



N° 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

Symposium 2003 : Défis reliés à la réalisation d'enquêtes pour la prochaine décennie

2003



Statistique
Canada

Statistics
Canada

Canada

Recueil du Symposium 2003 de Statistique Canada
Défis reliés à la réalisation d'enquêtes pour la prochaine décennie

MODÈLE CONCEPTUEL POUR LES MÉTADONNÉES DÉFINITIONNELLES D'UN ORGANISME STATISTIQUE

Kaveri Mechanda, Paul Johanis et Michael Webber¹

RÉSUMÉ

La présente communication porte sur le Cadre d'établissement de normes relatives aux domaines spécialisés, modèle conceptuel pour les métadonnées définitionnelles d'un organisme statistique. Ce cadre montre un tableau global des *unités statistiques* et des *variables* utilisées à Statistique Canada. Il a pour but de faire ressortir les domaines qui nécessitent de nouvelles normes. Il sert également à organiser les métadonnées et à structurer les définitions dans la Base de métadonnées intégrée (BMDI) de Statistique Canada. Ce modèle emprunte les concepts de la modélisation orientée objet et harmonise les métadonnées conceptuelles de Statistique Canada avec la norme ISO/IEC 11179. La présente communication décrit l'élaboration du Cadre d'établissement de normes relatives aux domaines spécialisés et comprend des graphiques montrant le modèle de base.

MOTS CLÉS : ISO/IEC 11179, métadonnées, unités statistiques, variables.

1. CONTEXTE

La nécessité de normes concernant les métadonnées, et en particulier l'uniformité des définitions dans la foule de domaines spécialisés d'un organisme statistique, est aujourd'hui plus urgente que jamais. Depuis que les données accessibles par machine sont devenues largement disponibles au cours des années 60, on réutilise et on traite de plus en plus les données, d'où la nécessité de métadonnées plus abondantes et de meilleure qualité (Norbotten, 1993). Plus récemment, grâce à la disponibilité de divers types de données et d'information sur Internet, les utilisateurs peuvent consulter et comparer des données d'un domaine spécialisé à celles d'un autre. Ceci a fait ressortir la nécessité de définitions uniformes dans les domaines spécialisés, dans le temps et à l'échelle des organismes. De plus, en raison des départs à la retraite et d'autres changements au sein du personnel d'un organisme statistique, la bonne organisation des métadonnées devient un important moyen de transfert du savoir. L'évolution rapide de la technologie nécessite un modèle conceptuel indépendant de la technologie. L'utilisation répandue de données statistiques, le roulement de personnel et l'évolution de la technologie rendent nécessaire la mise en place d'un modèle de métadonnées conceptuelles au sein de l'organisme statistique.

Les organismes statistiques créent des métadonnées statistiques; ces métadonnées sont nécessaires à la production et à l'utilisation pertinentes des données statistiques. Les utilisateurs de métadonnées statistiques ont des besoins divers en matière d'interprétation et d'utilisation des métadonnées. Sundgren définit trois principaux types d'information que les données rendent accessibles (Struijs 1999, Sundgren 1993).

Ce sont les suivants :

- Une information pragmatique qui renseigne l'utilisateur sur l'objet et l'utilisation des données. Exemple : des renseignements sur la façon d'utiliser un produit.
- Une information sémantique qui renseigne l'utilisateur sur le sens et le contenu des données. Exemple : des définitions des variables.
- Une information syntaxique qui renseigne l'utilisateur sur les aspects matériels et techniques des données. Exemple : des représentations des données et des processus de traitement des données.

¹ Kaveri Mechanda, analyste principal, Division de la géographie, Statistique Canada, Ottawa, Canada K2E 7P8; Paul Johanis, directeur, Division des normes, Statistique Canada, Ottawa, Canada K2E 7P8.

Pour les besoins du Cadre d'établissement de normes relatives aux domaines spécialisés, on considère les métadonnées définitionnelles comme un type d'information sémantique sur la terminologie des métadonnées statistiques.

Les organismes statistiques ont le mandat de produire des données pour le bien public. Les concepts doivent donc être en accord avec l'intérêt du public. L'approche présentée ici consiste à modéliser les métadonnées du point de vue de l'utilisateur ou, plus précisément, en fonction des données qui sont diffusées plutôt qu'en fonction de celles qui sont visées par le processus de collecte de données d'un organisme statistique. Le raisonnement est le suivant : les métadonnées définitionnelles doivent être axées sur l'utilisateur, tout en étant coordonnées avec le processus de collecte de données de l'organisme statistique.

Dans le cadre de leurs efforts de création de métadonnées statistiques, les organismes statistiques ainsi que des organismes nationaux et internationaux ont adopté des concepts et un cadre axés sur cette activité. Bien qu'elle ne porte pas expressément sur les métadonnées statistiques, la norme ISO/IEC 11179, Partie 1, *Cadre pour la spécification et la normalisation des éléments de données*, a inspiré les organismes statistiques, qui l'ont utilisée. De plus en plus répandue dans les systèmes de modélisation, la modélisation orientée objet est devenue un outil évolué pour les métadonnées statistiques. On a beaucoup travaillé à la conception du stockage et du traitement des métadonnées, et notamment des métadonnées pertinentes aux données recueillies au moyen d'enquêtes. On s'est moins attardé à la conception (globale) de métadonnées définitionnelles pour la diffusion des données.

Statistique Canada s'emploie à mettre dans la Base de métadonnées intégrée (BMDI) de l'information sur les enquêtes ainsi que les métadonnées définitionnelles connexes. Ce registre de métadonnées est mis à jour quotidiennement avec l'ajout de métadonnées concernant les données publiées par Statistique Canada le même jour. Il permet aux utilisateurs d'accéder directement aux métadonnées sur Internet. Comme cette base de données servira à stocker toutes les métadonnées concernant les données publiées sur Internet, on a besoin d'un cadre conceptuel uniforme pour soutenir et normaliser ses métadonnées définitionnelles.

La Politique concernant les normes de Statistique Canada a pour objectif de normaliser les métadonnées définitionnelles. Elle touche l'examen, la documentation, l'autorisation et le contrôle de l'utilisation des définitions normalisées des concepts, des *variables* et des *classifications* utilisés dans les domaines spécialisés ainsi que des *populations* et des *unités statistiques* auxquelles ils s'appliquent (Statistique Canada, 1988). Les normes sont établies conformément à cette politique. Une approche systématique et globale de l'établissement de nouvelles normes doit s'appuyer sur un modèle conceptuel pour toutes les métadonnées définitionnelles de Statistique Canada.

Le Cadre d'établissement de normes relatives aux domaines spécialisés offre un cadre conceptuel pour les métadonnées définitionnelles d'un organisme statistique. Il soutient la BMDI ainsi que l'établissement de nouvelles normes. Il sera enrichi par les métadonnées provenant de la BMDI. Puis, les métadonnées qu'il contient seront évaluées en fonction de la nécessité d'harmoniser et de structurer les définitions et du besoin d'élaborer des classifications normalisées. Les nouvelles normes seront enregistrées et rendues accessibles au moyen de la BMDI. Ainsi, par le biais de la BMDI, le Cadre d'établissement de normes relatives aux domaines spécialisés constituera un moyen systématique pour réaliser l'objectif de normalisation des métadonnées définitionnelles.

2. QU'EST-CE QUE LE CADRE D'ÉTABLISSEMENT DE NORMES RELATIVES AUX DOMAINES SPÉCIALISÉS?

Le Cadre d'établissement de normes relatives aux domaines spécialisés a pour but de constituer un modèle conceptuel cohérent et complet pour les métadonnées définitionnelles utilisées à Statistique Canada. Ce cadre sert à montrer dans quels domaines il faut des normes concernant les métadonnées définitionnelles à Statistique Canada. Il modélise l'information définitionnelle (les définitions dont les utilisateurs ont besoin) plutôt que l'information syntaxique (celle dont les machines ont besoin). Le fait d'offrir une vue d'ensemble complète des métadonnées définitionnelles est considéré comme une première étape dans l'harmonisation des métadonnées.

Le cadre a été créé en fonction des principes suivants :

- Efficacité – le modèle doit être utilisable et renseigner sur les métadonnées utilisées à Statistique Canada. Par exemple, il doit être aisément applicable au contexte de la BMDI et révéler dans quels domaines il y a lieu de poursuivre la normalisation.
- Exhaustivité – le modèle doit pouvoir comprendre toutes les *unités statistiques* et tous les concepts que couvre l'information publiée à Statistique Canada.
- Efficience – le modèle doit être aussi simple et aussi clair que possible.

2.1 Influences sur le cadre

La structure du cadre est inspirée des sources suivantes :

- Politique concernant les normes de Statistique Canada
- Ouvrages sur la modélisation orientée objet
- Norme ISO/IEC 11179

On a appliqué à l'organisation de l'information et des métadonnées statistiques les modes d'ordonnement systématique de l'information et des métadonnées selon le paradigme orienté objet et la norme ISO/IEC 11179. La modélisation orientée objet a été conçue pour les applications de systèmes, et la norme ISO/IEC 11179 a été établie à l'origine pour l'échange de données. Statistique Canada a emprunté des idées de ces deux modes de modélisation pour créer son cadre. L'annexe A présente les équivalences des termes utilisés dans la norme ISO/IEC 11179 et dans les ouvrages sur la modélisation orientée objet. Comme la BMDI de Statistique Canada suit le modèle de la norme ISO/IEC 11179, il importe de coordonner les équivalences des termes utilisés à Statistique Canada avec les normes et les concepts de la modélisation orientée objet.

Le cadre modélise les métadonnées définitionnelles identifiées dans la Politique concernant les normes de Statistique Canada. À l'heure actuelle, il modélise les *unités statistiques* et les *variables*. Par la suite, il englobera les *classifications* et les *populations*.

2.2 Contenu du cadre

L'organisation des métadonnées à l'intérieur du modèle est délibérément influencée par la conception et l'organisation des domaines statistiques spécialisés à Statistique Canada. On a procédé ainsi afin que le cadre repose sur une conception reconnue des domaines spécialisés et qu'il soit indépendant de la technologie.

À l'heure actuelle, le cadre comporte une répartition des *unités statistiques* et une répartition des *variables*. La répartition des unités statistiques comporte quatre points de vue qui divisent le modèle en quatre sections. Ces quatre sections sont fondées sur différents aspects du domaine spécialisé. Le cadre définit également différents types d'unités statistiques qui donnent une structure à la définition de la relation entre les unités statistiques. Quant à la répartition des variables, elle montre comment les variables sont (ou pourraient être) organisées.

À l'heure actuelle, cet exercice de modélisation se limite à l'information produite et publiée régulièrement par Statistique Canada. Par exemple, les données provenant d'autres organismes et hébergées sur le site Web de Statistique Canada, ainsi que l'information présentée dans des articles analytiques, n'ont pas été retenues dans cet exercice de modélisation. Toutefois, les données recueillies par d'autres organismes et qui sont traitées et publiées par Statistique Canada (c.-à-d. certaines données administratives et environnementales) sont comprises dans cet exercice.

3. UNITÉS STATISTIQUES ET POINTS DE VUE COMPRIS DANS LE CADRE

La Politique concernant les normes de Statistique Canada définit les *unités statistiques* comme suit : « unité d'observation ou mesure pour laquelle on recueille ou calcule des données » (Statistique Canada, 1988).

L'organisation de ces *unités statistiques* proposée par le cadre est présentée ci-dessous.

3.1 Points de vue compris dans le cadre

Pour couvrir la totalité des données et de l'information publiées par Statistique Canada, le cadre tient compte de différents points de vue. On a choisi quatre *macro-unités statistiques* pour tenir compte de quatre points de vue différents à l'intérieur du modèle :

- La société
- L'économie
- L'environnement
- L'État

Ces *macro-unités statistiques* sont très semblables aux quatre grands secteurs pour lesquels Statistique Canada diffuse régulièrement des données. De plus, elles offrent quatre perspectives différentes, puisque le point de vue « la société » présente la perspective des personnes, soit un point de vue sociologique, le point de vue « l'économie » présente la perspective des unités commerciales et des unités économiques, soit un point de vue économique; le point de vue « l'environnement » présente la perspective du capital naturel; le point de vue « l'État » vise à renseigner sur la gouvernance au Canada. Afin de représenter de manière holistique les activités de gouvernance, la section « l'État » comprend la prestation, par le secteur privé, de services autorisés par l'État.

On peut considérer ces *macro-unités statistiques* comme la base sur laquelle on publie des données sur les *unités statistiques*, ou encore comme des agrégations d'*unités statistiques fondamentales*.

3.2 Unités statistiques fondamentales et stéréotypes

Dans le but de rendre le modèle simple et robuste, on a limité le nombre d'*unités statistiques*. On a adopté l'approche de la modélisation orientée objet pour définir les types d'*unités statistiques*. Un objet est « un concept, une abstraction ou une chose dont l'étendue et le sens sont bien définis à l'égard du problème à régler » (Rumbaugh, Blaha, Premerlani, Frederick et Lorenson, 1991). Dans cet exercice, on a réduit au minimum le nombre d'objets utilisés pour caractériser les données publiées par Statistique Canada. À cette fin, on a d'abord cerné un petit nombre d'*unités statistiques fondamentales*. On a également modélisé des dérivations couramment utilisées (sous-ensembles, groupements, etc.) d'*unités statistiques fondamentales*.

On entend par *unités statistiques fondamentales* celles qui ne constituent pas des types d'une autre *unité statistique* et qu'on ne peut pas dériver sous forme de groupement d'une autre *unité statistique*.

On a stéréotypé les *unités statistiques fondamentales* pour faire ressortir les différences dans leur comportement en les répartissant comme suit :

- Les *unités statistiques* dont l'évolution ou le fonctionnement fait l'objet de rapports par Statistique Canada sont stéréotypées à titre d'« agents ». Dans le cas du point de vue « la société », une « personne » est un *agent*.
- Les *unités statistiques* qui représentent les actions des *agents* et faisant l'objet de rapports par Statistique Canada sont stéréotypées à titre d'« événements ». On entend par *événements* des faits qui sont discontinus dans le temps (ils surviennent au cours d'une période) et finis (on peut les compter). Dans le cas du point de vue « la société », une « naissance » est un *événement*.
- D'autres *unités statistiques* faisant l'objet de rapports par Statistique Canada, et généralement produites ou gérées par des *agents*, sont stéréotypées à titre d'« éléments ». Dans le cas du point de vue « l'économie », un « produit » est un *élément*.

Le tableau 1, Répartition des unités statistiques, montre des exemples d'*unités statistiques* et de leurs stéréotypes. L'identification des *agents* à l'intérieur de chacune des quatre *macro-unités statistiques* révèle la portée de chaque point de vue; ce dernier tient compte des métadonnées du point de vue de ses agents. Les agents d'une *macro-unité statistique* peuvent interagir avec ceux d'une autre *macro-unité statistique*.

3.3 Dérivations d'unités statistiques fondamentales couramment utilisées

Les *unités statistiques fondamentales* ont pour but de réduire au minimum le nombre d'unités. Toutefois, bon nombre d'*unités statistiques* sont si couramment utilisées qu'il faut les intégrer au cadre. Ce sont toutes des *unités statistiques* qu'on peut dériver d'une façon ou d'une autre des *unités statistiques fondamentales*. Voici trois dérivations courantes de l'*unité statistique fondamentale* :

- *Sous-catégories d'unité statistique* en fonction d'une propriété inhérente. Exemple : « hommes » ou « femmes » utilisés comme *unités statistiques* pour lesquelles on déclare des données sur les « personnes ».
- *Rôles* que peuvent jouer les *unités statistiques*. Exemples : « étudiant », « mère » et « employé » dans le cas des *unités statistiques fondamentales* « personne », ainsi que « producteur », « employeur » et « consommateur » dans le cas des *unités statistiques fondamentales* « unité commerciale ». Les *rôles* diffèrent des *sous-catégories* puisque la même *unité statistique* peut jouer plus d'un *rôle* à la fois. Elle peut jouer un *rôle* et, avec le temps, cesser de le jouer.
- *Agrégations* ou *hiérarchies d'agrégations d'unités statistiques fondamentales*. Par exemple, « famille de recensement » constitue une *agrégation* de « personnes ». « Entreprise », « compagnie », « établissement » et « emplacement » forment une *hiérarchie d'agrégations* d'« unités commerciales ».

Tableau 1
Répartition des unités statistiques

Macro-unité statistique	Unités statistiques fondamentales	Unités statistiques dérivées couramment utilisées
La société (du Canada)	<p><u>Agents</u></p> <ul style="list-style-type: none"> • Personne <p><u>Événements</u></p> <ul style="list-style-type: none"> • Naissance • Décès • Mariage • Divorce • Immigration • Émigration 	<p>Personnes</p> <p>Sous-ensembles courants</p> <ul style="list-style-type: none"> • En fonction de l'âge : Enfants, jeunes, adultes, personnes âgées • En fonction de l'âge et du sexe : Hommes, femmes <p>Rôles courants :</p> <ul style="list-style-type: none"> • Étudiant • Employé • Voyageur • Consommateur • Mère • Père <p>Groupements courants :</p> <ul style="list-style-type: none"> • Ménage • Famille de recensement • Famille économique
L'économie (du Canada)	<p><u>Agents</u></p> <ul style="list-style-type: none"> • Unité institutionnelle • Unité commerciale <p><u>Événements</u></p> <ul style="list-style-type: none"> • Opération économique • Faillite <p><u>Éléments</u></p> <ul style="list-style-type: none"> • Produit • Construction de logements • Emploi • Économie 	<p>Unités institutionnelles</p> <p>Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Secteur des sociétés non financières • Secteur des sociétés financières • Secteur des administrations publiques • Secteur des institutions sans but lucratif • Secteur des ménages <p>Rôles courants :</p> <ul style="list-style-type: none"> • Producteur • Employeur • Consommateur <p>Unités commerciales</p> <p>Rôles courants :</p> <ul style="list-style-type: none"> • Producteur • Employeur • Consommateur <p>Hiérarchies d'agrégations :</p> <ul style="list-style-type: none"> • Entreprise • Compagnie • Établissement • Emplacement

<p>L'environnement (du Canada)</p>	<p><u>Agent</u></p> <ul style="list-style-type: none"> • Nature <p><u>Événements</u></p> <ul style="list-style-type: none"> • Accident environnemental • Qualité des écosystèmes <p><u>Éléments :</u></p> <ul style="list-style-type: none"> • Ressources agricoles • Ressources forestières • Ressources marines • Réserves aquatiques • Ressources énergétiques • Ressources minérales • Ressources fauniques • Ressources atmosphériques 	<p>Accidents environnementaux – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Déversement de matières dangereuses • Incendie de forêt <p>Qualité des écosystèmes – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Niveau d'ozone au sol supérieur à l'objectif <p>Ressources minérales – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Nickel • Potasse <p>Ressources énergétiques – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Gaz naturel • Pétrole brut <p>Ressources forestières – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Bois d'œuvre
<p>L'État</p>	<p><u>Agents</u></p> <ul style="list-style-type: none"> • L'État <p><u>Événements</u></p> <ul style="list-style-type: none"> • Permis • Demande de prestations • Admission • Infraction • Accusation 	<p>L'État – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Organisme gouvernemental • Établissement d'enseignement • Établissement de justice • Établissement de santé <p>Permis – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Permis de bâtir <p>Demande de prestations – Sous-ensembles courants :</p> <ul style="list-style-type: none"> • Demandes de prestations d'assurance-chômage

La répartition des unités statistiques montre comment les *unités statistiques* sont organisées dans le Cadre d'établissement de normes relatives aux domaines spécialisés (voir le tableau 1). La première colonne (à gauche) montre les quatre points de vue correspondant aux *macro-unités statistiques* : la société, l'économie, l'environnement et l'État. La colonne suivante montre les *unités statistiques fondamentales* liées à chaque point de vue, avec des exemples pour les stéréotypes *agents*, *événements* et *éléments*. Si le graphique montre l'organisation des unités statistiques dans le cadre, il ne montre pourtant pas encore la totalité des *unités statistiques*. L'inclusion des données de la BMDI devrait permettre d'élargir et d'améliorer le contenu du modèle. Les *unités statistiques* pour lesquelles il existe des définitions normalisées sont en caractères gras; les autres *unités statistiques* ne font pas encore l'objet de définitions normalisées.

4. VARIABLES COMPRISES DANS LE CADRE

Dans ce cadre, les *variables* sont organisées selon une structure hiérarchique : on trouve la *variable* au bas de la hiérarchie, les *concepts* au milieu et le *groupe* au sommet de la hiérarchie. Ces termes sont expliqués ci-dessous. Un sujet d'études général comme les « caractéristiques ethno-culturelles » constitue un exemple de *groupe*. Un *concept* lié à ce *groupe* serait la « langue ». Pour une personne, les *variables* liées à ce concept seraient la « langue maternelle », la « langue parlée à la maison » et les « langue(s) officielle(s) parlée(s) ».

4.1 Variables

La Politique concernant les normes de Statistique Canada définit la *variable* comme suit : *caractéristique ou attribut particulier d'une unité statistique* (Statistique Canada, 1988). Le glossaire des termes de classification des Nations Unies en donne une définition un peu plus élaborée :

Caractéristique d'une unité observée qui peut supposer plus d'un ensemble de valeurs auquel on peut attribuer une mesure numérique ou une catégorie d'une classification (Groupe d'experts des classifications économiques et sociales internationales).

Dans la description d'une population, une *variable* est la représentation du *concept* par une classification ou une unité de mesure. L'*unité statistique* pour laquelle la *variable* est observée n'est pas présente dans le nom de la *variable*. Toutefois, l'*unité statistique* est implicite.

4.2 Concepts

Statistique Canada observe les propriétés des *unités statistiques* et publie des rapports à leur sujet. Ces propriétés sont considérées comme des *concepts*. Les *concepts* sont abstraits puisqu'ils ne sont liés à **aucune** classification ou unité de mesure. Du point de vue de la modélisation orientée objet, on peut les considérer comme les propriétés d'une catégorie d'objet pour l'*unité statistique*.

Certains *concepts* (propriétés) peuvent être considérés comme communs à toutes les *unités statistiques*. L'existence est une propriété de chaque *unité statistique* (elle est observée, donc elle existe). Lorsqu'on déclare uniquement le nombre (taux, pourcentage) d'une unité, il s'agit d'un *concept*. Le lieu et le temps sont les deux autres *concepts* communs aux *unités statistiques*. D'autres propriétés des *unités statistiques* sont souvent recoupées par le temps ou par le lieu dans les tableaux de données statistiques agrégées.

4.3 Groupes

Un *groupe* est un rassemblement de *concepts*. On peut grouper les *concepts* en différents ensembles afin d'obtenir différentes perspectives sur les phénomènes représentés par les *concepts*. Ces groupes ne sont pas nécessairement incompatibles. Par exemple, la langue peut être considérée comme une caractéristique du capital humain et comme une caractéristique ethnoculturelle. On peut recréer ces groupements en fonction d'une utilisation donnée.

On peut aussi coordonner ce mode de groupement des variables avec le mode de consultation des métadonnées sur le site Web de Statistique Canada. Ce dernier offre une fonction « Parcourir par sujet » qui permet à l'utilisateur de consulter le site Web par termes de la liste de thèmes. Il s'agit généralement de termes comme « commerce », « agriculture », « santé », « justice », etc. Pour consulter le site Web, on pourrait rassembler les *concepts* en ensembles de *groupes* représentés par les titres de la liste de thèmes. On peut également appliquer ce mode de classement thématique des domaines spécialisés à toutes les *unités statistiques* comprises dans chaque *macro-unité statistique* (point de vue). Par exemple, pour le thème « agriculture », les *unités statistiques* de la société (personnes) sont des « exploitants agricoles », les *unités statistiques* économiques (entreprises) sont des « exploitations agricoles » et les *unités statistiques* environnementales (terrains et sols) peuvent subir les effets de l'exploitation agricole; enfin, les *unités statistiques* de l'État (établissements d'enseignement) peuvent offrir des cours en « agriculture ».

5. APPLICATION DU CADRE À L'ÉLABORATION DE DÉFINITIONS

Le Cadre d'établissement de normes relatives aux domaines spécialisés est conçu pour organiser les métadonnées définitionnelles et pour évaluer la nécessité d'établir de nouvelles normes. Le cadre mis sur pied suppose également une approche structurée de l'élaboration de définitions. Des définitions bien structurées des métadonnées permettent de mieux comprendre et utiliser les données. Ces définitions sont aussi plus faciles à transcrire sous forme de code de programmation.

Voici quelques exemples de règles de structure qui pourraient guider l'élaboration des définitions d'*unités statistiques* :

- Pour une *unité statistique*, définir sa relation par rapport à d'autres *unités statistiques*. Par exemple, on pourrait poser les questions suivantes :
 - S'agit-il d'une *unité statistique fondamentale*?
 - Dans la négative, de quelles *unités statistiques fondamentales* est-elle dérivée? Quel est son stéréotype?
- Pour une *unité statistique*, quelles valeurs de la *variable* doivent faire partie de la définition?
 - Par exemple, quelle tranche d'âge peut correspondre à la *sous-catégorie* « personne âgée »?
- Pour les *unités statistiques* qui sont des *rôles*, il faut préciser à quelles conditions on joue le *rôle* et on cesse de jouer le *rôle*.
 - Par exemple, considère-t-on une personne comme un étudiant uniquement durant la période au cours de laquelle elle est inscrite à un certain nombre minimal de cours?

Compte tenu de ces règles, au moment d'élaborer la définition d'« enfant », on pourrait poser la question suivante : « S'agit-il d'une définition fondée sur l'âge (étendue des valeurs de la *variable*) ou sur la relation par rapport à une autre *unité statistique* (parent)? » Ces règles peuvent faciliter la conception et la formulation des définitions.

En perfectionnant le cadre, on établira des règles d'élaboration des définitions des *variables*, des *classifications* et des *populations*.

Le fait de structurer la réflexion concernant les définitions liées au domaine spécialisé d'un organisme statistique pourrait faciliter l'élaboration des définitions et l'harmonisation des définitions entre elles à l'intérieur du Cadre d'établissement de normes relatives aux domaines spécialisés.

6. FUTURS TRAVAUX

Le Cadre d'établissement de normes relatives aux domaines spécialisés offre une vue d'ensemble des métadonnées définitionnelles de Statistique Canada. À mesure qu'on lui ajoutera des métadonnées, le cadre devrait montrer dans quels domaines il y a lieu d'établir de nouvelles normes. Il constitue un cadre uniforme et complet pour les métadonnées définitionnelles de la BMDI.

La poursuite des travaux d'élaboration du Cadre d'établissement de normes relatives aux domaines spécialisés suppose les activités suivantes :

- Étendre la portée du cadre pour inclure et définir toutes les *unités statistiques* et les *variables* pour lesquelles Statistique Canada publie régulièrement des données.
- Créer de nouvelles *classifications* au besoin.
- Examiner les définitions des *unités statistiques* et des *variables* et les harmoniser d'un domaine spécialisé à l'autre, dans le temps et à l'échelle des organismes.
- Utiliser le cadre comme « carte routière » pour chercher des métadonnées définitionnelles sur Internet.

On envisage également d'utiliser le Cadre d'établissement de normes relatives aux domaines spécialisés dans d'autres applications, dont l'attribution de métadonnées pour le site Web de Statistique Canada.

Il y a lieu de perfectionner le Cadre d'établissement de normes relatives aux domaines spécialisés pour inclure les *classifications* et les *populations* et, peut-être, pour définir les *variables* de façon plus détaillée. Les travaux de mise à jour de la BMDI, notamment en ce qui concerne l'élaboration des *classifications*, permettront aussi de perfectionner le Cadre d'établissement de normes relatives aux domaines spécialisés (Johanis, Brooks, Dunstan et Lévesque, 2003).

La nécessité de normes uniformes concernant les métadonnées définitionnelles des domaines spécialisés d'un organisme statistique est aujourd'hui plus urgente que jamais. L'élaboration du Cadre d'établissement de normes relatives aux domaines spécialisés est essentielle pour organiser les métadonnées définitionnelles ainsi que pour rendre les métadonnées cohérentes, interprétables et facilement accessibles.

ANNEXE A Équivalences des termes

Les termes utilisés dans la norme ISO/IEC 11179 et dans les ouvrages sur la modélisation orientée objet sont semblables, et pourtant différents. Le tableau ci-dessous montre les équivalences de divers termes par rapport à ceux du Cadre d'établissement de normes relatives aux domaines spécialisés.

Modélisation orientée objet (Définitions tirées d'ouvrages sur la modélisation orientée objet)	Cadre d'établissement de normes relatives aux domaines spécialisés	ISO/IEC 11179 (Définitions tirées de la norme ISO/IEC 11179)
<p><u>Objet</u> (...) concept, abstraction ou chose dont l'étendue et le sens sont bien définis à l'égard du problème à régler (...) les objets ont une identité et on peut les distinguer (Rumbaugh, Blaha, Premerlani, Frederick et Lorensen, 1991)</p>	<p><u>Unité statistique</u> Unité d'observation ou mesure pour laquelle on recueille ou calcule des données (Statistique Canada, 1988). L'unité statistique est définie en fonction de sa relation à d'autres unités statistiques ou en fonction de l'étendue des valeurs de la variable.</p>	<p><u>Objet</u> Toute partie du monde imaginable ou perceptible (ISO/IEC, 1999).</p>
<p><u>Catégorie d'objet</u> (...) groupe d'objets ayant des propriétés (attributs) semblables, un comportement (fonctionnement) commun, des relations communes avec d'autres objets et une sémantique commune (...) (Rumbaugh, Blaha, Premerlani, Frederick et Lorensen, 1991).</p> <p><u>Catégorie</u> Élément de modèle servant à modéliser des choses intellectuelles ou matérielles. Une catégorie est un type (...) (Ericksson et Penker, 1998).</p>	<p><u>Population</u> Ensemble d'unités statistiques auxquelles se rapporte un ensemble de données (Statistique Canada, 1988). La population est définie en fonction d'un espace (région géographique) et d'une période.</p>	<p><u>Catégorie d'objet</u> Ensemble d'objets. Ensemble de notions, d'abstractions ou de choses du monde réel qu'on peut cerner et définir explicitement et dont les propriétés et le comportement suivent les mêmes règles (...) On peut former des catégories d'objet en combinant deux ou plusieurs autres catégories d'objet (ISO/IEC, 1999).</p>
<p><u>Propriété</u> État abstrait de l'objet (Norbotten, 1993).</p>	<p><u>Groupe</u> Idée générale ou abstraite qui exprime les phénomènes économiques ou sociaux à étudier ou à observer. Abstraction pouvant servir à modéliser un ensemble de concepts.</p> <p>Ce concept est indépendant de l'unité statistique.</p>	<p><u>Propriété</u> Particularité commune à tous les membres d'une catégorie d'objet (ISO/IEC, 1999).</p>
	<p><u>Concept</u> Idée générale ou abstraite qui exprime le phénomène social et (ou) économique qu'il faut évaluer.</p> <p>Propriété abstraite de la catégorie d'objet, indépendante de la représentation. Le concept est un terme de la taxonomie du domaine spécialisé d'un groupe. Il peut avoir une portée plus grande qu'un concept d'élément de données.</p>	<p><u>Concept d'élément de données</u> Concept qu'on peut représenter sous forme d'un élément de données décrit indépendamment d'une représentation particulière (...) La combinaison d'une catégorie d'objet et d'une propriété est un concept d'élément de données (ISO/IEC, 1999).</p>

<p><u>Attribut</u> (...) valeur que possèdent les objets d'une catégorie (...) Chaque attribut a une valeur pour chaque instance d'objet (...) chaque nom d'attribut est unique à l'intérieur d'une catégorie (...) différentes instances d'objet peuvent avoir des valeurs identiques ou différentes pour une catégorie d'objet donnée (Rumbaugh, Blaha, Premerlani, Frederick et Lorensen, 1991).</p> <p>Cette définition suppose le mappage d'une représentation.</p>	<p><u>Variable</u> Caractéristique ou attribut particulier d'une unité statistique (Statistique Canada, 1988).</p>	<p><u>Élément de données</u> Unité de données pour laquelle la définition, l'identification, la représentation et les valeurs admissibles sont précisées au moyen d'un ensemble d'attributs (...) on peut aussi considérer un élément de données comme composé de deux parties : un concept d'élément de données et une représentation (ISO/IEC, 1999).</p> <p><i>Une variable s'applique à des unités statistiques, mais son nom ne contient pas celui d'une unité statistique. On peut donc considérer l'élément de données générique, qui comprend la propriété et la représentation, comme l'équivalent de la variable.</i></p>
---	---	--

RÉFÉRENCES

- Colledge, M.J. (1999), "Statistical Integration Through Metadata Management", *International Statistical Review*, Vol.67, No. 1 pp. 79-98.
- Eriksson, H-E. et M. Penker (1998), *UML Toolkit*, John Wiley and Sons, Inc.
- Expert Group on International Economic and Social Classifications, *United Nations Glossary for Classification Terms*. Available at: http://unstats.un.org/unsd/class/family/glossary_short.htm
- Gilbert, N. G. (1981), *Modelling Society*, George Allen & Unwin (Publishers) Ltd.
- Grossmann, W. (1999), *Statistical Metadata Tutorial*, available at: http://www.univie.ac.at/dac/papers/Tutorial_gesamt.pdf
- ISO/IEC (1999), ISO/IEC 11179-1 *Framework for the specification and standardisation of data elements*, First Edition.
- Johanis, P., (2000), "Statistics Canada's Integrated Metadatabase, Current Status and Future Plans", paper presented at the Work Session on METIS, Statistical Commission and Economic Commission for Europe Conference of European Statisticians.
- Johanis, P. (2002), "Assessing the Quality of Metadata: The Next Challenge", paper presented at the Work Session on METIS, Statistical Commission and Economic Commission for Europe Conference of European Statisticians.
- Johanis, P., B. Brooks, T. Dustan, et J-S. Lévesque, (2003), "Statistics Canada's Implementation of the Data Element Model", paper for the Metadata Registries Open Forum.
- Muller, R. J. (1999), *Database Design for Smarties, Using UML for Data Modelling*, Morgan Kaufmann Publishers.
- Norbotten S. (1993), "Statistical meta-knowledge and -data", *Statistical Journal for the United Nations Economic Commission for Europe*, Vol. 10 n No. 2, pp. 101-112.
- Rumbaugh, J., M. Blaha, W. Premerlani, E. Frederick et W. Lorensen (1991), *Object Oriented Modelling and Design*, Prentice Hall.
- Statistics Canada (July 15, 1988), Policy on Standards, *Statistics Canada Policy Manual*.
- Struijs, P. (1999), "Metadata at Statistics Netherlands", paper presented at the 1999 FCSM Research Conference, Available at: <http://www.fcsm.gov/99papers/fcsm-def.html>
- Sundgren, B. (1993), "Statistical metainformation systems – pragmatics, semantics, syntactics", *Statistical Journal for the United Nations Economic Commission for Europe*, Vol. 10 n No. 2, pp. 121-142.