



---

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

Genève, 20 et 21 juin 2024

Point 10 c) de l'ordre du jour provisoire

**Rapports, directives et recommandations élaborés  
sous les auspices de la Conférence : nouvelle version  
du modèle générique d'informations statistiques****Nouvelle version du modèle générique d'informations  
statistiques****Document établi par le secrétariat***Résumé*

Le modèle générique d'informations statistiques (GSIM) décrit et définit un ensemble d'informations utilisées dans le processus de production statistique. En normalisant ces informations, le GSIM peut faciliter la réutilisation et le partage des méthodes, des outils et des processus au sein des organismes statistiques et entre eux.

La présente note donne un bref aperçu du GSIM, du processus d'examen mené dans le cadre du Groupe d'appui à la mise en œuvre des normes, créé par le Groupe de haut niveau sur la modernisation de la statistique officielle (ci-après, le Groupe de haut niveau), et des changements apportés dans la nouvelle version du modèle. La version précédente a été approuvée par la Conférence des statisticiens européens (ci-après, la Conférence) en 2020.

La Conférence est invitée à approuver la nouvelle version du GSIM, qui est présentée dans le document INF.2 sous une forme imprimable et sur [ce site GitHub](#).



## I. Contexte

1. Dans le monde entier, les organismes statistiques entreprennent des activités similaires, même si les processus qu'ils utilisent varient. Chacune de ces activités donne lieu à une utilisation et une production d'informations similaires (par exemple, tous les organismes définissent des populations aux fins de leurs observations statistiques, utilisent des classifications statistiques, créent des ensembles de données et diffusent des informations). Bien que les informations utilisées par les organismes statistiques soient fondamentalement les mêmes, chaque organisme a tendance à les décrire de manière légèrement différente des autres et parfois de différentes manières en son propre sein.
2. Le modèle générique d'informations statistiques (GSIM) est un cadre de référence pour les informations statistiques, mis au point pour la première fois en 2012. Il définit un ensemble de classes d'information normalisées et décrites de façon uniforme, qui peuvent être utilisées comme intrants et produits dans la conception et la production de statistiques. En tant que cadre de référence, le GSIM explique les liens importants qui existent entre les entités prenant part à la production de statistiques ; il peut d'ailleurs servir à encadrer l'élaboration et l'utilisation de normes ou de spécifications cohérentes à des fins de mise en œuvre.
3. En tant que langage commun pour décrire les informations statistiques, le GSIM peut faciliter la communication au sein des organismes statistiques et entre eux. Il peut servir de base à une collaboration approfondie, à la normalisation ou au partage d'outils et de méthodes et joue ainsi un rôle important dans la modernisation, la rationalisation et l'harmonisation des normes et de la production associées aux statistiques officielles, tant au niveau national qu'international. Le GSIM est l'une des pierres angulaires de la modernisation de la statistique officielle et de l'élimination des cloisonnements entre les différents domaines. Il s'agit d'un élément essentiel de la vision stratégique du Groupe de haut niveau sur la modernisation de la statistique officielle (ci-après, le Groupe de haut niveau)<sup>1</sup>.
4. Une modernisation de la production statistique est nécessaire pour que les organismes statistiques restent pertinents et flexibles dans un environnement d'information dynamique et compétitif. Nous espérons que les organismes statistiques adopteront et mettront en œuvre le GSIM et le langage commun qu'il représente.
5. La section II du présent document donne une vue d'ensemble du GSIM, tandis que la section III décrit les étapes de son développement. La section IV présente ensuite les principaux changements apportés dans la nouvelle version du modèle et la section V donne un bref aperçu du processus de révision. Enfin, la section VI met en évidence les avantages que présente la mise en œuvre du GSIM pour un organisme.

## II. Vue d'ensemble du modèle générique d'informations statistiques

### A. Champ d'application

6. Le GSIM s'applique aux processus de production statistique décrits dans le modèle générique du processus de production statistique (GSBPM)<sup>2</sup>, en donnant aux classes d'information des noms adoptés d'un commun accord, en les définissant, en précisant leurs propriétés essentielles et en indiquant les relations qu'elles entretiennent entre elles. Il n'émet pas toutefois de postulats, quant aux normes ou aux technologies utilisées pour son application.

---

<sup>1</sup> Wiki des statistiques de la CEE – Groupe de haut niveau (<https://statswiki.unece.org/display/hlgbas>).

<sup>2</sup> Wiki des statistiques de la CEE – GSBPM (<https://statswiki.unece.org/display/GSBPM>).

7. Le GSIM ne recouvre pas les classes d'information en rapport avec les activités d'un organisme comme les ressources humaines, les finances ou les fonctions juridiques, sauf si ces informations sont directement utilisées pour la production statistique. Pour plus de renseignements sur ces activités, voir le modèle générique d'activité des organismes statistiques (GAMSO)<sup>3</sup>.

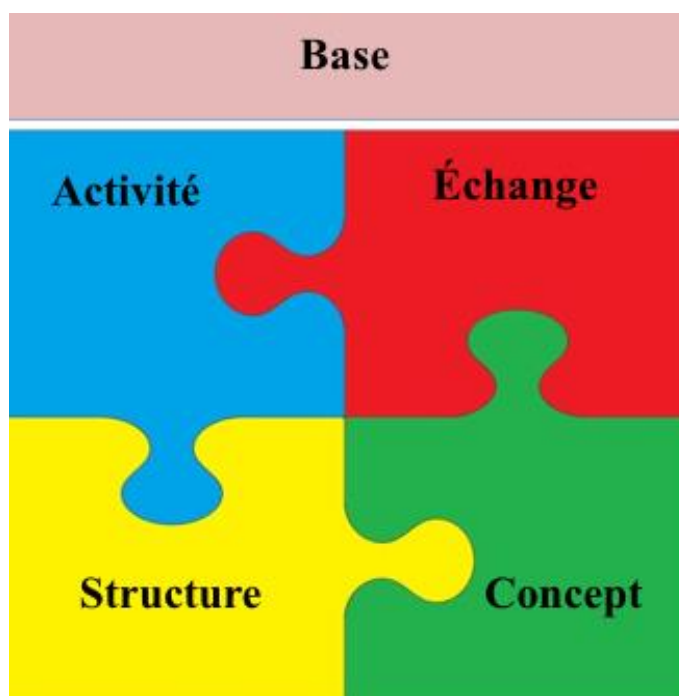
8. Le GSIM est un modèle théorique. Les organismes peuvent choisir des « normes de mise en œuvre » existantes, telles que Statistical Data and Metadata eXchange (SDMX) et Data Documentation Initiative (DDI) pour la mise en œuvre technique de normes dans les processus de production. Le GSIM fournit plutôt un point de référence commun sur le plan théorique, qui permet de comparer et de contextualiser les normes de mise en œuvre.

## B. Principaux éléments

9. Le modèle GSIM comporte des classes spécifiques qui définissent les informations sur le monde réel et peuvent être décrites comme des « classes d'information ». Il comprend près de 130 classes d'information qui sont les éléments d'information nécessaires à un processus de production statistique (tels que « variable théorique », « classification statistique » ou « intrant du processus »). Dans les versions précédentes du GSIM, ces éléments étaient appelés « objets d'information ».

10. Pour plus de simplicité, le GSIM regroupe ces différentes classes en cinq grands types en fonction de leur rôle dans la production statistique, comme le montre la figure suivante :

### Groupes de classes d'information de haut niveau du GSIM



11. Les cinq groupes de classes du GSIM peuvent être résumés comme suit :

- a) Le groupe de base fournit des caractéristiques réutilisables par les autres classes d'information à l'appui de fonctionnalités telles que l'identité et les variantes ;
- b) Le groupe des activités sert à prendre en compte la conception et l'exécution des programmes statistiques et des activités connexes ;
- c) Le groupe des concepts sert à définir la sémantique et la représentation des données ;

<sup>3</sup> Wiki des statistiques de la CEE – GAMSO (<https://statswiki.unece.org/display/GAMSO>).

d) Le groupe des échanges sert à modéliser le partage des données et des métadonnées, tant sur le plan interne qu'externe ;

e) Le groupe des structures sert à structurer les informations tout au long du processus statistique.

12. Chaque classe d'information comporte une définition, un texte explicatif et une liste d'attributs. Ces définitions, ainsi que les relations entre les classes, indiquent comment les organismes statistiques utilisent les informations au cours du processus de production.

### III. Étapes du développement du modèle générique d'informations statistiques

13. Le modèle a été mis au point pour la première fois en 2012, puis révisé à plusieurs reprises pour tenir compte des expériences des utilisateurs et de l'évolution du contexte économique. Il a été approuvé par la Conférence en 2017, en même temps que le GSBPM, le GAMS0 et l'Architecture commune de la production statistique. La version précédente du modèle (version 1.2) a été publiée en 2019 et approuvée par la Conférence en 2020.

14. En 2020, le groupe de modernisation du Groupe de haut niveau chargé de l'appui à la mise en œuvre des normes (ci-après dénommé Groupe d'appui à la mise en œuvre des normes) a mis sur pied une équipe spéciale chargée d'effectuer des mises à jour mineures supplémentaires du GSIM. Ces mises à jour ont consisté principalement à ajouter des classes d'information ou à modifier le nom d'une classe par rapport à la version précédente, à reformuler des définitions sans en modifier le sens principal ou à corriger des fautes de frappe<sup>4</sup>. Toutefois, les travaux menés par d'autres équipes spéciales du Groupe d'appui à la mise en œuvre des normes, comme l'équipe chargée de la mise en relation du GSBPM et du GSIM, ont donné lieu à des discussions sur la nécessité de faire davantage que des mises à jour mineures du GSIM pour en améliorer la cohérence et la convivialité, en particulier pour les utilisateurs du GSBPM. Par conséquent, le Groupe d'appui à la mise en œuvre des normes a demandé au Groupe de haut niveau d'élargir le mandat de l'équipe spéciale chargée de la révision du GSIM de manière à passer d'une mise à jour mineure à l'élaboration d'une nouvelle version complète. Le processus formel a donc eu lieu en 2022 et 2023.

15. Cette équipe spéciale était présidée par le Canada et se composait des membres suivants : Égypte, États-Unis d'Amérique, Finlande, France, Hongrie, Italie, Mexique, Norvège, Pays-Bas, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord et Commission économique pour l'Europe (CEE). Le Groupe de l'appui à la mise en œuvre des normes a supervisé les travaux de l'équipe spéciale afin de garantir que la nouvelle version soit conforme aux autres modèles du Groupe de haut niveau, comme le GSBPM et le GAMS0.

16. Conformément à l'objectif qui consiste à utiliser des méthodes modernes pour encourager la collaboration dans le domaine de la statistique officielle, la nouvelle version du GSIM (version 2.0) a été publiée dans un [répertoire GitHub](#) spécifique, ce qui est devenu ces dernières années un moyen très populaire de publier des ressources. L'hébergement du modèle sur une plateforme ouverte telle que GitHub devrait favoriser son adoption et permettre aux utilisateurs de proposer des améliorations qui pourraient être prises en compte dans les versions à venir.

### IV. Principaux changements depuis la version 1.2

17. Par rapport à la version précédente du GSIM, soit la version 1.2, les principaux changements sont les suivants :

a) Six classes d'information ont été ajoutées (par exemple, « Registre », « Document de référence », « Agent intelligent ») et huit supprimées (par exemple, « Registre administratif », « Registre statistique », « Point de données dimensionnelles ») et

<sup>4</sup> Directives relatives à la gouvernance des modèles ModernStats. <https://statswiki.unece.org/display/hlgbas/HLG-MOS+Outputs?preview=/120128748/338329756/ModernStats%20Model%20Governance.pdf>.

« Point de données unitaires ») afin de simplifier le modèle et de tenir compte de l'évolution des méthodes de travail des organismes statistiques ;

b) Cinq classes d'information ont été renommées (par exemple, « Intrans transformable » en « Intrans de base » et « Circuit d'échange » en « Instrument d'échange ») afin de les rendre applicables aux différents types de données (par exemple, les intrants non transformables tels que les registres) dans la production statistique et d'en préciser le sens ;

c) Les relations entre les classes d'information ont été actualisées, les principales mises à jour concernant les classes liées aux métadonnées référentielles fondées sur des cas d'utilisation réels tels que les rapports sur la qualité. L'accent a également été mis sur les classes d'information relatives à la conception des processus de production et des instruments d'échange, afin de mieux séparer ce qui relève de la conception de ce qui relève de la mise en œuvre.

18. Pour plus de détails sur les changements effectués, voir la [page de résumé](#) et le [journal des modifications](#) dans le répertoire GitHub.

## V. Processus de révision

19. La révision du GSIM a donné lieu à deux cycles de consultations : en 2022 et en 2023.

20. Le premier cycle a été réalisé par l'équipe spéciale chargée de la révision du GSIM, sur la base des observations des utilisateurs du modèle (recueillies en mars 2022) et des autres équipes spéciales du Groupe de l'appui à la mise en œuvre des normes, notamment l'équipe chargée de la mise en relation du GSBPM et du GSIM (2021).

21. La nouvelle version du modèle a été envoyée pour examen à tous les pays membres de la Conférence en juin et juillet 2023. Des observations ont été reçues du Bélarus, du Canada, de la Croatie, de l'Équateur, de la Finlande, de Malte, du Mexique et de la Norvège, soit par courriel, soit par l'intermédiaire d'un [répertoire des consultations GitHub](#), et ont été analysées par le secrétariat en concertation avec les dirigeants de l'équipe spéciale, ce qui a conduit à apporter quelques améliorations supplémentaires au modèle et a permis de publier celui-ci sur GitHub en décembre 2023.

22. Comme indiqué précédemment, les modifications apportées ont été publiées dans le répertoire GitHub sous la forme de définitions, de textes explicatifs et d'attributs (qui sont également décrits dans l'additif au présent document). Ce répertoire contient également des diagrammes interactifs qui illustrent les relations entre les classes d'information, que les utilisateurs peuvent explorer en cliquant sur une classe donnée pour visualiser les relations de cette classe avec ses plus proches voisins dans le modèle.

23. Le tableau ci-après résume les principales étapes processus de révision du GSIM.

Mars 2022	Premier retour d'informations (complété par les renseignements reçus précédemment sur les travaux du groupe chargé des supports d'information)
Juin 2023	Consultation à l'échelle de la Conférence sur une proposition de nouvelle version du modèle
Décembre 2023	Publication de GSIM 2.0 sur <a href="#">GitHub</a>

24. Outre la nouvelle version du modèle, l'équipe spéciale a également travaillé sur un guide de l'utilisateur du GSIM basé sur la spécification du GSIM (2013) et sur la mise à jour de la communication du GSIM (2020). Ces documents seront également disponibles dans le répertoire GitHub.

## VI. Avantages de la mise en œuvre du modèle générique d'informations statistiques

25. Différents degrés d'utilisation du GSIM sont envisagés. Dans certains cas, les organismes n'utiliseront le modèle qu'à titre de référence pour éclairer le débat, que ce soit en interne ou avec d'autres organismes. Dans d'autres cas, ils pourront choisir de mettre en œuvre le GSIM en tant que modèle d'information définissant leur environnement opérationnel. Différents scénarios d'utilisation du GSIM sont valables, bien que les organismes qui en font un usage complet puissent s'attendre à en tirer les plus grands avantages.

26. L'utilisation du GSIM présente un avantage important et immédiat : le modèle permet, en tant que langage commun, d'améliorer la communication à différents niveaux :

- Entre les différents acteurs du processus de production statistique (économistes et informaticiens) ;
- Entre les différents domaines statistiques ;
- Entre les organismes statistiques aux niveaux national et international.

27. L'amélioration de la communication peut entraîner un gain d'efficacité dans l'échange de données et de métadonnées au sein des organismes statistiques et entre eux, ainsi qu'avec les fournisseurs et les consommateurs d'informations extérieurs.

28. À plus long terme, la mise en œuvre du GSIM, en combinaison avec le GSBPM, pourrait apporter les avantages suivants :

- Créer un environnement propice à la réutilisation et au partage des méthodes, des éléments de programme et des processus ;
- Offrir la possibilité d'automatiser les processus et de réduire ainsi au minimum l'intervention humaine dans la production statistique ;
- Faciliter la réalisation d'économies d'échelle grâce à la mise au point d'outils communs par les organismes statistiques.

29. Le GSIM pourrait servir à orienter les investissements futurs vers les domaines de la production statistique où les besoins communs sont les plus importants. Il pourrait également permettre un certain degré de spécialisation au sein de la communauté statistique internationale, en permettant à certains organismes de se spécialiser dans des domaines particuliers et de partager leurs outils et leur expertise afin d'éviter les chevauchements avec les activités menées par d'autres organismes.

30. En définissant des classes d'information normalisées communes à toute la production statistique, quel que soit le sujet, le GSIM permet aux organismes statistiques de repenser l'organisation de leurs activités en vue de la rendre plus efficace. En proposant un contexte commun et des éléments de liaison, le GSIM et le GSBPM peuvent aider à décider quand et où utiliser des normes de mise en œuvre telles que SDMX et DDI afin de parvenir à une automatisation et une interopérabilité maximales. Ces considérations sont importantes pour tous ceux qui souhaitent construire des pipelines de données.

31. Il convient également de noter que le GSIM et d'autres modèles théoriques sont applicables au domaine nouveau de l'intelligence artificielle (IA). Bien que les progrès récents accomplis dans ce domaine offrent aux organismes statistiques la possibilité de rendre leur activité plus efficace, l'IA ne peut, à elle seule, transformer le processus de production statistique.

32. Dans le contexte de l'IA, les normes peuvent jouer 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 exploitables automatiquement), ce qui permet d'améliorer l'efficacité et l'automatisation des opérations.

33. Les normes telles que le GSIM constituent donc une base sur laquelle la production assistée par l'IA peut réaliser pleinement son potentiel. Étant donné que les normes améliorent généralement la qualité des résultats, les modèles issus de l'IA devraient en bénéficier, notamment sous forme d'une plus grande précision des résultats, d'une réduction de la part de conjecture nécessaire à l'IA pour interpréter les données ou d'une plus grande transparence des résultats obtenus.

---