Nations Unies ECE/CES/2014/2

Conseil économique et social

Distr. générale
28 janvier 2014
Français
Original: anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

Soixante-deuxième réunion plénière
Paris, 9-11 avril 2014
Point 5 de l'ordre du jour provisoire
Travaux du Groupe de haut niveau sur la modernisation de la production et des services statistiques

Modèle générique d'informations statistiques (GSIM): Communication à l'intention de tous les professionnels de la statistique

Document établi par le Groupe de haut niveau sur la modernisation de la production et des services statistiques

Résumé

Le présent document expose le modèle générique d'informations statistiques à l'intention des personnels des organismes de statistique. Il décrit brièvement les avantages offerts par le modèle et les incidences que son adoption peut avoir pour ces personnels. Il examine aussi les liens qui unissent ce modèle à d'autres cadres et normes comme le modèle générique du processus de production statistique, la Data Documentation Initiative et la norme SDMX (échange de données et de métadonnées statistiques). Des renseignements plus détaillés sont disponibles dans la spécification du GSIM dont des extraits sont reproduits en annexe et sur la page wiki[1] connexe qui pourrait être utile aux organismes qui adoptent ce modèle comme norme interne.

Le présent document est présenté à la Conférence des statisticiens européens pour information.

[1] http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model.

GE.14-20417 (F) 170314 180314

Merci de recycler

Table des matières

		Page
I.	Introduction	3
II.	Champ d'application	4
III.	Qu'est-ce que le modèle générique d'informations statistiques?	4
IV.	Avantages du modèle générique d'informations statistiques pour l'organisme dans son ensemble	7
	A. Avantages à long terme	7
	B. Avantages immédiats	7
V.	Le modèle générique d'informations statistiques et le modèle générique du processus de production statistique	8
VI.	À quoi le modèle peut-il me servir?	9
	A. Le GSIM vu sous l'angle de l'activité statistique	9
	B. Le GSIM, vu sous l'angle des technologies de l'information	11
VII.	Norme SDMX (échange de données et de métadonnées statistiques), Data Documentation Initiative et autres normes	12
Annexe		
	Extraits de la spécification du modèle générique d'informations statistiques	14
	A. L'information dans le processus de production statistique	14
	B. Information fondamentale	26

I. Introduction

1. Partout dans le monde, les organismes de statistique mènent des activités de même nature mais leurs méthodes peuvent varier. Toutes ces activités exploitent et produisent des informations analogues (par exemple, tous les organismes utilisent des classifications, créent des ensembles de données et diffusent de l'information). Bien que les informations utilisées par les organismes de statistique soient fondamentalement les mêmes, chacun opère un peu différemment (avec souvent des variations d'un service à l'autre). Par le passé, il n'existait pas de mode de description commun des informations utilisées. Cela ne facilitait pas la communication en interne et entre organismes de statistique et il n'existait donc pas de cadre de référence pour une collaboration approfondie, un effort de normalisation ou le partage d'outils et de méthodes.

2. Le modèle générique d'informations statistiques (GSIM) est le premier cadre de référence pour l'information statistique approuvé à l'échelon international. Ce cadre conceptuel global jouera un rôle important dans la modernisation, la simplification et l'harmonisation des normes et de la production associées aux statistiques officielles tant au niveau national qu'international.

3. Le GSIM constitue un cadre de référence d'objets d'information qui permet de faire une description générale de la définition, de la gestion et de l'utilisation des données et des métadonnées tout au long du processus de production de statistiques. Il offre un ensemble d'objets d'information normalisés, décrits de façon uniforme, qui sont les intrants et les produits dans le processus de conception et de production de statistiques. En tant que cadre de référence, le GSIM aide à mieux comprendre les liens importants qui existent entre les entités prenant part à la production de statistiques, et peut servir à encadrer l'élaboration et l'utilisation de normes ou de spécifications cohérentes en matière de mise en œuvre.

4. Le GSIM est l'une des pièces maîtresses de la modernisation des statistiques officielles et de l'abandon progressif du cloisonnement par domaine statistique. C'est un élément essentiel de la vision stratégique établie par le Groupe de haut niveau sur la modernisation de la production et des services statistiques et approuvée par la Conférence des statisticiens européens[2].

5. Il est nécessaire de moderniser la production statistique pour que les organismes de statistique restent utiles et adaptables dans un environnement de l'information dynamique et concurrentiel. On espère que ces organismes adopteront et mettront en œuvre le GSIM et le langage commun qu'il propose. Il reste qu'un modèle à lui seul ne peut pas transformer un organisme ou ses processus. Pour répondre aux besoins futurs des organismes de statistique, le GSIM est conçu pour permettre l'adoption d'approches novatrices de la production statistique dans toute la mesure possible. C'est l'un des principaux fondements de l'architecture commune de la production statistique[3], une initiative conjointe visant à concevoir des services communs et interchangeables dotés d'interfaces normalisées pour faciliter la normalisation et la modernisation. Le GSIM soutient dans le même temps les modes de production statistique en vigueur.

6. Le présent document donne une description sommaire du GSIM en résumant ses principales caractéristiques à l'intention d'un public de statisticiens relativement général. Pour plus de détails techniques, veuillez consulter la spécification et les documents connexes qui sont disponibles sur le site Web de la CEE[4].

---

[2] Voir l'adresse: www1.unece.org/stat/platform/display/hlgbas.

[3] Voir l'adresse: http://www1.unece.org/stat/platform/display/CSPA.

[4] Voir l'adresse: http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model+(GSIM).

II. Champ d'application

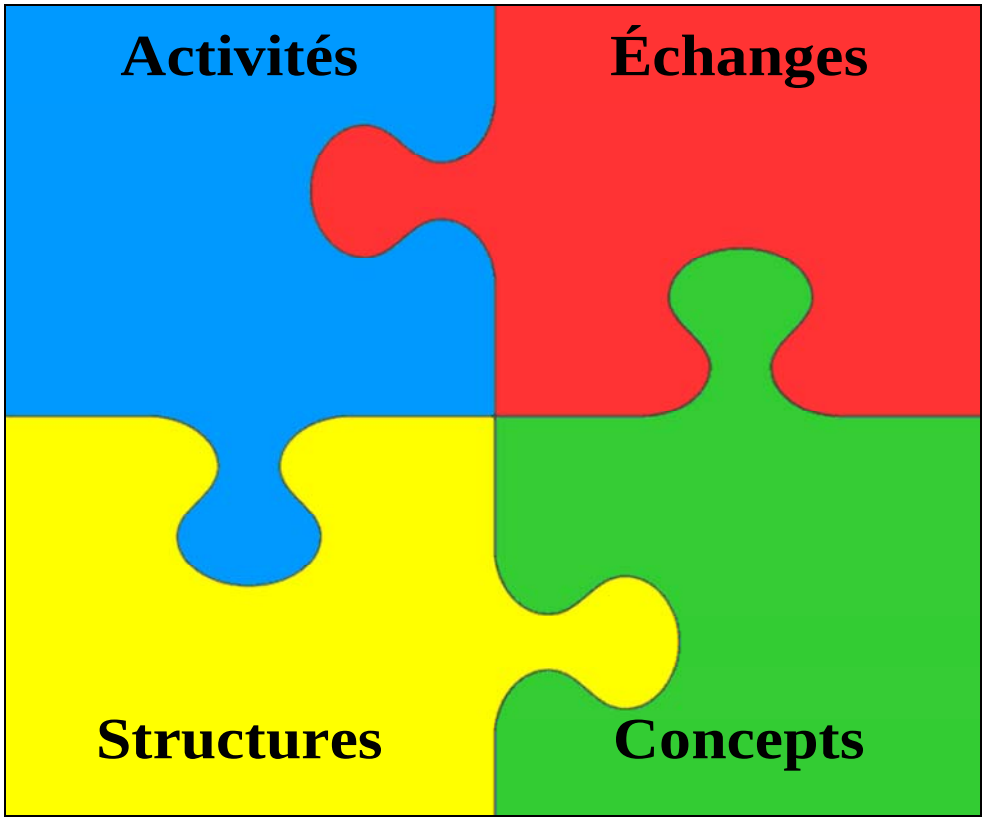
7. Le GSIM offre un cadre de référence pour les objets d'information à l'appui de tous les processus de production statistique tels que ceux décrits dans le modèle générique du processus de production statistique (GSBPM)[5], en donnant aux objets d'information des noms adoptés d'un commun accord, en les définissant, en précisant leurs propriétés essentielles et en indiquant les relations qu'ils entretiennent entre eux. Il n'émet pas toutefois de postulats, quant aux normes ou aux technologies utilisées pour son application.

8. Le GSIM ne recouvre pas les objets d'information en rapport avec les fonctions de gestion d'un organisme comme les ressources humaines, les finances ou les fonctions juridiques, sauf si ces informations sont directement utilisées pour la production statistique.

III. Qu'est-ce que le modèle générique d'informations statistiques?

9. Le GSIM contient des objets qui définissent l'information sur le monde réel − les «objets d'information». On peut citer en exemple les données et les métadonnées (comme les classifications) ainsi que les règles et paramètres nécessaires au déroulement des processus de production (par exemple, les règles d'édition des données). Le GSIM identifie environ 110 objets d'information classés dans quatre groupes de haut niveau et expliqués plus en détail dans les spécifications.

Figure 1
Groupes d'objets d'information de haut niveau du GSIM



[5] Voir l'adresse: www.unece.org/stats/gsbpm, document number (ECE/CES/2014/1).

10. Les quatre groupes de haut niveau sont décrits ci-après:

a) Le *groupe des activités* sert à appréhender les schémas et les plans des programmes statistiques, ainsi que les processus engagés pour exécuter ces programmes. Il permet notamment d'identifier les besoins statistiques, les processus d'activité qui constituent le programme statistique et les évaluations qui en sont faites;

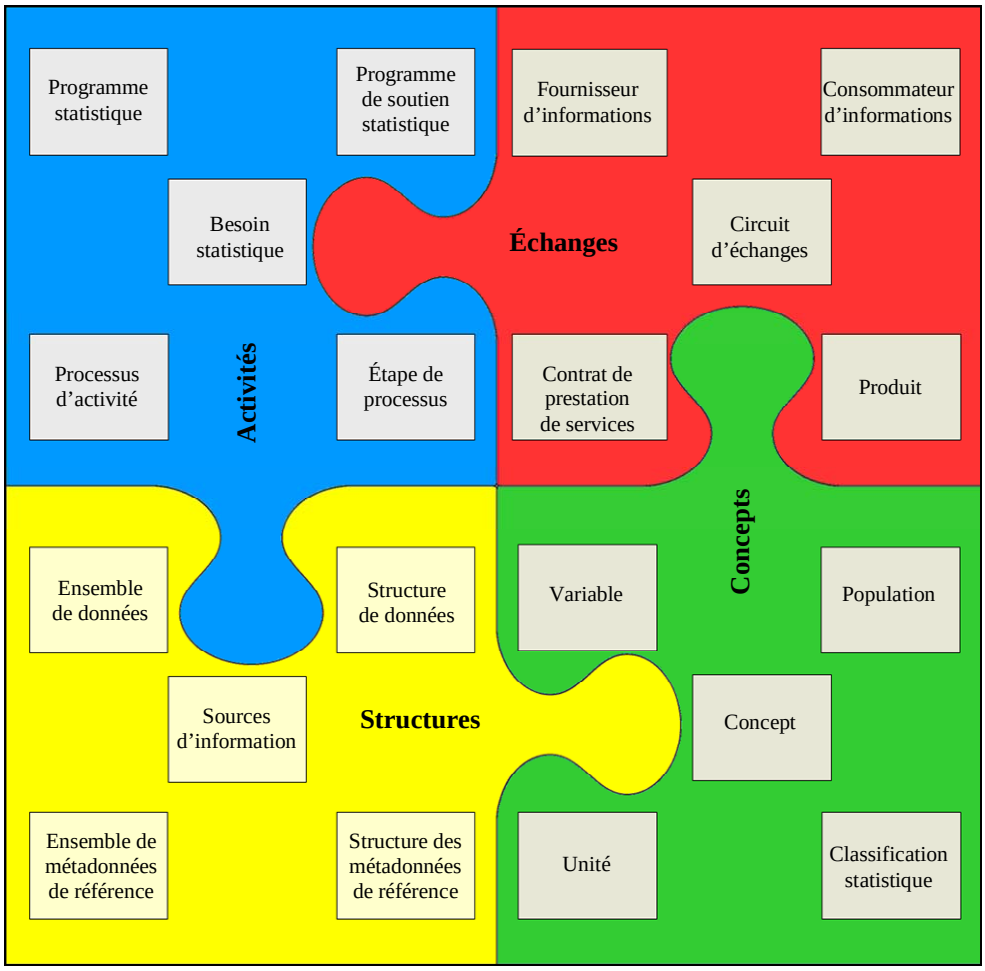
b) Le *groupe des échanges* sert à répertorier les informations qui entrent et sortent d'un organisme de statistique par le biais de circuits d'échanges. Il comprend les objets qui décrivent la collecte et la diffusion de l'information;

c) Le *groupe des concepts* sert à définir ce que les données veulent dire, en donnant à comprendre ce qu'elles mesurent;

d) Le *groupe des structures* sert à décrire et à définir les termes utilisés en rapport avec l'information et sa structure.

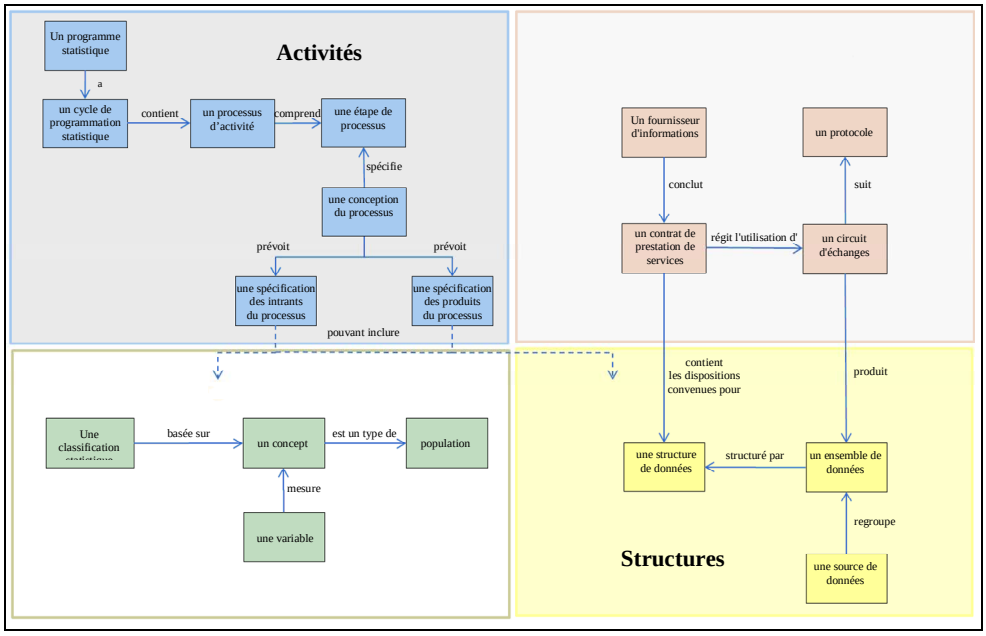
11. La figure 2 présente un schéma simplifié des objets d'information identifiés dans le GSIM. Elle donne aux utilisateurs des exemples d'objets appartenant à chacun des quatre groupes de haut niveau.

Figure 2
Schéma simplifié des objets d'information du GSIM



12. La figure 3 donne un autre aperçu de l'une des parties du GSIM. C'est une illustration un peu plus technique mais qui se veut tout de même compréhensible par un public relativement large. Les figures 2 et 3 peuvent l'une et l'autre être utilisées pour communiquer avec des utilisateurs souhaitant avoir des exemples d'objets et de relations figurant dans le GSIM.

Figure 3
Autre schéma simplifié des objets d'information du GSIM



13. La figure 3 donne un exemple d'objets d'information du GSIM illustrant certaines des informations importantes pour un organisme de statistique. Les objets d'information du GSIM figurent en italiques.

> «Un organisme de statistique lance un *programme statistique*. Ce dernier se rapporte à une activité en cours comme une enquête ou une série de produits statistiques et s'accompagne d'un *cycle de programmation statistique* (par exemple, il se répète à un rythme trimestriel ou annuel).
>
> Le cycle de programmation statistique contient plusieurs processus d'activité. Ces processus comprennent plusieurs étapes de processus spécifiées par la conception du processus. La conception de processus prévoit une spécification des intrants du processus et une spécification des produits du processus. Les spécifications sont souvent des éléments d'information qui se rapportent à des concepts et à des structures (par exemple, classification statistique, variable, population, structure des données et ensemble de données).
>
> Si, par exemple, le *processus d'activité* se rapporte à une collecte de données, il y aura un *fournisseur d'informations* qui convient de fournir des données à l'organisme de statistique (par le biais d'un *contrat de prestation de services*). Ce *contrat de prestation* contient les dispositions concernant une *structure de données* qui a été agréée et régit le *circuit d'échanges* utilisé pour les informations entrantes. Le *circuit d'échanges* peut être un *questionnaire* ou un *registre administratif*. L'information sera reçue par le biais d'un mécanisme particulier (*protocole*) comme un entretien ou un échange de fichiers de données.

L'*ensemble de données* produit par le *circuit d'échanges* sera stocké dans une *source de données* et organisé suivant une *structure de données*.»

IV. Avantages du modèle générique d'informations statistiques pour l'organisme dans son ensemble

14. Le GSIM est conçu pour être utilisé par les organismes à divers degrés. Dans certains cas, il peut seulement servir de modèle auquel les organismes se réfèrent pour mieux se comprendre lorsqu'ils communiquent en interne ou avec des homologues. Dans d'autres cas, un organisme peut choisir d'appliquer le GSIM comme modèle d'information lui permettant de définir son environnement opérationnel. Il existe plusieurs scénarios valables pour l'utilisation du GSIM, mais ce sont les organismes qui en font l'usage le plus complet qui peuvent en espérer les plus grands avantages.

A. Avantages à long terme

15. Le GSIM offre une série d'objets d'information normalisés qui sont les intrants et les produits dans la conception et la production de statistiques. En définissant des objets communs à tous les exercices de production statistique, quel qu'en soit le sujet, le GSIM permet aux organismes de statistique de revoir l'organisation de leur activité en vue d'une plus grande efficacité.

16. Le GSIM pourrait être utilisé pour orienter l'investissement futur vers les domaines de production statistique où les besoins communs sont les plus importants. Il pourrait aussi autoriser un certain degré de spécialisation parmi les organismes de statistique internationaux. Par exemple, certains organismes pourraient se spécialiser dans la correction des variations saisonnières, l'analyse de séries chronologiques ou la validation de données, et d'autres pourraient tirer parti de ces compétences techniques.

17. Si le GSIM est mis en œuvre en conjonction avec le GSBPM, il offrira de plus grands avantages. Le GSIM pourrait:

- Créer un environnement propice à la réutilisation et au partage de méthodes, de composantes et de processus;
- Offrir la possibilité de mettre en œuvre un contrôle du processus fondé sur des règles, réduisant ainsi au minimum l'intervention humaine dans le processus de production;
- Faciliter la réalisation d'économies d'échelle par la mise au point d'outils communs par la communauté des organismes statistiques.

B. Avantages immédiats

18. L'un des grands avantages présentés par le GSIM est qu'il fournit un langage commun pour améliorer la communication à différents niveaux:

- Entre les différents intervenants dans la production statistique (experts chargés des aspects commerciaux et des technologies de l'information);
- Entre les différents domaines statistiques;
- Entre les organismes de statistique aux échelons national et international.

19. L'amélioration de la communication se traduira par des échanges de données et de métadonnées plus efficaces au sein des organismes de statistique, entre ces organismes, et avec les utilisateurs et les fournisseurs externes.

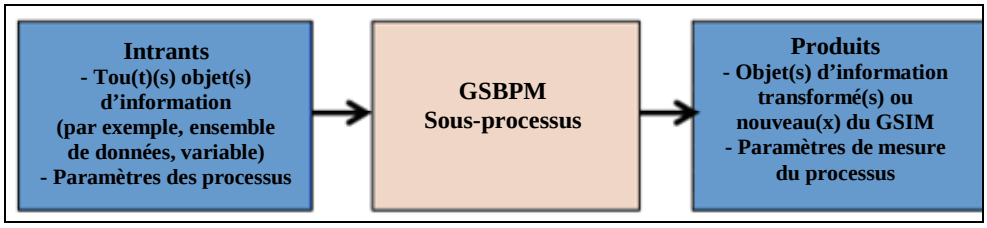
20. Le GSIM peut être désormais utilisé par les organismes pour:

- Renforcer les capacités du personnel en utilisant le modèle comme un outil didactique qui donne une description simple, facile à comprendre, d'informations complexes, assortie de définitions claires;
- Valider les systèmes d'information existants et les comparer avec les meilleures pratiques récemment apparues dans le monde et, le cas échéant, tirer parti des compétences techniques internationales;
- Aider à l'élaboration ou à l'actualisation des normes internationales ou locales afin qu'elles répondent aux besoins les plus divers de la communauté statistique internationale.

V. Le modèle générique d'informations statistiques et le modèle générique du processus de production statistique

21. Le GSIM et le GSBPM sont des modèles complémentaires pour la production et la gestion de l'information statistique. Le GSBPM est un modèle du processus de production statistique qui énumère les activités menées par les producteurs de statistiques officielles qui se traduisent par des produits d'information. Ces activités se divisent en sous-processus comme, par exemple, «imputation» et «calcul d'agrégats». Comme le montre la figure 6, le GSIM sert à décrire les sous-processus du GSBPM en définissant les objets d'information qui circulent entre eux, sont produits par eux et sont utilisés par eux pour produire des statistiques officielles.

Figure 4
Relation entre le GSIM et le GSBPM



22. On peut tirer un meilleur profit du GSIM s'il est appliqué en conjonction avec le GSBPM. Le GSBPM se montre lui aussi plus utile s'il est appliqué en conjonction avec le GSIM. Il est néanmoins possible (bien que ce ne soit pas l'idéal) d'appliquer l'un sans l'autre. De même que les processus de production statistique pris individuellement ne comportent pas tous les sous-processus prévus par le GSBPM, il est très peu probable que tous les objets d'information du GSIM soient nécessaires dans un processus de production statistique donné.

23. Une gestion avisée des métadonnées est indispensable à la bonne marche des processus d'activité statistique. Des métadonnées sont associées à chaque phase du GSBPM, qu'elles soient produites pendant cette phase, actualisées ou reportées sans changement d'une phase précédente. Dans le cadre de ce modèle, le processus global de gestion des métadonnées met l'accent sur la création, l'actualisation, l'utilisation et la réutilisation de métadonnées. Les stratégies et systèmes de gestion des métadonnées sont donc indispensables au fonctionnement du GSBPM et sont facilités par le GSIM.

24. L'application conjointe du GSIM et du GSBPM (ou d'équivalent spécifique à un organisme) peut:

- Faciliter l'élaboration de systèmes efficients de collecte, de traitement et de diffusion pilotés par les métadonnées;
- Aider à l'harmonisation d'infrastructures de calcul statistique.

25. Le GSIM sert d'appui à une approche cohérente des métadonnées, en facilitant le rôle premier que prévoit pour ces dernières la partie A du Cadre commun des métadonnées, relative aux Métadonnées statistiques dans un contexte professionnel[6], à savoir que les métadonnées *devraient seulement et formellement définir les contenus et les liens entre les objets et les processus dans le système d'information statistique*.

VI. À quoi le modèle peut-il me servir?

A. Le GSIM vu sous l'angle de l'activité statistique

26. Le GSIM vous aidera à mieux communiquer avec vos homologues (locaux et internationaux).

27. Il est souvent malaisé de communiquer un sujet d'un domaine à l'autre, de sorte que le partage de concepts, de variables et de composantes de conception est difficile sans exercice de mappage complexe. Le GSIM peut servir de langue commune et faciliter la communication entre:

- Les spécialistes de tel ou tel sujet, les méthodologues et les informaticiens;
- Les statisticiens chargés des différents domaines dans un organisme de statistique;
- Les statisticiens de divers organismes.

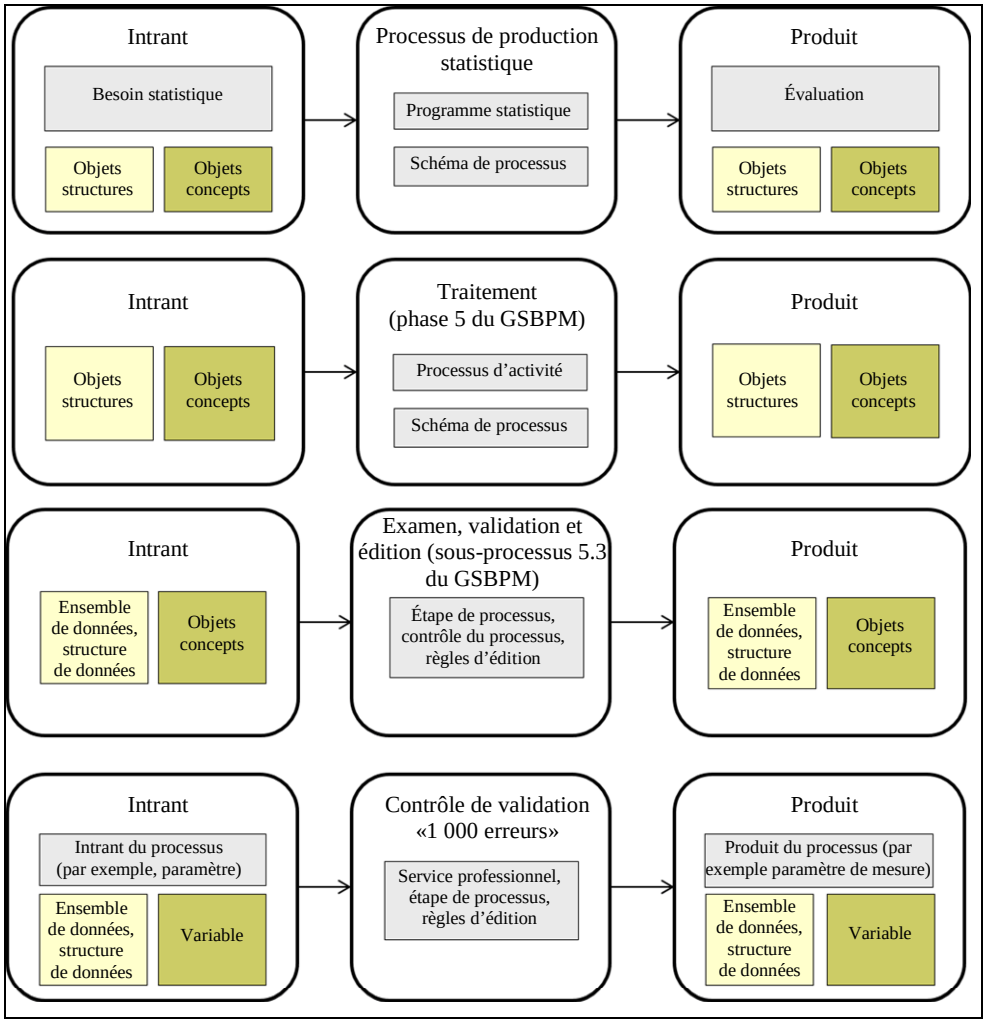
28. Le GSIM vous aidera à concevoir et mieux comprendre vos processus (ainsi que leurs intrants et produits).

29. Pour un cycle de production, un statisticien peut concevoir l'intrant et le produit, et le processus qui va de l'un à l'autre. Suivant le GSIM, l'intrant et le produit peuvent être conçus sous forme d'objets d'information «structures» et d'objets d'information «concepts» et le processus qui les relie peut être conçu en utilisant les objets d'information «activités». Les objets «structures» et les objets «concepts» sont fournis par les spécialistes du sujet à l'étude.

30. Comme on le voit à la figure 4, si le GSBPM est vu comme un cadre de référence pour les processus de production statistique, le premier niveau peut être jugé équivalent au processus de production statistique dans son ensemble. Le niveau suivant correspond à une phase du processus de production statistique (par exemple, la phase «traitement» du GSBPM). Le troisième niveau correspond à un sous-processus (par exemple, le sous-processus 5.3 du GSBPM − Examen et validation). Le quatrième niveau correspond aux divers modules entrant dans la composition du sous-processus, comme le repérage des valeurs financières qui peuvent s'exprimer en milliers plutôt qu'en unités.

[6] http://www1.unece.org/stat/platform/display/metis/The+Common+Metadata+Framework.

Figure 5
Objets d'information du GSIM dans le contexte du GSBPM



31. L'un des grands problèmes auxquels les statisticiens se heurtent est celui des éléments de conception non réutilisables qui sont souvent recréés ou tout au moins modifiés pour chaque cycle de production. Le GSIM facilite la description des intrants et des produits à chaque niveau du GSBPM suivant le même schéma, ce qui confère une structure cohérente à la conception de processus statistiques. Il appuie la conception, la spécification et la mise en œuvre de méthodes harmonisées et de technologies standard pour créer un système de production statistique qui puisse se répandre au sein de la profession.

32. L'utilisation du GSIM permettra d'élaborer des modules de processus réutilisables et souples que les statisticiens pourront employer pour élaborer des produits finals plus ou moins complexes, et donc d'obtenir une plus grande variété de produits et de répondre plus facilement à l'évolution des besoins des clients.

33. L'utilisation du GSIM permettra de réduire la charge de travail car de nombreux processus pourront servir plusieurs objectifs en succession et être réutilisés. Ainsi, on perdra moins son temps à répéter les mêmes travaux et l'on pourra davantage se consacrer à l'innovation.

34. À long terme, le GSIM rendra les statisticiens moins tributaires des technologies de l'information.

35. Les statisticiens s'inquiètent beaucoup aujourd'hui de la possibilité d'application et d'utilisation et de la stabilité de leurs méthodes et solutions techniques. Suivant l'approche «cloisonnée» de la production statistique, le sujet dépend lourdement des technologies de l'information pour la conception, l'élaboration et la production de systèmes statistiques.

36. Les statisticiens pourront davantage maîtriser l'agencement de leurs processus et seront donc plus autonomes pour la conception et la production de leurs statistiques.

37. La production s'appuiera sur des applications plus uniformes, mieux adaptables au changement et moins vulnérables face aux mouvements d'effectifs. Le recours plus systématique à des applications normalisées qui pourront plus facilement s'étendre à différents domaines statistiques permettra aux statisticiens de passer plus facilement d'un domaine à l'autre.

B. Le GSIM, vu sous l'angle des technologies de l'information

38. Les informaticiens s'inquiètent principalement des chevauchements d'activité résultant de l'organisation «cloisonnée» de la production statistique. Les besoins changeants et divergents liés à ce type d'organisation donnent naissance à des solutions sur mesure utilisables une seule fois, tandis que les rotations élevées des effectifs du secteur informatique peuvent se traduire par des applications mal documentées et non normalisées.

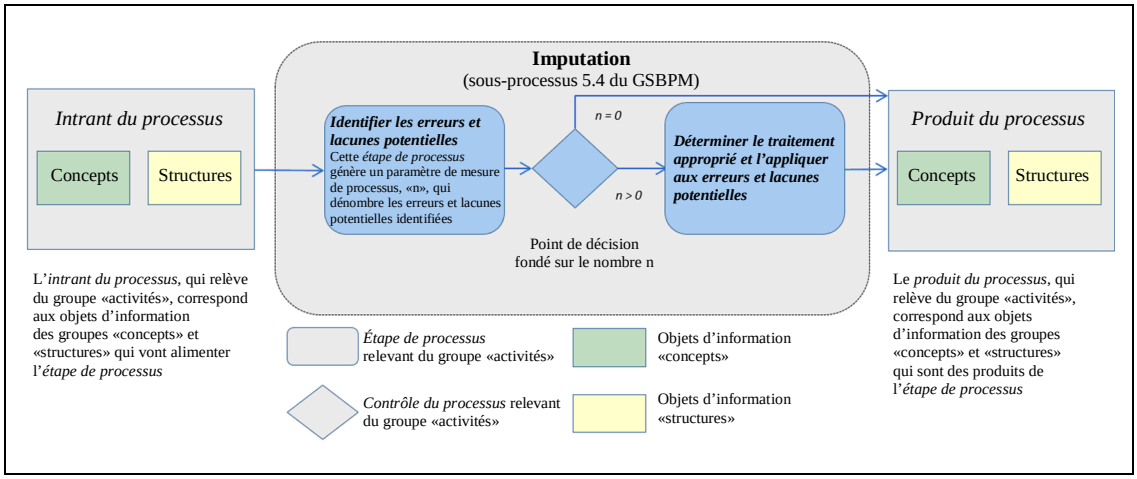
39. La mise en place du GSIM tant au niveau national qu'international peut déjà apporter des avantages à court terme pour les spécialistes des technologies de l'information. Le GSIM offrira aux informaticiens un langage commun leur permettant de dialoguer avec leurs clients et leurs homologues à l'échelle locale et internationale.

40. Au niveau national, les statisticiens deviendront plus autonomes pour la conception (voir la figure 6) et la production de leurs statistiques, et en réutilisant des composantes harmonisées ou en leur attribuant de nouveaux objectifs, le GSIM autorisera des systèmes de production plus flexibles et modulaires. La production s'appuiera sur des applications plus uniformes, mieux adaptables au changement et moins vulnérables face aux mouvements d'effectifs d'informaticiens. Le recours plus systématique à des applications normalisées qui pourront plus facilement s'étendre à différents domaines statistiques permettra aux informaticiens de passer plus facilement d'un domaine à l'autre.

41. L'utilisation du GSIM permettra de réduire la charge de travail car de nombreux éléments pourront servir plusieurs objectifs en succession et être réutilisés. Ainsi, on perdra moins son temps à répéter les mêmes travaux et l'on pourra davantage se consacrer à l'innovation.

42. Libérés de ces tâches, les informaticiens pourront travailler à des applications plus adaptables et voir comment mieux répondre aux nouveaux besoins de l'organisme de statistique et de ses clients en général. Ils auront aussi plus de temps pour créer des processus fiables, modulaires, harmonisés, bien documentés conformes aux prescriptions de l'architecture commune de la production statistique.

Figure 6
Concevez votre propre processus d'imputation



43. À l'échelle internationale, les possibilités seront plus grandes de coconception et de codéveloppement de composantes communes à même de répondre à des besoins plus détaillés émanant d'un plus large éventail d'utilisateurs. Les développeurs informatiques auront aussi accès à des homologues plus nombreux parlant tous la même langue pour décrire leurs informations statistiques.

VII. Norme SDMX (échange de données et de métadonnées statistiques), Data Documentation Initiative et autres normes

44. En tant que cadre de référence d'objets d'information, le GSIM entretient une relation de complémentarité avec des normes comme l'échange de données et de métadonnées statistiques (SDMX) et la Data Documentation Initiative (DDI), qui sont communément utilisées pour représenter et échanger des données et métadonnées statistiques.

45. Les objets d'information décrits par le GSIM sont conceptuels; aucune représentation physique spécifique de l'information n'est prescrite. On peut en donner une illustration simple: le nom d'un organisme se définira comme un même concept que l'information soit enregistrée dans une base de données, sur un tableur, dans un fichier au format CSV (pour «comma separated value» ou «valeurs séparées par des virgules»), dans un fichier XML ou écrite à la main sur une feuille de papier.

46. Le GSIM permet aux organismes de démarrer leurs travaux avec une langue commune concernant les données et les métadonnées utilisées tout au long du processus de production statistique. Dans ce contexte, les objets d'information du GSIM ont été définis suivant des représentations pertinentes conformes aux normes SDMX et DDI.

47. Le modèle aidera les organismes de statistique à décrire et gérer les informations statistiques en utilisant un langage commun tandis qu'au niveau des systèmes, l'information sera représentée et échangée sous un format technique adapté et normalisé.

48. S'il est possible de définir les objets d'information du GSIM suivant les normes SDMX et DDI (dont l'exploitation peut apporter des avantages opérationnels conséquents), le GSIM n'exige pas pour autant l'utilisation de ces normes. Certains producteurs et utilisateurs de statistiques peuvent opter pour différentes normes à des fins particulières.

Dans d'autres cas, les producteurs de statistiques veulent bien utiliser les normes SDMX et/ou DDI mais possèdent des systèmes d'information de la génération précédente qu'il n'est pas rentable d'actualiser pour pouvoir les utiliser avec ces normes.

49. En décrivant les informations statistiques au moyen du GSIM comme point commun de référence, les utilisateurs peuvent mieux corréler deux séries d'informations statistiques représentées différemment du point de vue technique.

50. Par exemple, un statisticien peut recevoir des données décrites suivant la norme DDI et d'autres sous un format établi par son organisme. Il peut mettre ces deux formats en correspondance avec le GSIM. Il sera à même de distinguer les différences d'ordre purement technique de celles qui trahissent des divergences conceptuelles plus profondes.

51. Une fois que l'on a compris la nature et l'ampleur de ces différences, il est souvent plus facile de donner à l'information une représentation technique commune (SDMX ou DDI par exemple) qui permet d'en intégrer et d'en étudier le contenu. Cette méthode permet de bien comprendre les résultats de cette conversion technique suivant une norme unique et d'assurer la fiabilité d'un point de vue conceptuel.

52. Il existe un certain nombre de synergies entre l'utilisation du GSIM comme cadre de référence et l'application de normes de représentation de type SDMX et DDI. Ces synergies ont été délibérément optimisées.

53. Par exemple, au moment de déterminer l'ensemble de définitions à utiliser pour les objets d'information dans le cadre du GSIM, les normes et modèles existants ont été utilisés comme principales sources de référence. Bien qu'aucune des sources existantes n'ait ni le même but ni le même champ d'application que le GSIM − qui est d'offrir un cadre de référence d'objets d'information recouvrant la totalité du processus de production statistique − la mise au point de chacune de ces sources a exigé l'examen et la prise en charge de besoins et scénarios spécifiques en rapport avec des types de données et de métadonnées statistiques particuliers.

54. Ainsi, le GSIM a bénéficié du temps investi dans les travaux d'analyse, de modélisation, de mise à l'essai et de perfectionnement exigés par le développement de ces normes et modèles à leur niveau actuel de maturité. Il faut aussi comprendre que ce n'est pas «sans raison» que le GSIM s'écarte des termes et définitions utilisés dans ces normes et modèles. Lorsqu'il s'en éloigne, c'est en raison de contradictions au sein même des normes et modèles pertinents existants, ou entre ces normes et modèles, et/ou du fait que de l'avis des statisticiens, d'autres termes et définitions répondent mieux à leurs besoins opérationnels. La révision du Modèle de classification de Neuchâtel, aux fins d'une harmonisation et d'une intégration complètes avec le GSIM, est la conséquence directe d'une telle approche.

Annexe

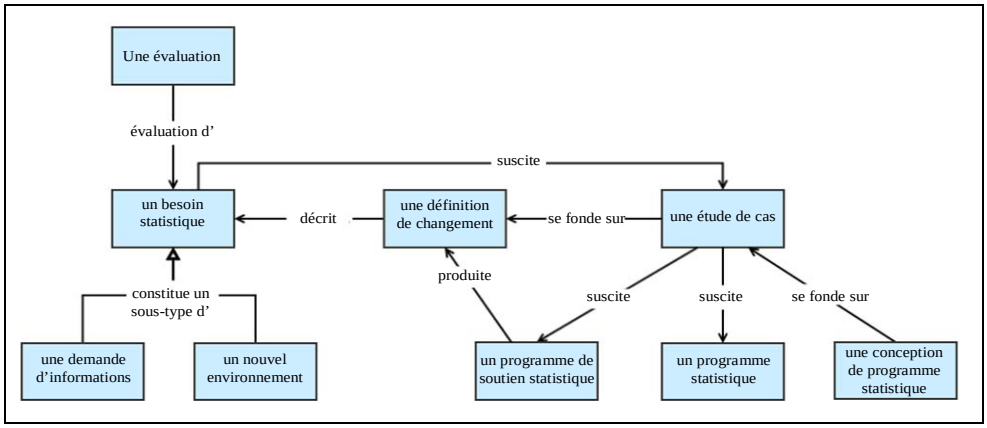
Extraits de la spécification du modèle générique d'informations statistiques[7]

A. L'information dans le processus de production statistique

1. La présente section examine différentes modalités selon lesquelles les objets d'information sont utilisés dans le processus de production statistique. Elle examine huit scénarios différents, en désignant les objets d'information utilisés et les relations entre ces objets.

1. Identification et évaluation des besoins statistiques

Figure 1
Identification et évaluation des besoins statistiques



2. Un organisme réagit et évolue par suite de divers besoins. Un besoin statistique se présente à un organisme de statistique sous la forme d'un nouvel environnement ou d'une demande d'informations.

3. Un nouvel environnement suppose l'existence d'un changement provoqué par des facteurs externes. Ce changement peut ne concerner que l'organisme en question, comme une réduction budgétaire ou de nouveaux besoins exprimés par les parties concernées, ou peut avoir plus d'ampleur, comme l'apparition d'une nouvelle méthode ou d'une nouvelle technologie.

4. Lorsqu'un organisme reçoit une demande d'informations, celle-ci précisera l'information dont un particulier ou un organisme parmi les utilisateurs a besoin à des fins particulières. Il peut s'agir d'utilisateurs au sein de l'organisme ou d'utilisateurs externes. Par exemple, l'équipe chargée d'établir la comptabilité nationale peut avoir besoin de lancer un nouveau processus d'activité pour alimenter en nouveaux intrants son exercice de compilation. La demande d'informations se définit généralement par rapport à un champ d'activité qui correspond à ce que l'utilisateur souhaite mesurer. Une fois reçue, la demande est examinée et son contenu précisé avec l'utilisateur. Suite à ces éclaircissements, des recherches sont faites pour vérifier si l'on dispose déjà des données voulues. La mise à jour de ces ensembles de données peut être facilitée par la recherche de concepts et de classifications. Chacune de ces activités est décrite par une étape de processus.

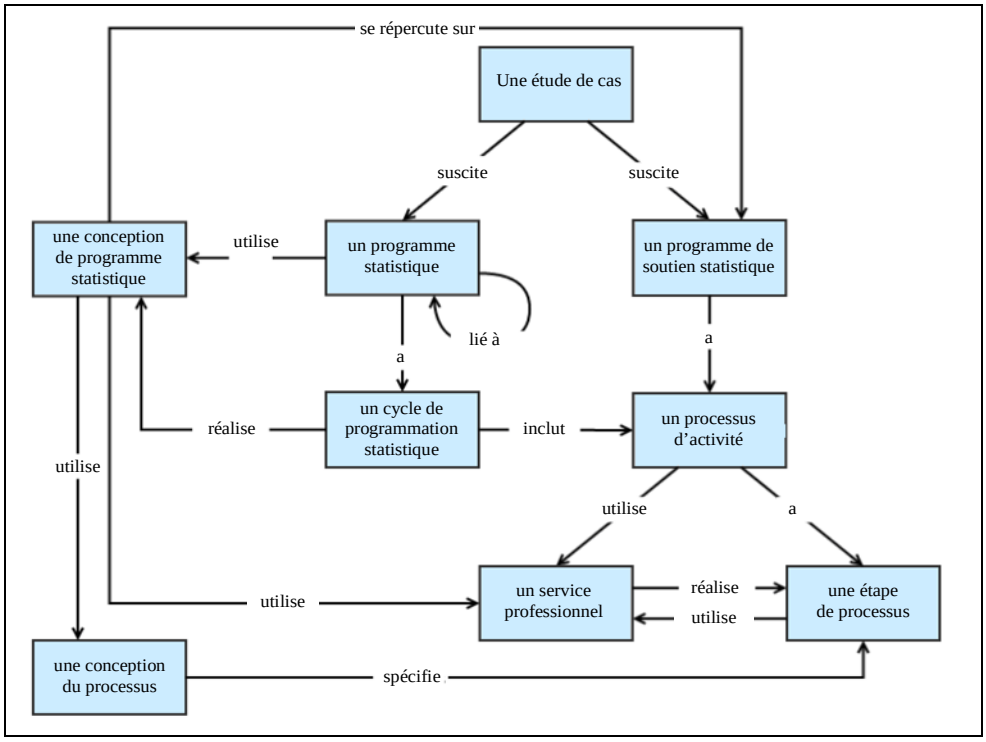
[7] Note: Le présent appendice n'est pas le texte complet de la spécification du GSIM, il ne comprend que des extraits de ce document.

5. Un besoin statistique − suite à une demande d'informations ou un nouvel environnement − est systématisé sous la forme d'une définition de changement, généralement suscitée par un programme de soutien statistique (un «programme de changement statistique»). La définition de changement indique le caractère spécifique du changement du point de vue de ses incidences sur l'organisme ou ses programmes statistiques ou programmes de soutien statistique spécifiques. Cette définition de changement est utilisée comme intrant pour une étude de cas. Si tout fonctionne bien, un nouveau programme statistique ou un nouveau programme de soutien statistique sera mis en œuvre qui donnera lieu à une nouvelle conception de programme statistique qui redéfinira la manière dont s'effectue un programme statistique existant.

6. Un besoin statistique peut aussi apparaître en interne. À tout moment du processus de production statistique, un organisme peut entreprendre une évaluation pour déterminer l'utilité ou l'efficacité du processus d'activité ou de ses intrants et produits. Une évaluation sera entreprise pour évaluer les ressources, les processus ou les produits, laquelle pourra porter sur tout objet décrit dans le modèle. Ces évaluations peuvent prendre la forme d'analyses des lacunes effectuées dans le cadre des études de cas et d'évaluations visant à déterminer si un produit statistique répond au besoin pour lequel il a été élaboré à l'origine.

2. Conception et gestion des programmes statistiques

Figure 2
Conception et gestion des programmes statistiques



7. Un organisme de statistique répond à un besoin statistique mis à jour en réalisant une étude de cas. Celle-ci peut avoir trois suites: l'élaboration d'un nouveau programme de soutien statistique, la création d'un nouveau programme statistique ou l'évolution d'une conception de programme statistique existante à mettre en œuvre au moyen d'un programme statistique existant.

8. Les programmes de soutien statistique prennent en charge les activités de l'organisme de statistique telles que les programmes relatifs au changement statistique, à la gestion des données, à la gestion des métadonnées, les programmes de recherche méthodologique, etc. Les programmes de gestion des classifications en constituent un bon exemple.

9. Les programmes statistiques sont ceux que lance un organisme pour produire des statistiques (par exemple, une étude sur le commerce de détail). Ces programmes sont cycliques − ils prévoient des cycles de collecte, de production et de diffusion de produits. Chacun de ces cycles est représenté par un objet de type cycle de programmation statistique. Le cycle de programmation statistique est une activité qui se répète pour produire des statistiques à un moment déterminé (pour reprendre le même exemple, l'étude du commerce de détail pour mars 2012).

10. Les programmes statistiques requièrent des conceptions de programmes statistiques pour atteindre leurs objectifs. Celles-ci recouvrent la conception de toutes les activités à entreprendre, notamment à l'échelle des processus d'activité. Un cycle de programmation statistique comprend généralement plusieurs processus d'activité que l'on peut corréler aux processus et sous-processus figurant dans le modèle générique du processus de production statistique (GSBPM). Ces processus d'activité peuvent se répéter au cours d'un cycle. Chaque itération peut se composer d'activités multiples du même type ou de types différents. Par exemple, au cours d'un même cycle, le programme statistique peut prévoir trois itérations de collecte et de traitement des données, puis une phase d'analyse et de diffusion des produits statistiques qui en résultent. Chacune de ces activités peut être considérée comme un processus d'activité distinct.

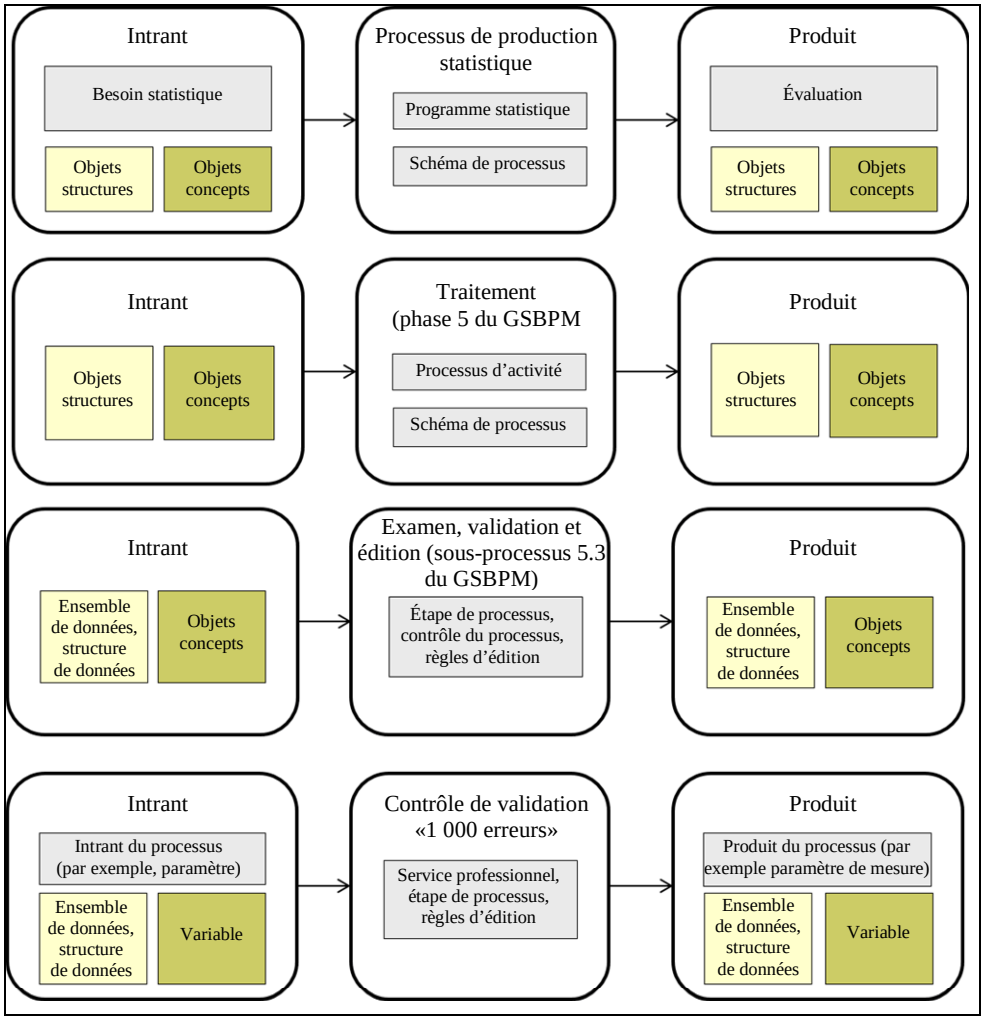
11. La conception de programme statistique définit la manière selon laquelle les processus d'activité seront conduits. Il est notamment possible dans ce cadre de recourir à des services professionnels réutilisables (qui peuvent être obtenus à l'extérieur de l'organisme de statistique) ou de concevoir et d'utiliser des processus plus traditionnels. Dans ce dernier cas, des objets de type conception de processus seront utilisés pour spécifier des étapes de processus (bien que des services professionnels réutilisables puissent aussi être spécifiés par des conceptions de processus et des étapes de processus, ceux-ci existent déjà et n'exigent pas de nouveau travaux de conception au titre de l'objet «conception de programme statistique»).

12. On notera que les conceptions de programme statistique spécifient lesquelles des étapes de processus nécessiteront des conceptions de processus et, en outre, quels types de services professionnels seront utilisés, mais ne vont pas jusqu'à préciser, à un niveau inférieur, les modalités d'exécution de ces étapes de processus et services professionnels. Ces spécifications relèvent de l'objet «conception de processus».

3. Conception des étapes de processus

13. Avant d'expliquer les objets que le GSIM utilise pour représenter la conception d'étapes de processus, il importe d'examiner la nature des processus de façon plus générale. Les types d'objets fournis par le GSIM assurent des fonctions spécifiques. Dans le GSIM, les processus d'activité se composent d'étapes de processus. Chacune de ces étapes peut être «à large échelle» ou «à petite échelle» au choix du concepteur du processus d'activité dont il est question (voir la figure ci-après).

Figure 3
Les étapes de processus peuvent être aussi longues ou courtes que nécessaire

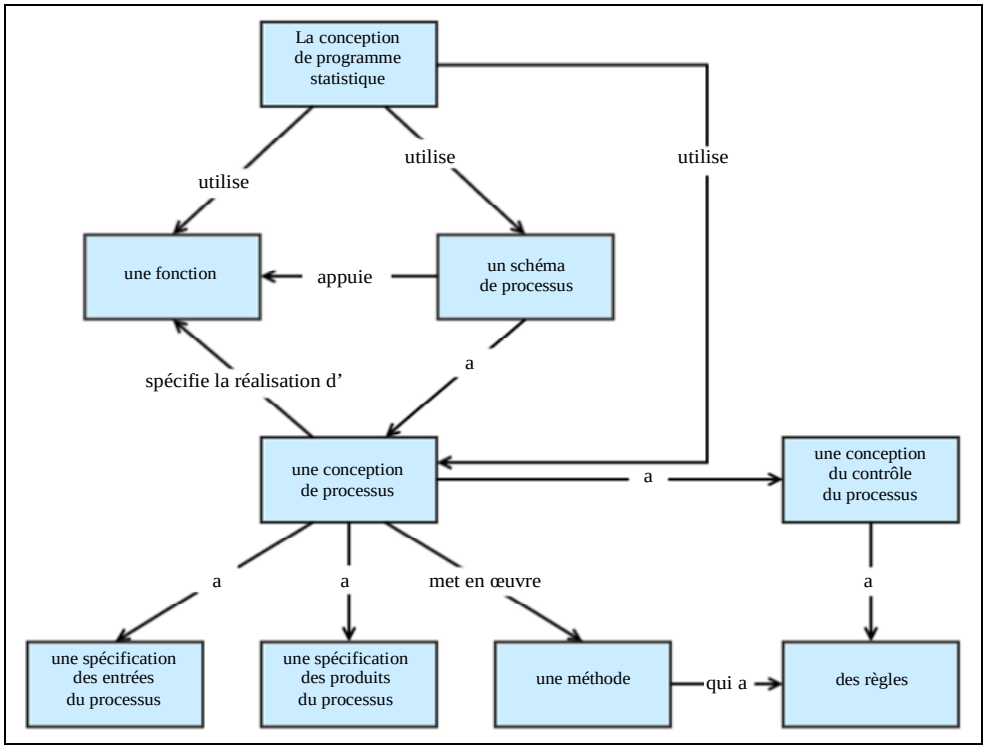


14. Les étapes de processus peuvent comprendre des «sous-étapes», ces «sous-étapes» peuvent en comprendre d'autres et ainsi de suite à l'infini. En règle générale, les produits d'une étape de processus deviennent des intrants pour l'étape suivante. On peut aussi appliquer une logique d'exécution conditionnelle à l'enchaînement des étapes de processus qui se fonde sur des paramètres adoptés ou des conditions remplies par les produits d'une étape de processus antérieure.

15. Il est donc entendu que la conception d'une étape de processus peut utiliser d'autres étapes de processus voire même d'autres services professionnels qui ont déjà été conçus et mis à disposition pour être réutilisés. Suivant un scénario plus traditionnel, l'étape de processus est d'abord conçue puis exécutée. À l'avenir, il est prévu d'utiliser de plus en plus couramment des services professionnels réutilisables conçus et mis en œuvre par un autre organisme externe. Les sections suivantes décrivent ces deux scénarios.

A. Conception des étapes de processus

Figure 4
Conception des étapes de processus



16. Une conception de programme statistique est associée à une étape de processus de haut niveau dont l'objet «conception de processus» comprend toutes les sous-étapes et tous les flux fonctionnels requis pour la mise en application de ce programme statistique. Chaque étape d'un processus d'activité statistique se voit attribuer un but particulier. C'est ce qu'indique la fonction associée à l'étape de processus. À titre d'exemple, une fonction pourrait consister, par exemple, à «imputer les valeurs qui manquent dans les données». Pour appuyer cette fonction, il faudra un processus d'imputation qui a son tour exigera une conception de processus.

17. Le GSIM part du principe d'une séparation de la conception et de la production, aussi suppose-t-il que les étapes de processus sont élaborées pendant la phase de conception. Une fois que l'on a divisé un processus de production statistique planifié en étapes de processus, il faut définir une conception de processus pour chaque étape. La conception de processus précise comment chaque étape de processus se déroulera. Une conception de processus peut exiger un schéma de processus qui est un ensemble désigné de conceptions de processus et de flux connexes (conceptions de contrôle du processus) dont il est décidé qu'ils seront réutilisés.

18. Les conceptions de processus définissent plusieurs éléments: elles identifient les différents types d'intrants et de produits représentés par la spécification des intrants du processus et la spécification des produits du processus. Parmi les exemples d'intrants de processus figurent les données, les métadonnées comme les classifications statistiques, l'imputation et les règles d'édition, les paramètres, etc. Les produits de processus peuvent être des rapports de divers types (paramètres de mesure de traitement, rapports sur la validation et la qualité des données, etc.), des ensembles de données édités, de nouveaux ensembles de données, des instances nouvelles ou révisées de métadonnées, etc.

19. Si l'on prend le même exemple, le concepteur du processus définit les intrants dans la spécification des intrants du processus sous forme de règles d'imputation et d'un ensemble de données que l'on cherche à imputer. La spécification des produits du processus comprendrait un ensemble de données édité contenant les valeurs imputées, plus un rapport indiquant lesquelles des valeurs ont été imputées.

20. La conception de processus définit la logique de contrôle, à savoir l'enchaînement et la logique d'exécution conditionnelle entre les différents sous-processus (étapes de processus). Le déroulement de ces opérations est décrit dans la conception du contrôle de processus. Lorsque l'on élabore une conception de processus, on spécifie une conception de contrôle du processus qui fournit des informations sur «l'étape suivante». Parfois, une étape de processus sera suivie par la même étape en toutes circonstances. Dans ces cas, la conception du contrôle du processus enregistre simplement laquelle des étapes de processus vient en suivant. Il arrivera toutefois que l'on puisse choisir l'étape à suivre ensuite. Dans ce cas, la conception de contrôle du processus énumèrera toutes les «étapes suivantes» possibles et les critères à appliquer afin de choisir celle(s) à engager par la suite.

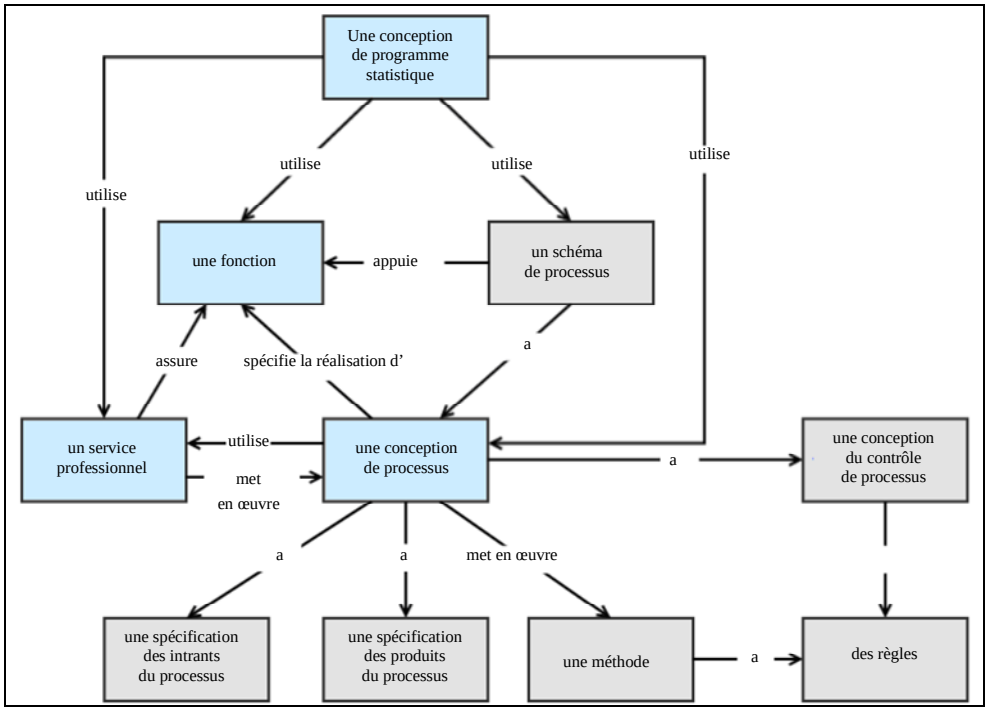
21. La conception de processus associée à cette étape de processus désignera la méthode qui sera utilisée pour assurer la fonction liée à l'étape de processus. Par exemple, si la fonction consiste à «imputer les valeurs qui manquent dans les données», la méthode à utiliser pourrait être l'«imputation des données voisines les plus proches».

22. Par méthode on entend la méthode à utiliser qui s'accompagne d'une série de règles à appliquer. Par exemple, chaque fois que l'on utilisera la méthode de l'«imputation des données voisines les plus proches», on appliquera une règle (paramétrée) pour déterminer ce que sont les «données voisines les plus proches». Dans cet exemple, la règle sera mathématique (par exemple, fondée sur une formule). Les règles peuvent aussi être logiques (par exemple, si la condition 1 est «fausse» et la condition 2 est «fausse», alors il faut qualifier la proposition «imputation requise» de «vraie», et dans les autres cas la qualifier de «fausse».

23. Les objets «conception de processus» et «conception de contrôle de processus» qui en résultent (ainsi que les spécifications des intrants du processus connexes et les spécifications des produits du processus) seraient utilisés lors de la mise en œuvre de l'étape de processus.

B. Utilisation des services professionnels réutilisables

Figure 5
Utilisation des services professionnels réutilisables



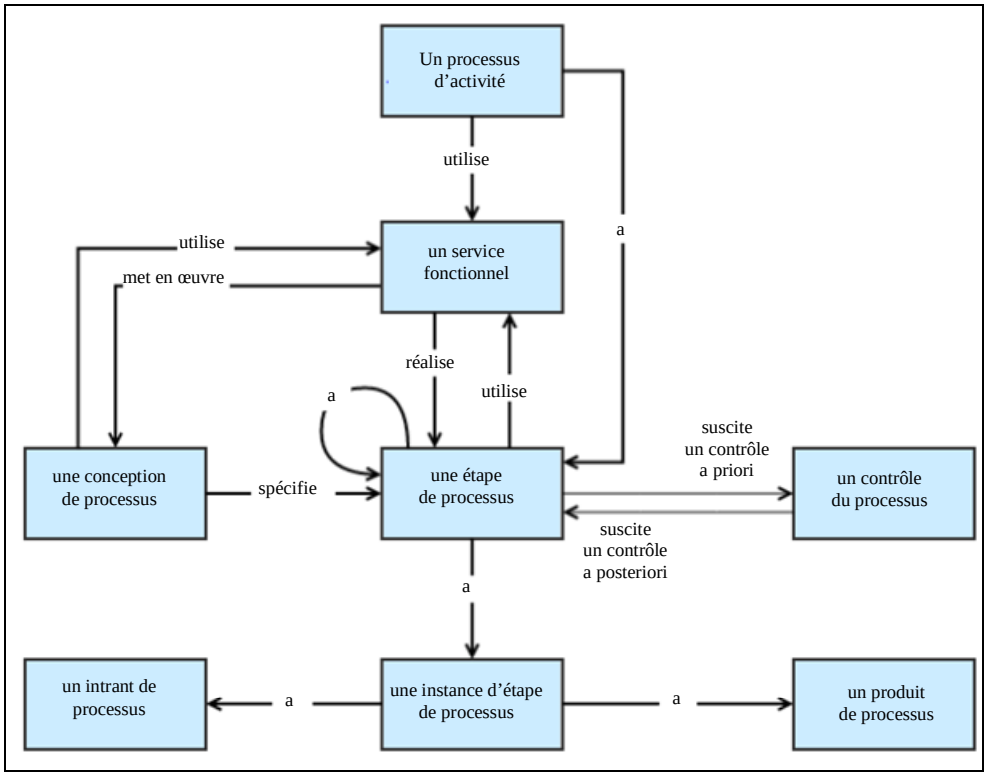
24. Il n'est pas toujours nécessaire de concevoir les étapes de processus d'un programme statistique dès le départ. L'architecture commune de la production statistique (CSPA) décrit la manière dont les organismes de statistique peuvent créer des services statistiques pouvant être facilement réutilisés chez leurs homologues. Du point de vue du GSIM, un service statistique est un service professionnel. Un service professionnel est un moyen d'assurer une fonction (une capacité que possède un organisme, qui habituellement s'exprime en termes généraux, et en termes de haut niveau, et qui, pour s'accomplir, exige une combinaison d'organisation, d'effectifs, de processus et de technologies).

25. Le partage et la réutilisation croissants des services fonctionnels ont pour effet une réduction considérable des ressources nécessaires pour répondre aux nouveaux besoins de production statistique, et la diminution du temps nécessaire pour l'élaboration de nouveaux produits statistiques. Pour faciliter ce processus, la CSPA a introduit le concept de catalogue de services statistiques selon lequel différents organismes de statistique peuvent répertorier les services statistiques qu'ils ont mis au point pour les partager avec des homologues.

26. Des services fonctionnels ont déjà été conçus, assortis de tous les types d'intrants, types de produits, conception de contrôle du processus normaux et d'autres propriétés ont déjà été spécifiées. Ainsi, un service fonctionnel peut agir de façon similaire à une étape de processus conçue au sein de l'organisme mais sans exiger autant d'effort qu'un scénario traditionnel.

4. Déroulement des processus

Figure 6
Déroulement des processus



27. Un programme statistique doit exécuter des processus tendant à réaliser un certain nombre de fonctions. Il existe deux façons d'opérer: une étape de processus peut être directement exécutée par un processus d'activité, ou un service fonctionnel réutilisable peut être mis au service du processus d'activité comme déclencheur intermédiaire de l'exécution de l'étape de processus.

28. Pour comprendre comment cela fonctionne, il convient de caractériser la nature des étapes de processus de façon plus détaillée. Les étapes de processus sont des ressources spécifiées dans le cadre d'une conception de processus, et qui peuvent être exécutées à de multiples reprises. Les étapes de processus peuvent se tenir à de nombreux niveaux de granularité et peuvent faire appel à d'autres étapes de processus et de sous-processus. Au cours de l'exécution, la navigation entre sous-processus s'effectue comme indiqué par un contrôle du processus qui met lui-même en œuvre une conception de contrôle du processus.

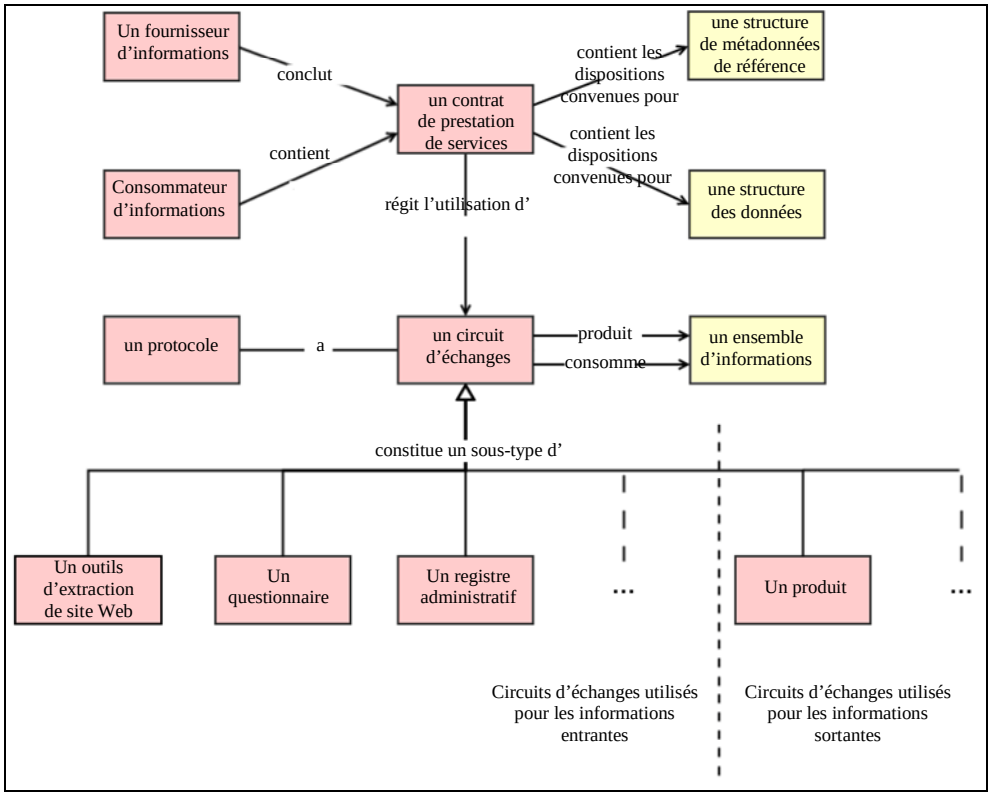
29. Chaque exécution d'une étape de processus est représentée par une instance d'étape de processus. C'est à ce niveau qu'interviennent les instances spécifiques d'intrants et de produits. Les types d'intrants et de produits sont spécifiés dans la conception du processus (spécification des intrants du processus et spécification des produits du processus) − les instances effectives des intrants et des produits sont associées à l'instance d'étape de processus et représentées par les objets «intrants du processus» et «produits du processus». Tout type d'information peut constituer un intrant − règles, paramètres, ensembles de données, métadonnées de nombreuses catégories, etc. Les produits peuvent aussi être très variés, et l'on trouve fréquemment parmi eux des paramètres de mesure de processus, divers types de rapports, ainsi que des données et des métadonnées.

30. Lors de l'exécution de la conception de processus, il faut que quelqu'un ou quelque chose applique la méthode et les règles désignées. La conception de processus peut désigner le service professionnel qui mettra en œuvre la méthode au moment de l'exécution. Un service professionnel représente un service rendu par un logiciel (comme il est décrit dans la section ci-dessus) ou une personne. L'affichage d'une publication sur le site Web d'un institut de statistique ou le regroupement de formulaires de réponses recueillis dans une source de données partagée aux fins d'un traitement ultérieur sont deux exemples de services professionnels.

31. Il convient de noter que ce modèle appuie à la fois les processus automatisés et manuels et les processus susceptibles de contenir des sous-processus de l'un ou l'autre type.

D. Échanges d'informations

Figure 7
Circuits d'échanges



32. Les organismes de statistique recueillent des données et des métadonnées de référence auprès de fournisseurs d'informations qui peuvent être par exemple des personnes interrogées par sondage ou des fournisseurs de registres administratifs et diffusent des données auprès de consommateurs d'informations, qui peuvent être des organismes publics, des entreprises et des membres du public. Chacun de ces échanges de données et de métadonnées de référence emprunte un circuit d'échanges, qui décrit les moyens de recevoir (collecte) ou d'envoyer (diffusion) des informations. Les fournisseurs et les consommateurs d'informations peuvent être des organismes ou des particuliers opérant à l'intérieur ou à l'extérieur de l'organisme de statistique.

33. Différents circuits d'échanges sont utilisés pour la collecte et la diffusion. On peut citer comme exemples de circuits d'échanges de collecte un questionnaire, un outil

d'extraction de sites Web et un registre administratif. Le seul exemple de circuit d'échanges de diffusion figurant actuellement dans le GSIM est constitué par les produits. Des circuits d'échanges additionnels peuvent être ajoutés par les organismes en fonction de leurs besoins.

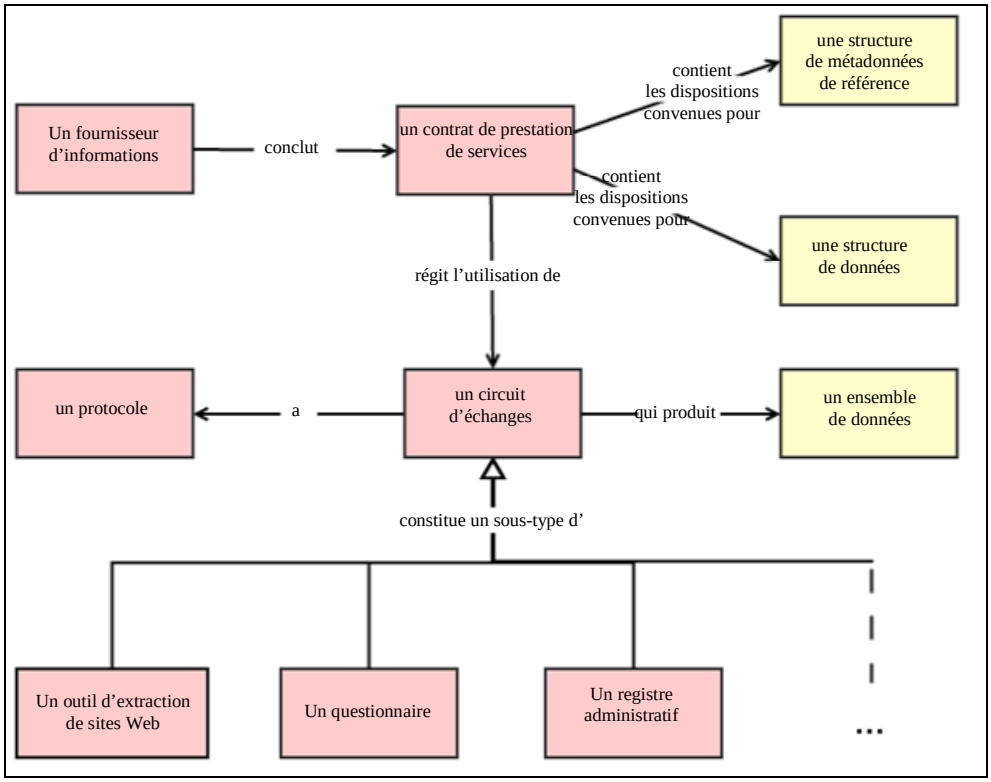
34. L'utilisation d'un circuit d'échanges est régie par un contrat de prestation de services conclu entre le bureau de statistique et le fournisseur d'informations (collecte) ou le consommateur d'informations (diffusion). Le contrat de prestation, qui peut être formel ou tacite, apporte les bases juridiques ou autres suivant lesquelles les deux parties conviennent d'échanger des données. Les parties utilisent aussi le contrat de prestation pour convenir de la structure de données et de la structure de métadonnées de référence des informations à échanger.

35. Le mécanisme d'échange d'informations par le circuit d'échanges est défini dans un protocole (par exemple, service Web SDMX, échange de fichiers de données, interrogatoire direct).

36. Pour collecter des données, un organisme de statistique reçoit des données et des métadonnées de référence du fournisseur d'informations conformément au protocole et au contrat de prestation de services et le circuit d'échanges produit un ensemble d'informations. Pour diffuser des données, le circuit d'échanges absorbe un ensemble d'informations qui est alors fourni au consommateur d'informations d'une manière conforme au protocole et au contrat de prestation. On trouvera plus de renseignements sur la collecte et la diffusion dans les sections ci-après.

5. Collecte d'informations

Figure 8
Circuits d'échanges pour la collecte d'informations



37. Le GSIM modélise trois exemples de circuit d'échanges de collecte: questionnaire, outil d'extraction de sites Web et registre administratif. Chacun de ces trois exemples est détaillé dans l'annexe A. Les organismes de statistique peuvent collecter des données et des métadonnées de référence auprès de fournisseurs d'informations en utilisant des circuits d'échanges additionnels comme le transfert de fichiers, les services Web et la numérisation de données. Les organismes de statistique peuvent élargir le GSIM aux circuits intéressant leur domaine d'intervention.

38. L'utilisation d'un circuit d'échanges pour la collecte est régie par un contrat de prestation de services entre l'organisme de statistique et le fournisseur d'informations. Les deux parties utilisent le contrat de prestation pour convenir de la structure de données et de la structure de métadonnées de référence. Le mécanisme de collecte d'informations via le circuit d'échanges est défini par un protocole (par exemple, interrogatoire direct, échange de fichiers de données, robot de recherche sur le Web). L'organisme de collecte utilise l'information recueillie pour produire l'ensemble d'informations qui peut contenir des données ou des métadonnées de référence.

7. Traitement et analyse de l'information

39. Le GSIM offre une grande souplesse dans la description du traitement et de l'analyse de l'information.

40. Le processus de production statistique peut se comprendre du seul point de vue des données[8]. Les organismes de statistique s'efforcent de produire des données fiables de haute qualité appuyées par les métadonnées permettant d'optimiser leur utilité. C'est pourquoi il faut prendre en compte l'évolution des données tout au long du processus de production. La spécificité de nombreuses activités est pilotée par les métadonnées, mais à l'issue du processus de production, les métadonnées sont un appui pour les données et en dernier ressort un produit statistique. D'où l'importance qu'il y a à comprendre la relation entre les données et les métadonnées.

41. Les données recueillies parviennent à l'organisme de statistique par le biais d'un circuit d'échanges. Quels que soient le mode de collecte des données et leur origine, c'est une ressource qui engagera un processus d'évolution passant par de nombreuses phases différentes. Les données initiales sont décrites comme un ensemble de données assorti des structures de données pertinentes. Les ensembles de données sont stockés de façon structurée dans une source de données. Les ensembles de données sont les principaux intrants et produits d'une série d'étapes de processus entrant dans le cadre d'un programme statistique.

42. Au fur et à mesure que l'organisme de statistique progresse de données d'entrées brutes à un ensemble de données toujours plus fines, on comprendra que chaque phase de traitement ajoute de nouveaux ensembles de données à la source de données. Il existe de nombreuses méthodes différentes à même d'influer sur ces activités. Celles-ci sont mises en œuvre par le biais des différentes étapes de processus entreprises par l'organisme de statistique.

43. À un certain stade (qui peut intervenir à différents moments du processus de production, suivant le type de mises en forme à opérer) les données sont analysées en vue de l'élaboration de produits statistiques. L'analyse des données consiste à utiliser des ensembles de données provenant de la source de données comme intrants à des processus tels que des mécanismes de confidentialité, ou pour la production des explications des données. Les opérations effectuées au cours de l'analyse varieront selon ce que l'on veut

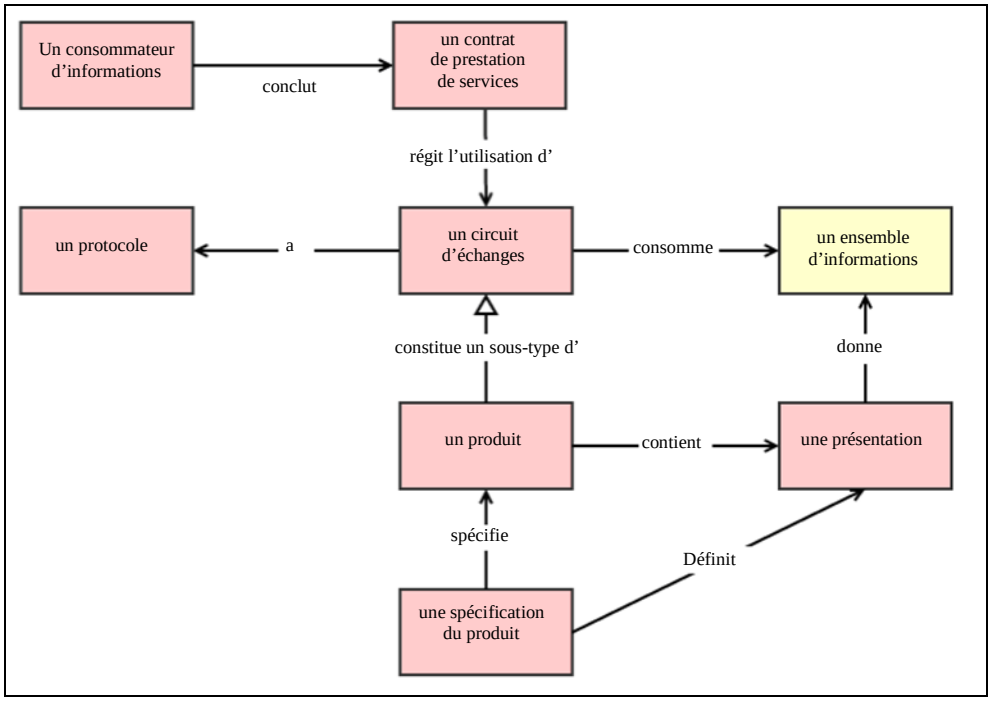
[8] Bien que les paragraphes qui suivent portent l'accent sur les données, leur contenu peut s'appliquer aux métadonnées de référence.

obtenir comme produits à terme − des unités statistiques dont la confidentialité est protégée peuvent constituer un produit, ou des indicateurs et des tableaux agrégés peuvent être publiés pour traiter de questions d'intérêt général, et ceux-ci appellent différents types d'analyses − mais il s'agit toujours ici de continuer à mettre au point l'information figurant dans la source de données.

44. Par le passé, on partait de l'hypothèse qu'une collecte de données serait suivie d'un traitement, d'une analyse et d'une diffusion de produits statistiques. C'est un processus qui exige du temps et des ressources. L'un des moyens par lesquels un organisme de statistique peut fonctionner plus efficacement consiste à réutiliser des données pour l'élaboration des nouveaux produits qui lui sont demandés, ce qui lui permet d'abaisser les coûts et de raccourcir le temps nécessaire à la production. Vue sous cet angle, la source de données peut être considérée comme une richesse institutionnelle qui doit être gérée et exploitée dans la plus grande mesure possible.

8. Diffusion de l'information

Figure 9
Circuit d'échanges pour la diffusion de l'information



45. Un organisme de statistique diffuse des informations statistiques à un consommateur d'informations.

46. Le consommateur d'informations accède à un ensemble d'informations par le biais d'un produit (ou potentiellement par un autre circuit d'échanges) qui contient une ou plusieurs présentations. Chaque présentation donne généralement une idée des données et des métadonnées connexes permettant de définir et de décrire la structure des données présentées, et éventuellement des métadonnées de référence sous la forme de supports écrits, par exemple des rapports sur la qualité.

47. Une présentation peut prendre diverses formes − par exemple, une visualisation à l'écran d'un tableau de données sous forme graphique affiché sur une page HTML, un fichier PDF téléchargeable, ou un fichier SDMX au format XML.

48. Une spécification du produit définit le contenu de la présentation. Par produit, qui englobe les présentations, on peut comprendre un produit spécifique normalisé de l'organisme de statistique, comme par exemple:

- Un bulletin statistique périodique (par exemple, une publication mensuelle de l'indice des prix de détail);
- Un ensemble de contenus statistiques générés dynamiquement produit en réponse à la demande d'un consommateur d'informations qui souhaite accéder aux données de l'organisme par le biais d'une interface de programmation d'application (API), ou;
- Un outil d'exploration de données qui peut être intégré au site Web de l'organisme de statistique.

49. La spécification du produit définit aussi les éléments requis de l'ensemble d'informations pour la présentation. Les spécifications sont fréquemment déterminées par un processus interne (à l'organisme) qui aurait des produits statiques expressément standardisés à produire (comme les bulletins statistiques susmentionnés). Pour les produits dynamiquement générés, certains aspects de la spécification pourraient être déterminés par le consommateur d'informations pendant l'exécution, par une dynamique de machine à machine comme l'illustre plus haut l'exemple de l'API. Dans l'un comme dans l'autre cas, les demandes auront pour effet que la spécification du produit précise les données et/ou les métadonnées de référence de l'ensemble d'informations qui figureront dans chaque présentation.

50. Le mécanisme de fourniture d'un produit est spécifié dans un protocole (par exemple, SDMX-ML, DDI XML, PDF, etc.). Cette information de formatage s'intègre dans la spécification du produit de façon à ce que le produit et ses présentations soient élaborés dans le format voulu.

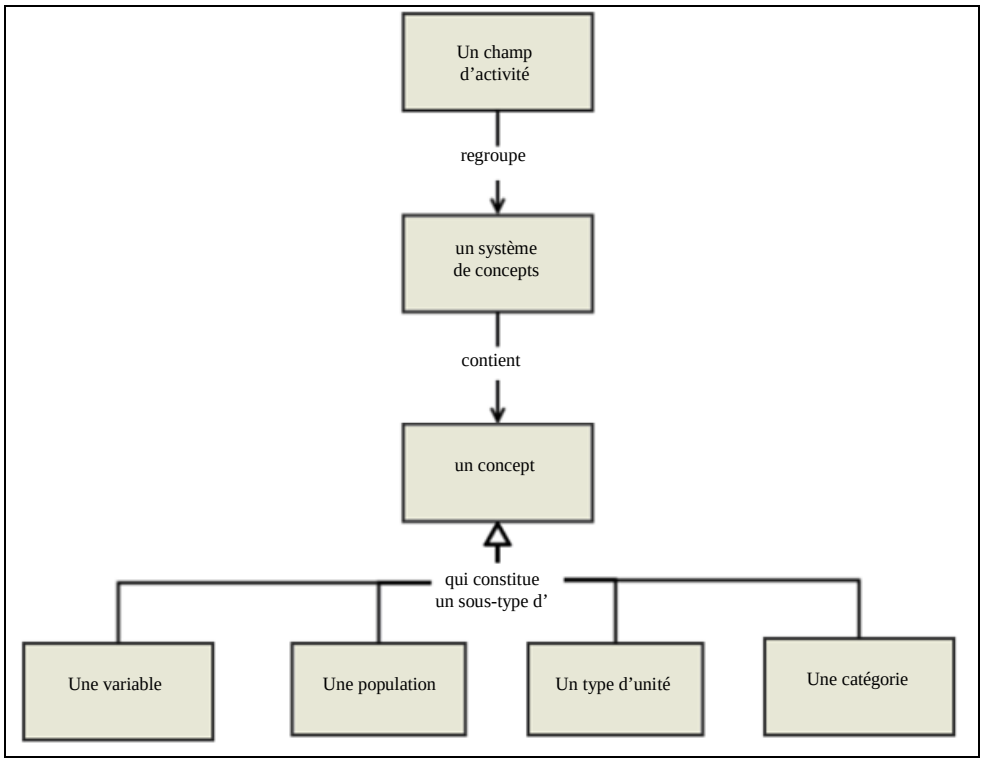
51. Il peut y avoir divers types de consommateurs d'informations suivant la forme que prend la demande. Le consommateur d'informations peut être une personne qui consulte le site Web de l'organisme de statistique et prend connaissance du contenu d'une page Web, ou ce peut être un programme informatique qui demande l'information via une API, en faisant une recherche SDMX. L'accès offert au consommateur d'informations est assujetti à un contrat de prestation de services qui fixe les conditions d'accès et d'utilisation. Cela peut prendre la forme d'une acceptation tacite des conditions d'utilisation des données provenant d'un site Web auquel le consommateur d'informations a accès, ou dans le cas d'un accès beaucoup plus détaillé via une API, d'un processus d'enregistrement plus actif.

B. Information fondamentale

52. Les groupes «concepts» et «structures» du GSIM comprennent les objets d'information au cœur du processus de production statistique. Autrement dit, il s'agit d'objets conceptuels et structurels qui sont utilisés comme intrants du processus et produits du processus. Le domaine du GSIM relatif aux concepts comprend les ensembles d'objets d'information qui décrivent et définissent les termes utilisés lorsque l'on parle des phénomènes du monde réel que les statistiques mesurent dans leur application pratique. Le domaine relatif aux structures comprend l'ensemble des objets d'information utilisés en rapport avec les données et les métadonnées de référence et leurs structures. Les objets décrits dans la présente section du document servent à fournir des informations qui aident les utilisateurs de données et de métadonnées à comprendre les résultats des processus d'activité et des programmes statistiques.

1. Concepts

Figure 10
Concepts



53. Un concept se définit dans le GSIM de façon abstraite comme une «unité de pensée qui se différencie par ses caractéristiques». Les concepts sont utilisés de diverses façons tout au long du cycle de vie statistique et chaque rôle différent d'un concept est décrit au moyen d'un objet d'information différent (les objets sont des sous-types de concept). Un concept peut être utilisé dans les cas suivants:

a) Comme caractéristique: Le concept est utilisé par une variable pour décrire la caractéristique particulière qui doit être mesurée concernant une population. Par exemple, pour mesurer le concept de genre dans une population d'adultes des Pays-Bas, la variable associe ce concept au type d'unité «personne»;

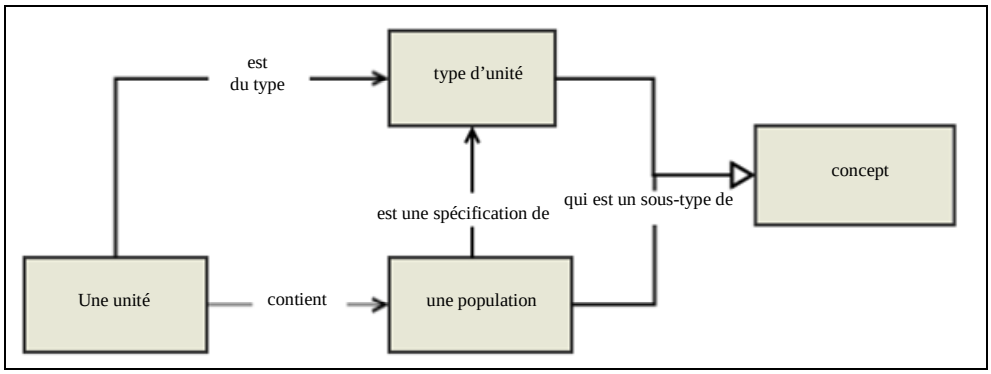
b) Comme type d'unité ou population: Pour décrire l'ensemble d'objets au sujet desquels une information doit être obtenue dans une étude statistique. Par exemple, la population d'adultes aux Pays-Bas, qui se fonde sur le type d'unité «personnes»;

c) Comme catégorie pour décrire un concept de manière plus détaillée. Par exemple, femme et homme pour le concept de genre. Des codes peuvent être liés à une catégorie par un nœud (par exemple un élément de code ou de classification) utilisable dans une liste de codes ou une classification statistique;

54. Les systèmes de concepts sont des ensembles de concepts qui sont structurés par les relations qui les unissent. Un champ d'activité regroupe des systèmes de concepts en fonction du domaine ou de la connaissance spécialisée auquel ils appartiennent (par exemple, le marché du travail, le tourisme).

2. Population

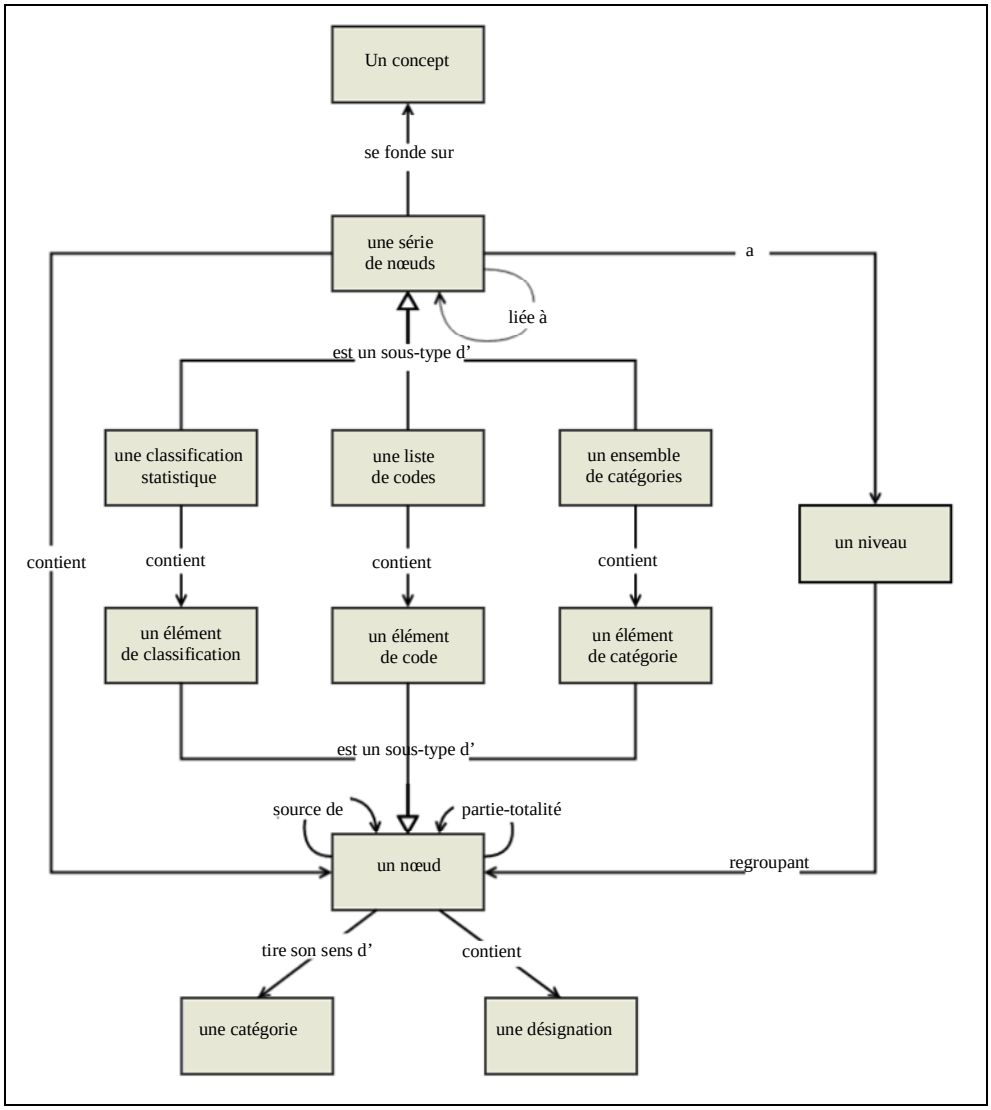
Figure 11
Populations et unités



55. Il existe plusieurs catégories de populations qui dépendent de l'étape de processus dans laquelle elles sont utilisées. Par exemple, un organisme de statistique peut se reporter à une cible, une étude, un cadre ou une population analysée. Les objets qui intéressent un processus statistique sont les unités (par exemple, une personne ou une entreprise donnée). Des données sont recueillies sur ces unités. Une unité est spécifiée dans le modèle de deux façons. C'est une entité individuelle associée à une population sur laquelle une information peut être obtenue. Un type d'unité (par exemple, des personnes ou des entreprises) est une façon d'identifier un type abstrait d'unité mesuré par une variable.

3. Nœuds et séries de nœuds

Figure 12
Succession de nœuds et de séries de nœuds



56. Une catégorie est un type particulier de concept dont le rôle est de définir une caractéristique. Une catégorie peut être utilisée de trois façons. Dans le GSIM, celles-ci sont décrites sous la forme de trois sous-types de nœuds − un élément de catégorie, un élément de code et un élément de classification. Les catégories se regroupent dans des séries de nœuds selon la manière dont elles peuvent être utilisées. Il existe trois sous-types de ces groupes (séries de nœuds) − ensembles de catégories, listes de codes et classification statistiques.

57. Un ensemble de catégories est un ensemble d'éléments de catégorie qui renferment le sens d'une catégorie sans représentation associée. On peut donner comme exemple d'ensemble de catégories: de sexe masculin, de sexe féminin.

58. Dans une liste de codes, les éléments de code renferment le sens des catégories associé à une représentation de code. On peut donner comme exemple de code de représentation: de sexe masculin, de sexe féminin

59. Une classification statistique est semblable à une liste de codes. Elle associe le sens d'une catégorie avec une représentation de code. Le contenu d'une classification statistique doit toutefois remplir certains critères et avoir un certain statut. Les éléments de classification doivent être mutuellement exclusifs et conjointement exhaustifs pour le niveau auquel ils figurent dans la classification statistique. On peut citer comme exemple de classification statistique 1. de sexe masculin, 2. de sexe féminin, 3. transsexuel.

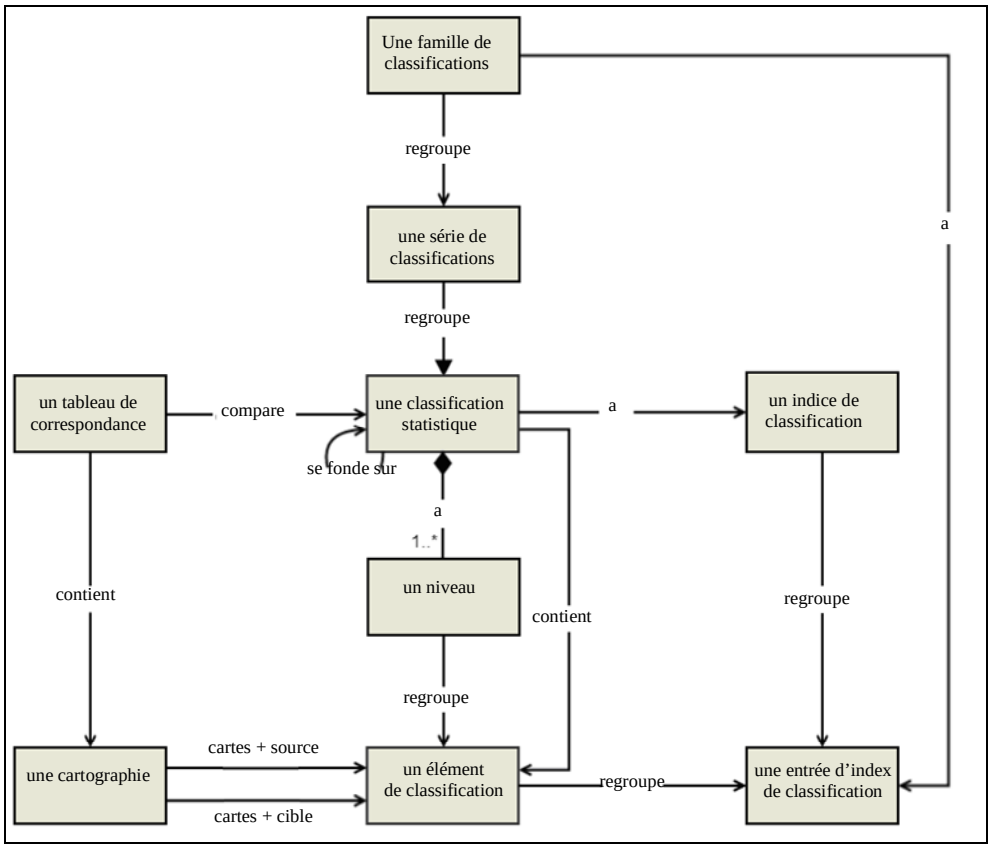
60. Une liste de codes n'a pas à satisfaire les mêmes critères que la classification statistique. La liste de codes peut aussi contenir des éléments de code additionnels pour appuyer une utilisation particulière de la liste de codes, comme l'incorporation de valeurs manquantes.

61. Les similitudes entre les classifications statistiques, les listes de codes et les ensembles de catégories sont intégrées à leur lien (en tant que sous-types) à la série de nœuds. De même, les trois types d'élément qui constituent chaque groupe (élément de classification, élément de code et élément de catégorie respectivement) sont des sous-types de nœud.

4. Classification statistique

62. La présente section décrit une classification statistique et ses objets de gestion connexes pour montrer un aspect particulier de la partie du GSIM qui intéresse la série de nœuds. On trouvera des informations plus détaillées sur les classifications statistiques en particulier dans le modèle de classification statistique du GSIM:

Figure 13
Classifications statistiques



63. La figure ci-dessus donne un aperçu des objets liés aux classifications statistiques.

64. Une famille de classifications est un groupe de séries de classifications connexes fondées sur un concept commun (par exemple, une activité économique). Une série de classifications est un ensemble d'une ou de plusieurs classifications statistiques qui se fondent sur le même concept. Les classifications statistiques d'une série de classifications se rattachent l'une à l'autre sous forme de versions ou d'actualisations. En général, ces classifications statistiques ont le même nom, par exemple la Classification internationale type, par industrie, de toutes les branches d'activité économique (CITI) ou la Classification internationale type des professions (CITP).

65. Une classification statistique est un ensemble de catégories qui peuvent être assignées à une ou plusieurs variables représentées utilisées dans la production et la diffusion de statistiques. Les catégories à chaque niveau de la structure de classification doivent être mutuellement exclusives et recouvrir conjointement tous les aspects de tous les objets/unités de la population visée. On peut citer comme exemple de classification statistique la CITI Rev.4.

66. Les catégories se définissent par rapport à une ou plusieurs caractéristiques d'une population visée particulière. Une classification statistique peut avoir une structure plane, linéaire, ou peut être structurée hiérarchiquement, de sorte que toutes les catégories à des niveaux inférieurs sont des sous-catégories d'une catégorie du niveau supérieur.

67. Une classification statistique a des catégories qui sont représentées par des éléments de classification. Ces éléments de classification sont organisés en niveaux déterminés par la hiérarchie. Un niveau est une série de concepts qui sont mutuellement exclusifs et conjointement exhaustifs; par exemple, une section, une division, un groupe et une classe de la CITI Rev.4.

68. Un élément de classification associe le sens d'une catégorie, sa représentation (par exemple, un code) et des informations additionnelles afin de satisfaire aux critères de classification statistique, par exemple, «A − agriculture, sylviculture et pêche» et un texte explicatif d'accompagnement comme une information sur ce qui est inclus et exclus.

69. Des classifications statistiques peuvent être des versions ou des variantes. Un type de variante de classification statistique se fonde sur un type de version de classification statistique. Dans une variante, les catégories de la version peuvent être fractionnées, agrégées ou regroupées pour fournir des compléments ou des alternatives à l'ordre et à la structure normalisés de la classification statistique initiale.

70. Un tableau de correspondance est un ensemble de cartes. Ces cartes relient un élément de classification dans une classification statistique à l'élément de classification qui lui correspond dans une autre classification statistique par l'intermédiaire du concept qui est commun aux deux éléments de classification. Par exemple, dans un tableau de correspondance qui illustre la relation entre la CITI Rev.4 et le Système de classification des industries de l'Amérique du Nord (SCIAN) 2007, «0112 − Culture du riz» dans la CITI Rev.4 correspond à «111160 − Culture du riz» dans le SCIAN par le biais du concept commun de «culture du riz».

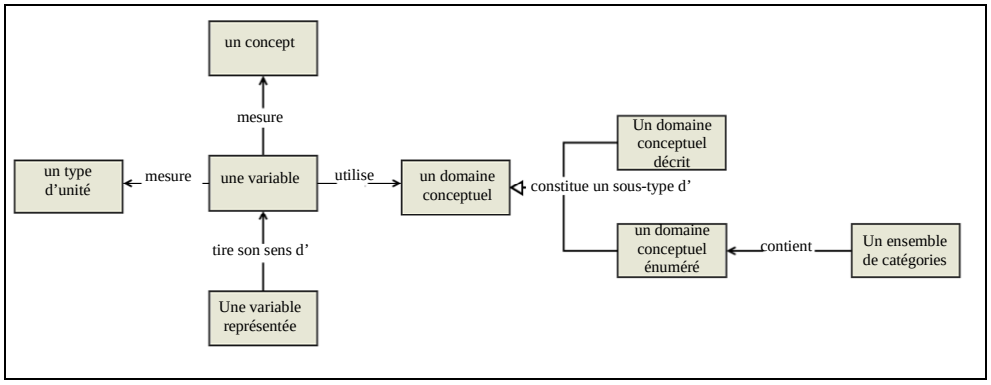
71. Un indice de classification montre la relation entre le texte extrait de sources de données statistiques (réponses à des questionnaires d'enquête, fichiers administratifs) et une ou plusieurs classifications statistiques. Un indice de classification peut être utilisé pour assigner les codes des éléments de classification aux observations dans les programmes statistiques.

72. Une entrée d'indice de classification est un mot ou un bref intitulé (par exemple, le nom d'une localité, d'une activité économique ou un titre fonctionnel) qui décrit un type de concept auquel s'applique un élément de classification, ainsi que le code de l'élément de

classification correspondant. Chaque entrée d'indice de classification se rapporte généralement à un élément de la classification statistique. Bien qu'une entrée d'indice de classification puisse être associée à un élément de classification à n'importe quel niveau d'un classement statistique, elle est normalement associée à des éléments de classification du plus bas niveau.

5. Variable

Figure 14
Variable



73. Lorsqu'il est utilisé dans le cadre d'un processus d'activité, un type d'unité définissant une population est associé à une caractéristique. L'association du type d'unité et d'un concept jouant le rôle d'une caractéristique est appelée variable (voir la figure 14). Par exemple, si la population est celle des adultes aux Pays-Bas, alors une variable pertinente pourrait être le concept de degré d'instruction associé au type d'unité «personne».

74. La variable (le degré d'instruction de la personne) ne donne aucune information sur la manière dont la valeur résultante peut être représentée. Cette information (le domaine de valeur) est associée à la variable représentée. Cette distinction favorise la réutilisation d'une définition de variable lorsque ce que l'on mesure est conceptuellement similaire mais représenté d'une façon différente.

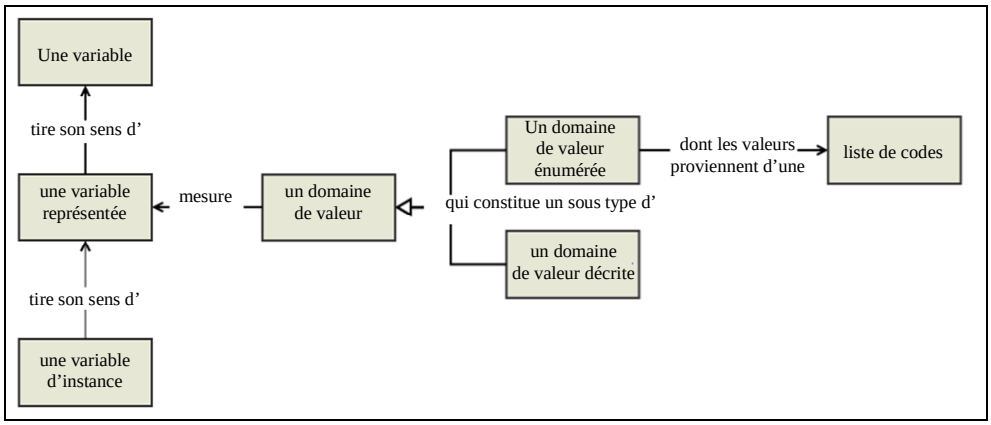
75. Une variable dérivée est créée par une étape de processus qui applique une méthode à un ou plusieurs intrants de processus (variables). Le produit de processus de l'étape de processus est la variable dérivée.

76. Un domaine conceptuel est associé à une variable. Il a deux sous-types: le domaine conceptuel décrit et le domaine conceptuel énuméré. Un domaine conceptuel énuméré, en conjonction avec un ensemble de catégories, contient des informations sur la sémantique des catégories utilisées par la variable.

6. Variable représentée

77. Le GSIM aide les utilisateurs à comprendre à la fois le sens et la représentation concrète sous forme de données de l'objet. Ainsi, il distingue le niveau conceptuel du niveau de la représentation dans le modèle pour faire la différence entre les objets utilisés pour décrire conceptuellement l'information et les objets qui assurent la fonction de représentation.

Figure 15
Variable représentée



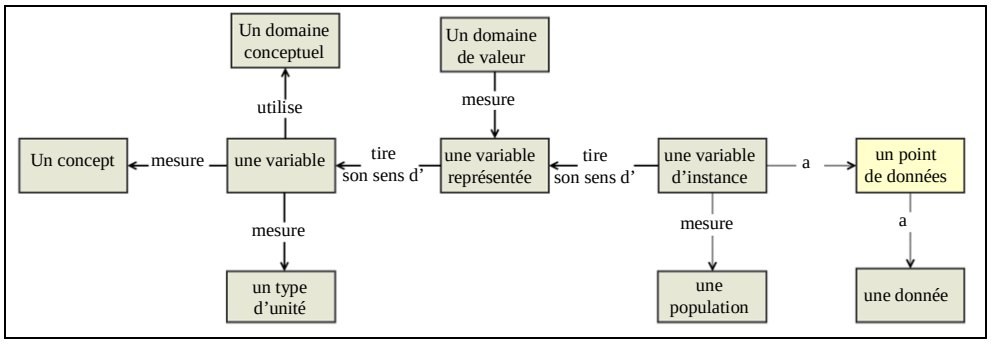
78. La variable représentée (voir la figure 15) ajoute des informations qui décrivent la manière dont les valeurs résultantes peuvent être représentées en association avec un domaine de valeur. Si les domaines conceptuels sont associés à une variable, les domaines de valeur sont associés à une variable représentée. Ces deux domaines sont différenciés en ce sens que le GSIM sépare l'aspect sémantique (domaine conceptuel) et la fonction de représentation (domaine de valeur).

79. Le domaine de valeur énumérée et le domaine de valeur décrite (les deux sous-types du domaine de valeur) donnent des informations sur la manière dont la variable représentée est représentée. Le domaine de valeur énumérée le fait en conjonction avec une liste de codes, tandis que le domaine de valeur décrite définit comment former les valeurs autrement que par une liste expresse.

80. Le domaine de valeur comprend les informations sur le type de donnée et l'unité de mesure. Le type de donnée contient des informations sur les calculs que l'on a le droit d'effectuer sur la donnée (donnée nominative, donnée ordinale, donnée d'intervalle, etc.) tandis que l'unité de mesure (tonnes, nombre de …, dollars, etc.) précise la mesure du domaine de valeur. Par exemple, les codes relatifs au sexe donnent lieu à des données statistiques nominatives tandis que l'âge exprimé en années donne lieu à des données d'intervalle.

6. Variable d'instance

Figure 16
Variable d'instance

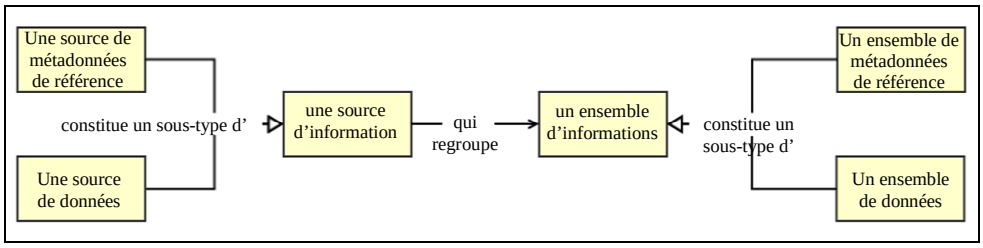


81. Une variable d'instance (voir la figure 16) est une variable représentée qui a été associée à un ensemble de données. Cela peut correspondre à une colonne de données dans une base de données. Par exemple, l'«âge de tous les présidents des États-Unis soit maintenant (s'ils sont vivants) soit l'âge qu'ils avaient à leur mort» est une colonne de données décrite par une variable d'instance, qui est la combinaison de la variable représentée décrivant «l'âge de la personne» et du domaine de valeur des «nombres naturels décimaux (en années)».

82. Une donnée est contenue par un point de données dans un ensemble de données. Elle peut se définir par la mesure d'un domaine de valeur associé à une variable d'instance descriptive combinée au lien avec une unité (pour les données unitaires) ou une population (pour les données dimensionnelles).

8. Sources d'information

Figure 17
Sources d'information



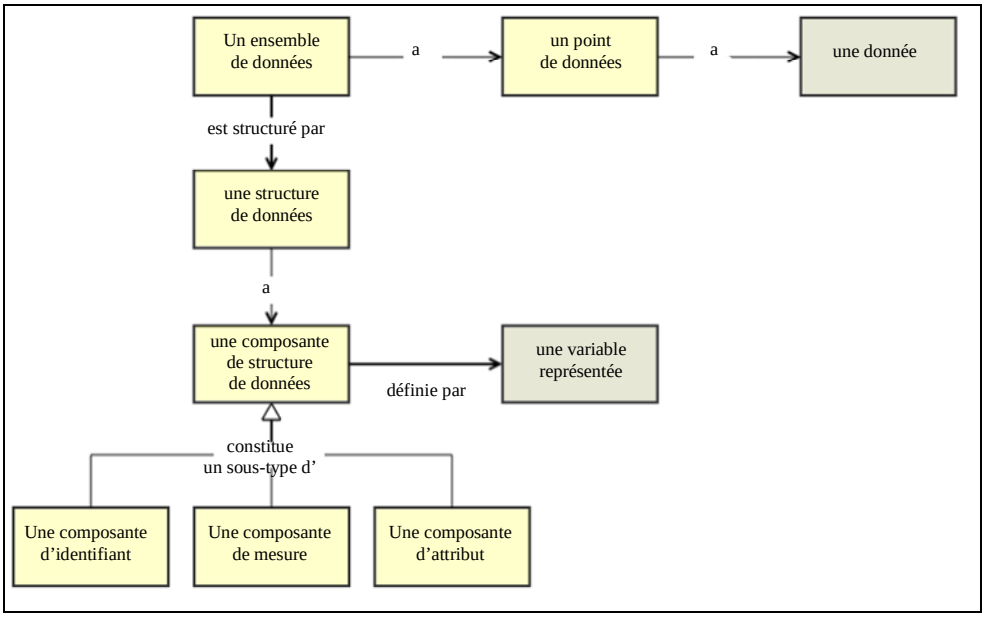
83. Les organismes de statistique recueillent, traitent, analysent et diffusent des ensembles d'informations, qui sont soit des données (ensembles de données) soit des métadonnées de référence (ensembles de métadonnées de référence).

84. Chaque ensemble de données doit être structuré suivant une structure de données (par exemple, une structure pour la balance des paiements, la démographie, le tourisme, l'éducation, etc.). De la même manière, un ensemble de métadonnées de référence doit être structuré suivant une structure de métadonnées de référence (par exemple, le cadre de qualité d'une organisation).

85. Les sources d'information contiennent des ensembles d'informations. Ces derniers ont pour principal objectif d'aider à trouver et gérer les ensembles d'informations en fournissant des informations sur la localisation et autres intéressant ces tâches. Il existe deux types de sources d'information. Les sources de données contiennent des ensembles de données, et les sources de métadonnées de référence contiennent des ensembles de métadonnées de référence.

9. Ensembles de données

Figure 18
Ensembles de données



86. Un ensemble de données a des points de données. Un point de données est un espace vacant (par exemple, une case vide dans un tableau) dans un ensemble de données pour une donnée. La donnée est la valeur qui vient remplir cet espace (par exemple, un élément d'information factuel obtenu par une mesure ou créé par un processus de production). Une structure de données décrit la structure d'un ensemble de données au moyen de composantes de structure de données (composantes d'identifiant, composantes de mesure et composantes d'attributs). Celles-ci sont toutes des variables représentées dotées de rôles spécifiques.

87. Les ensembles de données prennent différentes formes, par exemple, celles de registres administratifs, de séries chronologiques, de données de panel, ou de données de survie, pour ne citer que quelques exemples. Le type d'ensemble de données détermine la série d'attributs spécifiques à définir, le type de structure de données requis (structure de données unitaires ou structure de données dimensionnelles) et les méthodes applicables aux données.

88. Par exemple, un registre administratif est caractérisé par une structure de données unitaires, avec des attributs comme son but initial ou la date de mise à jour la plus récente de chaque enregistrement. Il renferme une variable d'identification des enregistrements, et peut être utilisé pour définir une population utilisée comme cadre, pour remplacer ou compléter les enquêtes existantes, ou comme intrant auxiliaire de l'imputation. La mise en correspondance des enregistrements est un exemple de méthode particulièrement adaptée aux registres.

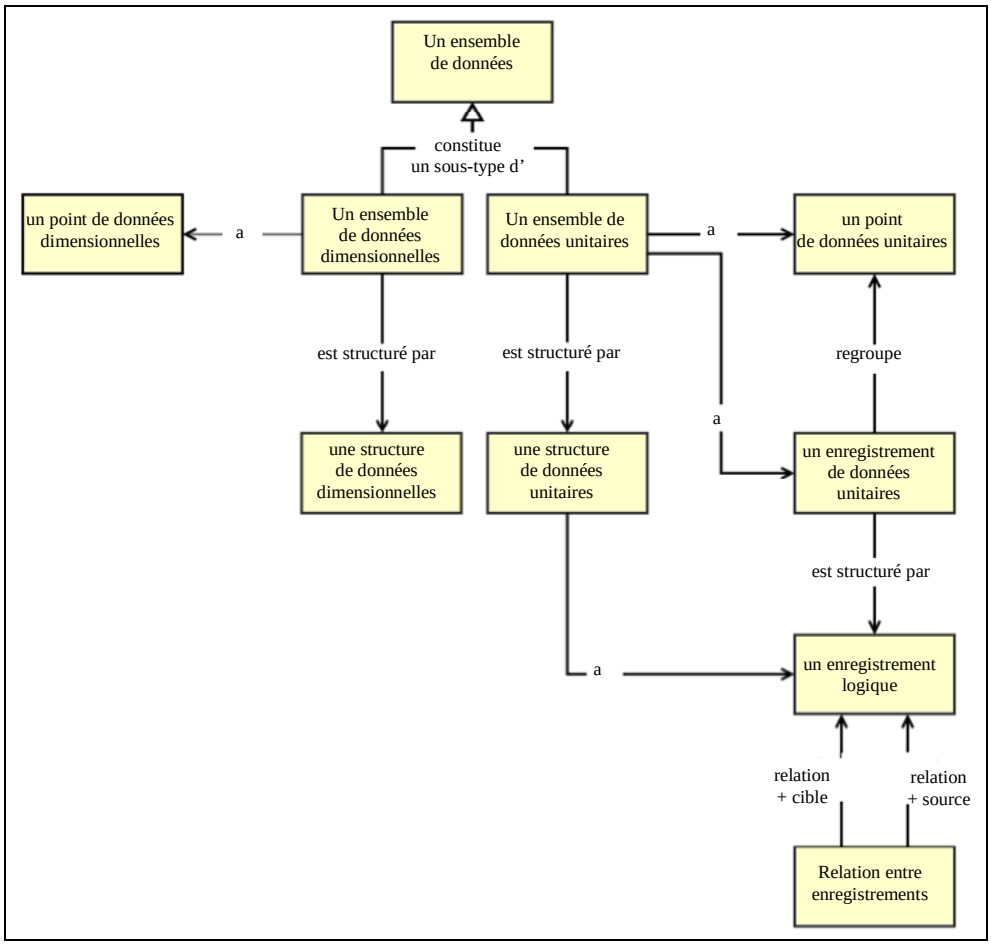
89. La série chronologique peut être citée comme exemple de type d'ensemble de données défini par une structure de données dimensionnelles. Elle a des attributs spécifiques comme la fréquence et le type d'agrégation temporelle et des méthodes spécifiques, par exemple, la correction des variations saisonnières, et doit contenir une variable temporelle.

90. Les données unitaires et les données dimensionnelles sont des perspectives sur les données. Bien que ce ne soit pas généralement le cas, le même ensemble de données pourrait être décrit des deux façons. Parfois, ce qu'un organisme considère comme des données dimensionnelles (par exemple, un bureau national de statistique) peut être considéré comme des données unitaires par d'autres (par exemple, Eurostat où l'unité est un État membre). Une collecte de données spécifique n'a pas besoin d'être considérée par principe comme faisant partie de l'une ou l'autre catégorie. Cette question de perspective est conceptuelle. La distinction opérée dans le GSIM tient au fait que l'ensemble de données unitaires contient des données sur les unités et l'ensemble de données dimensionnelles contient des données sur les unités ou sur les populations.

91. Le GSIM indique que tous les ensembles de données doivent avoir une structure associée. Il existe toutefois des cas où un ensemble de données n'en a pas − parce qu'elle n'a pas été archivée, ou parce qu'elle a été perdue, ou que l'on ne la connait pas. Les données de ce type pourraient devenir plus courantes pour les organismes de statistique à l'avenir. Pour qu'un organisme de statistique utilise ces données, celles-ci devront subir un processus de structuration. Par exemple, lorsque l'on recherche de nouvelles sources de données potentielles pour un nouveau besoin statistique, ou un besoin statistique qui a changé, il faudra mettre en place un processus par lequel ces nouvelles données seront analysées pour déterminer leur contenu et leur structure. Ce n'est qu'après ce processus que ces nouveaux ensembles de données pourront être décrits au moyen des objets de la structure de données. Ces données non structurées sont actuellement décrites par le GSIM sous la forme d'intrants de processus. Les organismes pourraient étendre le GSIM à cette utilisation en créant un nouveau sous-type de l'objet ensemble d'informations.

10. Structures des données dimensionnelles et unitaires

Figure 19
Structures de données



92. Une structure de données dimensionnelles décrit la structure d'un ensemble de données dimensionnelles au moyen de variables représentées dotées de rôles spécifiques.

93. La combinaison de dimensions figurant dans la structure de données dimensionnelles crée une clef ou un identifiant des valeurs mesurées. Par exemple, le pays, l'indicateur, l'unité de mesure, la fréquence et le facteur temps identifient ensemble les articles d'une série chronologique internationale (par exemple, le produit intérieur brut, l'endettement intérieur brut) mesurée dans des unités différentes (par exemple, devises différentes, changements en pourcentage) et à différents intervalles (par exemple, annuels, trimestriels). Les cases d'un tableau aussi pluridimensionnel contiennent des valeurs d'observation.

94. Une mesure est la variable qui fournit un contenant à ces valeurs d'observation. Elle emprunte sa sémantique à un sous-ensemble des dimensions de la structure des données dimensionnelles. Dans l'exemple précédent, on peut considérer que l'indicateur et l'unité de mesure sont les dimensions qui fournissent cette sémantique, alors que la fréquence et le temps sont les dimensions temporelles et le pays la dimension géographique. Outre la simple «valeur d'observation», on peut citer en exemple de mesure une «valeur de l'observation avant rupture» dans le cas des séries chronologiques. Les dimensions décrivent généralement les variables représentées au moyen de domaines

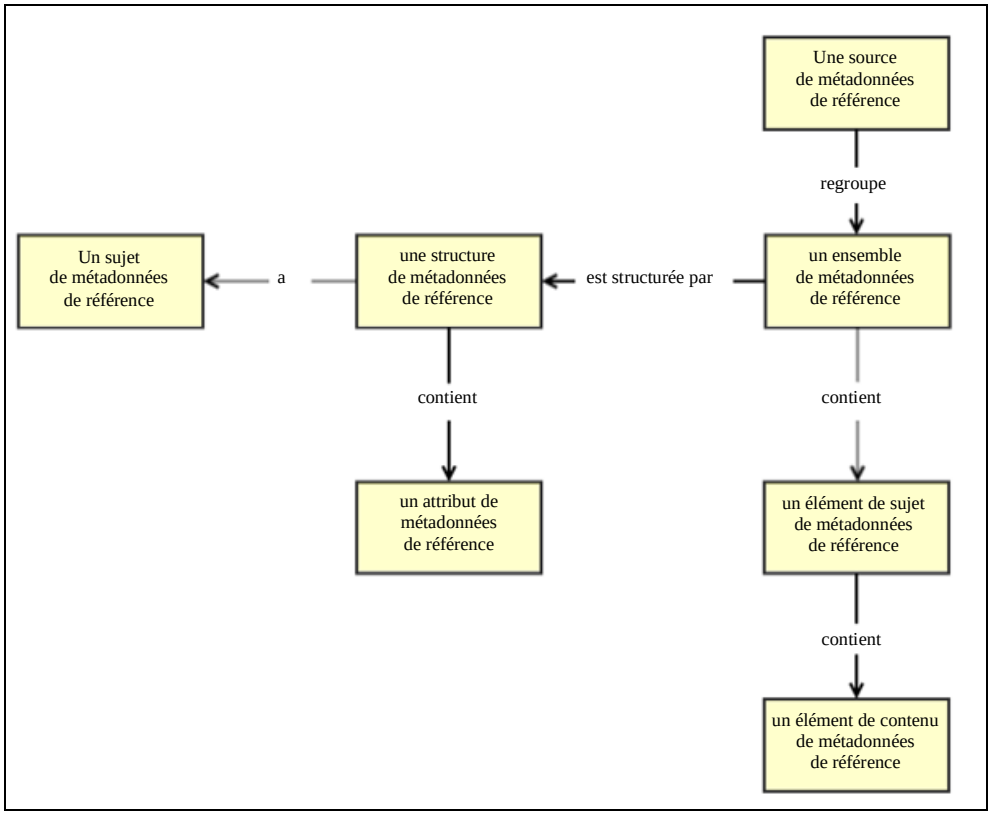
de valeurs codés (domaines de valeur énumérée) et les mesures des variables représentées au moyen de domaines de valeurs non codées (domaines de valeur décrite).

95. Une structure de données unitaires décrit la structure d'un ensemble de données unitaires au moyen de variables représentées dotées de rôles spécifiques. Elle établit une distinction entre la structure logique et la structure physique de l'ensemble de données. Un ensemble de données unitaires peut contenir des données sur plus d'un type d'unité, chacune représentée par son propre type d'enregistrement.

96. Les enregistrements logiques décrivent la structure de ces types d'enregistrement, indépendamment des caractéristiques physiques, en se rapportant aux variables représentées qui peuvent comprendre une identification d'unité (par exemple, le numéro du ménage). Une relation d'enregistrement définit les relations source-cible entre les enregistrements logiques.

11. Ensembles de métadonnées de référence

Figure 20
Ensembles de métadonnées de référence



97. Les informations qui décrivent les caractéristiques des statistiques sont des «métadonnées de référence». Ces métadonnées peuvent être étendues, comme celles décrivant un programme de statistique tout entier, ou étroites, comme celles décrivant un point de données particulier. Les sources de métadonnées de référence, un type spécial de source d'information, fournissent des contenants de haut niveau pour les métadonnées de référence.

98. Un ensemble de métadonnées de référence organise les métadonnées de référence dont la structure est définie dans une structure de métadonnées de référence, laquelle

spécifie à la fois le sujet des métadonnées de référence pour lequel des métadonnées de référence peuvent être incorporées, et une liste structurée d'attributs de métadonnées de référence qui peut être rapporté ou conçu pour le sujet de métadonnées de référence dont il est question.

99. Ces sujets peuvent être n'importe quel type d'objet du GSIM, ou n'importe quel point de données, ou un ensemble de points de données créé à partir d'une structure de données particulière.

- Exemple de type d'objet du GSIM qui peut être un sujet de métadonnées de référence: un produit pour lequel une liste est spécifiée dans un domaine de valeur. Le domaine de valeur spécifie la liste des produits effectifs pour lesquels des métadonnées de référence peuvent être rapportées ou conçues au moyen de cette structure de métadonnées de référence;
- On peut citer comme exemples d'attributs de métadonnées de référence le rang, le champ d'application, la description de la méthodologie et l'indicateur de qualité.

100. Un ensemble de métadonnées de référence contient les métadonnées de référence effectivement rapportées ou conçues. L'élément de sujet de métadonnées de référence identifie l'objet dont il est effectivement question, par exemple le produit effectif comme la balance des paiements et la position de l'investissement international, Australie, juin 2013, ou des points de données effectifs comme les points de données pour une seule région dans un ensemble de données visant toutes les régions d'un pays.

101. L'élément de contenu de métadonnées de référence est la métadonnée effective pour l'élément de sujet de métadonnée de référence. Chaque élément de contenu de métadonnée de référence contient la métadonnée de référence rapportée pour un attribut de métadonnée de référence spécifié dans la structure de métadonnée de référence.

Tableau 1
Exemple d'utilisation d'objets «métadonnées de référence» du GSIM

Objet du GSIM	Bulletin statistique de l'ONS: finances du secteur public, octobre 2013: **Tableau 1**
Structure de métadonnées de référence	Implicite
Sujet de métadonnées de référence	Composante des métadonnées de référence
Attribut de métadonnées de référence	Note de bas de page du tableau
Ensemble de métadonnées de référence	Notes de bas de page
Élément de sujet de métadonnées de référence	*Composante de structure de données*: milliard de livres sterling, budget ordinaire du secteur public; budget courant du secteur public ex APF
Élément de contenu de métadonnées de référence	Texte figurant dans les notes de bas de page