Module d’assurance qualité lors de l’utilisation des données administratives et autres sources de données pour produire les statistiques officielles

Préparé par le Sous-groupe sur les données administratives et autres sources des données du Groupe des experts des Nations Unies sur les cadres nationaux d’assurance de la qualité (GE-CNQA)[[1]](#footnote-2)

Version du 23 avril 2024

Introduction du module

1. Ce module distingue trois sources de données en fonction de leur objectif et de l’entité responsable de leur compilation : les sources de données statistiques telles que les enquêtes, les sources de données administratives et d’autres sources de données. En général, les autres sources de données comprennent les sources de données associées au terme « big data», à moins qu’elles ne soient déjà incluses, dans certains cas, dans des sources de données statistiques ou administratives. Les nouvelles sources de données peuvent souvent être associées à d’autres sources de données. Toutefois, elles peuvent également être considérées comme faisant partie de sources de données statistiques ou administratives en fonction des circonstances nationales.[[2]](#footnote-3)
2. L’utilisation de **sources de données administratives** offre de nombreux avantages potentiels aux organismes statistiques, tels que le rapport coût-efficacité, la réduction du fardeau des répondants, l’amélioration des délais, de l’actualité, de la pertinence, de l’exactitude et de la fiabilité, en raison de leur capacité à obtenir des données fortement désagrégées. L’utilisation des **autres sources de données**, telles que celles disponibles dans le secteur privé, peut être rentable, en plus de permettre la production des rapports beaucoup plus fréquents et opportuns et d’illustrer des phénomènes difficiles ou impossibles à saisir avec les sources de données statistiques et administratives traditionnelles. Cela pourrait conduire à une meilleure pertinence.
3. Ce module complète les cadres nationaux d’assurance de la qualité (CNAQ) génériques existants et fournit des recommandations plus précises et détaillées lors de l’utilisation des données administratives et d’autres sources de données pour produire les statistiques officielles. Ce module ne peut être utilisé comme cadre d’assurance de la qualité pour les données administratives et autres sources de données. Il est destiné à être utilisé en conjonction avec un CNAQ existant, puisqu’il n’intègre pas, par exemple, des aspects clés d’un CNAQ tels que la coordination et l’environnement institutionnel.[[3]](#footnote-4)
4. Ce module s’adresse aux organismes statistiques[[4]](#footnote-5) qui utilisent ou souhaitent utiliser les données administratives et autres sources de données pour produire les statistiques officielles. Il repose sur un examen des pratiques nationales respectives, des lignes directrices, des cadres et des outils disponibles[[5]](#footnote-6) et sur la mise en cohérences de leurs éléments avec le Cadre national d’assurance de la qualité pour les statistiques officielles des Nations Unies (CNAQ des Nations Unies). Le module se compose de deux parties à savoir : une approche conceptuelle et une liste de contrôle de dix exigences essentielles. Il comprend également plusieurs annexes contenant des informations complémentaires.
5. L’approche conceptuelle de **la partie 1** fournit une structure globale des considérations de qualité typiques lors de l’utilisation des données administratives et autres sources de données. Elle porte sur les conditions préalables, ainsi que les actions et les considérations générales relatives à la qualité. Ces actions et considérations générales sont parfois qualifiées d’«hyper-dimensions» dans diverses pratiques nationales et lignes directrices disponibles qui ont servi de base à ce module. La **partie 2** fournit une liste de contrôle de dix exigences essentielles pour garantir la qualité ainsi qu’un ensemble de pratiques suggérées («meilleures») qui fournissent des orientations plus précises et plus détaillées lors de l’utilisation des données administratives et autres sources de données que celles contenues dans un CNAQ générique. Bon nombre de ces exigences essentielles se chevauchent avec des parties du CNAQ des Nations Unies et d’autres CNAQ. Les pratiques suggérées dans ce module sont ambitieuses. Elles peuvent ne pas s’appliquer en toutes circonstances et peuvent être difficiles ou pas toujours possibles à suivre pour les organismes de statistique en fonction de leur contexte.[[6]](#footnote-7) Toutefois, lorsqu’elles sont applicables, elles permettent d’identifier les domaines d’amélioration dans l’esprit du cycle «Planifier-Faire-Vérifier-Agir» pour l’amélioration continue rendu populaire par W. Edwards Deming.
6. **L’annexe 1** fournit un sous-module pour la validation des données sources et une liste d’indicateurs de qualité supplémentaires. **L’annexe 2** fournit un glossaire des termes utilisés dans le présent document. **L’annexe 3** contient une liste des exigences du CNAQ des Nations Unies qui sont pertinentes pour garantir la qualité des statistiques officielles lors de l’utilisation données administratives et autres sources de données. **L’annexe 4** présente une cartographie des dix exigences essentielles avec les exigences pertinentes du CNAQ des Nations Unies énumérées à l’annexe 3.
7. Le présent module synthétise les pratiques nationales existantes, les lignes directrices, les cadres et les outils disponibles pour assurer la qualité des statistiques lors de l’utilisation de données administratives et autres sources de données. Il se veut pratique et concis. Les pays qui ont déjà des pratiques bien établies en matière d’assurance qualité lors de l’utilisation des données administratives et autres sources de données peuvent considérer ce module comme un point de référence supplémentaire qui soutient ce qu’ils font déjà. Le module a été élaboré en référence au CNAQ des Nations Unies, mais il est censé être applicable indépendamment du CNAQ ou du code de bonne pratique pour l’assurance qualité suivi par un organisme statistique.
8. Ce module a été élaboré par le sous-groupe sur les données administratives et autres sources de données (ci-après dénommé «sous-groupe»). Ce sous-groupe a été créé par le Groupe des experts des Nations Unies sur les cadres nationaux d’assurance de la qualité (GE-CNAQ) afin de fournir des orientations pratiques et concises ainsi que des bonnes pratiques aux organismes statistiques pour assurer la qualité des statistiques officielles lorsque les données administratives et autres sources de données sont utilisées pour la production de statistiques officielles.[[7]](#footnote-8) Le Groupe des experts et son Sous-Groupe sont composés d’experts en assurance qualité des États Membres et des organisations internationales et régionales. Le Sous-Groupe a commencé ses travaux en examinant les pratiques et lignes directrices nationales existantes afin d’identifier les principes de qualité, les exigences, les éléments à garantir et les indicateurs qui sont essentiels ou d’une importance particulière lors de l’utilisation des données administratives et autres sources de données. Ces éléments ont ensuite été mis en correspondance avec les principes et exigences du CNAQ des Nations Unies et synthétisés en un ensemble de dix exigences essentielles et un ensemble de pratiques suggérées pour chaque exigence.

**Encadré 1:** Définitions de deux termes importants[[8]](#footnote-9)

**Cadre national d’assurance de la qualité (CNAQ):** un système cohérent et holistique de gestion de la qualité statistique. C’est un outil pour tous ceux qui officient dans le domaine des statistiques officielles. Son objectif est d’améliorer la qualité au niveau du système statistique, notamment en ce qui concerne la gestion, la coordination et les dispositions institutionnelles, les processus et les produits statistiques afin de répondre aux besoins des utilisateurs. Il propose une norme de qualité pour assurer la confiance dans les statistiques officielles.

**Cadre** **national d’assurance de la qualité des Nations Unies (CNAQ des Nations Unies):** le cadre national d’assurance de la qualité générique des Nations Unies, qui figure au chapitre 3 et à l’annexe du Manuel. Le Manuel et les recommandations qui y figurent ont été adoptés par la Commission statistique des Nations unies en mars 2019. Le CNAQ des Nations Unies se compose de principes, d’exigences et d’éléments à garantir. Le CNAQ des Nations Unies ne vise pas à remplacer les cadres et lignes directrices existants en matière d’assurance qualité des statistiques officielles. Les pays et les producteurs individuels des statistiques officielles qui sont déjà pleinement engagés dans l’assurance qualité et qui suivent l’un des cadres de qualité existants peuvent considérer le CNAQ des Nations Unies et le Manuel uniquement comme un point de référence supplémentaire qui appuie ce qu’ils font déjà, et comme une source d’information sur l’application de l’assurance qualité dans différentes situations.

Partie 1: Approche conceptuelle pour assurer la qualité lors de l’utilisation des données administratives et autres sources de données

1. L’approche conceptuelle fournit une structure globale des éléments à prendre en considération pour garantir la qualité lors de l’utilisation des administratives et autres sources de données. Elle se compose de conditions préalables importantes, de quatre actions et de deux considérations générales, comme illustré à **la figure 1**. Ces actions et considérations générales sont appelées «hyper-dimensions» dans certaines pratiques nationales et lignes directrices disponibles. Les quatre actions peuvent être comprises comme des activités ou des étapes qui doivent généralement être réalisées lors de l’utilisation des données administratives et autres sources de données. Elles peuvent se succéder ou être entreprises individuellement, en fonction des circonstances spécifiques. Les deux considérations générales reflètent les aspects qui doivent être pris en compte au cours des quatre actions et concernent les besoins des utilisateurs, la qualité des données sources et des métadonnées.
2. L’approche conceptuelle repose sur l’identification et l’analyse des exigences et pratiques pertinentes dans un large ensemble de pratiques nationales disponibles et de lignes directrices disponibles pour l’assurance qualité lors de l’utilisation des données administratives et autres sources de données, ainsi que sur l’identification et l’analyse des exigences et pratiques pertinentes dans le cadre du CNAQ des Nations Unies. En conséquence, les actions et les considérations générales et les conditions préalables présentées dans l’approche conceptuelle de la figure 1 sont étroitement liées aux exigences pertinentes du CNAQ des Nations Unies, comme indiqué par les chiffres entre parenthèses.[[9]](#footnote-10) En même temps, ces deux concepts sont liés à la liste de contrôle des dix exigences essentielles de la partie 2, comme illustré à l’annexe 4. Les différentes parties de l’approche conceptuelle sont également cohérentes et se chevauchent avec les phases et sous-processus du modèle de générique du processus de production statistique (GSBPM), qui s’applique à l’utilisation de toute source de données.[[10]](#footnote-11)
3. Une brève description des quatre actions identifiées dans l’approche conceptuelle est fournie ci-après. Les deux considérations générales relatives aux besoins des utilisateurs, à savoir la qualité des données source et des métadonnées, sont incluses dans la description des quatre actions, car elles en font partie intégrante et ne sont pas séparées. En outre, certaines conditions préalables doivent être prises en compte avant ou pendant le processus d’exploration de l’utilisation des données administratives et autres sources de données, à savoir l’accès aux données et la confidentialité.

Action 1: Identification des besoins en statistiques et sélection de la source de données

1. La première action opérationnelle lors de l’utilisation d’une source de données administrative ou d’une autre source de données est souvent la sélection de la source de données elle-même (voir l’exigence 10.3 du CNAQ des Nations Unies) sur la base des besoins des utilisateurs en matière de statistiques pertinentes (illustrée par la Considération générale 1). La sélection de la source de données nécessite également une évaluation préliminaire des métadonnées et des données de base de la source de données potentielle (Considération générale 2). Les considérations éthiques, la nécessité de sélectionner une source de données sur une base objective et une analyse coûts-avantages initiale sont également importantes lors de la sélection d’une source de données.

Action 2: Coopération avec les fournisseurs de données

1. Pour une source de données sélectionnée, un accord de coopération ou un protocole d’accord avec le fournisseur de données[[11]](#footnote-12) est une bonne pratique pour faciliter l’accès et l’utilisation du niveau des unités d’observation et des données agrégées (voir l’exigence 2.5 ou 2.6 du CNAQ des Nations unies). Cet accord contiendra des informations sur les modalités de mise à disposition des données sollicitées, notamment en ce qui concerne le mécanisme de transfert technique, l’identification des jeux de données et des données, le calendrier de transmission et les exigences de qualité. La mise à disposition d’un rapport de qualité ou une déclaration de qualité par le fournisseur des données à l’organisme statistique et référencé ou joint dans l’accord de coopération est une bonne pratique. L’Institut National de Statistique (INS) donne également des orientations aux fournisseurs de données sur les normes statistiques et coopère en permanence avec eux. L’accord de coopération devrait reposer sur une évaluation approfondie des métadonnées (considération générales 2).

Action 3: Acquisition et traitement des données

1. Les organismes statistique ont besoin de procédures appropriées pour l’acquisition et le traitement des données administratives et autres sources de données (voir l’exigence 12.3 du CNAQ des Nations Unies). Parmi les autres considérations figurent l’assurance de la solidité méthodologique, l’amélioration du potentiel des données administratives et autres sources de données, l’assurance des procédures statistiques appropriées et la gestion de la charge des répondants par le biais du partage des données, du couplage des données et de l’utilisation des données administratives et autres sources de données. L’utilisation des données de base nécessite également une évaluation approfondie de leur qualité et des métadonnées complètes (considération générales 2). L’action 3 est étroitement liée aux phases de «Collecte» et de «Traitement» du GSBPM.

Action 4: Diffusion

1. Les utilisateurs et leurs besoins sont identifiés, et des statistiques basées sur des sources de données nouvelles et existantes sont élaborées en réponse aux besoins émergents de la société en matière d’information (considération générale 1), et les besoins des utilisateurs sont satisfaits. La diffusion des résultats statistiques qui en résultent tient compte des caractéristiques particulières des données administratives et autres sources de données.

**Figure 1:** Assurer la qualité des statistiques officielles lors de l’utilisation des données administratives et autres données de base — une approche conceptuelle identifiant les actions et les aspects pertinents à prendre en considération.

**Conditions préalables importantes:**

* **Base juridique (2.5 et 2.6)\***
* **Confidentialité et sécurité des données (7.1-7.6)\***

***Action 1. Besoin en statistique et sélection de la source de données (10.3, 10.5, 5.2, 5.3, 11.1, 11.2, 11.5)\****

* **Identification et évaluation des sources de données**
* **Evaluation de la conformité de la population aux exigences de résultats, notamment les concepts, les nomenclatures, etc.)**
* **Sélection de la source de données**

*Considérations générales 1*

**Besoins des utilisateurs en matière de statistiques pertinentes (14.1,14.3)\***

*Considérations générales 2*

**Qualité des données de base et des métadonnées**

(2.7, 11.5,

12.3, 12.5)\*

***Action 2. Coopération avec les fournisseurs de données (2.5, 2.6, 2.7, 3.2, 7.5, 10.3, 11.5,16.2)\****

* **Accord de coopération**
* **Rapport qualité et renforcement des capacités**
* **Orientation sur le développement des jeu de données administratives**
* **Orientation sur les normes et classifications statistiques, notamment les mises à jour**

***Action 3. Acquisition et traitement des données (****8.5, 10.1, 10.3****, 10.5,*** 11.5, 12.1, 12.2, 12.3, 12.5, 13.4, 15.1, 18.2)\*

* **Mise en relation**
* **Imputation**
* **Édition, conversion**
* **Registre statistique, Archivage**

***Action 4. Diffusion (14.1,14.4,15.1,17.4,19.2)\****

* **Satisfaction des utilisateurs**
* **Métadonnées standardisées et rapports de qualité orientés vers utilisateurs**

\* Les chiffres entre parenthèses indiquent le lien avec les exigences ou principes pertinents du CNAQ des Nations Unies. Les exigences pertinentes des Nations Unies sont énumérées à l'annexe 3.

Partie 2: Liste de contrôle des dix exigences essentielles

1. La partie 2 présente une liste de contrôle de dix exigences essentielles qui fournissent des orientations spécifiques et détaillées pour assurer la qualité lors de l’utilisation des données administratives et autres sources de données pour produire les statistiques officielles. La liste de contrôle est complémentaire à l’utilisation du CNAQ et ne s’y substitue pas. Les dix exigences sont considérées comme essentielles car, sans elles, la qualité des statistiques officielles produites à partir des données administratives et autres sources de données peut ne pas être suffisamment garantie.
2. Les dix exigences essentielles sont étayées par un ensemble de pratiques suggérées (ou «meilleures») qui ont été identifiées lors d’un examen approfondi d’un vaste ensemble de pratiques nationales et de documents d’orientation.[[12]](#footnote-13) Ces pratiques suggérées sont souvent plus détaillées que les éléments à garantir dans le CNAQ des Nations Unies. La raison en est que le CNAQ des Nations Unies est générique, tandis que la présente liste de contrôle vise à fournir des orientations plus spécifiques sur l’utilisation des données administratives et autres sources de données. Les utilisateurs qui ont déjà pris l’habitude d’assurer la qualité de leurs statistiques officielles lorsqu’ils utilisent les données administratives et autres sources de données peuvent trouver cette liste de contrôle utile en tant que point de référence supplémentaire pour valider leurs pratiques existantes.
3. La liste de contrôle de dix exigences essentielles est étroitement liée aux actions et aux considérations générales de l’approche conceptuelle de la figure 1. L’approche conceptuelle fournit une structure et une compréhension globales pour l’utilisation des données administratives et autres sources de données pour la production des statistiques officielles. La liste de contrôle des dix exigences essentielles fournit un outil concret et des orientations pratiques à cet égard et complète l’utilisation d’un CNAQ. Elle contient des exigences qui ne sont pas abordées ou qui ne sont pas suffisamment reflétées dans les CNAQ lors de l’utilisation des données administratives et autres sources de données, et pourraient être utilisées pour mettre à jour les CNAQ ou CNAQ des Nations Unies.[[13]](#footnote-14) Elle ne fait aucune distinction entre une source de données nouvellement explorée et une sources de données déjà acquise et utilisée régulièrement.

Encadré 2: Cas particulier des données publiques

La majorité des dix exigences essentielles font référence aux fournisseurs de données en tant que détenteur ou propriétaire des données d’une source de données particulière qui les rendront disponibles à des fins statistiques. Cependant, il existe également des sources de données telles que les médias sociaux et les méthodes d’acquisition de données telles que le « web-scraping », où les données publiques sont directement exploitées par l’organisme statistique sans nécessairement s’engager dans une relation avec leurs détenteurs ou le propriétaires. Des considérations juridiques et éthiques particulières peuvent s’appliquer dans de tels cas, tandis que certaines des exigences essentielles ne le sont pas.

Les Dix Exigences essentielles

1. **L’utilisation des données administratives et autres sources de données doit répondre à un ensemble de conditions préalables et tenir compte des besoins des utilisateurs.** Cette exigence essentielle résume les conditions préalables d’accès aux données et à la confidentialité, ainsi que la prise en compte des besoins des utilisateurs qui sont bien reflectés dans le CNAQ des Nations Unies et d’autres cadres d’assurance de la qualité couramment utilisés, mais nécessitent une attention particulière lors de l’utilisation des données administratives et autres sources de données. Cette exigence essentielle est reflétée dans les conditions préalables importantes et la Considération générale 1 de l’approche conceptuelle présentée à la figure 1.

Pratiques suggérées

1. L’accès aux données est légal.
2. L’accès aux données est effectif.
3. La source de données est conforme aux lois et réglementations en vigueur (notamment, le consentement des propriétaires de données, le cas échéant) et ses données peuvent légalement être utilisées pour produire les statistiques officielles.
4. La confidentialité des données personnelles, les informations commerciales et la sécurité des données sont assurées.
5. Les besoins des utilisateurs sont pris en compte.
6. **Les nouvelles sources de données, les nouveaux fournisseurs de données ainsi que l’utilisation de sources de données multiples sont explorés de manière proactive**: Les organismes statistiques explorent constamment l’utilisation de nouvelles sources de données ou l’utilisation simultanée de sources de données multiples pour produire les statistiques officielles. Cette exigence essentielle est prise en compte avant et lors de la sélection d’une source de données (Action 1 de l’approche conceptuelle).

Pratiques suggérées

1. Les politiques, lignes directrices et procédures pratiques pour étudier et tester le potentiel des nouvelles sources de données pour la production des statistiques existent et cela s’étend à l’utilisation possible de multiples sources de données par l’intégration des données.[[14]](#footnote-15)
2. Les organismes statistiques collaborent avec les organismes gouvernementaux, les institutions publiques, les entreprises privées, les entités non gouvernementales et les organisations de la société civile, les institutions universitaires et de recherche et autres innovateurs technologiques pour identifier et explorer l’utilisation de nouvelles sources de données et l’intégration de sources de multiples données.
3. Un centre d’innovation ou une unité institutionnelle similaire existe de préférence à l’INS et appuie systématiquement l’exploration et le test de nouvelles sources de données et l’intégration des sources de multiples données pour la production des statistiques officielles au sein de l’institut national de statistique et du système statistique national.
4. **Il existe des informations de base sur les fournisseurs de données et des informations générales sur la source de données**: Il existe des informations de base sur le fournisseur des données administratives et autres sources de données. Les organismes statistiques disposent d’informations générales sur la source des données, notamment la façon dont elles sont générées. Cette exigence essentielle est généralement prise en compte lors de la sélection d’une source de données (Action 1 de l’approche conceptuelle).

Pratiques suggérées

1. Les informations sur le nom et l’adresse du fournisseur des données et sur la manière de le contacter existent (personne de contact, unité organisationnelle au sein du fournisseur de données).
2. Les informations sur le statut juridique, le type d’organisation, la résidence, l’objet ou le mandat et la gestion ou la direction du fournisseur de données existent.
3. Les informations générales sur la source des données, telles que l’objectif, la méthode et la fréquence de la collecte des données, la gestion/le stockage des données ainsi que la population qu’elle vise à couvrir (population cible) sont disponibles.
4. Une évaluation préliminaire de l’utilité de la source de données pour la production des statistiques officielles est réalisée.
5. Les informations générales sur les limites de la source de données lorsqu’elles sont utilisées pour produire les statistiques officielles sont fournies.
6. Une première évaluation coûts-avantages est réalisée, car l’utilisation des données administratives et autres sources de données peut entraîner des coûts importants (voir également 7.b).
7. **Le fournisseur et la source de données sont évalués en fonction des risques qu’ils présentent**: Les fournisseurs de données administratives et autres sources de données, ainsi que la source de données elle-même, sont évalués par rapport aux risques que l’utilisation de la source de données peut poser, et les mesures d’atténuation possibles sont envisagées. Cette exigence essentielle est prise en compte lors de la sélection d’une source de données ou lors de l’engagement avec le fournisseur de données (Action 1 et action 2 de l’approche conceptuelle).

Pratiques suggérées

1. Les informations sur le fournisseur de données, la source de données et les données sont suffisantes pour permettre une évaluation des risques que leur utilisation pour produire des statistiques officielles.
2. Le fournisseur des données est évalué pour savoir si l’utilisation de ses sources de données et de ses données pour les statistiques officielles présente des risques tels que le manque de continuité et de reproductibilité de la compilation des données, le manque de fiabilité, les considérations éthiques, les risques liés à la vie privée et à la sécurité des données, les risques liés à la réputation, etc.
3. Les mesures visant à atténuer les risques potentiels sont identifiées, y compris les options d’urgence ou de repli lorsque les données ne sont pas fournies comme convenu.
4. **Les accords de coopération avec les fournisseurs de données existent et le cas échéant, cette coopération est permanente**: Il existe des accords de coopération entre les agences statistiques et les fournisseurs de données administratives et autres sources de données, portant sur les conditions d’accès et de livraison, la confidentialité, la sécurité des données et la coopération permanente. Cette exigence essentielle est prise en compte lors du l’engagement avec le fournisseur des données (Action 2 de l’approche conceptuelle).

Pratiques suggérées

1. L’Institut National de Statistique a conclu des accords de coopération avec tous ses fournisseurs de données administratives et autres sources de données en ce qui concerne l’utilisation d’une source de données particulière couvrant à la fois les données au niveau de l’unité d’observation et les données agrégées.
2. Les accords de coopération avec les fournisseurs de données précisent les conditions d’accès aux données, notamment le contenu, la couverture, la fréquence, la ponctualité et le format de transmission, la sécurité et la confidentialité. Ils prévoient des options alternatives lorsque les données ne sont pas fournies comme convenu.
3. Les accords de coopération précisent les modalités de coopération et de communication (retour d’information et suivi) avec les fournisseurs de données en ce qui concerne la qualité des données et d’autres questions pertinentes telles que l’évolution des données au fil du temps (couverture, concepts et définitions, méthode d’acquisition, etc.). Cette pratique est également liée à l’exigence essentielle 7 qui porte sur la qualité des données de base.
4. Les organismes de statistique participent à la conception, à l’élaboration et au traitement des données administratives et des données provenant d’autres sources de données, selon le cas et dans la mesure du possible, afin de les rendre plus adaptées à des fins statistiques et autres, notamment aux fins propres du fournisseur de données.
5. Les fournisseurs de données sont tenus de répondre à toute question et à tout problème de qualité dans un délai raisonnable, tout en respectant les exigences en matière de protection de la vie privée.
6. Les organismes statistiques fournissent des orientations aux fournisseurs de données, notamment sur les normes et nomenclatures statistiques, ainsi que toute modification de celles-ci.
7. L’accord de coopération spécifie le mécanisme approprié pour fournir un retour d’information sur les questions de qualité, tout partage de données (dans le cadre du respect de la législation et la réglementation nationales) et d’autres avantages ou services que l’organisme statistique peut fournir au fournisseur de données.
8. Il y a une unité central à l’INS ou au sein du SSN qui soutient l’établissement des accords de coopération avec les fournisseurs de données.

1. **Le fournisseur de données assure la qualité de ses données et produit un rapport de qualité (ou une déclaration de qualité) en coopération avec l’organisme statistique, le cas échéant**: Le fournisseur de données assure la qualité des données de base aux fins prévues et fournit un rapport de qualité (ou similaire) à l’organisme statistique. Cette exigence essentielle est prise en compte lors de l’engagement avec le fournisseur des données (Action 2 de l’approche conceptuelle) et fait partie de la Considération générale 2 de l’approche conceptuelle.

Pratiques suggérées

1. Les fournisseurs de données administratives et autres sources de données ont fourni une description de leurs procédures d’assurance qualité lors de la compilation et du traitement des données (y compris toute transformation, édition, estimation, traitement des unités et valeurs manquantes, et valeurs aberrantes, etc.) dans la mesure du possible.
2. Les procédures d’assurance qualité des fournisseurs de données sont évaluées et les données qui en résultent sont, en générale, adéquates en ce qui concerne leur utilisation pour la production des statistiques officielles.
3. Un rapport de qualité, une déclaration de qualité ou une note décrivant l’exactitude, l’exhaustivité, l’actualité, la ponctualité et d’autres caractéristiques pertinentes des données est élaboré par l’organisme statistique et le fournisseur de données.
4. Les résultats des audits de qualité et autres formes d’évaluation de la qualité menés par le fournisseur de données ou auprès de celui-ci sont communiqués à l’organisme statistique, s’il y a lieu et si possible.
5. **La qualité des données d’entrée est systématiquement évaluée par l’organisme statistique:** La qualité des données d’entrée est systématiquement et régulièrement évaluée et évaluée à toutes les étapes pertinentes du processus de production statistique. Cette exigence essentielle est prise en compte lors de la sélection d’une source de données (action 1 de l’approche conceptuelle), lors de l’utilisation des données (action 3 de l’approche conceptuelle) et dans le cadre de la considération générale 2 de l’approche conceptuelle.

Pratiques suggérées

1. Les données administratives et autres données sont systématiquement évaluées au cours de la phase de sélection, en coopération avec le fournisseur de données, en fonction d’un ensemble de critères objectifs prédéfinis concernant leur utilisation potentielle et leur utilité pour produire des statistiques officielles, notamment en termes d’exactitude, d’exhaustivité, de couverture des différents groupes et d’éventuels biais dus à la sous-représentation ou de la surreprésentation de groupes spécifiques, de la cohérence conceptuelle et de la comparabilité, de la dimension temporelle (actualité, périodicité et période de référence) et de l’accessibilité, y compris en termes de coût et de confidentialité.
2. La sélection d’une source de données fait suite à une analyse coûts-avantages tenant compte des ressources humaines, des infrastructures et d’autres coûts et de la durabilité, étant donné que l’utilisation des données administratives et d’autres sources de données peut entraîner des coûts importants (par exemple, dans le cas de jeux de données brutes très volumineux).
3. Les données administratives et autres sources de données sont systématiquement et régulièrement évaluées avant et pendant l’utilisation pour vérifier l’exactitude, l’exhaustivité, la couverture des différents groupes et les biais éventuels en raison de la sous-représentation ou de la surreprésentation de groupes spécifiques, de la cohérence conceptuelle et de la comparabilité, de la dimension temporelle (temps, périodicité et période de référence), des contrôles techniques/accessibilité, ainsi que de l’intégrabilité/liaison, et les résultats sont reflétés dans les rapports périodiques sur la qualité (voir annexe 1, tableau 1: Sous-module pour la validation des données d’entrée).
4. **Il existe des métadonnées complètes concernant les données de base**: Il existe des métadonnées complètes et standardisées sur les données administratives et autres sources de données concernant les concepts, définitions et nomenclature utilisés, la structure des données, les unités et les variables (et les valeurs possibles), la couverture (population), la zone de référence, la période de référence, l’actualité, la méthode de collecte, le traitement des données, y compris pour les valeurs aberrantes, les erreurs, les estimations et les imputations, le nombre d’enregistrements, le format des données et des fichiers, les changements passés et futurs au fil du temps (stabilité), etc. Cette exigence essentielle est prise en compte lors de la sélection d’une source de données (action 1 de l’approche conceptuelle), lors de l’utilisation et de la diffusion des données (Action 3 et action 4 de l’approche conceptuelle) et dans le cadre de la considération générale 2 de l’approche conceptuelle.

Pratiques suggérées

1. Les informations complètes sont fournies sur les concepts, définitions et nomenclatures utilisées.
2. Les informations complètes sont disponibles sur les jeux de données, notamment la structure des données, les unités et les variables (et les valeurs et ensembles de codes possibles), la couverture (population), la zone de référence, la période de référence et l’actualité.
3. Les informations détaillées sont fournies sur la base et la méthode de collecte, de traitement et d’édition des données, y compris pour les valeurs aberrantes, les erreurs, les estimations, les imputations et l’utilisation d’algorithmes, de logiciels et d’intelligence artificielle chez le fournisseur de données, ainsi que toutes les règles et réglementations relatives à la production des données, le cas échéant.
4. Les informations sont produites sur le nombre d’enregistrements, les données et le format de fichier.
5. Les métadonnées suivent un format standard.[[15]](#footnote-16)
6. Les informations complètes sont fournies sur les modifications pertinentes passées et prévues de la compilation des données au fil du temps (stabilité).
7. Les limites éventuelles de la source de données pour la production des statistiques officielles sont identifiées, notamment celles liées à la couverture de la population, à l’exhaustivité de l’information, à l’éventuel biais et à l’alignement des unités administratives avec les unités statistiques d’intérêt (p. ex. unité fiscale par rapport à l’établissement ou à l’entreprise).
8. **Le traitement des données de base par l’organisme statistique respecte les normes, les directives et les meilleures pratiques:** Le traitement et l’édition des données suivent les normes, les lignes directrices et les meilleures pratiques, et sont documentés et contrôlés. Il existe une collaboration permanente avec le fournisseur de données. Cette exigence essentielle est prise en compte lors de l’utilisation des données (Action 3 de l’approche conceptuelle) et d’une partie de la Considération générale 2 de l’approche conceptuelle.

Pratiques suggérées

1. Les lignes directrices sont disponibles pour l’assurance et la gestion de la qualité au sein de l’organisme statistique lors de l’utilisation des données administratives et autres sources de données. Ces lignes directrices devraient être standardisées autant que possible et reflétées également les circonstances particulières.
2. Le modèle générique du processus de production statistique est utilisé pour structurer et documenter le processus statistique.
3. Les processus de prétraitement et de traitement des données, y compris les contrôles de qualité, la gestion des erreurs, la standardisation, la transformation et l’agrégation des données, l’estimation et l’imputation au sein de l’organisme statistique sont bien documentés, testés et contrôlés, et suivent les meilleures pratiques.
4. L’impact de tout changement conceptuel ou autre affectant les données administratives et autres sources de données est systématiquement évalué par l’organisme statistique.
5. Les fournisseurs de données sont consultés par l’organisme statistique en cas de questions et de problèmes de qualité.
6. La qualité des produits statistiques, notamment des registres statistiques fondés sur les données administratives et autres sources de données, est systématiquement évaluée par l’organisme statistique.
7. L’utilisation des données administratives et autres sources de données est encouragée et pratiquée au sein de l’organisme de statistique, notamment par l’échange des données avec d’autres unités administratives et d’autres organismes statistiques, et par l’établissement d’un couplage des données, en utilisant des outils, des technologies et des méthodes de pointe, y compris le SDMX.
8. **La diffusion des résultats statistiques répond aux besoins des utilisateurs.** Les caractéristiques particulières des données administratives et autres sources de données sont prises en compte lors de la diffusion des produits statistiques.Cette exigence essentielle est prise en compte dans la diffusion des produits statistiques (Action 4 de l’approche conceptuelle) et dans le cadre des Considérations générales 1 et 2 de l’approche conceptuelle.

Pratiques suggérées

1. L’utilisation des données administratives et autres sources de données est fondée sur les nouveaux besoins de la société en matière d’information, notamment la consultation des utilisateurs; la satisfaction des utilisateurs est régulièrement mesurée et suivie.
2. Les utilisateurs reçoivent des métadonnées standardisées et des rapports sur la qualité ou des déclarations de qualité qui informent sur l’utilisation des données administratives et d’autres sources de données et sur les aspects pertinents de la qualité, ainsi que sur les limites de la production statistique, notamment l’exactitude, la fiabilité, la cohérence et la comparabilité.
3. Les indicateurs de qualité sont adaptés pour informer sur la qualité des statistiques basées sur les données administratives et autres sources de données.
4. L’accès aux microdonnées provenant de sources administratives et autres sources de données est soumis aux règles et protocoles de l’organismes statistique en matière de confidentialité statistique, sauf si des dispositions différentes ont été convenues avec les autorités compétentes et le fournisseur de données et sont mises à la disposition du public, le cas échéant.

Annexe 1: Sous-module de validation des données de base et liste des indicateurs supplémentaires pour évaluer les dix exigences essentielles

**Le tableau 1** fournit un sous-module pour la validation des données de base une fois qu’elles sont reçues par l’organisme statistique du fournisseur.[[16]](#footnote-17) Les indicateurs de qualité proposés peuvent faire partie d’un rapport sur la qualité que l’organisme statistique produit chaque fois qu’il reçoit des données. La plupart des contrôles peuvent être automatisés et font partie de ce que l’on appelle généralement la validation structurelle d’un jeu de données.[[17]](#footnote-18) Ce rapport de validation des données de base peut être partagé avec le fournisseur de données. En fonction de la situation spécifique et de l’accord de coopération, de nouvelles mesures pourraient être prises en vue d’une intégration de l’assurance qualité au sein des processus de production du fournisseur de données et de l’organisme statistique, ce qui pourrait également inclure une coopération sur la manière de traiter des problèmes spécifiques de qualité. Les indicateurs d’exactitude, d’exhaustivité et de représentativité et l’indicateur 23 sur la comparabilité des unités sont susceptibles d’être inclus dans les rapports sur la qualité destinés aux utilisateurs.

**Tableau 1:** Sous-module pour la validation des données de base: Indicateurs (numériques) possibles\*

| Dimensions de qualité/objet | Indicateurs | Mesures |
| --- | --- | --- |
| Contrôles techniques/accessibilité |  |  |
| * Jeu de données | 1. Lisibilité/accessibilité: Le jeu de données de la source est accessible et lisible par machine | Oui/Non |
|  | 2. Conformité de la structure des fichiers et des métadonnées: Le jeu de données contient tous les champs attendus | Oui/Non/Inconnu |
|  | 3. Nombre d’enregistrements: Nombre d’enregistrement reçus divisés par nombre d’enregistrement attendus (par exemple, le nombre d’enregistrement de la dernière transmission) | Pourcentage |
| * Variables | 4. Nombre de variables: Nombre d’enregistrements ayant une valeur valide pour une variable divisé par le nombre d’enregistrements | Pourcentage |
| Précision (mesures numériques) |  |  |
| * Unités (objets) d’observation | 5. Légitimité/authenticité: unité sans clé (s) d’identification autorisée(s) (mauvaise (s) ou invalides), le cas échéant, avec un pourcentage calculé sur la base du nombre d’observations dans le jeu de données | Pourcentage |
|  | 6. Incohérence dans le jeu des données: Unités ayant des relations non logiques avec d’autres unités, le cas échéant, avec un pourcentage calculé en fonction du nombre d’observations dans le jeu de données (la correction est obligatoire avant de poursuivre) | Pourcentage |
|  | 7. Inconsistance dans le jeu de données: Unités ayant des relations non plausibles ou suspectes avec d’autres unités, le cas échéant, avec un pourcentage calculé en fonction du nombre d’observations dans le jeu de données (la correction est facultative avant de poursuivre) | Pourcentage |
| * Variables | 8. Erreur de mesure: Valeurs pour lesquelles une erreur de mesure est indiquée par le fournisseur de données, le pourcentage étant calculé sur la base du nombre d’observations dans le jeu de données | Pourcentage |
|  | 9. Valeurs incohérentes: Valeurs ayant une relation non logique avec d’autres informations pour l’unité d’observation ou en dehors de l’intervalle possible (erreurs), le pourcentage étant calculé sur la base du nombre d’observations dans le jeu de données (la correction est obligatoire avant de poursuivre) | Pourcentage |
|  | 10. Valeurs invraisemblables: Valeurs ayant des relations non plausible ou suspectes avec d’autres informations de l’unité d’observation ou en dehors de la fourchette prévue (valeurs aberrantes), avec un pourcentage calculé sur la base du nombre d’observations dans le jeu de données  (la correction est facultative avant de poursuivre) | Pourcentage |
| Exhaustivité/Représentativité |  |  |
| * Unités (objets) d’observation | 11. Unités manquantes: Unités non incluses, le pourcentage étant calculé sur la base du nombre d’observations dans la population cible | Pourcentage |
|  | 12. Unités en dehors de la population cible: Unités n’appartenant pas à la population cible avec le pourcentage calculé sur la base du nombre d’observations dans le jeu de données | Pourcentage |
|  | 13. Redondance/doublons: Unités dupliquées lorsqu’un doublon est observé, le pourcentage étant calculé sur la base du nombre d’observations dans le jeu de données | Pourcentage |
|  | 14. Sélectivité/représentativité: Écarts (sur-couverture ou sous-couverture) dans les caractéristiques pertinentes entre le jeu de données et la population cible, par exemple le pourcentage des caractéristiques dans le jeu de données moins le pourcentage dans la population cible | Pourcentage |
| * Variables | 15. Valeurs manquantes, le pourcentage étant calculé sur la base du nombre d’observations dans le jeu de données | Pourcentage |
|  | 16. Valeurs imputées: Valeurs pour lesquelles une valeur imputée est indiquée par le fournisseur de données, le pourcentage étant calculé sur la base du nombre d’observations dans le jeu de données | Pourcentage |
| Dimension temporelle |  |  |
| * Jeu de données | 17. Actualité: Date de réception — Date de fin de la période de référence | Jours |
|  | 18. Délai moyen d’inscription: Date d’enregistrement des données administratives — Date de fin de la période de référence | Jours |
|  | 19. Ponctualité: Date de réception — Date de livraison convenue | Jours |
| * Unités (objets) d’observation | 20. Dynamique: Nouvelles unités dans le jeu de données, anciennes unités qui n’existent plus dans le jeu de données, les pourcentages étant calculés en fonction du nombre d’observations dans le jeu de données | Pourcentage |
| * Variables | 21. Stabilité: Modification des valeurs des variables des unités existantes au fil du temps | Pourcentage |
| Liaison/intégrabilité |  |  |
| * Ensemble de données | 22. Couplage: Unités pouvant être clairement liées aux unités du registre ou à d’autres jeux de données (par exemple, par l’intermédiaire d’une variable ou d’un identifiant unique), le pourcentage étant calculé en fonction du nombre d’observations dans le jeu de données | Pourcentage |
| * Unités (objets) d’observation | 23. Comparabilité des unités dans la source: Unités ayant la définition du concept souhaitée, le pourcentage étant calculé sur la base du nombre d’observations dans le jeu de données | Pourcentage |
| * Variables | 24. Comparabilité des valeurs: Différences dans la valeur des variables des unités liées | Pourcentage |

\* Adapté de Daas P, Ossen S., BLUE-ETS (2011): Livrable 4.2: Rapport sur les méthodes préférées pour les indicateurs de qualité des sources de données administratives. Disponible à l’adresse suivante: http://www.pietdaas.nl/beta/pubs/pubs/BLUE-ETS\_WP4\_Del2.pdf.

La plupart des pratiques suggérées («meilleures») dans la liste de contrôle des dix exigences essentielles sont formulées de manière à permettre une réponse oui/non/partiellement et peuvent donc servir d’indicateurs de qualité ou de liste de contrôle des indicateurs de qualité. **Le tableau 2** fournit une liste indicative d’exemples d’indicateurs supplémentaires tirés principalement d’une liste d’indicateurs de qualité élaborés pour le modèle générique du processus de production statistique (GSBPM). De nombreux indicateurs du tableau 2 se chevauchent avec les pratiques suggérées («meilleures») dans la liste de contrôle des dix exigences critiques. Des travaux supplémentaires sont nécessaires pour élaborer une liste de contrôle consolidée des indicateurs pour évaluer l’état d’avancement de la mise en œuvre des dix exigences essentielles.

**Tableau 2**: Liste indicative des indicateurs supplémentaires pour les exigences essentielles\*

|  |  |  |
| --- | --- | --- |
| Exigence essentielles | Indicateurs\* | Mesures possibles |
| CR 1 | 1. Dans quelle mesure les parties prenantes ont-elles confirmé les besoins statistiques détaillés (quoi, quand, comment et pourquoi) tels que documentés par l’INS? | Oui/Non/Partiellement |
|  | 2. Dans quelle mesure la source de données répond-elle à la demande d’information? | Oui/Non/Partiellement |
| CR 4 | 3. Durée prévue des séries chronologiques comparables. | Nombre de périodes |
| CR 5 | 4. Existence d’un plan de notification préalable des modifications à venir de la source de données. Un plan d’urgence pour les changements apportés aux données ou à la source de données est-il disponible? | Oui/Non |
|  | 5. Pourcentage de données transmises conformément aux accords conclus avec les propriétaires de données administratives (par exemple, format, calendrier), le pourcentage étant calculé en fonction du nombre d’observations dans le jeu de données. | Pourcentage |
| CR 7 | 6. Mesure dans laquelle les concepts, définitions et classifications associés aux variables et aux populations (clés) suivent les normes internationales ou nationales. | Échelle |
|  | 7. Degré de conformité conceptuelle de l’unité réelle (objet) d’observation avec l’unité d’observation souhaitée et impact de toute différence sur l’exactitude des données\*\* | Qualitative |
|  | 8. Degré de conformité conceptuelle des variables avec les variables souhaitées et impact de toute différence sur l’exactitude des données\*\* | Qualitative |
| CR 8 | 9. Quand la technique de collecte des données a-t-elle été révisée/améliorée pour la dernière fois? | Années |
| CR 9 | 10. Mesure dans laquelle les techniques d’intégration des données administratives sont comprises et spécifiées, tant pour l’utilisation directe qu’indirecte des jeux de données administratives. | Oui/Non/Partiellement |
|  | 11. Dans quelle mesure le processus opérationnel utilise-t-il des méthodes standardisées ou bien connues pour les phases suivantes (p. ex. codage, vérification et imputation, intégration des données, pondération, estimation, révision), de manière transparente? | Oui/Non/Partiellement |
|  | 12. Pourcentage de processus (et sous-processus) GSBPM identifiés et documentés avec leurs flux | Pourcentage |
|  | 13. Les spécifications des systèmes de production et du flux de travail tiennent compte du type de données traitées (données des répondants ou données administratives ou combinaison de celles-ci). | Oui/Non/Partiellement |
|  | 14. Les composants du processus de couplage de données ont-ils été testés et affinés? | Oui/Non/Partiellement |
|  | 15. Degré de cohérence avec d’autres sources, avec des données provisoires, avec des estimations rapides, et avec les résultats antérieurs du même processus. | Échelle |
|  | 16. Comparabilité des résultats dans le temps et avec les résultats d’autres sources de données\*\* | Qualitative |
| CR 10 | 17. Dans quelle mesure des contraintes juridiques ont-elles été prises en compte en ce qui concerne les produits statistiques, par exemple, mais sans s’y limiter, pour garantir la confidentialité des données et prévenir la divulgation des informations sensibles? | Oui/Non/Partiellement |
|  | 18. Les règles de confidentialité et les procédures d’accès aux microdonnées ont-elles été conçues? | Oui/Non/Partiellement |
|  | 19. Dans quelle mesure les données sont-elles réellement protégées? Quel est le risque résiduel de divulgation? | Qualitative |

\* Sélectionné du document « Indicateurs de qualité pour le modèle générique du processus de production statistique (GSBPM) — Pour les statistiques dérivées des enquêtes et des sources de données administratives, version 2.0, octobre 2017, de la Commission économique des Nations Unies pour l’Europe .

\*\* Indicateurs ajoutés.

Annexe 2: Glossaire des définition de termes importants

Ce glossaire fournit les définitions de termes importants utilisés dans ce module pour l'assurance qualité lors de l'utilisation de sources de données administratives et autres sources de données. Les définitions présentées ci-dessous sont utilisées tout au long de ce module, mais les pays peuvent avoir leurs propres définitions. Certaines des définitions décrivant les aspects de la qualité doivent être appliquées de manière spécifique dans certaines situations, par exemple, l’actualité n'est pas le même pour les producteurs de statistiques officielles qui utilisent des données d'entrée et pour les utilisateurs de produits statistiques. Les définitions ci-dessous sont tirées du Manuel des cadres nationaux d'assurance de la qualité des Nations Unies en statistiques officielles (ci-après dénommé le Manuel) (voir paragraphe 1.14), sauf indication contraire.

Termes qui décrivent les aspects de qualité

* **Exactitude**: Proximité des estimations aux valeurs exactes ou réelles que les statistiques étaient censées mesurer.
* **Accessibilité**: Facilité et conditions d’accès des informations statistiques.
* **Clarté**: Disponibilité de la documentation appropriée relative aux statistiques et assistance supplémentaire que les producteurs mettent à la disposition des utilisateurs.
* **Cohérence et homogénéité**: Capacité de combiner de manière fiable des statistiques et des ensembles de données de différentes manières et pour différentes utilisations. La cohérence est souvent utilisée comme synonyme d’homogénéité.[[18]](#footnote-19)
* **Comparabilité**: Mesure dans laquelle les différences entre les statistiques de zones géographiques différentes, de domaines non géographiques ou dans le temps peuvent être attribuées à des différences entre les valeurs réelles des statistiques.
* **Exhaustivité et couverture**: L'exhaustivité fait référence à la mesure dans laquelle toutes les statistiques nécessaires sont disponibles. La mesure de la disponibilité des statistiques nécessaires se réfère normalement à des ensembles de données [ensemble d'observations] et compare l'ensemble de données requis à celui disponible. La couverture est la définition du champ d’application des données compilées. Cet élément de métadonnées est utilisé pour décrire les dimensions délimitant les statistiques produites, par exemple géographiques, les produits, les secteurs économiques et autres, l’industrie, la profession, les transactions, les groupes démographiques. etc., ainsi que les exceptions et exclusions pertinentes. Il peut également spécifier la période pour laquelle les données sont fournies

(voir le glossaire SDMX — version 2.1 — décembre 2020).

* **Confidentialité et vie privée**: propriété des données indiquant si elles font l'objet de restrictions de diffusion. Les données sont protégées par la confidentialité dans les cas où une divulgation non autorisée pourrait être préjudiciable ou dommageable aux intérêts de la source ou d'autres parties concernées. Par exemple, les données permettant l’identification d’une personne physique ou morale, directement ou indirectement, peuvent être qualifiées de confidentielles selon la législation nationale ou internationale applicable (voir pour plus de détails, le Glossaire SDMX – Version 2.1 – Décembre 2020). Les données confidentielles désignent les données qui permettent d'identifier des unités statistiques individuelles, directement ou indirectement, divulguant ainsi des informations individuelles. Pour déterminer si une unité statistique est identifiable, il convient de tenir compte de tous les moyens pertinents qui pourraient raisonnablement être utilisés par un tiers pour identifier l'unité statistique (voir le manuel du système statistique européen pour les rapports sur la qualité et les métadonnées, édition 2020). La vie privée est l’état d’une personne non surveillé ou dérangé par d’autres personnes (Oxford Dictionary).
* **Intégrabilité / Liaison :** la facilité avec laquelle les données de la source peuvent être intégrées dans le système de production statistique. Pour les unités statistiques (objets), il s'agit de la comparabilité et de la facilité de relier les unités de la source à celles couramment utilisées par les organismes statistiques. Pour les variables, il s'agit de la proximité des valeurs de la source avec les faits de variables similaires (voir Daas P, Ossen S., BLUE-ETS (2011).
* **Pertinence**: Mesure dans laquelle les statistiques répondent aux besoins des utilisateurs.
* **Fiabilité**: Proximité de la (des) valeur(s) initialement estimée(s) avec la (les) valeur(s) estimée(s) ultérieurement si des chiffres préliminaires sont diffusés.
* **Actualité** : Délai entre la fin d’une période de référence (ou date) et la diffusion des statistiques.
* **Ponctualité**: Délai entre la date de publication et la date cible à laquelle les données ou statistiques auraient dû être fournies.

Conditions générales

* **Fournisseurs de données et producteurs de statistiques**: Ce Manuel fait la distinction entre les fournisseurs de données, qui contribuent au processus de production de statistiques (tels que les répondants et les détenteurs ou propriétaires de données statistiques, administratives ou autres), et les producteurs de statistiques, qui produisent un résultat statistique. Selon le contexte, le terme fournisseur de données ne s’entend dans ce Manuel que des détenteurs ou propriétaires de donnée.
* **Sources de données**: Ce Manuel distingue trois sources de données en fonction de leur objectif et de l’entité responsable de leur compilation : les sources de données statistiques, telles que les enquêtes; les sources de données administratives; et autres sources de données. En général, les autres sources de données incluent celles qui sont associées à l’expression « mégadonnées », à moins qu’elles soient déjà incluses, dans certains cas, dans des sources de données statistiques ou administratives. Les nouvelles sources de données peuvent souvent être associées à d’autres sources de données. Cependant, elles peuvent également être considérées comme faisant partie de sources de données statistiques ou administratives, en fonction de la situation nationale.
* **Ensemble de données, structure des données, observations, unités et variables**: un ensemble de données fait référence à une collection organisée d'observations qui partagent la même structure de données (dimensions, attributs et mesures). Les observations peuvent également être appelées « enregistrements » ou « enregistrements de données ». Les observations contiennent des informations sur l'unité ou l'objet de l'observation qui peut être une personne, un ménage, une entreprise, un emplacement, une transaction, etc. Une variable est une caractéristique d'une unité observée qui peut prendre plus d'un ensemble de valeurs (internes). définition interne partiellement basée sur le glossaire SDMX – Version 2.1 – décembre 2020).
* **Données d’entrée et données sources**: fait référence aux données utilisées dans le processus de production statistique. Dans le contexte de ce module, les données d'entrée et les données sources ont la même signification (définition de travail interne).
* **Métadonnées**: Données qui définissent et décrivent d’autres données. On peut faire la distinction entre métadonnées structurelles et métadonnées de référence. Les métadonnées structurelles définissent et accompagnent les données. Elles sont constituées des identificateurs et descripteurs nécessaires pour la manipulation, l’organisation, la récupération et le traitement d’un ensemble de données statistiques (titres, sous-titres, brèves descriptions, noms de dimension, noms de variables, etc.). Les métadonnées de référence sont plus générales et décrivent les concepts et méthodologies statistiques utilisés pour la collecte et la production de données. Elles fournissent des informations sur la qualité des données, aidant ainsi les utilisateurs à interpréter ces dernières. Contrairement aux métadonnées structurelles, les métadonnées de référence peuvent être découplées des données (c’est-à-dire qu’elles peuvent être générées, collectées ou diffusées séparément des statistiques auxquelles elles se rapportent).
* **Cadre national d’assurance de la qualité (CNAQ)**: Système cohérent et global de gestion de la qualité statistique qui assure la confiance et la qualité des statistiques officielles.
* **Institut national de statistique (INS)**: Premier organisme statistique au sein d’un système statistique national. Les expressions office national de la statistique et institut national de la statistique sont synonymes. En général, l’INS assume un rôle de coordination au sein du système statistique national et est responsable du développement, de la production et de la diffusion des statistiques officielles dans de multiples domaines statistiques.
* **Système statistique national (SSN)**: Ensemble des organismes et unités statistiques (organismes statistiques) d’un pays qui développent, produisent et diffusent des statistiques officielles pour le compte du gouvernement national (et d’autres niveaux de gouvernement). Il incombe à chaque pays de définir la portée de son SSN (voir aussi organismes statistiques, fournisseurs de données et producteurs de statistiques et écosystème de données).
* **Statistiques officielles**: Statistiques décrivant, de manière représentative, les phénomènes économiques, démographiques, sociaux et environnementaux d’intérêt public. Les statistiques officielles sont développées, produites et diffusées en tant que bien public par les membres du SSN, conformément aux Principes fondamentaux de la statistique officielle et aux cadres de qualité acceptés tels que le CNAQ des Nations Unies, ainsi que d’autres normes et recommandations statistiques reconnues au plan international. Dans de nombreux pays, les statistiques officielles sont définies et décrites dans les programmes statistiques.
* **Rapport qualité**: une manière typique d’enregistrer les résultats d’une évaluation de la qualité. On peut distinguer un rapport orienté vers le producteur comprenant des métadonnées pour enregistrer les problèmes de qualité et les améliorations par rapport à un rapport orienté utilisateur comprenant des métadonnées destinées aux utilisateurs des produits statistiques, leur permettant d'évaluer si les produits sont appropriés aux objectifs fixés. (voir le manuel du système statistique européen pour les rapports sur la qualité et les métadonnées, édition 2020).
* **Organismes statistiques**: Les membres du SSN sont appelés « organismes statistiques » et englobent l’INS et autres producteurs de statistiques officielles. Les organismes statistiques autres que l’INS n’ont généralement pas pour principal objectif ou tâche de produire des statistiques officielles et seule une section ou un petit groupe de personnes y produisent des statistiques. Les exigences de qualité pour les processus et la production sont les mêmes pour toutes les statistiques officielles. Toutefois, pour un ministère ou un organe administratif dont une partie seulement produit des statistiques, les exigences liées à l’environnement institutionnel s’appliquent uniquement à l’entité qui produit des statistiques officielles. Par exemple, bien que le ministère ou l’organe administratif ne soit généralement pas indépendant, l’unité au sein des ministères/organes administratifs responsable de la production de statistiques devrait décider de manière indépendante de la manière de produire des statistiques et du calendrier de leur diffusion

Annexe 3: Exigence pertinentes du CNAQ des Nations unies

L’annexe 3 contient le texte complet des exigences du CNAQ des Nations unies (contenues dans l’annexe du Manuel) qui sont pertinentes ou contiennent des pratiques suggérés («meilleures») ou des éléments à garantir qui sont pertinents pour assurer la qualité des sources de données administratives et autres sources de données. Les exigences et les éléments pertinents à garantir les plus pertinents sont mis en évidence. Les exigences sont énumérées telles qu’elles sont et dans leur séquence dans le CNAQ des Nations unies sans aucune modification. L’annexe 3 est fournie uniquement pour faciliter la consultation, en tenant ainsi compte des utilisateurs dans les pays ou les organismes de statistique qui n’utilisent pas le CNAQ des Nations unies.

Exigence 2.5 : L’institut national de la statistique et, le cas échéant, d’autres organismes

statistiques ont l’autorité légale ou une autre disposition officielle pour obtenir des données administratives et un accès adéquat à ces données auprès d’autres organismes publics

à des fins statistiques.

* La loi statistique prévoit des dispositions appropriées pour garantir à l’INS et, le cas échéant, à d’autres organismes statistiques, le droit d’obtenir ou d’accéder aux données administratives en temps utile.
* Dans les cas où les organismes statistiques n’ont pas légalement le droit d’obtenir des données administratives, des protocoles d’entente sont en place pour permettre un tel accès.
* L’accès des organismes statistiques aux données administratives est gratuit.
* Des accords avec les propriétaires de données administratives sont en place pour permettre l’accès aux données. Ils décrivent les conditions techniques d’accès et les possibilités d’intégration de ces données avec celles provenant d’autres sources de données administratives.
* Les organismes statistiques participent à la conception et au développement des ensembles de données administratives afin de les adapter à des fins statistiques; cette implication s’étend à toute interruption du processus pour ces ensembles de données.

Exigence 2.6 : L’institut national de la statistique et, le cas échéant, d’autres organismes statistiques ont l’autorité légale ou une autre disposition officielle et des accords connexes permettant l’accès aux données (y compris les mégadonnées) gérées par des sociétés privées ou d’autres organisations non gouvernementales à des fins statistiques sur une base régulière, notamment pour des tests et des expérimentations.

* La loi statistique prévoit des dispositions appropriées pour garantir à l’INS et, le cas échéant, aux autres organismes statistiques, le droit d’obtenir ou d’accéder, en temps utile, aux données détenues par des sociétés privées ou autres organisations non gouvernementales à des fins statistiques (comme les sociétés qui fournissent des services aux personnes physiques et morales résidant dans le pays).
* La loi statistique prévoit des sanctions adéquates pour garantir l’accès aux données privées, le cas échéant (telles que des amendes pour n’avoir pas donné un tel accès).
* Dans les cas où les organismes statistiques ne sont pas légalement autorisés à accéder aux données gérées par des sociétés ou autres organisations non gouvernementales, des protocoles d’accord sont en place pour accorder un tel accès.
* Les organismes statistiques examinent la pertinence et la portée des données demandées.
* L’accès aux données privées et leur utilisation sont conformes aux procédures convenues entre les organismes statistiques et les propriétaires ou détenteurs des données.

Exigence 2.7 : L’institut national de la statistique coopère avec les fournisseurs de données et leur donne un appui et des conseils.

* L’INS consulte régulièrement les fournisseurs de données et entretient une coopération avec les fournisseurs de données administratives ainsi qu’avec les sociétés, entreprises et autres organisations qui détiennent des données afin de renforcer la valeur statistique et l’utilisation de ces sources.
* Les rapports sur la qualité des données administratives sont élaborés en coopération avec l’INS et le propriétaire des données, et décrivent notamment l’exactitude, l’exhaustivité, l’actualité et la ponctualité.90
* Les détenteurs de données administratives, les entreprises et les autres organisations reçoivent des commentaires sur la qualité des données fournies pour des améliorations futures.
* Des accords de partenariat avec les fournisseurs de données sont signés.

Exigence 3.2 : L’institut national de la statistique donne un appui et des conseils à tous les fournisseurs de données et producteurs de statistiques officielles pour l’application des normes statistiques.

* L’INS veille à l’utilisation des normes statistiques par les fournisseurs de données et les producteurs de statistiques officielles.
* Des rapports périodiques sont produits sur le respect des normes statistiques internationales, régionales et nationales.
* Les normes statistiques sont communiquées et mises à la disposition de tous les fournisseurs de données et producteurs de statistiques officielles.
* Les plans et les calendriers d’élaboration et d’application des nouvelles normes sont communiqués à l’avance.
* L’INS aide les autres producteurs de statistiques et fournisseurs de données à mettre en œuvre les normes statistiques internationales, régionales et nationales, le cas échéant.

Exigence 5.2 : Les organismes statistiques mettent en œuvre une déclaration ou un code de conduite ou d’éthique qui régit les pratiques statistiques, dont le respect fait l’objet d’un suivi.

* Il existe des directives éthiques ou un code de conduite pour assurer l’impartialité et l’objectivité.
* Les directives sont à la disposition du public.
* La mise en œuvre des directives fait l’objet d’un suivi.

Exigence 5.3 : Les sources de données et les méthodologies sont choisies de manière objective.

* Les sources, concepts, méthodes et processus pour le développement, la production et la diffusion de données sont choisis sur la base de considérations statistiques, de principes nationaux et internationaux, et de bonnes pratiques.

Exigence 7.1 : Le secret statistique est garanti par la loi.

* Il existe une loi ou autre disposition officielle claire en vigueur qui exige la gestion appropriée des informations reçues des répondants et des fournisseurs de données afin de garantir la confidentialité des statistiques et la sécurité des données.

Exigence 7.2 : Des normes, directives, pratiques et procédures appropriées sont en place pour garantir la confidentialité des statistiques.

* Des directives et instructions sur la garantie de la confidentialité des statistiques tout au long du processus statistique sont fournies à tout le personnel des organismes statistiques.
* Il existe des programmes de formation réguliers et continus pour tout le personnel sur le concept de confidentialité statistique et les meilleures pratiques pour garantir la confidentialité des informations fournies.
* La structure organisationnelle et les dispositions prises pour l’élaboration et la mise en œuvre de pratiques garantissant la confidentialité statistique permettent de répondre aux besoins.
* Les membres du personnel signent des accords de confidentialité lors de leur nomination, lesquels sont également valables après leur départ de l’organisme.

Exigence 7.3 : Des protocoles stricts permettant de garantir la confidentialité des données s’appliquent aux utilisateurs ayant accès aux microdonnées à des fins de recherche ou statistiques.

* La loi sur la statistique ou autre disposition légale fixe des conditions claires permettant aux chercheurs d’accéder aux données confidentielles à des fins scientifiques.
* Des règles de confidentialité, un contrôle de la divulgation et des procédures d’accès aux microdonnées s’appliquent tout au long du processus statistique.
* Les organismes statistiques suivent l’utilisation des ensembles de microdonnées pour identifier les cas où la confidentialité des données peut être violée (par exemple, par l’appariement de fichiers) et prennent des mesures correctives immédiates pour y remédier.

Exigence 7.4 : Des sanctions sont prévues pour toute violation délibérée du secret statistique.

* Des dispositions légales ou autres sont en place pour permettre des sanctions administratives, pénales et disciplinaires en cas de violation du secret statistique.
* Les informations sur les dispositions permettant des sanctions pour violation du secret statistique sont partagées avec tout le personnel et mises à la disposition du public.

Exigence 7.5 : La sécurité et l’intégrité des données et leur transmission sont garanties par des politiques et pratiques appropriées.

* Une politique de sécurité informatique est en place et connue du personnel.
* Conformément à la stratégie informatique, des mesures et processus de sécurité physiques appropriés sont en place pour garantir la sécurité des données et bases de données, conformément aux meilleures pratiques et aux normes internationales.
* Des audits réguliers du système de sécurité des données sont effectués.
* Tous les accès aux entrepôts de données et canaux de transmission sont surveillés.
* Pendant le transfert de données, le risque de violation est évalué et des procédures appropriées sont appliquées pour éliminer ou limiter le plus possible ce risque.

Exigence 7.6 : Le risque que les répondants puissent être identifiés est évalué et géré.

* Il existe un équilibre entre le niveau de risque acceptable d’identification des répondants et la facilité d’utilisation des données.
* Des processus appropriés sont en place pour évaluer le risque de divulgation d’informations sensibles et le risque que des répondants puissent être identifiés à partir de la publication de statistiques ou microdonnées, et des procédures sont appliquées conformément à la politique de diffusion des données afin de limiter le plus possible ce risque.
* Toutes les procédures mises en œuvre pour réduire de manière adéquate le risque d’identification sont correctement documentées et mises à disposition dans les métadonnées liées à l’ensemble de données statistiques.
* Les utilisateurs sont informés que des procédures visant à réduire le risque d’identification ont été mises en œuvre et que de telles procédures pourraient entraîner une perte d’informations.

Exigence 8.5 : Des lignes directrices pour la mise en œuvre de la gestion de la qualité sont définies et mises à la disposition du public.

* Des lignes directrices pour la mise en œuvre de la gestion de la qualité sont produites et publiées, et :
  + Décrivent les principes de qualité et le cadre suivi
  + Décrivent l’ensemble du processus statistique et identifient la documentation pertinente pour chaque étape de la production
  + Décrivent les méthodes de contrôle de la qualité à chaque étape du processus de production statistique
  + Identifient les indicateurs (mesures de qualité) permettant d’évaluer la qualité des principales étapes de la production, notamment les indicateurs pour les données sources
* Les lignes directrices, les manuels méthodologiques et les manuels sur les pratiques recommandées en matière d’assurance de la qualité sont mis à la disposition du public.
* Des mécanismes sont en place pour assurer la qualité de la collecte de données (notamment l’utilisation de données administratives et d’autres sources) et de l’édition des données.

Exigence 10.1 : Les méthodologies appliquées par les organismes statistiques sont conformes aux normes, lignes directrices et bonnes pratiques internationales et sont régulièrement revues et révisées selon les besoins.

* Les structures organisationnelles adéquates sont mises en place pour le développement et l’application de méthodes statistiques solides.
* Des processus d’évaluation et production de rapports sont en place, permettant à la direction de l’organisme statistique de s’assurer que des approches méthodologiques judicieuses ont été adoptées et appliquées tout au long du processus de production.
* Les méthodologies d’enquêtes et l’utilisation de données administratives et d’autres sources de données sont évaluées périodiquement.
* Le plan d’échantillonnage repose sur une méthodologie solide.
* Des procédures de suivi appropriées sont prévues et mises en œuvre en cas de non-réponse.
* Les procédures d’édition statistique et les méthodes d’imputation reposent sur une méthodologie solide.
* Lorsque la modélisation statistique est utilisée dans le processus de production statistique (par exemple, pour la désaisonnalisation), la validité des hypothèses du modèle est soigneusement prise en compte et l’impact sur les estimations finales est évalué.
* Les organismes statistiques examinent les méthodes utilisées par les partenaires externes pour la compilation des données et la production des statistiques.

Exigence 10.3 : Les organismes statistiques choisissent les sources de données en tenant compte de l’exactitude et de la fiabilité, de l’actualité, des coûts, de la charge pour les répondants et d’autres considérations nécessaires.

* L’utilisation de sources de données alternatives, y compris les enquêtes et recensements existants, les données administratives, les mégadonnées ou autres sources de données, est constamment évaluée.
* La qualité doit être évaluée lors de l’utilisation de données administratives91 ou d’autres sources de données. Idéalement, lors de l’utilisation de données administratives, il convient de s’assurer que :
  + La population répond aux exigences en matière de production statistique
  + Les classifications sont appropriées
  + Les concepts sous-jacents sont appropriés
  + Les dossiers sont complets et à jour
  + La couverture géographique est complète et les unités de mesure sont définies/identifiées de manière appropriée
* Lors de l’utilisation d’autres sources de données (comme les mégadonnées), les défis méthodologiques spécifiques tels que ceux liés à la population statistique et la véracité et la volatilité de telles données doivent être pris en compte.

Exigence 10.5 : Les organismes statistiques coopèrent avec la communauté scientifique pour améliorer les méthodes et promouvoir l’innovation dans l’élaboration, la production et la diffusion des statistiques.

* La collaboration avec la communauté scientifique est en place, par exemple dans le cadre de conférences, d’ateliers, de groupes de travail et de formations/cours, afin d’examiner les évolutions méthodologiques et technologiques pertinentes (par exemple concernant l’exploitation de nouvelles sources de données).
* Des accords ont été conclus avec des établissements universitaires en matière de coopération et d’échange de personnel qualifié.
* Le personnel collabore sur des questions méthodologiques avec des collègues au niveau international.
* Une participation et des présentations régulières aux conférences nationales et internationales pertinentes sont encouragées pour l’échange de connaissances et d’expériences.
* Les organismes statistiques organisent des conférences, séminaires, ateliers ou manifestations analogues aux niveaux national et international, avec la participation de la communauté scientifique.

Exigence 11.1 : Les coûts de production de toutes les statistiques officielles sont mesurés et analysés, et des mécanismes sont en place pour assurer le rapport coût-efficacité des activités ou processus statistiques.

* Il existe un système d’enregistrement des coûts et des délais utilisés pour tous les produits statistiques, et l’estimation de la durée des principaux processus devrait être possible.
* Les coûts de la production de statistiques sont bien documentés à chaque étape du processus et sont régulièrement examinés et analysés à travers les produits statistiques afin d’évaluer l’efficacité de leur production.
* Des analyses coûts-avantages sont effectuées pour déterminer les compromis appropriés en termes de qualité des données.
* Le rapport coût-efficacité de chaque enquête statistique est évalué.
* La nécessité de collecter chaque variable d’enquête est justifiée.
* Il existe un processus d’examen continu qui permet de déterminer si un programme particulier fonctionne toujours de la manière la plus rentable pour répondre aux exigences énoncées.
* Les instruments de collecte de données sont conçus pour minimiser les coûts et les délais de codage et d’édition

Exigence 11.2 : Des procédures existent pour évaluer et justifier les demandes de nouvelles statistiques par rapport à leur coût.

* Les demandes de nouvelles statistiques sont régulièrement enregistrées et évaluées par les experts en statistiques sur les approches méthodologiques proposées et les coûts associés, et sont partagées avec la direction en fonction des informations fournies par les utilisateurs et en coopération avec d’autres parties prenantes.
* Avant d’envisager une nouvelle collecte de données, il existe des mécanismes permettant de vérifier si les sources de données déjà disponibles peuvent être utilisées avec un impact minimal sur leur objectif et leur qualité.
* Lors de l’introduction de nouvelles statistiques, une analyse coûts-avantages est réalisée.

Exigence 11.5 : Des efforts proactifs sont déployés pour améliorer le potentiel statistique des données administratives et d’autres sources de données.

* Les organismes statistiques contribuent au processus législatif pour obtenir et conserver un accès aux sources de données administratives et autres sources de données à des fins statistiques, le cas échéant.
* Des accords appropriés (accords de prestation de services ou législation nationale, par exemple) avec les propriétaires ou détenteurs de données administratives et d’autres collectes de données sont conclus et mis à jour en fonction des besoins, en précisant l’accès aux données et aux métadonnées, ainsi que d’autres aspects pertinents.
* Une évaluation des sources de données administratives possibles est effectuée avant le lancement de toute nouvelle enquête.
* Les méthodes d’appariement et d’intégration des données sont appliquées de manière proactive tout en garantissant la sécurité et la confidentialité des données.
* Les rapports sur la qualité des données administratives et autres sources de données utilisées pour les statistiques officielles sont élaborés par l’organisme statistique responsable en coopération avec le ou les propriétaires des données.

Exigence 12.1 : Les processus statistiques sont testés avant leur mise en œuvre.

* La stratégie de test est développée dans le cadre de la phase de conception du modèle générique des processus de production statistique.
* Les procédures de saisie des données et les outils et instruments de collecte des données tels que les systèmes électroniques sont testés pour garantir la simplicité et une intrusion minimale dans la vie privée, et sont ajustés si nécessaire avant leur mise en œuvre.
* Les questionnaires d’enquête sont testés à l’aide de méthodes appropriées (enquête pilote, groupes de discussion, etc.)
* Les systèmes de collecte de données administratives et autres sources de données sont testés avant d’être utilisés.
* Le traitement des données et les procédures de traitement des données sont testés et ajustés, le cas échéant, avant leur application effective.
* Les résultats des tests sont pris en compte dans la mise en œuvre du processus de production et sont approuvés.
* Dans le cas de l’intégration de données provenant d’une ou de plusieurs sources, la qualité des procédures de couplage est testée.

Exigence 12.2 : Les processus statistiques sont bien établis et régulièrement contrôlés et révisés selon les besoins.

* Les organismes statistiques disposent de procédures et de lignes directrices documentées contenant des recommandations sur les méthodologies appropriées aux différentes étapes du processus de production statistique.
* La documentation des processus de production devrait suivre le modèle GSBPM.
* Une politique d’archivage des données et statistiques est en place et suivie.
* Les procédures statistiques utilisent des techniques statistiques reconnues sur le plan international.
* Les données de toutes sources sont examinées et validées pour identifier les problèmes, erreurs et divergences potentiels comme les valeurs aberrantes, les données manquantes et les erreurs de codage.
* Lorsque le codage est effectué selon un processus automatisé, une équipe de codeurs bien formés est chargée de vérifier le codage automatisé et de traiter les cas non codés.
* Les effets de l’édition et de l’imputation des données sont analysés dans le cadre de l’évaluation de la qualité de la collecte de données.
* Toutes les bases de données statistiques sont conçues et organisées de manière à permettre et à faciliter le couplage des données, en utilisant des identifiants uniques pour les unités statistiques, le cas échéant, tout en garantissant la sécurité et la confidentialité des données.

Exigence 12.3 : Des procédures sont en place pour utiliser efficacement les sources de données administratives et autres sources de données à des fins statistiques.

* Les organismes statistiques utilisent des outils et lignes directrices pour évaluer la qualité des données de sources administratives et autres sources.
* Des processus et progiciels appropriés pour la collecte, le traitement et l’analyse des données de sources administratives et autres sources sont développés et mis en œuvre.
* Les propriétaires ou détenteurs des données de sources administratives et autres sources informent les organismes statistiques de tout changement dans le processus de production des données.
* Les métadonnées relatives aux sources de données administratives ou autres sources à la disposition des organismes statistiques, notamment les concepts et définitions, les classifications, la couverture par rapport à la population cible et d’autres aspects qualitatifs.
* Il existe une documentation décrivant comment les données provenant de sources de données administratives et autres sources répondent aux exigences statistiques en termes de définitions, de concepts et de couverture, entre autres.

Exigence 12.5 : Les métadonnées et la documentation des méthodes et des différents processus statistiques sont gérées tout au long des processus et partagées comme il convient.

* Il existe une politique sur la documentation des métadonnées liée aux processus de production statistique.
* Les politiques et normes de révision et de mise à jour des métadonnées sont suivies.
* Les travaux de préparation des statistiques et des métadonnées associées devraient être effectués en parallèle.
* Les métadonnées sont saisies tout au long du processus opérationnel suivant le GSBPM et stockées dans un système de gestion de métadonnées.
* Les méthodes et processus statistiques sont documentés de manière à permettre de recréer l’ensemble du processus de production statistique.

Exigence 13.4 : Le partage de données, le couplage de données et l’utilisation de sources de données administratives et autres sources de données sont encouragés afin de limiter au minimum la charge pour les répondants.

* La documentation des données déjà disponibles dans le SSN, notamment les données archivées, existe et est partagée.
* Il existe des procédures et des outils techniques pour le partage et le couplage de données au sein du SSN (accords formels, services Web, bases de données communes, par exemple).
* Les entrepôts de données sont partagés entre les organismes statistiques pour la production de statistiques officielles et conformément aux politiques de confidentialité.
* Des informations sur la qualité des données à coupler existent (par exemple, sur la couverture et les possibilités de couplage).
* L’utilisation des données administratives et autres données comme alternative aux données d’enquête pour la production de statistiques officielles est encouragée dans tout le SSN

Exigence 14.1 : Des procédures sont en place pour identifier les utilisateurs et leurs besoins et pour les consulter sur le contenu du programme de travail statistique.

* Il existe une législation ou autre disposition officielle prévoyant l’obligation de consulter les principaux utilisateurs de statistiques.
* Des processus de consultation structurés et périodiques (tels que conseils consultatifs, comités ou groupes de travail) avec les principales parties prenantes et les utilisateurs sont en place pour examiner le contenu du programme statistique et l’utilité des statistiques existantes, ainsi que pour identifier les nouveaux besoins en statistiques.
* Le retour d’information des utilisateurs par un service d’assistance aux utilisateur, d’un centre ou d’une ligne d’assistance sont analysés pour comprendre et identifier les besoins des utilisateurs.
* Les données sur l’utilisation des statistiques (par exemple, analyse Web, nombre et types de téléchargements, abonnés aux rapports) sont collectées et analysées pour améliorer les résultats statistiques.

Exigence 14.3 : Des statistiques basées sur des sources de données nouvelles et existantes sont élaborées pour répondre aux nouveaux besoins en informations de la société.

* Un laboratoire d’innovation est établi pour examiner et expérimenter de nouvelles sources de données afin de répondre aux nouveaux besoins en informations.
* Une coopération avec la communauté scientifique et les propriétaires ou détenteurs de nouvelles sources de données est établie pour expérimenter et exploiter ces sources de données.
* Les possibilités d’exploitation de nouvelles sources de données sont régulièrement examinées par la direction.

Exigence 14.4 : La satisfaction des utilisateurs est régulièrement mesurée et fait l’objet d’un suivi systématique.

* Des enquêtes de satisfaction et des études sur les utilisateurs sont régulièrement réalisées et analysées.
* Les actions d’amélioration découlant des enquêtes de satisfaction des utilisateurs et des études sur les utilisateurs sont identifiées et mises en œuvre.
* Les enquêtes de satisfaction des utilisateurs incluent des questions sur leur opinion concernant la disponibilité des métadonnées.
* Des mesures sont en place pour évaluer la satisfaction des principaux utilisateurs à l’égard de produits particuliers (par exemple, des enquêtes et des indicateurs spécifiques de la satisfaction des utilisateurs, notamment les délais, etc., au niveau des produits).

Exigence 15.1 : Les données sources, les données intégrées, les résultats intermédiaires et les produits statistiques sont régulièrement évalués et validés.

* Des systèmes d’évaluation et de validation des données sources, des données intégrées, des résultats intermédiaires et des produits statistiques sont développés et gérés.
* Les données sont systématiquement vérifiées et comparées aux données provenant d’autres sources et au fil du temps.
* Les produits statistiques sont comparés à d’autres informations existantes afin de s’assurer de leur validité.

Exigence 15.3 : Des études et analyses de révisions sont effectuées et utilisées pour améliorer les sources de données, les processus statistiques et les résultats.

* Les données et statistiques préliminaires et révisées sont clairement identifiées.
* Des explications sur le calendrier, les raisons et la nature des révisions sont disponibles.
* La politique de révision suit des procédures standard et transparentes.
* Les informations sur l’ampleur et l’orientation des révisions des indicateurs clés sont utilisées pour améliorer les processus statistiques.
* Les informations sur l’ampleur et l’orientation des révisions des indicateurs clés sont fournies et rendues publiques.

Exigence 16.2 : La relation avec les fournisseurs de données est gérée en tenant compte des besoins d’actualité et de ponctualité.

* Des accords sont signés avec les fournisseurs de données sur les dates prévues et le format de livraison.
* Des procédures sont en place pour assurer la circulation efficace et opportune des données des fournisseurs vers les organismes statistiques.
* Des procédures de suivi sont en place pour garantir la réception des données des fournisseurs en temps indiqué.

Exigence 17.4 : L’accès aux microdonnées est autorisé à des fins de recherche, sous réserve des règles et protocoles spécifiques en matière de confidentialité des statistiques qui sont publiés sur le site Web de l’organisme statistique.

* L’organisme statistique contrôle ou surveille l’accès des chercheurs aux microdonnées en fournissant celles-ci dans un environnement sécurisé.
* Les chercheurs sont régulièrement consultés sur l’efficacité des modalités d’accès aux microdonnées.
* Des installations d’accès à distance sont disponibles pour accéder aux microdonnées, avec les contrôles appropriés.

Exigence 18.2 : Des procédures ou lignes directrices sont en place pour assurer et surveiller la cohérence et l’homogénéité internes, intrasectorielles et intersectorielles.

* Les statistiques dérivées de différentes sources ou ayant des périodicités différentes (par exemple, mensuelle, trimestrielle, annuelle) sont comparées et toute différence est expliquée et réconciliée, le cas échéant.
* La coopération et l’échange de connaissances entre les différents programmes et domaines statistiques sont encouragés.
* Des procédures et des lignes directrices spécifiques aux processus sont disponibles pour garantir la cohérence interne des produits.
* Avant de commencer de nouvelles statistiques ou de nouveaux programmes statistiques, l’analyse de la relation conceptuelle et méthodologique avec les statistiques existantes est faite.
* Les produits statistiques sont comparés à ceux d’autres sources statistiques ou administratives fournissant des informations identiques ou similaires sur le même sujet, et les divergences sont identifiées et expliquées aux utilisateurs.
* Des procédures ou lignes directrices internes sont élaborées afin d’assurer et de contrôler la cohérence et l’homogénéité internes.
* Des procédures et des lignes directrices sont élaborées afin que les résultats de différentes sources puissent être combinés. La conformité est évaluée périodiquement.

Exigence 19.2 : Les métadonnées sont documentées, archivées et diffusées conformément aux normes reconnues au niveau international.

* Les normes internationales, régionales, nationales ou internes sont utilisées pour la documentation, la gestion et l’archivage des métadonnées.
* Des procédures sont en place pour garantir que les métadonnées sont documentées conformément aux systèmes de métadonnées normalisés et régulièrement mises à jour.
* Les métadonnées sont mises à disposition en même temps que les données et statistiques auxquelles elles se rapportent.
* La diffusion des métadonnées est adaptée aux différents besoins, tels que ceux des producteurs et utilisateurs de statistiques.
* Un moyen systématique d’archivage des métadonnées est disponible, garantissant également que les métadonnées sont accessibles pour une réutilisation ultérieure.
* Un glossaire des concepts statistiques est disponible pour le public.

Annexe 4: Cartographie des dix exigences essentielles pour l’assurance qualité lors de l’utilisation des données administratives et autres sources de données avec l’approche conceptuelle et les exigences pertinentes du CNAQ des Nations Unies

**Le tableau 3** montre les liens existants entre les dix exigences essentielles pour l’assurance qualité (AQ) lors de l’utilisation des données administratives et autres sources de données (AOS) et les exigences pertinentes du CNAQ des Nations Unies, selon l’approche conceptuelle. Le tableau montre ce qui est déjà couvert par le CNAQ des Nations unies et contribue ainsi d’éviter les doubles emplois. Il montre également des lacunes dans le CNAQ des Nations Unies en ce qui concerne l’AQ pour les données administratives et autres sources de données et peut donc être utilisé pour mettre à jour le CNAQ des Nations Unies et les cadres nationaux respectifs. En outre, le tableau montre comment les dix exigences essentielles se rapportent aux actions et aux considérations générales de l’approche conceptuelle.

**Tableau 3:** Cartographie des dix exigences essentielles pour l’assurance qualité lors de l’utilisation des données administratives et autres sources de données et des exigences pertinentes du CNAQ des Nations Unies selon l’approche conceptuelle

| Exigence essentielles pour l’assurance qualité des AOS | Exigence pertinente du CNAQ des Nations Unies |
| --- | --- |
| Conditions préalables importantes: Accès aux données et confidentialité | |
| CR 1: L’utilisation des données administratives et autres sources de données doit répondre à un ensemble de conditions préalables et tenir compte des besoins des utilisateurs. | Exigences 2.5, 2.6  Exigences 7.1-7.6  Exigence 14.1, 14.3 |
| Action 1: Identification des besoins en statistiques et sélection de la source de données | |
| CR 2: De nouvelles sources de données, des fournisseurs de données ainsi que l’utilisation de sources de données multiples sont explorés de manière proactive | Exigence 10.3 et 10.5  Exigences 11.1, 11.2, 11.5  Exigence 14.1, 14.3  Exigence 5.3 |
| CR 3: Il y a des informations de base sur le fournisseur de données et des informations générales sur la source de données. | Exigence 10.3 |
| CR 4: Le fournisseur de données et la source de données sont évalués en fonction des risques qu’ils présentent | Exigence 10.3  Exigences 5.2, 5.3  Exigence 7.5 |
| CR 7: La qualité des données de base est systématiquement évaluée par l’organisme statistique. | Exigence 10.3  Exigence 12.3  Exigence 15.1 |
| CR 8: Il existe des métadonnées complètes sur les données de base | Exigence 10.3  Exigence 12.3, 12,5 |
| Action 2: Coopération avec le fournisseur de données | |
| CR 4: Le fournisseur des données et la source de données sont évalués en fonction des risques qu’ils présentent | Exigence 10.3  Exigence 7.5 |
| CR 5: Il existe des accords de coopération avec les fournisseurs de données et une coopération permanente, le cas échéant. | Exigences 2.5, 2.6, 2.7  Exigence 3.2  Exigence 16.2 |
| CR 6: Le fournisseur de données assure la qualité des données et produit un rapport de qualité (ou une déclaration de qualité) en coopération avec l’organisme statistique, le cas échéant. | Exigence 2.7  Exigence 10.3  Exigence 11.5 |
| Action 3: Acquisition et traitement des données | |
| CR 7: La qualité des données de base est systématiquement évaluée par l’organisme statistique. | Exigence 10.3  Exigence 12.3  Exigence 15.1 |
| CR 8: Il existe des métadonnées complètes sur les données de base | Exigence 10.3  Exigence 12.3, 12,5 |
| CR 9: Le traitement des données de base par l’organisme statistique respecte les normes, les lignes directrices et les meilleures pratiques | Exigences 12.1, 12.2, 12.3  Exigence 8.5  Exigences 10.1, 10.5  Exigence 11.5  Exigence 13.4  Exigence 15.3  Exigence 18.2 |
| Action 4: Diffusion | |
| CR 8: Il existe des métadonnées complètes sur les données de base | Exigence 12.3, 12,5 |
| CR10: La diffusion des résultats statistiques répond aux besoins des utilisateurs | Exigence 14.1, 14.3  Exigence 15.1  Exigence 17.4  Exigence 19.2 |
| Considération générale 1: Besoins des utilisateurs en matière de statistiques pertinentes\* | |
| CR 1: L’utilisation des données administratives et autres sources de données doit répondre à un ensemble de conditions préalables et tenir compte des besoins des utilisateurs. | Exigence 14.1, 14.3 |
| CR10: La diffusion des résultats statistiques répond aux besoins des utilisateurs | Exigence 14.1, 14.3  Exigence 17.4  Exigence 19.2 |
| Considération générale 2: Qualité des données d’entrée et des métadonnées\* | |
| CR 6: Le fournisseur des données assure la qualité de ses données et produit un rapport de qualité (ou une déclaration de qualité) en coopération avec l’organisme statistique, le cas échéant. | Exigence 2.7 |
| CR 7: La qualité des données de base est systématiquement évaluée par l’organisme statistique. | Exigence 10.3  Exigence 12.3  Exigence 15.1 |
| CR 8: Il existe des métadonnées complètes sur les données de base | Exigence 10.3  Exigence 12.3, 12,5 |
| CR 9: 9. Le traitement des données de base par l’organisme statistique respecte les normes, les lignes directrices et les meilleures pratiques | Exigences 12.1, 12.2, 12.3  Exigence 8.5  Exigences 10.1, 10.5  Exigence 11.5  Exigence 13.4  Exigence 15.3  Exigence 18.2 |
| CR10: La diffusion des résultats statistiques répond aux besoins des utilisateurs | Exigence 14.1, 14.3  Exigence 17.4  Exigence 19.2 |

\* Toutes les exigences du CNAQ de l’ONU qui peuvent être allignées à une exigence critique spécifique ne s’appliquent pas aux considérations générales.

\*\*\*\*\*

Current version: <https://forms.gle/G1y1s4GLbFytK5v16>.

**Consultation Globale sur un projet de Module d’Assurance Qualité lors de l’utilisation de Sources de Données Administratives et Autres pour produire des Statistiques Officielles**

Veuillez répondre aux questions suivantes avant le 4 juin 2024

[\*Obligatoire]

Nom :\*

Organisation :\*

Pays :\*

Courrier électronique  :\*

***Veuillez fournir votre réponse ici : ..***

1. Veuillez fournir vos commentaires ou suggestions sur l'Introduction du module (paras. 1-8)
2. Veuillez fournir vos commentaires ou suggestions sur la Partie 1 (paras. 9-15)
3. Veuillez fournir vos commentaires ou suggestions sur la Partie 2 (paras. 16-18)
4. Veuillez fournir vos commentaires ou suggestions sur la Partie 2 – Exigences Critiques 1 et 2
5. Veuillez fournir vos commentaires ou suggestions sur la Partie 2 – Exigences Critiques 3 et 4
6. Veuillez fournir vos commentaires ou suggestions sur la Partie 2 – Exigences Critiques 5 et 6
7. Veuillez fournir vos commentaires ou suggestions sur la Partie 2 – Exigences Critiques 7 et 8
8. Veuillez fournir vos commentaires ou suggestions sur la Partie 2 – Exigences Critiques 9 et 10
9. Veuillez fournir vos commentaires ou suggestions sur l'Annexe 1 – Sous-module et liste supplémentaire
10. Veuillez fournir vos commentaires ou suggestions sur l'Annexe 2 – Glossaire
11. Veuillez fournir vos commentaires ou suggestions sur l'Annexe 3 – Exigence pertinentes du CNAQ des Nations unies
12. Veuillez fournir vos commentaires ou suggestions sur l'Annexe 4 – Tableau de correspondance
13. Veuillez indiquer si vous êtes intéressé à tester le projet de Module : OUI/NON
14. Veuillez nous faire part de vos suggestions pour améliorer ce projet de Module\*
15. Veuillez nous faire part d’autres commentaires que vous pourriez avoir, ou si vous avez des pratiques spécifiques ou des expériences que vous pouvez partager\*

1. Voir https://unstats.un.org/unsd/methodology/dataquality/about/. [↑](#footnote-ref-2)
2. Le Manuel des cadres nationaux d’assurance de la qualité des statistiques officielles de l’ONU (ci-après dénommé *Manuel*) contient une liste d’autres sources de données au paragraphe 7.6. [↑](#footnote-ref-3)
3. Les pays qui n’ont pas de CNAQ peuvent se référer à la feuille de route élaborée par le GE-CNAQ, qui fournit des orientations détaillées sur l’élaboration et la mise en œuvre d’un CNAQ, voir <https://unstats.un.org/unsd/methodology/dataquality/roadmap/>. [↑](#footnote-ref-4)
4. Voir la définition de «agence statistique» dans le glossaire de l’annexe 2. [↑](#footnote-ref-5)
5. Veuillez consulter le répertoire des pratiques nationales, des lignes directrices, des cadres et des outils disponibles sur l’assurance de la qualité lors de l’utilisation de sources de données administratives et autres, disponible à l’adresse suivante: <https://unstats.un.org/wiki/pages/viewpage.action?pageId=224264259>. [↑](#footnote-ref-6)
6. Par exemple, aucune distinction n’est faite quant à savoir si une source de données est explorée pour la première fois ou déjà acquise et utilisée régulièrement. [↑](#footnote-ref-7)
7. Voir le mandat et la composition du sous-groupe, disponible à l’adresse suivante: <https://unstats.un.org/wiki/display/EGNQAFSA/EG-NQAF+Subgroup+on+admin+and+other+data+sources>. Le mandat du Sous-Groupe comprend également la question de l’utilisation de sources de données multiples qui ne sont toutefois pas couvertes par le présent module et qui seront examinées séparément. [↑](#footnote-ref-8)
8. Sur la base du *manuel,* disponible à l’adresse suivante : <https://unstats.un.org/UNSDWebsite/data-quality/user-manual>. [↑](#footnote-ref-9)
9. Ces prescriptions du CNAQ des Nations unies sont énumérées à l’annexe 3 par souci de commodité. [↑](#footnote-ref-10)
10. Il est possible de mettre en évidence les sous-processus du GSBPM qui sont particulièrement concernés pour assurer la qualité des statistiques lors de l’utilisation des données administratives et autres sources de données. Par exemple, les sous-processus 1.1 Identifier les besoins, 1.2 Consulter et confirmer les besoins, 1.3 Établir des objectifs de réalisation et 1.4 Identifier les concepts correspondent à l’Action 1 de l’approche conceptuelle qui concerne la sélection de la source de données en fonction des besoins des utilisateurs. Toutefois, l’assurance qualité est un processus global dans le cadre du GSBPM et il n’existe actuellement pas de correspondance détaillée entre l’assurance qualité et les sous-processus GSBPM. [↑](#footnote-ref-11)
11. Voir la définition de «fournisseur de données» dans le glossaire de l’annexe 2. Les fournisseurs de données sont censés fournir des données de niveau de l’unité d’observation et des données agrégées, selon le cas. [↑](#footnote-ref-12)
12. <https://unstats.un.org/wiki/pages/viewpage.action?pageId=224264259>. [↑](#footnote-ref-13)
13. L’annexe 4 présente une cartographie des dix exigences essentielles en matière d’assurance qualité lors de l’utilisation des données administratives et autres sources de données avec les exigences pertinentes du CNAQ des Nations Unies, conformément à l’approche conceptuelle. Elle permet de voir ce qui est déjà couvert dans le CNAQ des Nations Unies et où il y a des lacunes. [↑](#footnote-ref-14)
14. L’intégration des données peut être réalisée, par exemple, par l’utilisation des méthodes de couplage des observations. Voir Manuel, par. 7.18. pour plus d’informations et de références. [↑](#footnote-ref-15)
15. Un format standard pour les métadonnées de données base peut être élaboré sur la base des normes connexes existantes, telles que la norme du système statistique européen sur les métadonnées de référence et les rapports de qualité des processus et les produits statistiques; voir le manuel du Système Statistique Européen (SSE) pour les rapports sur la qualité et les métadonnées — réédition 2021, disponible à l’adresse suivante: <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-21-021>. [↑](#footnote-ref-16)
16. La liste des indicateurs proposés est déjà largement utilisée pour les données administratives, et davantage d’expérience et de rétroaction sont nécessaires pour mieux évaluer leur applicabilité à d’autres données et pour élaborer des indicateurs supplémentaires, si nécessaire. [↑](#footnote-ref-17)
17. Pour de plus amples informations sur la mise en œuvre des validations automatisées, veuillez consulter les travaux du groupe de travail du Secrétariat SDMX pour le langage de validation et de transformation (VTL) qui a élaboré un langage standard pour définir les règles de validation et de transformation, voir <https://sdmx.org/?page_id=5096>. [↑](#footnote-ref-18)
18. Cependant, certains organismes statistiques font une distinction entre les deux termes, la cohérence se référant à l’utilisation de normes et l’homogénéité se référant à la logique interne d’un ensemble de données. [↑](#footnote-ref-19)