

Distr.
GÉNÉRALE

CES/AC.71/2003/9
31 janvier 2003

ANGLAIS et FRANÇAIS seulement

COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE POUR L'EUROPE
(ONU)

COMMISSION EUROPÉENNE

CONFÉRENCE DES STATISTICIENS
EUROPÉENS

OFFICE STATISTIQUE DES
COMMUNAUTÉS EUROPÉENNES
(EUROSTAT)

ORGANISATION DE COOPÉRATION ET DE
DÉVELOPPEMENT ÉCONOMIQUES (OCDE)
DIRECTION DES STATISTIQUES

Réunion mixte CEE/Eurostat/OCDE sur la gestion des systèmes d'information statistique
(Genève, 17-19 février 2003)

Point III: Gestion efficace d'une complexité technique croissante

LE GUIDE STATISTIQUE

Communication sollicitée

émanant de Statistique Canada¹

I. INTRODUCTION

1. Le guide statistique présente un aperçu théorique d'une infrastructure automatisée de gestion de l'information dans un bureau de la statistique. Il s'agit d'un modèle générique structurel qui simplifie les phases de *réalisation* et d'*exécution* des projets de système d'information à Statistique Canada. Il sert de guide aux planificateurs des opérations, aux architectes de systèmes d'information et aux concepteurs de systèmes et, moyennant adaptation, il peut être utilisé par d'autres bureaux de la statistique. Le champ d'application du document se limite aux projets dont l'exécution suppose la mise au point, la tenue à jour et l'exploitation de systèmes d'information statistique.

2. On trouvera à la section II la description de l'infrastructure du programme et de l'infrastructure qui existent actuellement à Statistique Canada.

- L'infrastructure du programme facilite la *réutilisation des processus* et comprend des logiciels et des outils polyvalents pouvant être réutilisés pour élaborer et tenir à jour des systèmes d'information sur des enquêtes précises.
- L'infrastructure facilite la *réutilisation des données* et des *métadonnées*. Elle comprend des bases de données (Registre des entreprises, Base de métadonnées intégrée, etc.) et des services (codage, par exemple) intégrés conçus pour plusieurs enquêtes.

3. La section III est à la fois descriptive et interprétative. On y examine le guide statistique en fonction de deux modèles formels.

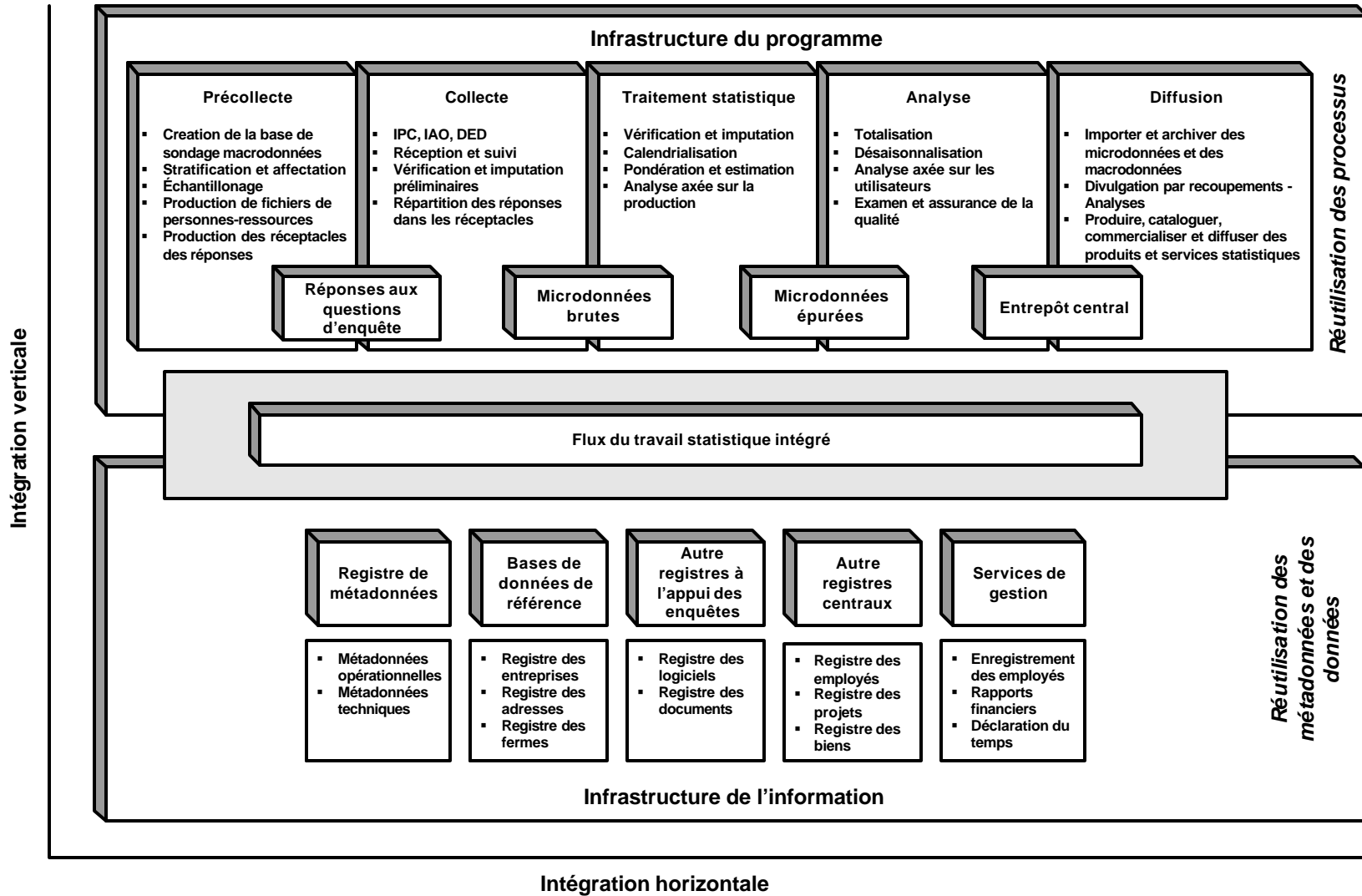
¹ Établi par Ron Graves (ron.graves@statcan.ca) et Ted Hutton (tedhutton@rogers.com).

- La version adaptée d'un diagramme de la hiérarchie des fonctions est présentée en premier lieu. Ce diagramme montre les principales phases associées à chaque projet de système d'information statistique ainsi que les magasins de données intégrés les plus susceptibles d'être consultés à chaque phase.
- Un modèle théorique de Statistique Canada est ensuite présenté sous la forme d'un diagramme entités-relations. Le modèle général ne représente que les entités et les relations de niveau supérieur et aborde leur rôle dans le contexte du guide statistique.

4. Les auteurs proposent de rationaliser l'élaboration des systèmes d'information dans un bureau de la statistique à tous les stades des enquêtes en encourageant, dans la mesure du possible, la réutilisation des informations existantes (logiciels, données, métadonnées opérationnelles et techniques ainsi que documentation). En fin de compte, cette réutilisation se traduit également par une plus grande standardisation et simplification. Le guide statistique peut être utilisé à d'autres fins.

- Il peut être utilisé pour évaluer la mesure dans laquelle le bureau de la statistique a rendu compte des risques et des possibilités définis dans chacun de ses principaux composants automatisés.
- Plus récemment, le guide a permis à l'Office of National Statistics (ONS) du Royaume-Uni de connaître le degré d'harmonisation entre le Cadre des technologies de l'information (CTI) de Statistique Canada, le guide statistique et les principes d'architecture de l'ONS.
- Enfin, la direction peut examiner le projet, décider quels sont les buts (c'est-à-dire les fins) les plus importants et élaborer un plan d'action (c'est-à-dire les moyens) pour atteindre les objectifs stratégiques à l'aide de la technologie de l'information.

Figure 1 - Infrastructure du programme et infostructure



I. INFRASTRUCTURE DU PROGRAMME ET INFOSTRUCTURE

A. Description

5. La figure 1 montre l'infostructure et l'infrastructure du programme de Statistique Canada (flux du travail statistique intégré non mis en œuvre). L'infrastructure du programme indique les phases (précollecte, collecte, etc.) et les processus génériques (création de la base de sondage, stratification et affectation, pondération et estimation, etc.) utilisés à l'occasion d'une enquête type (ou d'un recensement). Les observations suivantes peuvent être formulées relativement à l'infrastructure du programme.

- Chaque processus générique comprendra un certain nombre d'algorithmes standards pouvant être réutilisés. Il sera possible d'utiliser un algorithme existant dans de nombreux cas, tout comme il sera possible d'adapter un algorithme afin d'atteindre les objectifs précis d'une enquête. Il faudra élaborer un nouvel algorithme uniquement dans un nombre limité de cas.

Dans le contexte d'un système de traitement réparti, chaque processus générique doit avoir une *interface de services* désignant les paramètres requis par le processus, le service à utiliser et un renvoi à l'ensemble de données à traiter. Une deuxième méthode de réutilisation, moins souhaitable celle-là, consisterait à copier le code source d'un processus générique dans un programme existant. Bien que cette méthode soit utile lorsque l'algorithme original doit être *adapté* et qu'une solution d'urgence s'impose, l'algorithme source doit être modifié une fois que les changements ont été apportés.

- Quatre bases de données (réponses aux questions d'enquête, microdonnées brutes, microdonnées épurées et entrepôt de données de l'organisation) peuvent être utilisées pour faire passer les données statistiques d'une phase à l'autre. Statistique Canada utilise CANSIM II comme entrepôt de données à des fins d'analyses transversales et de séries chronologiques de macrodonnées d'enquête. Cette base de données peut être mise à jour à chaque enquête à l'aide d'opérations uniformisées. Aucune stratégie commune de stockage des microdonnées d'enquête n'a toutefois été établie.

6. L'infostructure fournit une *liste partielle* des métadonnées et des ensembles de données réutilisables de Statistique Canada (qui ne sont pas tous utilisés à l'heure actuelle). Cette réutilisation peut être effectuée de l'une ou l'autre des façons suivantes.

- Des experts peuvent passer en revue les métadonnées opérationnelles pendant les phases de formulation et d'élaboration des projets. Deux des principaux objectifs seraient de s'assurer que les concepts sont harmonisés d'une enquête à l'autre et que les valeurs des variables sont uniformisées.
- Les métadonnées techniques peuvent être utilisées par des architectes en technologie de l'information (TI) pendant les phases d'élaboration et de mise à jour des projets. Il devrait être possible, par exemple, de réutiliser un schéma conceptuel d'une application à une autre si les applications font partie du même domaine spécialisé.
- Les bases de données de référence sont utilisées dans le contexte de familles d'applications faisant partie du même domaine spécialisé. Par exemple, à Statistique Canada, le Registre des entreprises sert de fichier source afin de créer les bases de sondage des enquêtes-entreprises. Cela signifie que chaque enquête-entreprise pourra tirer parti des activités de tenue à jour du Registre, d'établissement de profils et de mise à jour consécutives aux enquêtes entreprises par la Division du Registre des entreprises.
- Le registre des logiciels est utilisé pour administrer les composants des logiciels réutilisables et stocker leur code source. Il est possible de consulter des métadonnées opérationnelles décrivant le ou les algorithmes (c'est-à-dire les méthodes) pris en charge par chaque

composant de logiciel réutilisable. L'interface des services pour chaque composant est également décrit.

- Le registre des documents fait office d'entrepôt central des importants documents produits par le Bureau. Les documents de référence (exposés de principes généraux, plans d'activités du Bureau, documents concernant certaines enquêtes et pouvant s'appliquer à d'autres enquêtes semblables, etc.) peuvent être examinés par des experts, des méthodologistes et des architectes en TI pendant la phase de formulation d'une enquête. Dans le même ordre d'idée, les importants documents concernant une enquête (documentation sur les systèmes, descriptions des méthodes, procédures opérationnelles, etc.) et élaborés pendant la phase de réalisation peuvent être consignés et stockés dans le registre à des fins de mise à jour et d'examen ultérieurs.

7. Enfin, la base de données intégrée sur le flux de travail statistique enregistre les procédures opérationnelles (tant les processus automatisés que les mesures exigeant une intervention humaine) qu'un système d'information statistique doit exécuter. Dans certains cas, le processus d'un système d'information statistique ne peut pas être exécuté tant que le processus d'un autre système n'est pas terminé. Ainsi, les procédures de mise à jour du registre source d'une enquête doivent être terminées avant que la base de sondage d'une enquête précise (ou d'un ensemble d'enquêtes) ne soit créée. Dans d'autres cas, une personne doit approuver le résultat d'une étape de traitement précise avant que les autres étapes ne puissent être exécutées.

II. LE GUIDE STATISTIQUE

8. Le principe premier de la présente communication est que les modèles architecturaux au niveau des entreprises sont essentiels à la gestion de la complexité technique. On trouvera à la sous-section D la justification de ce principe, basée sur l'interprétation des figures 2 et 3 ainsi que sur les leçons tirées de la révision, de l'amélioration et de la validation de ces diagrammes.

9. La figure 2 est un diagramme de la hiérarchie des fonctions dont le principal objectif est de dresser la liste des différentes fonctions accomplies par les travailleurs du savoir dans un bureau de la statistique. La figure 3 présente quant à elle un diagramme entités-relations dont l'objectif principal est de désigner les entités (objets) importantes pour les projets supposant l'élaboration et l'exploitation de systèmes d'information statistique. Les importantes associations entre ces entités sont également décrites. Ces deux types de modélisation sont utilisés dans le cadre de projets qui ont recours à la méthode structurée de conception et d'analyse des systèmes.

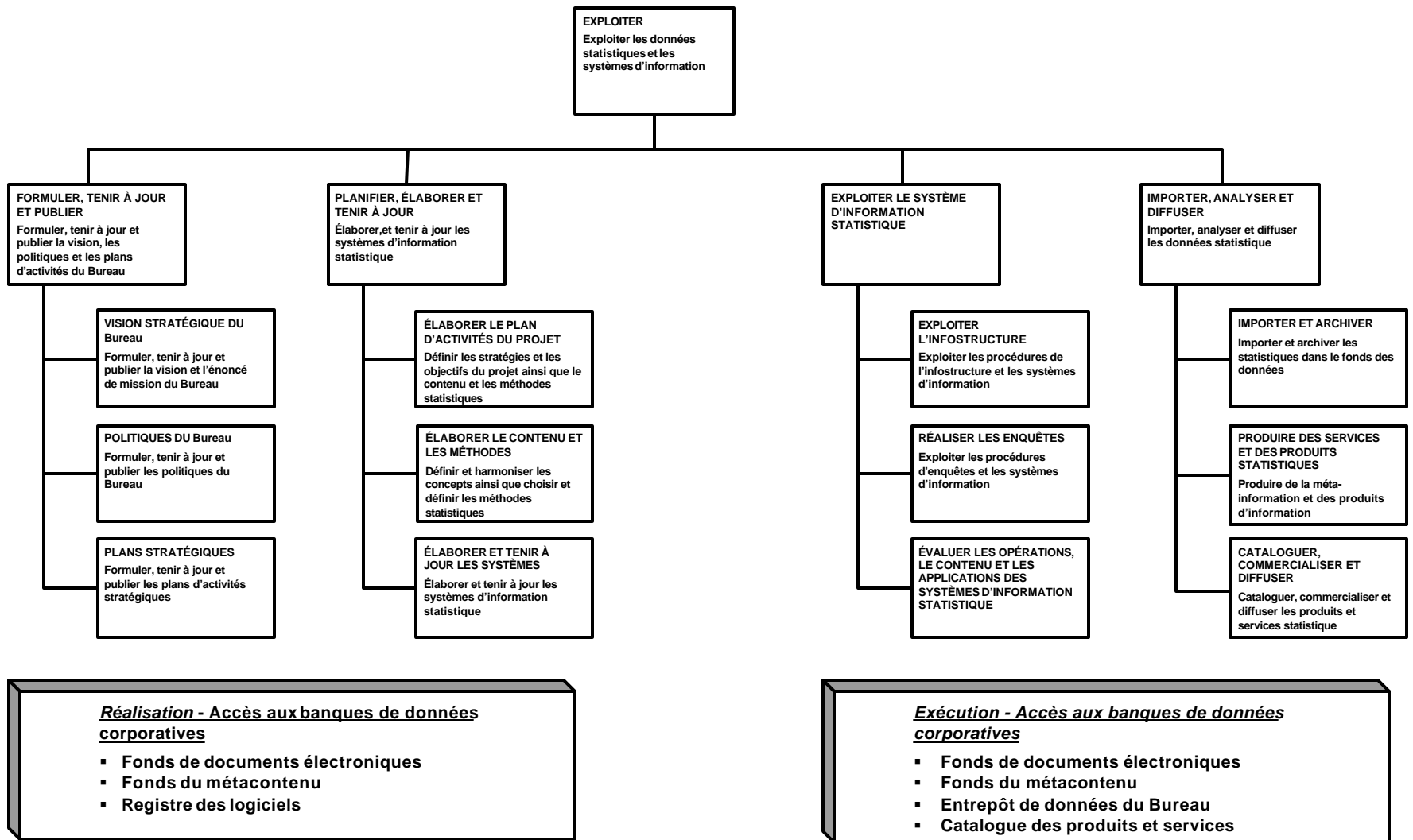
A. Le diagramme de la hiérarchie des fonctions du cycle de vie des enquêtes

10. On trouvera à la figure 2 un *très bref aperçu* des principales fonctions remplies pendant le cycle de vie des enquêtes. La liste des magasins de données centraux normalement consultés pendant les phases de *réalisation* et d'*exécution* du projet se trouve au bas du diagramme. Les observations suivantes peuvent également être formulées.

- Les deux colonnes de gauche indiquent les fonctions (activités) accomplies pendant les étapes de lancement et d'approbation du projet ainsi que pendant l'élaboration des systèmes (phase de réalisation).

Les fonctions citées à la colonne FORMULER, TENIR À JOUR ET PUBLIER sont indépendantes des projets. Elles portent sur la formulation et la publication i) de l'énoncé de vision stratégique du Bureau, ii) des politiques du Bureau et iii) des plans d'activités stratégiques du Bureau. Étant donné qu'ils influent sur les exigences du projet, ces documents doivent être passés en revue pendant la phase de formulation du projet. Il est possible que les besoins opérationnels d'un projet entraînent des changements à une politique ou une stratégie existante.

Figure 2 - Diagramme de la hiérarchie des fonctions du cycle de vie des enquêtes



La colonne PLANIFIER, ÉLABORER ET TENIR À JOUR porte tout d'abord sur les fonctions associées à l'élaboration du plan d'activités du projet. Le plan d'activités englobe les fonctions suivantes : définition des caractéristiques cibles de l'enquête, désignation des principaux partenaires, définition des résultats escomptés de l'enquête de façon à atteindre les objectifs fixés et à répondre aux besoins des partenaires, définition des principaux intrants de l'enquête, définition des méthodes statistiques à utiliser de façon à transformer les intrants en extrants, estimation du financement requis et établissement d'un échéancier. Le plan d'activités doit être approuvé par le Bureau avant de passer à la phase ÉLABORER LE CONTENU ET LES MÉTHODES. Des travaux poussés portant notamment sur la définition et l'harmonisation des concepts, la conception du questionnaire ainsi que le choix et la définition des méthodes statistiques doivent être terminés en grande partie avant de commencer l'élaboration du système d'information statistique. Les décisions préliminaires prises dans le contexte de l'établissement du plan d'activités peuvent être modifiées pendant les phases de réalisation et/ou d'exécution du projet.

Il est nécessaire d'avoir accès à différents magasins de données centraux pendant la phase de réalisation. Par exemple, le fonds de documents électroniques permet aux travailleurs du savoir d'examiner des documents généraux et détaillés sur les enquêtes pouvant avoir des caractéristiques semblables à l'enquête envisagée. D'autre part, on consulte le fonds du métacontenu pour s'assurer que les principaux concepts spécialisés sont uniformisés d'une enquête à l'autre. Le registre des logiciels est consulté pour savoir si des logiciels n'existent pas déjà au regard de certains processus statistiques propres à l'enquête. Chaque magasin de données présente des possibilités de *réutilisation* dans le cadre d'un projet et, partant, la possibilité de réaliser des économies de temps et d'argent.

- Les deux colonnes de droite renferme la liste des fonctions associées à la réalisation (phase de l'exécution) des projets.

La case EXPLOITER LE SYSTÈME D'INFORMATION STATISTIQUE comprend la production des systèmes d'information du programme (enquête) et de l'infrastructure. Il faut que les opérations relatives au champ d'application et à la mise à jour du Registre des entreprises soient terminées avant que ne soit créée la base de sondage d'une enquête-entreprise type à Statistique Canada. De même, nombre des *mises à jour consécutives à l'enquête* doivent être mises en œuvre. De cette façon, les flux de travail associés à ces deux types de systèmes d'information statistique sont interdépendants. On trouve à la figure 1 une liste détaillée des processus génériques associés à un logiciel d'enquête type.

La case ÉVALUER LES OPÉRATIONS, LE CONTENU ET LES APPLICATIONS DES SYSTÈMES D'INFORMATION STATISTIQUE représente la dernière étape du cycle d'une enquête donnée. Les activités comprennent la production des totalisations requises, l'exécution des méthodes de désaisonnalisation et la production de rapports analytiques. L'évaluation porterait sur des sujets comme le contrôle de la qualité et des activités liées au codage, la saisie et la vérification des données, l'efficacité du questionnaire, la stratégie d'échantillonnage, les taux de réponse ainsi que l'évaluation de la qualité statistique du programme d'enquête et la production de rapports connexes. Sundgren (1999) traite de l'évaluation de façon plus substantielle.

Les fonctions IMPORTER, ANALYSER ET DIFFUSER permettent de mettre à la disposition du public les données statistiques (fichiers de microdonnées *anonymisés*, macrodonnées et métadonnées) produites dans le cadre d'une enquête. Le fichier de microdonnées épurées produit dans le cadre d'une enquête est exporté vers un entrepôt de données central, où il peut être utilisé pour produire des totalisations personnalisées à l'intention de nos partenaires de l'extérieur. Le personnel du centre d'échange s'assure que les ensembles de valeurs des

variables du fichier de microdonnées épurées répondent aux normes du Bureau définies dans le fonds du métacontenu général. Les macrodonnées sont stockées dans une base de données multidimensionnelles (CANSIM 2 est la base de données multidimensionnelles de Statistique Canada) servant à l'établissement d'analyses de séries chronologiques et transversales.

Il est également nécessaire d'avoir accès à des magasins de données centraux pendant la phase d'exécution. Ainsi, l'entrepôt de données central est mis à jour pour chaque enquête lorsque des microdonnées et des macrodonnées propres à une enquête sont exportées vers le centre d'échange du Bureau. Les fonds de documents électroniques et du métacontenu sont mis à jour en réponse aux besoins changeants des enquêtes.

B. Le modèle général

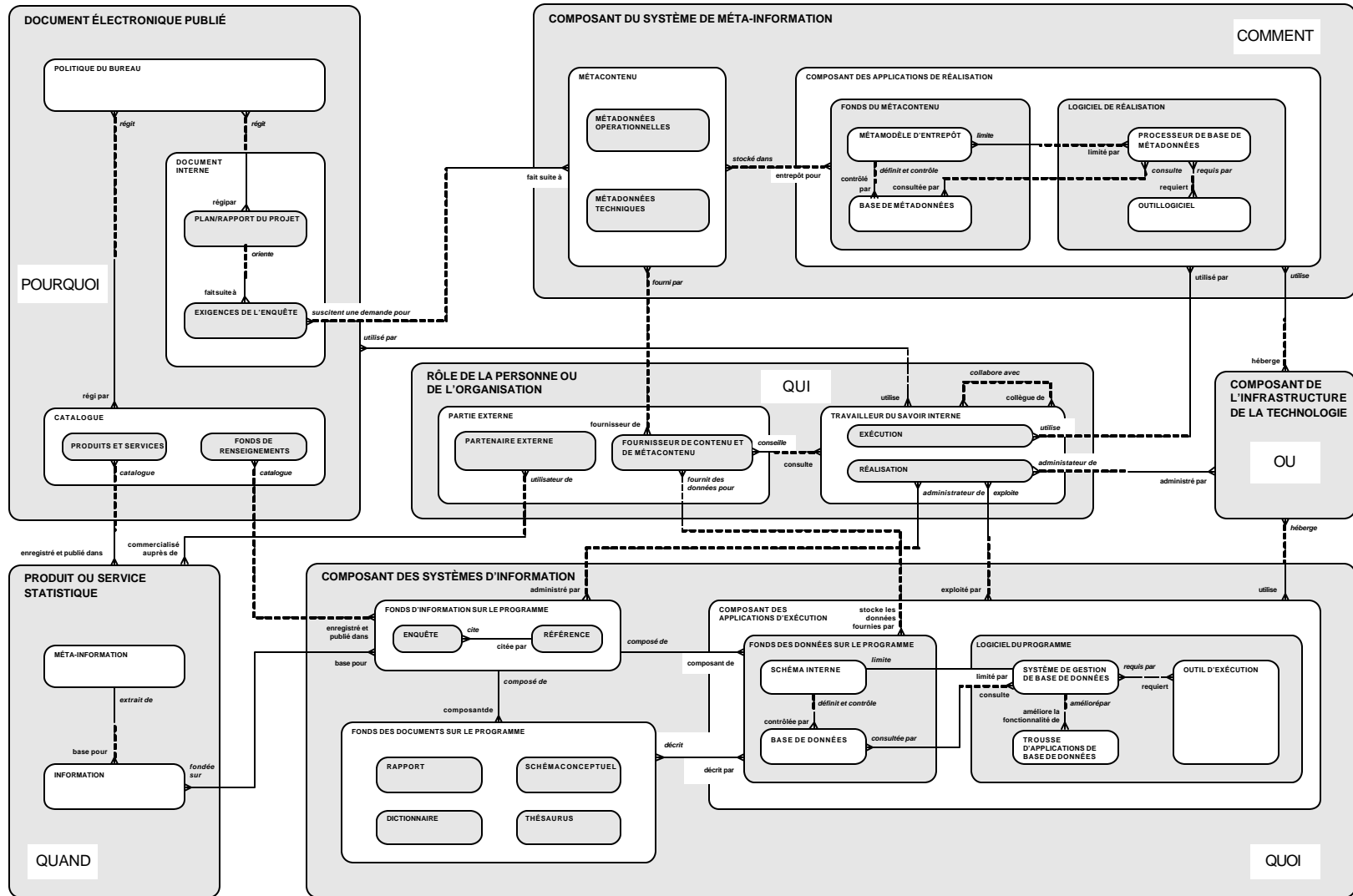
11. Le diagramme entités-relations (DER), à la figure 3, constitue un aperçu conceptuel de haut niveau du guide statistique. Six entités primaires (supertypes) «*hébergent*» un certain nombre d'entités de niveau inférieur (sous-types). Voici une brève description du diagramme.

- Les liens importants (relations) entre deux entités sont représentés par des lignes continues (*liens obligatoires*) ou pointillées (*liens facultatifs*) ou par une combinaison de lignes continues et pointillées (liens facultatifs dans une direction et obligatoires dans l'autre).
- Les phrases décrivant tous les liens ont un sujet, un verbe et un complément. Exemples : une politique générale *peut* régir un ou plusieurs plans/rapports de projet; un plan/rapport de projet *doit* être régi par une ou plusieurs politiques générales. Du point de vue graphique, un segment de ligne continue à l'entité qui joue le rôle de sujet symbolise un lien obligatoire, tandis qu'un segment de ligne pointillée représente un lien facultatif.
- Des *pattes d'oie* sont utilisées comme terminaisons de ligne pour symboliser les liens *multiples* (un ou plusieurs liens). Une terminaison de ligne simple signifie un lien *unique* (un et un seul lien). Ainsi, un métamodèle d'entrepôt peut définir et contrôler *une ou plusieurs* bases de métadonnées, mais une base de métadonnées ne doit être contrôlée que par *un et seulement un* métamodèle d'entrepôt.

12. Deux autres observations s'imposent. Premièrement, bien que le DER indique un certain nombre d'associations, les auteurs de la présente communication ont dû, *en raison du manque d'espace*, i) omettre de préciser pourquoi certaines associations sont facultatives et d'autres sont obligatoires et ii) ne pas faire état d'associations moins importantes. Seul le modèle général est présenté pour la même raison. Un ensemble de modèles plus détaillés est actuellement disponible, mais les descriptions ne seront pas prêtes avant le printemps 2003.

13. Deuxièmement, le modèle général présente, *du point de vue des auteurs*, une infrastructure d'automatisation de la gestion de l'information dont l'utilisation serait indiquée à Statistique Canada. Les entités et les relations figurant dans le modèle général sont celles qui ont été jugées les plus importantes. Dans certains cas, ces entités n'existent pas encore. Des discussions plus poussées avec d'autres employés de Statistique Canada et d'autres bureaux de la statistique entraîneraient assurément des changements au modèle. En choisissant une méthode de modélisation formelle, il sera possible de tenir le modèle à jour grâce à plusieurs exercices de *validation*.

Figure 3 - Le modèle général



14. Les entités primaires de la figure 3 représentent les principales activités d'un bureau de la statistique. Le champ d'application de cette version du guide statistique se limite aux projets dont la mise en œuvre suppose l'élaboration, la tenue à jour et l'exploitation de systèmes d'information statistique. Les principaux éléments sont les suivants.

- L'élément rôle de la personne ou de l'organisation joue un rôle central et comprend deux sous-types, soit partie externe et travailleur du savoir interne. Une partie externe peut être un partenaire externe et/ou un fournisseur de contenu et de métacontenu. Les partenaires externes sont des utilisateurs de nos produits et services statistiques, tandis que les fournisseurs de contenu et de métacontenu sont des fournisseurs de métadonnées opérationnelles et de métadonnées techniques et que les données sont stockées dans un fonds de données sur le programme.

Les travailleurs du savoir internes ont un rôle à jouer sur le plan de la *réalisation* et/ou de l'*exécution*. Du côté de la réalisation, ils utilisent un *outil logiciel* [les outils de génie logiciel assisté par ordinateur (CASE, Computer Assisted Software Engineering) sont couramment utilisés pour élaborer et tenir à jour des métadonnées techniques, les formulaires peuvent être utilisés pour recueillir des métadonnées opérationnelles, une partie du code pour une application peut être générée, etc.] pour créer et réutiliser des métadonnées opérationnelles et techniques. Pour ce qui est de l'exécution, ils exploitent un composant d'exécution, administrent une infrastructure de technologie ou gèrent un fonds d'information sur le programme. Un employé peut jouer plus d'un rôle.

- L'élément *Document électronique publié* symbolise un fonds de documents électroniques généraux. Les travailleurs du savoir internes du bureau de la statistique ont accès à ces documents. Les documents publics peuvent être consultés en effectuant une recherche dans le fonds de documents électroniques mais, règle générale, ils sont consultés sur le site Internet du bureau de la statistique. Quant aux documents internes, on les consulte sur le réseau intranet du bureau.
- Le *composant du système de méta-information* comprend deux éléments : un *métacontenu* et un *composant d'applications de réalisation*. Les métadonnées opérationnelles et techniques forment le métacontenu. Les métadonnées opérationnelles sont consultées par des parties externes et des travailleurs du savoir internes, tandis que les métadonnées techniques sont examinées, réutilisées/adaptées et élaborées par des architectes en TI pendant les phases de formulation, d'élaboration et de tenue à jour des projets de systèmes d'information statistique.

Les deux types de métadonnées peuvent *logiquement* être stockées dans une *base de métadonnées* centrale. Il est toutefois plus probable que ces deux types de métadonnées soient stockées dans des bases de données distinctes, lesquelles sont alimentées à l'aide de différents types d'outils.

- Le composant des systèmes d'information comprend un fonds d'information sur le programme, un fonds de documents sur le programme et un composant d'applications d'exécution. Le fonds d'information sur le programme est formé du fonds de documents sur le programme et du fonds de données sur le programme. Les fonds d'information sur le programme sont enregistrés et publiés dans le catalogue des fonds de renseignements. Un fonds d'information sur le programme est soit un fonds d'information sur l'enquête, soit un fonds d'information de référence. Ce dernier fait l'objet de renvois (ou est réutilisé) par un ou plusieurs fonds d'information sur l'enquête.

Un fonds de documents sur le programme contient les différents types de documents décrivant un fonds de données sur le programme. Le modèle contient des documents uniformisés à des fins d'analyse et de recherche (rapport) ainsi que d'établissement des

concepts et de la terminologie (dictionnaire et thésaurus). Il contient aussi un schéma conceptuel. À noter que le dictionnaire et le thésaurus sont des sous-ensembles propres à des applications extraits du fonds de métacontenu général.

Les données recueillies dans le cadre d'une enquête sont stockées dans un fonds de données sur le programme après avoir été traitées par le logiciel du programme dans le contexte d'un composant précis d'applications d'exécution. Tous les fonds de données d'enquête proposés contiennent deux types de fonds de données : un fonds de microdonnées et un fonds de macrodonnées. De plus, les variables des fonds de données de référence peuvent être reproduites ou citées. À Statistique Canada, les macrodonnées sont stockées dans CANSIM 2, notre entrepôt de données central.

- L'élément *produit ou service statistique* comprend les sous-éléments *méta-information* et *information*. Les produits et services statistiques sont commercialisés auprès des partenaires de l'extérieur. Ils peuvent en outre être publiés dans le *Catalogue des produits et des services* du bureau de la statistique. L'*information sur les produits ou services statistiques* est fondée sur le fonds d'information sur le programme. Le fonds du métacontenu est utilisé pour produire la *méta-information sur les produits et les services statistiques*.
- Le *composant de l'infrastructure de la technologie* représente l'infrastructure de TI du bureau de la statistique. Tant le composant de la réalisation que le composant de l'exécution peuvent utiliser un ou plusieurs composants de l'infrastructure de la technologie (compris dans un diagramme de niveau inférieur) : *logiciel d'infrastructure, sécurité, matériel et communications*.

15. Chacune des six entités primaires formant le modèle général (DER) peut être considérée comme un important thème architectural pour décrire un bureau de la statistique. Ces thèmes ont un point central pouvant correspondre aux six colonnes (QUOI, COMMENT, OÙ, QUI, QUAND, POURQUOI) de la deuxième rangée du cadre Zachman. En voici la justification.

- **POURQUOI** - La motivation organisationnelle est fournie par les documents de planification opérationnelle exposant la vision, la mission, les objectifs et les politiques. Les technologies d'automatisation prenant en charge le fonds de documents électroniques sont fondées sur le langage XML.
- **QUAND** - Les « échéanciers directeurs » d'un bureau de la statistique sont établis d'après les délais de diffusion. Les technologies liées au flux du travail contribuent à la rationalisation des procédures opérationnelles et de l'exploitation des systèmes d'information statistique. Il s'en suivra une amélioration globale de la rapidité d'exécution.
- **QUI** - Les rôles et les responsabilités des travailleurs du savoir internes sont définis en fonction de la technologie. Les travailleurs du savoir devront continuellement se tenir au courant des nouvelles technologies.
- **QUOI** - D'un point de vue technologique, il s'agit des composants d'un système automatisé d'information statistique. L'élément du système d'information a été décomposé en fonds d'information sur le programme et en composants d'applications d'exécution. Cette façon de procéder nous permet d'associer plus facilement les rôles et les responsabilités des travailleurs du savoir à l'exploitation des systèmes d'information statistique.
- **COMMENT** - Les caractéristiques architecturales déterminant l'automatisation de la production statistique sont le reflet du *savoir-faire* du bureau de la statistique en la matière. Elles sont créées dans le contexte d'un système de méta-information. La réutilisation de ces éléments contribuera à la gestion de la complexité en limitant la diversité improductive.

- **OÙ** - Les éléments technologiques sont situés dans le composant de l'infrastructure de la technologie et ils sont gérés par ce composant. Ce composant autorise la transparence d'emplacement et, par la suite, le traitement réparti.

C. Quelques cadres et modèles pouvant être combinés

Le cadre Zachman classe les caractéristiques architecturales d'une entreprise donnée. Il peut s'agir, par exemple, de l'administration publique fédérale du Canada, de Statistique Canada ou d'une unité organisationnelle au sein de Statistique Canada. Le cadre Zachman a été utilisé à chacun de ces niveaux d'élaboration progressive et il continue d'être utilisé. La figure 4 montre les deux rangées supérieures du cadre Zachman, Ce sont celles qui présentent le plus d'intérêt pour la haute direction d'une entreprise.

Le Cadre des technologies de l'information (CTI) de Statistique Canada est un cadre stratégique assurant l'adéquation de la TI et des objectifs opérationnels de l'organisme. C'est un cadre de référence prenant la forme d'un assemblage cohérent de « perspectives » (bureau, programme, information, communauté, plateforme et administration publique) et présentant Statistique Canada comme une « entreprise ». C'est la raison pour laquelle il doit être harmonisé avec les stratégies opérationnelles de l'organisme et de toute l'administration publique.

Le guide statistique est un autre instrument de référence pouvant être utilisé efficacement dans le contexte de l'architecture d'entreprise à Statistique Canada. Il s'agit d'un modèle formel par opposition à un cadre. Il montre les relations structurelles entre les entités formant un bureau de la statistique.

Figure 4 - Adéquation entre le cadre Zachman, le CTI et le guide statistique

	QUOI Données	COMMENT Fonction	OÙ Réseau	QUI Personnes	QUAND Temps	POURQUOI Motivation	
Portée (Planificateur)	Perspective - Information	Perspective - Programme	Perspective - Plateforme	Perspective - Communauté	Perspective - Administration publique	Perspective - Bureau	Perspectives - CTI
Modèle d'entreprise (Possesseur)	Composant des systèmes d'information	Composant du système de meta-information	Composant de l'infrastructure de technologie	Rôle de la personne ou de l'organisation	Produit ou service statistique	Document électronique publié	Composants du modèle général

16. La valeur du cadre Zachman pour une entreprise donnée tient dans son rôle sur le plan des caractéristiques architecturales au regard d'une utilisation et réutilisation efficaces par des groupes de l'entreprise. L'investissement effectué sert à mettre au point les caractéristiques originales, habituellement dans le contexte d'un projet précis. L'entreprise réalise des économies lorsque ces caractéristiques sont réutilisées dans le cadre de projets ultérieurs. Chaque rangée de caractéristiques présente une vue complète d'une entreprise ou un composant d'entreprise d'un point de vue précis. À la figure 4, les groupes retenus sont les planificateurs et les possesseurs. Hutton et Graves (2002) font une étude plus approfondie de ce sujet dans la série de documents de Statistique Canada sur l'architecture d'entreprise.

17. La figure 4 montre les caractéristiques que sont les « perspectives » relatives au CTI et les « composants » du modèle général du guide statistique dans le contexte du cadre Zachman. L'occupation d'une case (cellule) particulière est fondée sur la ou les caractéristiques dominantes de la perspective ou du composant. Cette concordance symbolique est fournie pour faciliter l'utilisation de l'importante documentation sur l'interprétation d'applications d'entreprise fondée sur le cadre Zachman ainsi que la comparaison des architectures entre les organisations. Elle garantit également, **grâce au CTI**, que la

conception, la réalisation et la tenue à jour des composants de TI du système statistique demeurent en adéquation avec les besoins opérationnels de Statistique Canada.

18. Le CTI et le guide statistique sont *dynamiques* en ce sens qu'il faudra les mettre à jour si les besoins opérationnels changent (des exigences accrues en matière de sécurité à l'appui du processus de déclaration électronique des données, par exemple) et si de nouvelles technologies sont mises au point. Le diagramme de la hiérarchie des fonctions et le modèle général (DER) sont des modèles formels (des diagrammes de cheminement des données et des modèles de flux de travaux sont d'autres exemples) pouvant être créés et tenus à jour à l'aide d'un outil de génie logiciel assisté par ordinateur (CASE). Cette capacité de s'adapter aux besoins changeants est importante pour que les modèles tiennent compte des priorités et des besoins opérationnels.

D. De quelle façon le CTI, le guide statistique et le cadre Zachman contribuent-ils à la gestion de la complexité?

« Il existe une loi générale dans l'histoire de la science et des idées selon laquelle la simplicité suit la complexité et ne la précède pas. Les théories et les modèles sont perçus comme étant extrêmement complexes au moment de leur création et avant qu'ils ne soient largement acceptés. Par la suite, les théories sont simplifiées et ramenées à leurs éléments utilisables fondamentaux. Des parties seulement des théories sont utilisées, et les concepts sous-jacents de ces théories deviennent des parties de la culture commune et universelle. La théorie de la gravité de Newton et la théorie de l'électromagnétisme de Maxwell constituent des exemples de ce phénomène. » (Biezunski, 2002)
(Traduction)

19. À l'origine, le guide statistique a été mis au point en guise de modèle à l'appui de l'analyse architecturale dans le cadre du projet de connectivité et d'intégration géographique (Graves, Murray, Rushforth, 1998), parrainé par la Division de la géographie de Statistique Canada. Il a été utilisé en conjonction avec le cadre Zachman dans des projets d'architecture subséquents (Hutton, Graves, Meagher, 2000, et Graves, Hutton, 2001). Il a servi plus récemment permis à l'Office of National Statistics (ONS) du Royaume-Uni de connaître le degré d'harmonisation entre le CTI de Statistique Canada, le guide statistique et les principes d'architecture de l'ONS. Chacun de ces projets nous a permis de mieux comprendre le guide statistique. La présente communication était le catalyseur de l'élaboration d'une version moins complexe du modèle général (DER).

20. Le modèle général montre sur une page les plus importants composants de l'infrastructure d'automatisation. Combiné au CTI, il jette les bases de l'étude d'importantes questions liées aux TI. Voici trois exemples des questions générales pouvant être soulevées.

- Les travailleurs du savoir internes sont les plus importants atouts d'un bureau de la statistique. Mettons-nous à leur disposition les outils généraux qui leur permettront d'être efficaces sur le plan de la réalisation (services de génie logiciel assisté par ordinateur, services de consultation et de mise à jour en ligne des entrepôts de documents et de métadonnées, etc.) et de l'exécution (service de codage, service de gestion du contenu, etc.)?
- La réutilisation du matériel existant (documentation, métadonnées opérationnelles et techniques, données et logiciels) occupe une place centrale dans la présente communication. La réutilisation judicieuse se traduit par des systèmes de meilleure qualité, d'importantes économies et des délais de développement plus courts. Les magasins de données centraux qui favorisent la réutilisation sont indiqués de façon explicite (fonds de documents électroniques, fonds du métacontenu, etc.) ou implicite (Registre des entreprises, entrepôt de données, etc.) dans le diagramme. Ces magasins de données sont-ils utilisés efficacement par le bureau? Doit-on en ajouter d'autres (registre et dépôt de métadonnées techniques, etc.) pour simplifier davantage l'élaboration de systèmes?

- Un certain nombre d'entités figurant dans le modèle général représentent des registres généraux. Tous les registres (Gallagher et Carnahan, 2000) offrent un ensemble commun de services, et chaque registre offre un ensemble unique de services. Le bureau a-t-il eu recours à une stratégie concertée pour élaborer ses nombreux registres, ou ces registres ont-ils été établis de façon autonome? Possédons-nous un registre général de nos employés répondant à nos besoins en données administratives et à nos besoins en sécurité, par exemple?

21. Une fois le débat terminé, la direction pourra décider, en se fondant sur les priorités générales, quelles initiatives doivent être mises en œuvre pendant un exercice précis de planification.

22. Le guide statistique peut faire office de document architectural permanent et faire l'objet d'examens périodiques comme le CTI. Ces examens peuvent être effectués pour donner suite aux changements apportés au CTI. L'établissement de documents sur le guide statistique et l'harmonisation de celui-ci avec le cadre Zachman et le CTI :

- permettront à la direction d'évaluer le degré de conformité des futurs projets d'élaboration de systèmes en fonction des modèles;
- mettront à la disposition des architectes et des réalisateurs de systèmes des lignes directrices sur l'exécution des futurs projets d'élaboration de systèmes;
- favoriseront l'indexation dans le corpus de documents sur l'architecture d'entreprise;
- permettront de cerner le composant de l'infrastructure qui, une fois achetés ou mis au point, réduiront les échéanciers d'élaboration des projets nouveaux et remaniés.

23. Enfin, les documents sur le guide statistique seront établis à un niveau permettant dans une grande mesure la transmission des connaissances et des termes spécialisés à divers employés (recrues, employés expérimentés se joignant à Statistique Canada, entrepreneurs, spécialistes, etc.). Le document, s'il est utilisé en conjonction avec d'autres documents de référence (CTI, cadre Zachman, aperçu du traitement des données des enquêtes-entreprises, etc.) et méthodes, permettra à ces groupes cibles de mieux comprendre les dossiers et d'acquérir un vocabulaire favorisant l'efficacité de la communication. Ces avantages se traduiront ensuite par des systèmes de meilleure qualité.

III. CONCLUSION

24. Statistique Canada tient déjà compte d'un certain nombre d'initiatives mentionnées aux figures 1 à 4 inclusivement.

- Nous avons mis au point des logiciels généralisés pour nombre de nos processus statistiques génériques (échantillonnage, vérification et imputation, estimation, etc.), et nous utilisons d'autres outils génériques (Blaise, SAS, etc.) pour élaborer des types précis de systèmes.
- Les activités relatives à la plupart de nos bases de données de référence relèvent de la Direction des systèmes de classification. Ainsi, le Registre des entreprises relève de la Division du Registre des entreprises. Il est permis d'établir des bases de sondage pour des enquêtes-entreprises, de réaliser des enquêtes et de définir l'information en retour. La base de métadonnées intégrée est exploitée par la Division des normes et sert de dépôt central pour stocker des *métadonnées opérationnelles* et des *classifications d'entreprises*.
- Les activités de collecte relèvent de trois divisions : la Division des opérations et de l'intégration, la Division de la recherche et développement des opérations et la Division des opérations des enquêtes. La Division de la diffusion est responsable de l'exploitation et de la tenue à jour de notre entrepôt de données central, CANSIM II, lequel prend en charge les analyses transversales et chronologiques de nos macrodonnées.
- Le cadre Zachman est couramment utilisé par les architectes en TI du Bureau. Les documents (Graves, Hutton, 2002) de la série sur l'architecture d'entreprise montrent de

quelle façon le guide statistique, conjointement avec le cadre Zachman, peut servir de *modèle* pour l'établissement de plans d'activités concernant des projets de systèmes nouveaux ou remaniés.

25. Le principe premier de la présente communication est que les modèles architecturaux au niveau des entreprises sont essentiels à la gestion de la complexité technique. Les modèles que comprend le guide statistique ont été mis au point dans ce contexte.

26. Un deuxième principe commande que les modèles architecturaux au niveau des entreprises soient validés en fonction des modèles formels utilisés pour concevoir et élaborer les systèmes existants. Cette activité ne doit pas être négligée à l'avenir. Lors de l'élaboration de la version courante de ces modèles, une certaine forme de validation, dans le contexte de nos enquêtes-entreprises (Hutton, 1999), a été effectuée par les auteurs. En fait, il serait avantageux que d'autres bureaux de la statistique adoptent le guide statistique afin d'élaborer un ensemble de modèles décrivant tous les bureaux de la statistique.

Bibliographie

- BIEZUNSKI, MICHAEL, "Introduction to the Topic Maps Paradigm", XML Topic Maps, Addison-Wesley, July 2002.
- GALLAGHER, L., CARNAHAN, L., "A General Purpose Registry/Repository Information Model", National Institute of Standards and Technology, October 2000.
- GRAVES, R., MURRAY, S., RUSHFORTH, P., "Geographic Integration and Connectivity", Statistics Canada Working Paper, April 1998.
- GRAVES, R., HUTTON, T., "Business Architecture: General Topics and Strategies", Statistics Canada Working Paper, June 2002.
- GRAVES, R., HUTTON, T., "Business Architecture at Statistics Canada", Statistics Canada Working Paper, June 2002.
- HUTTON, T., "An Overview of Business Surveys Processing", Statistics Canada Working Paper, July 1999.
- HUTTON, T., GRAVES, R., MEAGHER, G., "An Architecture to Support Input-Output Table Production and SNA Branch Reconciliation", Statistics Canada Working Paper, June 2000.
- HUTTON, T., GRAVES, R., "Coherence of Reference Databases Architecture Study", Statistics Canada Working Paper, June 2001.
- INFORMATICS COMMITTEE, "Information Technology Framework", Statistics Canada Reference Document, June 2001.
- SUNDGREN, B., "Information Systems Architecture for National and International Statistical Offices", United Nations, Geneva 1999.
