuel de la recherche dans ce domaine. L'objectif est de produire un rapport détaillé contenant des recommandations pratiques et des lignes directrices à l'intention des spécialistes de l'apprentissage automatique, afin d'aider les organismes statistiques à appliquer efficacement cette méthode dans le cadre de la statistique officielle et à mettre au point des algorithmes d'apprentissage automatique. Grâce à la quantification de l'incertitude, les organismes statistiques peuvent améliorer la fiabilité des statistiques officielles produites au moyen de l'apprentissage automatique et, par conséquent, avoir une plus grande assurance lorsqu'ils communiquent avec les utilisateurs et le grand public au sujet de l'utilisation qu'ils font de l'apprentissage automatique et de l'intelligence artificielle ;

c) **Arguments économiques en faveur de la modélisation graphique et des bases de données graphiques tout au long du cycle de vie des données granulaires.** Cette activité sera centrée sur l'utilisation de la modélisation graphique et des bases de données graphiques pour répondre à la demande croissante de données granulaires de la part des chercheurs. Les objectifs sont notamment de cartographier les cas d'utilisation, de recenser les obstacles et de mieux comprendre les avantages potentiels des méthodes fondées sur les graphes pour la fourniture de services d'intelligence artificielle améliorés. Cela implique, du point de vue de la modernisation, de recueillir des informations et d'en faire des ressources qui peuvent être partagées à des fins de renforcement des capacités.

25. Le Groupe collaborera étroitement, dans les domaines pertinents, avec la Réunion d'experts sur l'édition des données statistiques, le Réseau de recherche fondamentale et d'autres initiatives internationales, le cas échéant.

## B. Réseau de recherche fondamentale

26. Le Réseau de recherche fondamentale, présidé par Statistics Netherlands, est le laboratoire d'idées de la communauté ModernStats qui est à l'origine de la réussite de projets du Groupe de haut niveau tels que ceux qui portent sur l'apprentissage automatique, les données synthétiques et l'informatique en nuage. Ce groupe est composé de membres de diverses organisations nationales et internationales ayant une large connaissance des sujets liés à l'innovation dans le cadre de la statistique officielle. Les activités prévues en 2024 sont les suivantes :

a) **Évaluation des propositions de projets soumises au Groupe de haut niveau :** tout au long de l'année, le Réseau de recherche fondamentale sélectionne de nouveaux sujets potentiels en collaborant activement avec la communauté statistique pour trouver de nouvelles idées et en évaluant les propositions de projets en fonction de la vision et des priorités du Groupe de haut niveau, sous la supervision de la Commission de contrôle. Le Réseau peut organiser des activités temporaires pour assurer le suivi des sujets prometteurs ou des propositions de projets qui n'ont pas été retenues ;

b) **Équipe de terrain en charge des jumeaux numériques :** l'objectif est d'étudier ce que recouvre le concept de « jumeaux numériques », où et comment la communauté de la statistique officielle pourrait contribuer à cette étude et si une normalisation internationale des nouvelles notions correspondantes pourrait être utile. Dans le domaine de la collecte de données, les modèles fondés sur les jumeaux numériques peuvent être utilisés pour simuler le recueil des données et tester les hypothèses. Les travaux se poursuivront pour dresser le bilan des initiatives existantes et déterminer les cas d'utilisation des données et des méthodes statistiques ;

c) **Connexion avec la réunion de 2024 des experts chargés d'étudier la collecte de données :** le groupe se mettra en relation avec le comité d'organisation de la réunion d'experts de 2024 sur la collecte de données, afin de discuter de la manière dont les nombreux arguments avancés lors des réunions tenues en 2023 par le Réseau de recherche fondamentale sur la collecte de données pourraient être examinés ou présentés lors de la réunion d'experts.

27. Étant donné que les initiatives sur la modernisation et l'innovation dans le domaine de la statistique officielle se multiplient, une coordination étroite entre ces initiatives sera d'autant plus importante en 2024. Le Réseau de recherche fondamentale continuera de participer aux autres activités du Groupe de haut niveau et de collaborer avec des acteurs extérieurs, tels que le sous-groupe d'analyse prospective du Réseau d'innovation du système statistique européen (SSE).

## C. Groupe des capacités et de la communication

28. Le Groupe des capacités et de la communication axe ses travaux sur les changements organisationnels à opérer et les problèmes de communication à surmonter à l'appui de la modernisation des organismes de statistique. Il est présidé conjointement par le Bureau central de statistique polonais et l'Office central irlandais de la statistique.

29. Ce groupe a adapté son programme de travail pour aider les organismes de statistique à faire face à l'évolution des modalités de travail (par exemple, après la pandémie) et à la nécessité d'intensifier la communication sur les plans interne et externe. Il se répartira en trois équipes spéciales et mettra en place des sous-équipes chargées de diverses activités au sein de chaque équipe spéciale :

a) **L'emploi de demain** : les employés des nouvelles générations ont un regard différent sur la vie, le travail et le poste qu'ils occupent. Les organismes nationaux de statistique ont et auront de plus en plus de mal à attirer du personnel nouveau et à retenir ce personnel en lui offrant simplement un emploi et un salaire. En fonction du temps et des ressources disponibles et de l'expertise des membres de l'équipe, l'équipe spéciale choisira parmi les sujets ci-dessous ceux sur lesquels cette dernière travaillera :

i) **Extension des travaux sur le modèle générique de croissance** : estimant que le modèle de croissance applicable aux questions organisationnelles complexes et les outils correspondants (2023) ont un grand potentiel, l'équipe spéciale des emplois de demain souhaite stimuler l'utilisation de ce modèle par les organismes nationaux de statistique, soutenir cette utilisation et encourager ces organismes à partager leurs données d'expérience et les enseignements tirés de cette expérience. Cette activité consistera à : i) demander aux organismes nationaux participants de partager leurs données d'expérience sur l'utilisation du modèle de croissance et des outils correspondants ; ii) promouvoir activement l'utilisation du modèle de croissance et des outils correspondants au sein des organismes nationaux de statistique ; iii) encourager la collaboration (bi- ou multinationale) dans ce domaine ;

ii) **Stratégie d'image de l'employeur** : l'image de marque de l'employeur est très importante pour tous les organismes nationaux de statistique, en particulier en raison de la forte concurrence sur le marché de l'emploi, principalement avec le secteur privé. Une bonne image de marque peut attirer les futurs employés dont les compétences sont indispensables aux organismes nationaux de statistique. L'activité sera axée sur le recensement des pratiques et le partage de données d'expérience entre les spécialistes des ressources humaines et de la communication, à l'aide du modèle générique de croissance. L'équipe élaborera ensuite un document succinct contenant une analyse des premiers enseignements tirés et des recommandations à l'intention des organismes de statistique, en tenant compte des différents niveaux de maturité et des mesures nécessaires à prendre ;

iii) **Analyse des données relatives aux ressources humaines** : cette activité consistera à examiner les outils de communication d'informations tels que Power BI, qui permettent de rassembler des données provenant de diverses sources, comme les données sur l'érosion naturelle des effectifs, les absences, le délai de recrutement et le pourcentage de déclarations en faveur de la diversité. L'utilisation de ces informations dans le cadre du cycle d'évaluation permet aux employés et aux équipes d'analyser les domaines d'intérêt de leur organisation ou de leur section ;

b) **Éthique** : L'équipe spéciale sera divisée en deux modules de travail – éthique des affaires et éthique des données – et se concentrera en 2024 sur les activités suivantes : i) organiser un atelier sur l'éthique (26-28 mars 2024 à Genève (Suisse))<sup>6</sup> ; ii) parachever l'analyse des résultats de l'enquête sur l'éthique ; iii) achever un ouvrage de référence sur l'éthique à l'intention des organismes nationaux de statistique ; iv) intégrer l'éthique dans divers domaines du Modèle générique d'activité des organismes statistiques (GAMSO) et du Modèle générique du processus de production statistique (GSBPM) en collaboration avec le Groupe d'appui à la mise en œuvre des normes ; v) élaborer une définition internationale commune de l'éthique des données et une conception commune de l'intérêt d'une telle éthique pour les organismes nationaux de statistique ; vi) inscrire l'éthique des données dans un cadre international qui soit fondé sur des principes et qui puisse être utilisé par les organismes nationaux de statistique ; vii) rassembler du matériel de formation et d'information sur l'éthique et les meilleures pratiques en matière de définition, de mise en œuvre et de diffusion de l'éthique des données au sein des différents organismes nationaux de statistique ; viii) présenter des études de cas sur les bonnes pratiques et leurs conséquences.

c) **Communication – utilisation de l'IA pour les statistiques officielles du point de vue de la communication** : l'équipe spéciale s'intéressera au domaine de l'IA générative et examinera les points suivants : i) comment utiliser l'IA pour stimuler la productivité des spécialistes de la communication ; ii) comment faire savoir que les organismes statistiques utilisent l'IA afin de maintenir la confiance dans les statistiques officielles ; iii) comment l'IA influe sur le paysage de l'information à mesure que les services fondés sur l'IA générative deviennent progressivement la principale source d'informations et de connaissances et comment les organismes statistiques devraient faire connaître leurs données et leurs services.

## D. Groupe d'appui à la mise en œuvre des normes

30. Le Groupe d'appui à la mise en œuvre des normes, présidé par Statistique Canada, apporte son soutien à la mise en application des modèles « ModernStats » (par exemple, les modèles GAMSO et GSBPM) dans le cadre d'une série d'activités comprenant l'élaboration, l'amélioration, l'intégration, la promotion et la mise à jour des modèles. Le Groupe de haut niveau étant le dépositaire de ces modèles, qui ont une portée mondiale, un soutien continu est essentiel.

31. Dans les domaines statistiques, les normes de mise en œuvre se sont multipliées (par exemple SDMX, DDI, VTL, PROV, SKOS, DCAT et schema.org)<sup>7</sup>, ce qui rend de plus en plus difficile pour les organismes statistiques (et en particulier pour les non-experts) de décider quand, où et comment utiliser efficacement ces normes dans l'ensemble du processus de production de manière interopérable. Étant donné que l'efficacité opérationnelle repose à la fois sur l'automatisation et l'interopérabilité, qui à leur tour reposent souvent sur des normes internationales de mise en œuvre, il est primordial de fournir à la communauté statistique des conseils clairs sur la manière d'utiliser ensemble ces normes afin d'atteindre les objectifs de modernisation avec efficacité par rapport au coût. Le Groupe d'appui à la mise en œuvre des normes est dans une position idéale pour donner des conseils de qualité sur la manière d'utiliser et d'intégrer de multiples normes et outils dans l'infrastructure des plateformes de données. Les modèles ModernStats peuvent : i) fournir le contexte adéquat pour rendre possible l'interopérabilité ; ii) être associés à la plupart, voire à la totalité, des normes mentionnées ci-dessus. Le Groupe d'appui à la mise en œuvre des normes travaille depuis un certain temps déjà à la concrétisation de cette vision et prévoit de poursuivre ses efforts dans ce sens. Dans ce contexte, il a donné la priorité aux activités suivantes :

<sup>6</sup> Voir la section IV (Réunions d'experts et ateliers) pour de plus amples détails.

<sup>7</sup> Statistical Data and Metadata eXchange (SDMX) (échange de données et de métadonnées statistiques), Data Documentation Initiative (DDI) (initiative de documentation des données), Validation and Transformation Language (VTL) (langage de validation et de transformation), PROV, Simple Knowledge Organization System (SKOS) (système simple d'organisation des connaissances), Data Catalog Vocabulary (DCAT) (vocabulaire des catalogues de données).



a) **Poursuite de la révision des modèles GSBPM et GAMS0** : l'objectif de cette révision est de veiller à ce que les modèles restent pertinents et continuent à servir de cadre de référence pour les organismes statistiques. Comme ces modèles sont étroitement liés (le GSBPM fait en réalité partie du GAMS0), ils sont révisés par une seule équipe spéciale dans le cadre de la même activité. Les expériences des autres groupes d'experts internationaux concernés (par exemple, l'équipe spéciale de l'éthique du Groupe des capacités et de la communication) seront prises en compte, de même que les résultats de la consultation des utilisateurs. Comme indiqué dans le programme de travail pour 2023, ces travaux seront achevés en 2024 ;

b) **Révision et mise à jour de l'Architecture commune des données statistiques (CSDA)** : il est indispensable de mettre à jour ce modèle pour tenir compte des versions actualisées du GSIM, du GSBPM et du GAMS0 et d'étendre la CSDA aux nouveaux domaines d'intérêt apparus au cours des cinq dernières années (par exemple, l'IA générative et les données FAIR<sup>8</sup>) et aux domaines qui n'ont pas été inclus dans la spécification de 2018 (par exemple, l'apprentissage automatique et la préservation de la vie privée) ;

c) **Application des normes SDMX, VTL et DDI à la mise en œuvre du modèle GSIM** : s'appuyant sur les travaux de l'équipe spéciale SDMX-DDI-GSBPM (2022-2023), cette activité vise à approfondir l'examen des éléments des normes SDMX, VTL et DDI afin d'étudier leur correspondance avec les classes du modèle GSIM, ainsi qu'à définir des cas d'utilisation susceptibles de faciliter la mise en œuvre du modèle GSIM par les organismes statistiques. Ce travail devrait également permettre de déterminer les entrées et les sorties des sous-processus du modèle GSBPM, sur la base de travaux antérieurs sur la relation entre les modèles GSBPM et GSIM (2021) ;

d) **Version 2 de l'Ontologie de base de la statistique officielle** : l'activité « Ontologie de base de la statistique officielle » (COOS), lancée en 2021, sert de modèle d'intégration pour l'ensemble des normes ModernStats, complété par des éléments de vocabulaire normalisés externes bien connus. Elle a permis d'obtenir des résultats importants à cet égard : une spécification de l'ontologie, un document de gouvernance, une politique relative aux identifiants uniformes de ressources (URI) et une ontologie OWL. La nécessité de mettre à jour la COOS s'explique en partie par la nécessité d'envisager des mises à jour des modèles GSIM, GSBPM et GAMS0, ainsi que de prendre en compte les commentaires reçus au sujet de la version initiale de la COOS. Cette mise à jour est également l'occasion d'examiner les dernières pratiques en matière de données ouvertes corrélées. Cette activité permettra de définir des cas d'utilisation, d'établir des lignes directrices et d'élaborer des documents en vue de poursuivre la diffusion et la promotion de la COOS.

32. Si les récentes avancées en matière d'IA offrent aux organismes statistiques une occasion importante d'agir avec plus d'efficacité, l'IA ne peut pas, à elle seule, transformer le processus de production. Les normes jouent un rôle essentiel dans la structuration des processus et l'amélioration de la qualité des informations (par exemple, en enrichissant les données d'une sémantique sous la forme de métadonnées normalisées automatiquement exploitables), ce qui permet d'accroître l'efficacité et l'automatisation des activités, base sur laquelle la production assistée par l'IA peut libérer tout son potentiel. Le Groupe de l'appui à la mise en œuvre des normes a entamé des discussions internes sur les avantages des normes à l'ère de l'IA et compte élaborer un document succinct résumant ces discussions afin d'aider à définir le champ d'application du prochain cycle d'activités de modernisation.

33. Le Groupe poursuivra également ses discussions sur la manière de soutenir l'utilisation des modèles ModernStats et d'autres normes et renforcera sa collaboration avec les organismes intéressés par l'intermédiaire d'une série de séminaires en ligne. Dans la mesure du possible, il s'efforcera d'améliorer la visibilité des modèles ModernStats, par exemple en organisant des sessions et en présentant des exposés lors des conférences et des réunions d'experts sur le sujet. Le travail restera coordonné avec d'autres groupes

<sup>8</sup> Faciles à trouver, accessibles, interopérables et réutilisables.

internationaux intéressés (par exemple, ESSnet CoP, CODATA)<sup>9</sup> et avec les autres activités du Groupe de haut niveau.

#### IV. Réunions d'experts et ateliers

34. Comme par le passé, des réunions d'experts seront organisées sous les auspices du Groupe de haut niveau dans plusieurs domaines de fond. Les réunions d'experts et les ateliers en présentiel suivront un cycle de vingt-quatre mois. Des réunions en ligne, qui pourront porter plus spécifiquement sur les questions émergentes et sur les projets et activités connexes du Groupe de haut niveau, pourront être organisées entre deux réunions en présentiel.

35. Les réunions d'experts et les ateliers sont toujours axés sur les innovations et la modernisation. Les thèmes et les sessions correspondent à la mission, à la vision et aux thèmes prioritaires du Groupe de haut niveau. Le cas échéant, les réunions ont pour but de récolter et de partager des informations sur les activités et les travaux du Groupe de haut niveau. L'un des principaux résultats de ces événements est le recensement des futurs domaines de travail et de collaboration entre les organismes. Les réunions d'experts s'adressent aux cadres intermédiaires et supérieurs. Il est prévu d'organiser les réunions suivantes en 2024 :

a) **Atelier sur l'éthique dans les organismes statistiques modernes** (26-28 mars, Genève). L'atelier sera organisé autour des thèmes suivants : i) l'éthique dans le contexte institutionnel ; ii) l'éthique dans la vie professionnelle quotidienne ; iii) l'éthique appliquée aux nouvelles sources de données et aux nouvelles technologies ; iv) l'éthique et la communication dynamique ;

b) **Réunion d'experts sur la collecte de données statistiques et les sources de données** (22-24 mai, Genève). La réunion, qui mettra l'accent sur les méthodes innovantes, y compris l'IA, sera organisée autour des thèmes suivants : i) méthode de collecte et d'acquisition de données auprès de sources multimodales et mixtes ; ii) autres sources de données et automatisation des processus ; iii) avenir des modes d'enquête et des enquêteurs ;

c) **Réunion d'experts sur la gestion et la formation des ressources humaines** (14-16 octobre, Genève) ;

d) **Réunion d'experts sur la diffusion et la communication des données statistiques** (à confirmer, en ligne) ;

e) **Atelier mondial ModernStats** (21 et 22 octobre, Genève) ;

f) **Atelier sur la modernisation de la statistique officielle organisé par le Groupe de haut niveau** (à confirmer, prévu en principe du 4 au 6 novembre à Genève).

36. Ces réunions sont organisées par des comités directeurs, qui sont composés de membres d'organismes nationaux de statistique, d'organisations internationales et d'universités.

37. Le Groupe de haut niveau, la Commission de contrôle, les équipes de projet et les groupes de modernisation du Groupe de haut niveau peuvent aussi demander, le cas échéant, qu'un séminaire en ligne ponctuel, court et ciblé ou un atelier en ligne soient organisés dans des domaines nouveaux.

#### V. Suivi des progrès et coordination

38. Un compte rendu des travaux relevant des groupes et des projets de modernisation est présenté chaque mois à la Commission de contrôle du Groupe de haut niveau. Celle-ci examine les éléments nouveaux et évalue les progrès accomplis, de concert avec les présidents de groupe et les chefs de projet, à l'occasion de réunions mensuelles. Si nécessaire,

<sup>9</sup> Communauté de pratique (CoP), Comité des données du Conseil international des sciences (CODATA).

des modifications sont apportées au programme de travail. Tous les deux mois, un bilan des activités de modernisation est mis à la disposition du public sur le [wiki ModernStats](#).<sup>10</sup>

39. Le Groupe de haut niveau s'efforce de veiller à harmoniser et à coordonner en permanence l'ensemble de ses activités avec les autres initiatives internationales, telles que les programmes menés dans le cadre du système statistique européen et de la Plateforme mondiale de l'ONU. Cela passe notamment par l'échange d'informations, la coordination et la collaboration. On outre, dans la plupart des cas, les participants aux activités du Groupe participent également à une partie des activités connexes menées par d'autres entités.

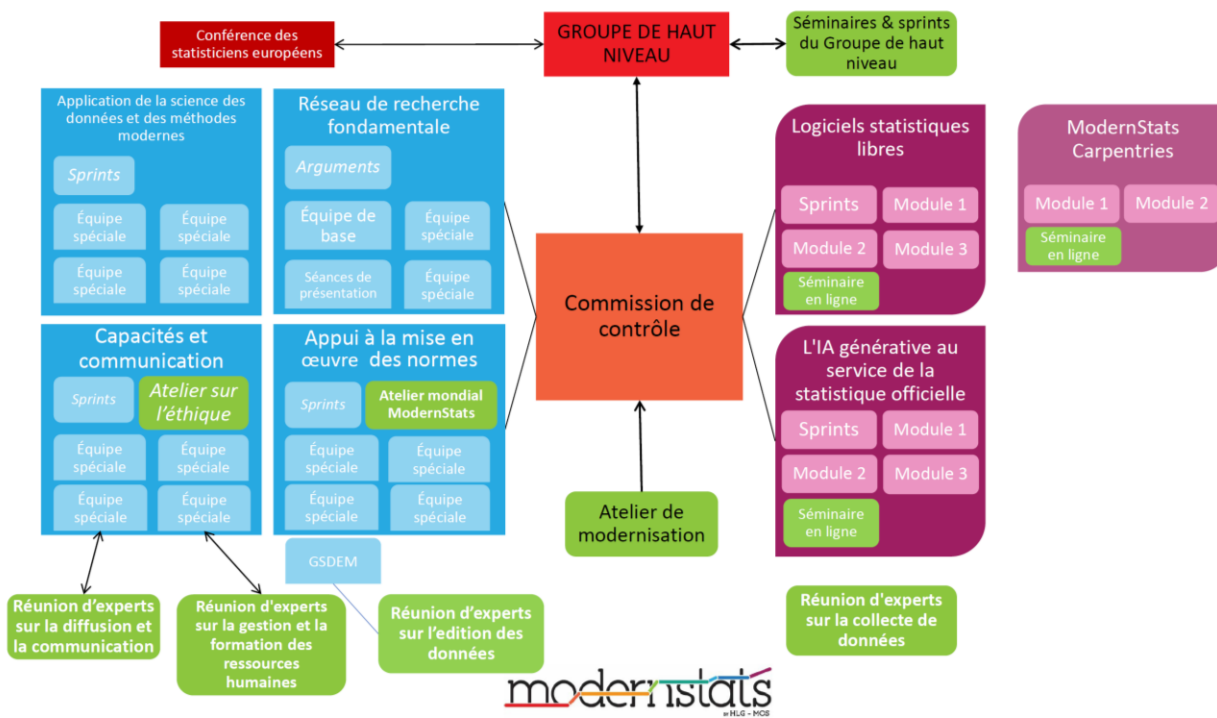
40. Les bons résultats obtenus par le Groupe de haut niveau n'ont été possibles que grâce aux efforts collectifs et à l'expertise des membres de la communauté statistique. Le personnel des organismes statistiques nationaux et internationaux ainsi que les universitaires et le secteur privé sont invités à participer à ces efforts conjoints pour stimuler l'innovation et le progrès en vue de moderniser la statistique officielle.

---

<sup>10</sup> <https://statswiki.unece.org/display/hlgbas/Modernisation+updates>.

# Annexe I

## Structure du Groupe de haut niveau en 2024



## Annexe II

## Propositions de projets pour 2024

## Descriptif d'un projet relatif aux logiciels statistiques libres

Le présent descriptif a été réalisé par Barteld Braaksma (Statistics Netherlands) ; il est soumis au Groupe de haut niveau pour approbation.

Type d'activité	
<input checked="" type="checkbox"/> Nouveau projet	<input type="checkbox"/> Extension d'un projet existant
Objectif	
<p>La communauté de la statistique officielle s'intéresse de plus en plus aux possibilités offertes par les logiciels statistiques libres en remplacement des solutions informatiques existantes (traditionnelles) basées sur des logiciels fermés (protégés). Les raisons de cet intérêt sont multiples : rentabilité, possibilités de collaboration, absence de dépendance à l'égard d'un seul prestataire, amélioration de la qualité des produits et rapidité de mise au point des nouveaux systèmes. En outre, le passage aux logiciels libres correspond au fait que les membres du personnel nouvellement recrutés ont souvent une connaissance approfondie de langages tels que R et Python, qui sont souvent utilisés pour élaborer des solutions statistiques (et relatives à la science des données) fondées sur des sources ouvertes. Cependant, ce changement peut poser des problèmes en ce qui concerne l'assistance, la maintenance, la formation, les conditions de partage, les aspects juridiques, etc. Compte tenu de la nature générique et universelle des questions soulevées et de l'importance considérable des gains qui peuvent être obtenus, une collaboration au sein de la communauté de la statistique officielle et la formulation de politiques communes s'imposent. Une série de séances de partage des connaissances et de discussion, organisées dans le cadre d'une activité du Réseau de recherche fondamentale en 2023, ont déjà montré que différents aspects de la question des logiciels libres suscitaient un intérêt certain. Parallèlement, la Conférence des statisticiens européens et le système statistique européen (SSE) ont organisé des activités autour du thème des logiciels libres.</p> <p><b>L'objectif du projet relatif aux logiciels statistiques libres est de parvenir, en s'appuyant sur des expériences concrètes liées à des cas d'utilisation présentant un large intérêt, à une meilleure compréhension commune des avantages et des inconvénients du passage à une utilisation plus poussée et plus complète des logiciels libres en tant que pierre angulaire de la production de statistiques officielles, ainsi que des choses à faire ou à ne pas faire dans ce domaine.</b></p> <p>L'expérience de l'Architecture commune de la production statistique (CSPA) a montré qu'il était très difficile de définir des cas de figure communs. Même au sein d'un même organisme, il n'est pas facile de trouver une base commune sur laquelle fonder une large utilisation des logiciels libres. Il existe toutefois des exemples concrets, ainsi que de nouvelles possibilités, à la fois dans les domaines de la saisie (collecte de données), du traitement (analyse et traitement) et de la production (diffusion). Des cas d'utilisation appropriés peuvent être trouvés dans tous ces domaines.</p> <p>Dans le domaine de la saisie, les organismes statistiques étudient la question du recours à des applications pour compléter ou remplacer les systèmes actuels de collecte de données. Un exemple est donné par les systèmes d'enquête rapide, qui permettent de collecter des données dans des domaines où des informations supplémentaires sont nécessaires rapidement (et souvent de manière urgente). Plusieurs instituts ont mis au point leurs propres systèmes d'enquête rapide. Il serait peut-être utile d'examiner s'il existe un terrain d'entente dans ce domaine. Autre exemple, le réseau ESSnet Smart Surveys a mené des recherches sur la mise au point d'applications destinées à faciliter la réalisation des enquêtes du système statistique européen (SSE) sur le budget des ménages et sur l'emploi du temps. Un modèle de collaboration prévoyant l'utilisation de sources ouvertes faciliterait considérablement le passage de la recherche à la production.</p> <p>Dans le domaine du traitement des données, de nombreux logiciels libres R ont été mis au point par plusieurs instituts nationaux de statistique et sont utilisés dans l'ensemble du SSE. Voir la liste de logiciels <a href="http://www.awesomeofficialstatistics.org">www.awesomeofficialstatistics.org</a> pour de nombreux exemples.</p>	

<p>Dans le domaine de la production, la suite .Stat<sup>11</sup> est utilisée par un grand nombre d'instituts statistiques et ce nombre ne cesse de croître. Cette suite bénéficie du soutien de la communauté SIS-CC. De nombreux utilisateurs créent leur propre code à partir de la fonctionnalité de base de .Stat, afin de mieux intégrer le logiciel dans leur propre environnement. Il pourrait être utile de partager le code, les données d'expérience et les pratiques dans ce domaine et de déterminer le rôle qu'une communauté telle que la SIS-CC pourrait jouer. Un autre exemple intéressant est celui du logiciel PxWeb (anciennement PCAxis), qui compte plusieurs utilisateurs depuis de nombreuses années, a été conçu à l'origine par Statistics Sweden et est désormais publié en tant que source ouverte avec le soutien d'un réseau ESSnet. Deux autres exemples de logiciels libres largement utilisés dans le domaine de la production sont ceux de Demetra pour la correction des variations saisonnières et d'Argus pour le contrôle de la divulgation des données statistiques.</p>
<p><b>Description du projet et des modules de travail ou des activités subsidiaires</b></p> <p>Le projet de logiciel statistique libre proposé pour 2024 pourrait comporter plusieurs modules de travail. La définition exacte peut rester plus ou moins ouverte à ce stade et se concrétiser lorsque le projet sera lancé et que les intérêts des participants seront mieux connus.</p> <p>Le <b>module de travail 0</b> est une activité préliminaire qui permet de décider de la <b>portée</b> et du <b>niveau d'ambition</b> du projet. Quels sont les aspects génériques à prendre en considération ? Quels sont les cas d'utilisation particuliers à traiter ? Une question plus spécifique est de savoir si les logiciels libres mis au point dans d'autres environnements (Pandas, PySyft, Spark, TensorFlow, ...) sont à prendre en considération.</p> <p>Le <b>module de travail 1</b> est axé sur les <i>aspects génériques</i> de l'utilisation systématique de méthodes fondées sur les logiciels libres pour établir les statistiques officielles et suit une approche descendante. Plusieurs modules secondaires peuvent être définis, couvrant des aspects tels que l'organisation de la maintenance, de l'assistance et de la formation ; les normes et les principes ; les aspects juridiques et les responsabilités ; le système de licence et la répartition équitable des coûts ; le développement communautaire, la communication et la collaboration avec, par exemple, la communauté scientifique et le secteur privé.</p> <p>Le <b>module de travail 2</b> traite des <i>cas d'utilisation</i> concrets liés aux sources ouvertes dans les domaines de la saisie, du traitement ou de la production, d'une manière ascendante. Les cas d'utilisation à étudier peuvent être traités séparément, mais il est important de définir un moyen de tirer des enseignements mutuels et d'interagir avec le module 1.</p> <p>Le <b>module de travail 3</b> traite des activités courantes du projet telles que la <i>gestion</i> et la <i>communication</i> (interne et externe), mais il a également un rôle important à jouer pour faciliter et stimuler l'interaction entre les modules 1 et 2 et leurs modules secondaires ou les cas d'utilisation qui y sont associés, afin de permettre une fertilisation croisée entre les expériences pratiques et les idées théoriques. En outre, ce module servira de lien avec d'autres activités liées aux logiciels statistiques libres, telles que le groupe <a href="#">OS4OS</a> du SSE, qui a notamment défini un ensemble de principes relatifs à l'utilisation de logiciels statistiques libres dans le SSE, et la conférence sur <a href="#">l'utilisation de R dans la statistique officielle</a>.</p>
<p><b>Produits et calendrier</b></p> <p>Les principaux résultats attendus du <b>module de travail 1</b> seront des modèles, des cadres, des listes de contrôle et des lignes directrices qui permettront de définir une approche normalisée de l'utilisation de logiciels statistiques libres. Le contenu exact du module ne pourra être décidé qu'en cours de route. Il pourrait être nécessaire d'inclure une stratégie de mise à jour des produits.</p> <p>Les principaux résultats attendus du <b>module de travail 2</b> sont des logiciels libres concrets correspondant aux cas d'utilisation envisagés, y compris (les étapes vers) la constitution d'un réseau d'utilisateurs.</p> <p>Le principal résultat externe attendu du <b>module de travail 3</b> est un plan de communication visant à diffuser les résultats du projet, qui pourra inclure de la documentation ou des lignes directrices, des cours de formation (liés par exemple aux projets Meta Academy 2022 ou Carpentries 2023) et un ou plusieurs ateliers.</p>

<sup>11</sup> Extrait du site de la SIS-CC : La suite .Stat est une plateforme en source libre, fondée sur des normes et intégrant divers composants, qui permet de produire et de diffuser efficacement des données statistiques de haute qualité. Elle repose sur le Modèle générique du processus de production statistique ([GSBPM](#)) et sur les normes d'échange de données et de métadonnées statistiques ([SDMX](#)).

<b>Organismes ou pays engagés</b>	
<p>Pour l'instant, aucun organisme ni aucun pays ne s'est formellement engagé, mais nous comptons sur un intérêt suffisant, compte tenu des activités récentes menées dans ce domaine. En 2023, le Réseau de recherche fondamentale a organisé une série de réunions en ligne sur les questions liées (au passage) aux sources ouvertes, qui ont attiré un large public. En outre, au sein du système statistique européen, un groupe qui travaille sur la question des sources ouvertes (OS4OS) a notamment élaboré des principes en la matière. Par ailleurs, les sources ouvertes ont été l'un des sujets abordés lors de la réunion plénière de juin 2023 de la Conférence des statisticiens européens. L'utilisation de systèmes de type ChatGPT pour transformer un code protégé en code ouvert (SAS-to-R) suscite beaucoup d'intérêt. Il est important de choisir une approche qui ne laisse personne de côté : les pays sont à des niveaux de maturité différents et les ressources peuvent varier en fonction, par exemple, de la taille des pays.</p>	
<b>Autres voies envisagées</b>	
<p>Une autre façon de procéder consisterait à poursuivre les activités dans le cadre du Réseau de recherche fondamentale ou à attendre d'autres initiatives du système statistique européen, mais 1) ce sujet suscite beaucoup d'intérêt au-delà de l'Union européenne et 2) cela signifie que la visibilité que confère un projet formel ne serait pas assurée et que nous aurions manqué l'occasion de définir des lignes directrices internationales et des perspectives communes largement partagées et approuvées.</p>	
<b>Quel est le lien avec la vision et les autres activités du Groupe de haut niveau ?</b>	
<p>Une utilisation plus poussée et plus complète des logiciels (statistiques) libres contribue directement à la modernisation de la production de statistiques officielles et correspond donc directement à la vision du Groupe de haut niveau. Il existe des synergies potentielles avec des projets passés ou présents, ainsi qu'avec les comités de modernisation et les réunions thématiques existants.</p>	
<b>Dates de début et de fin proposées</b>	
<b>Début</b> : janvier 2024	<b>Fin</b> : novembre 2024
<p>Un plan de travail plus détaillé devra être élaboré par le chef de projet et l'équipe de projet, sous la direction du Groupe de haut niveau et de la Commission de contrôle.</p>	

## Descriptif d'un projet relatif à l'IA générative au service de la statistique officielle

Le présent descriptif a été réalisé par le Groupe de l'application de la science des données et des méthodes modernes, le Réseau de recherche fondamentale et la CEE ; il est soumis au Groupe de haut niveau pour approbation.

Type d'activité	
<input checked="" type="checkbox"/> Nouveau projet	<input type="checkbox"/> Extension d'un projet existant
Objectif	
<p>Au cours des dernières années, les organismes statistiques ont utilisé l'apprentissage automatique pour améliorer l'efficacité des tâches à forte intensité de ressources, telles que la classification et le codage, l'édition et l'analyse d'images. L'apprentissage automatique a également permis d'élaborer de nouveaux produits statistiques, comme les indicateurs d'opinion dérivés des messages publiés sur les médias sociaux. Certains organismes statistiques ont commencé à étudier la possibilité de recourir à l'apprentissage automatique dans d'autres domaines du GSBPM, mais il est encore trop tôt pour comprendre tout le potentiel de cette technologie.</p> <p>Bien que les utilisations de l'apprentissage automatique soient principalement axées sur l'amélioration de l'efficacité de la prise de décision (par exemple, la classification des descriptions textuelles des emplois sous forme de codes de profession ou le repérage des valeurs aberrantes), l'introduction de ChatGPT à la fin de l'année 2022 a marqué un changement important dans ce domaine et a augmenté le nombre de cas potentiels d'utilisation de l'IA dans la statistique officielle.</p> <p>Les grands modèles de langage (GML), sur lesquels est fondé ChatGPT, ont des capacités uniques qui les distinguent des applications traditionnelles de l'apprentissage automatique, comme la capacité de générer des textes et de tenir des conversations semblables à celles d'un être humain. Les applications traditionnelles de l'apprentissage automatique, quant à elles, sont principalement axées sur l'assistance aux humains dans les tâches de prévision plutôt que sur la création de contenu. En outre, l'introduction de services conviviaux fondés sur les grands modèles de langage (par exemple, ChatGPT proposé par OpenAI et Bard par Google) a permis à l'IA générative de s'adresser non seulement à la communauté des chercheurs, mais aussi au grand public. Cette évolution a des conséquences importantes pour les organismes statistiques, notamment sur l'organisation de leur travail et sur les méthodes de diffusion de leurs produits.</p> <p>En réponse à cette évolution récente, les groupes de modernisation du Groupe de haut niveau (Réseau de recherche fondamentale et Groupe de l'application de la science des données et des méthodes modernes) ont lancé conjointement vers le milieu de l'année 2023 un livre blanc sur l'utilisation des GML dans la statistique officielle. Dans ce rapport, actuellement en phase de finalisation, des questions essentielles sont abordées, comme les conséquences pour les statistiques officielles, les cas d'utilisation au sein des organismes statistiques et les risques associés, et des recommandations ainsi que des considérations stratégiques sont formulées. Parallèlement à la rédaction du livre blanc, une série de sessions de démonstration ont été organisées pour présenter des exemples concrets d'application des GML et de ChatGPT (par exemple, le passage de SAS à R, l'utilisation des GML pour la diffusion ou l'utilisation de ChatGPT pour mettre à jour le système de classification).</p> <p>S'appuyant sur les travaux actuels décrits dans le livre blanc sur les GML, le projet relatif à l'IA générative au service de la statistique officielle vise à approfondir l'examen des questions qui se posent quand les organismes statistiques veulent utiliser l'IA de manière efficace et responsable (par exemple, la gouvernance ou les modèles ouverts), ainsi qu'à concevoir en commun des solutions concrètes.</p>	



Description du projet et des modules de travail ou des activités subsidiaires
<p>Étant donné l'évolution rapide qui caractérise ce domaine, il est important de faire preuve d'agilité. On trouvera ci-dessous quelques exemples d'activités considérées comme particulièrement importantes à l'heure actuelle. Ces activités s'appuient généralement sur les travaux menés par l'équipe de rédaction du livre blanc du Groupe de haut niveau. Toutefois, au stade actuel, le choix des activités reste plus ou moins ouvert ; il devra être arrêté si le projet est lancé.</p> <p><b>Activité 0. Cadrage et réorientation si nécessaire.</b></p> <p><b>Activité 1. Partage des cas d'utilisation :</b> des réunions seront régulièrement organisées en ligne afin de mettre en commun des cas concrets d'utilisation et des idées se rapportant non seulement à l'utilisation des GML sur la base de textes, mais aussi à d'autres types d'utilisation de l'IA générative multimodale (par exemple, la création d'images). La participation n'est pas limitée aux organismes statistiques ; elle inclut les administrations publiques et les instituts de recherche ou les établissements universitaires, à condition qu'ils soient intéressés par la statistique officielle.</p> <p><b>Activité 2. Élaboration concertée de solutions applicables aux cas d'utilisation les plus courants :</b> le projet permettra de recenser les cas d'utilisation qui présentent un intérêt commun pour les organismes statistiques (par exemple, les dialogueurs) et d'élaborer des solutions fondées sur des sources ouvertes.</p> <p><b>Activité 3. Compilation de pratiques et de recommandations concrètes :</b> sur la base des activités 1 et 2 ainsi que du livre blanc du Groupe de haut niveau, le projet consistera à élaborer des recommandations concrètes relatives à l'utilisation de l'IA générative dans la statistique officielle. Il est essentiel de se concentrer sur quelques thèmes qui intéressent particulièrement les organismes statistiques, plutôt que de chercher à élaborer un guide général.</p>
Produits et calendrier
<p><b>Activité 0 :</b> document actualisé sur la proposition de projet (janvier 2024)</p> <p><b>Activité 1 :</b> compilation d'exposés et de diapositives dans un répertoire GitHub partagé des services d'IA générative mis au point par les organismes statistiques, à mettre à jour en permanence, le cas échéant, pendant toute la durée du projet (de janvier à décembre 2024)</p> <p><b>Activité 2 :</b> solutions ouvertes fondées sur l'IA générative (janv.-déc. 2024)</p> <p><b>Activité 3 :</b> recommandations concrètes relatives à l'utilisation de l'IA générative dans la statistique officielle (janv.-déc. 2024)</p>
Organismes ou pays engagés
<p>Pour l'instant, aucun organisme ni aucun pays ne s'est formellement engagé, mais nous comptons sur un intérêt suffisant, compte tenu des activités récentes menées dans ce domaine.</p>
Autres voies envisagées
<p>Une autre solution serait de ne rien faire. D'autres initiatives internationales pourraient voir le jour sur le sujet, mais nous manquerions une occasion en ne tirant pas parti des travaux existants et de notre réseau d'experts pour répondre à une demande qui croît rapidement.</p>

<b>Quel est le lien avec la vision et les autres activités du Groupe de haut niveau ?</b>	
<p>L'IA générative est une nouvelle technologie qui pourrait changer radicalement la façon dont les organismes statistiques travaillent, aussi le projet correspond-il tout à fait à la vision du Groupe de haut niveau. En outre, le projet est étroitement lié à d'autres activités du Groupe de haut niveau, telles que le projet mené par ce dernier sur l'apprentissage automatique (2019-2020), la collaboration avec l'Office for National Statistics (ONS) du Royaume-Uni de Grande-Bretagne et d'Irlande du Nord dans le cadre du Groupe de l'apprentissage automatique (2021-2022) et l'atelier sur l'apprentissage automatique (2023).</p>	
<b>Dates de début et de fin proposées</b>	
<b>Début</b> : janvier 2024	<b>Fin</b> : décembre 2024

## Annexe III

### Résumé de la discussion qui s'est tenue en groupes restreints sur les propositions de projets pour 2024 lors de l'atelier sur la modernisation

1. Au cours de la session de discussion en groupes restreints, les participants ont donné leur avis sur les propositions de projets (relatifs aux logiciels statistiques libres et à l'IA générative au service de la statistique officielle). Les observations suivantes ont été formulées pour les deux projets :

- Il faudrait consacrer du temps, au début de chaque projet, à la création d'un répertoire des outils, des cas d'utilisation et de la documentation dont on dispose actuellement. Ce répertoire pourra être enrichi tout au long des projets ;
- Les problèmes liés au passage de l'expérimentation ou de l'idéation à la production doivent être pris en considération dès le début des projets. Les enseignements tirés d'autres projets tels que le projet sur l'apprentissage automatique pourraient s'avérer utiles. Il serait intéressant de disposer de lignes directrices ou de manuels à cet égard ;
- Il s'agit dans les deux cas de projets techniques et il est donc important que divers talents participent à ces projets (par exemple au niveau de la direction et, de manière horizontale, en lien avec les entreprises et les autres branches de l'organisme) ;
- Les projets devraient d'abord être axés sur les fondements nécessaires dans ces domaines, en même temps que sur l'élaboration et l'intégration de solutions pertinentes à petite échelle afin de démontrer l'utilité et les insuffisances de ces fondements.

2. En ce qui concerne le projet relatif aux logiciels statistiques libres :

- Il est important de faire la distinction entre l'utilisation et la production de logiciels libres. Si certains aspects sont communs, d'autres sont plus spécifiques à une perspective (par exemple, du point de vue du producteur, les questions relatives à la maintenance, à la gouvernance et à l'assistance sont particulièrement importantes). Le projet pourrait s'intéresser d'abord à une perspective, puis à l'autre ;
- Quelle que soit la perspective envisagée, de nombreux outils de formation et de collaboration sont nécessaires ;
- Il est important de prendre en compte la culture et le changement d'état d'esprit pour que les différentes branches de l'organisme sortent de leur zone de confort et considèrent les logiciels libres comme une option dès le départ ;
- La conception des invites pourrait être un cas d'utilisation visé par le projet relatif à l'IA ;
- Analyse du coût réel de la mise en œuvre de logiciels libres ;
- Il est important de s'intéresser aux autres solutions fondées sur des sources ouvertes qui sont mises en œuvre dans un autre contexte que celui de la statistique officielle, afin d'adapter ces solutions et de collaborer ;
- Différents niveaux de partage (par exemple, public, semi-public ou entre organismes statistiques) ;
- Le projet doit tenir compte des enseignements tirés des expériences précédentes (par exemple, la CSPA).

3. En ce qui concerne le projet relatif à l'IA générative au service de la statistique officielle :

- Il serait utile de disposer d'un répertoire des cas d'utilisation de l'IA générative. Trois à cinq cas d'utilisation pourraient être sélectionnés à partir de ce répertoire et servir de base à une collaboration en vue d'élaborer une solution concrète ;

- Le champ d'application des cas d'utilisation ne se limite pas à la partie production (GSBPM) mais s'étend au-delà (GAMSO) à des domaines tels que les ressources humaines et les finances ;
- Grâce à des expériences communes, le projet peut aider à déterminer ce qui peut être fait par d'autres (par exemple, le secteur privé) et ce que nous devons faire nous-mêmes ;
- Domaines pouvant faire l'objet d'une collaboration : conception des invites (par exemple, à l'aide de Carpentries) et copilotage ;
- Il serait important, tout en recherchant rapidement les domaines d'application dans lesquels l'IA peut présenter des avantages, d'envisager aussi « ce qu'il faut faire » et « ce qu'il ne faut pas faire » avec l'IA ;
- La confidentialité et la sécurité sont des préoccupations majeures pour les organismes statistiques. L'ouverture des algorithmes et des données de formation est également un aspect important pour les organismes publics ;
- Un processus commun d'assurance de la qualité (par exemple, comment évaluer et valider la solution élaborée à l'aide de l'IA) est nécessaire ;
- Collaborer avec les entreprises technologiques afin qu'elles puissent appliquer des pondérations différentes aux statistiques officielles lors de l'apprentissage des modèles ;
- Élaborer un manuel sur la mise au point des prochains modèles et la mise en production des applications (par exemple, prise en compte des biais et de l'éthique) ;
- Il existe des liens étroits avec d'autres initiatives (par exemple, le guichet unique ESSnet) et d'autres groupes de modernisation (par exemple, le module qui, au sein du Groupe de l'application de la science des données et des méthodes modernes, travaille sur la question de l'IA responsable). La coordination sera importante ;
- Il est essentiel d'expliquer aux utilisateurs comment nous utilisons l'IA générative ;
- Compte tenu de l'évolution rapide du domaine, il serait utile d'informer la communauté de l'état d'avancement du projet plus fréquemment que d'habitude.

---