

Entrepôt de données et outils d'analyse pour faciliter l'intégration des comptes macroéconomiques canadiens

Alistair Macfarlane¹

Résumé

L'entrepôt de données a permis de moderniser la façon dont le Système canadien des comptes macroéconomiques (CME) est produit et analysé aujourd'hui. En raison de son évolution continue, la quantité et les types de travaux analytiques qui sont effectués à l'intérieur des CME augmentent. Il apporte l'élément nécessaire d'harmonisation et de confrontation pendant l'évolution des comptes macroéconomiques vers une intégration complète. Les améliorations de la qualité, de la transparence et de l'actualité ont permis le raffermissement des statistiques qui sont diffusées.

1. Introduction

Les CME font partie d'un système de comptabilité macroéconomique hautement intégré, qui repose sur l'intégration de concepts, de méthodes et de données pour une série de comptes économiques portant sur la production, la consommation, les revenus, les flux financiers et les stocks. Cette intégration s'ajoute à la qualité globale des statistiques produites à l'intérieur du système, mais elle comporte des défis.

Statistique Canada a notamment pu relever ces défis en élaborant un entrepôt de données, qui facilite ce processus d'intégration. Le présent document fait état de la nature intégrée des CME et précise comment l'entrepôt de données facilite la compilation des données et agit comme outil d'analyse multidimensionnelle, fournissant des renseignements commerciaux analytiques aux responsables de la compilation du Système de comptabilité nationale du Canada (SCNC) à Statistique Canada.

2. Les comptes macroéconomiques canadiens

Les comptes macroéconomiques canadiens fournissent des renseignements économiques au public canadien et aux intervenants du domaine économique, grâce à un cadre central hautement intégré, qui valide les rapports entre la production, la consommation et l'accumulation de revenu et de richesse². En soit, chaque composante révèle un aspect important de l'économie canadienne à un moment particulier de l'histoire ou fournit un point de vue dans la série chronologique. Toutefois, la valeur ajoutée dans le contexte canadien est l'approche de rapports intégrés, qui exige que les données sous-jacentes soient synchronisées avec chaque composante. Par conséquent, un dollar de dépense équivaut à un dollar de revenu à n'importe quel moment donné dans le temps.

Il est loin d'être vrai que l'on peut prétendre que ce processus est relativement facile ou qu'il est considéré comme la norme pour un organisme statistique national. L'engagement de Statistique Canada à améliorer la qualité et l'actualité de ses statistiques macroéconomiques, grâce à l'intégration et au rapprochement du cadre, est unique en son genre. Il

¹ Alistair Macfarlane, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario), Canada, K1A 0T6 (alistair.macfarlane@canada.ca)

² Tebrake, J. (2008) – Directeur général, Direction des comptes macroéconomiques, Statistique Canada, CSMA Foundations_Module 1_Essential Concepts_3-Powerpoint Presentation

a vraiment commencé avec la révision historique du SCN en 1986, alors que l'agrégat du PIB commun a vu le jour. Selon Kishori Lal, ancien directeur général des Comptes macroéconomiques : « L'intégration exige un engagement formel et à long terme des cadres supérieurs, la collaboration volontaire des diverses parties du SCNC et la ferme résolution d'élaborer et de mettre en œuvre un ensemble commun de règles de classification et d'agrégation... »³ Fidèle à ces propos, 30 ans plus tard, l'intégration des comptes macroéconomiques est toujours en pleine évolution.

3. Les avantages d'une plus grande intégration

Au départ, bon nombre des différentes données d'entrée macroéconomiques produites par le Système de comptabilité nationale de l'époque (maintenant le CME) étaient produites d'une manière relativement indépendante les unes des autres. Même s'il existait des intrants communs sous-jacents, le travail en vase clos ou à un moins grand niveau d'intégration a donné des résultats différents. Les calendriers de production, les normes comptables et les cycles de révision étaient variés. On notait des différences entre les secteurs et les univers, ainsi que dans les niveaux de granularité et les fréquences de déclaration et de disponibilité des données. Moins le cadre était intégré, plus il était probable que des écarts statistiques soient présents. Lorsque deux produits mesurent conceptuellement le même phénomène, mais aboutissent à des valeurs différentes, on se demande : lequel est le bon? Le rapprochement des données des comptes fait partie intégrante de la validation et est essentiel à tout processus statistique. Toutefois, la capacité de rapprocher et d'expliquer les différences, au moment opportun, a augmenté en parallèle avec la taille et la complexité des comptes macroéconomiques. Le moment était venu d'adopter une approche plus intégrée pour les comptes macroéconomiques et il était plus essentiel d'avoir un outil intégré pour rapprocher les différences. Par conséquent, l'entrepôt de données a été élaboré afin d'alléger l'effort conceptuel et des TI requis pour confronter les données de différentes sources.

Les avantages de comptes macroéconomiques plus intégrés sont nombreux, le plus important étant le fait que la qualité des données est renforcée. L'interrelation entre la production, la consommation, et l'accumulation de revenus et de richesses est plus évidente avec l'intégration. Un mouvement dans l'un des agrégats se reflète immédiatement dans toutes les autres composantes, en raison du lien entre les comptes.

Davantage de données réelles seront disponibles pour être utilisées dans chaque cycle de production. Grâce à l'intégration des comptes macroéconomiques, il est moins nécessaire d'estimer les composantes macroéconomiques qui étaient auparavant désynchronisées en raison d'un système moins intégré. Les données réelles sont plus facilement disponibles et sont utilisées dans les produits statistiques, ce qui suscite un plus grand niveau de confiance à l'égard de l'information.

La méthode intégrée permet de confronter d'autres sources de données, confirmant aux utilisateurs que le message statistique est transmis comme il se doit. En théorie, plus le nombre de sources de données que l'on peut utiliser dans une confrontation ou comme indicateur est élevé, plus la validité des données est grande.

Le caractère interchangeable des sources de données est un autre avantage. Un seul fournisseur peut fournir plusieurs produits qui peuvent être traités à nouveau ou adaptés aux besoins particuliers du consommateur.

4. Une plus grande intégration signifie davantage de codépendance

Or, l'intégration évolue en même temps que l'interdépendance et la complexité de la production de statistiques en temps opportun. Pour élaborer un cadre plus intégré, il faut une plateforme relationnelle qui permet de transformer les données sources en plusieurs cadres de classification et de faire un rapprochement entre ses cadres. De même, toutes les sources de données ne sont pas directement produites par les comptes macroéconomiques de manière à être couplées et visibles pour tous.

³ Kishori, L. (2003), « Le Système de comptabilité nationale du Canada – Un cadre intégré », dans Recueil d'articles de Kishori Lal, Bibliothèque de Statistique Canada, n° 13-598-XIF au catalogue, page 315.

Quand les données sont de plus en plus partagées ou qu'elles deviennent les intrants d'autres estimations, il existe un niveau accru de codépendance. Afin de produire des statistiques en temps plus opportun pour les intervenants, il faut qu'il y ait un degré élevé d'harmonisation à l'intérieur des calendriers de production des comptes macroéconomiques. Cela permet aux données de circuler le plus rapidement possible dans le système de production et d'être diffusées de manière plus opportune.

Un peu comme on conçoit le transfert le plus efficace d'un témoin entre des coureurs à relais de manière à réduire le plus possible la durée de la course, l'intégration des composantes macroéconomiques doit être efficace et transparente afin de produire des statistiques actuelles.

5. L'entrepôt de données facilite l'intégration

L'entrepôt de données permet la compilation et la production de statistiques macroéconomiques. Les données sources sont présentées dans de nombreux cadres de classification où un rapprochement est possible. Des sources de données telles que l'EDI⁴, les données sur les salaires et les avantages et les données fiscales sont mises à la disposition des analystes des CME qui les utilisent pour confronter les données.

Par conséquent, l'entrepôt de données des CME représente, au niveau conceptuel, l'équivalent des « routes et aqueducs »⁵ et permet la comparaison et la confrontation de flux de données de plusieurs divisions et de différents systèmes de traitement (p. ex., INCOMM, FAME, KEYNES)⁶, ce qui mène à un ensemble plus intégré de comptes macroéconomiques. L'entrepôt permet aussi d'explorer les données jusqu'au plus faible niveau de collecte, ce qui facilite les travaux analytiques ou l'exercice de rapprochement.

6. Étapes en vue de l'intégration

Le premier défi pour l'entrepôt de données a été d'harmoniser les questions conceptuelles. L'équipe de l'entrepôt de données (cadres supérieurs, méthodologistes, comptables, ressources des TI et experts spécialisés) a travaillé avec ardeur afin d'harmoniser les cadres, les concepts comptables et les calendriers de production. Si, en théorie, les mêmes données étaient utilisées pour élaborer deux produits macroéconomiques ou plus et que seul le système de classification différait, les chiffres devraient être égaux au niveau agrégé, par exemple, une fois par industrie et une fois par produit. Même si le système de classification était différent, la somme de l'agrégation devrait être égale. Toutefois, dans de nombreux cas, cela ne s'est pas concrétisé pour des raisons assez valables. Certains des défis qui se sont posés pendant les premières étapes de l'harmonisation des CME sont énoncés ci-après.

L'harmonisation de l'entrepôt de données a été le résultat des efforts suivants :

1. Coordination des calendriers de production. Diverses divisions de Statistique Canada produisaient et révisaient les données selon des calendriers très différents. Afin de pouvoir confronter les données, il est devenu nécessaire de synchroniser les produits. Il s'agissait non seulement du produit final, mais aussi du produit sous forme d'entrée statistique intermédiaire dans un autre processus. Du temps et des efforts considérables, pendant plusieurs cycles, ont été nécessaires pour harmoniser les calendriers d'enquête et de production, afin que le modèle d'entrepôt de données fonctionne.
2. Alignement des cadres et de la classification, et mise à jour des versions. Il fallait rationaliser et uniformiser les cadres des divisions. À mesure que l'intégration avançait, l'interdépendance dans chaque classification qui devait être uniformisée progressait également. L'uniformisation des classifications entre les divisions des

⁴ L'EDI est l'abréviation de l'Enquête sur les dépenses en immobilisations, qui est produite par Statistique Canada.

⁵ Sabourin, J.D. (2015). Référence faite par un ancien gestionnaire de contenu d'entrepôt de données des CME et employé de Statistique Canada, Jordan Sabourin, pour décrire son point de vue de la structure de l'entrepôt de données.

⁶ INCOMM est le système de production utilisé par la Division de l'analyse des industries (DAI), FAME est le système de production utilisé par la Division des comptes économiques nationaux (DCEN), et KEYNES est le système de production utilisé par la Division de la statistique du secteur public (DSSP).

CME était devenue essentielle. En outre, il a fallu assurer la synchronisation des versions, puisqu'on ne peut pas comparer les données d'une division codées pour le SCIAN⁷ 2012 à celles du SCIAN 2007.

3. Les concepts comptables devaient être harmonisés. Malheureusement, le monde de la finance ne supporte pas toujours les comparaisons. Les entreprises et les administrations publiques peuvent avoir et ont des dates de fin d'exercice différentes (civil par opposition à financier). Les cadres comptables peuvent être répartis entre la méthode de la comptabilité de caisse et celle de la comptabilité d'exercice. Il était nécessaire d'assurer l'uniformité pour pouvoir faire des comparaisons.
4. Comparabilité de la fréquence et de la granularité des rapports. On ne peut pas comparer directement des données trimestrielles et des données mensuelles ou annuelles. Il ne convient pas non plus de comparer des données provisoires et des données finales et de s'attendre à ce qu'elles soient équivalentes.
5. Établissement de concordances pour relier les classifications. Au bout du compte, la plus grande réalisation a été la capacité pour l'entrepôt d'afficher deux sources de données ou plus sur la même page et de permettre une analyse microéconomique. Cette capacité vient du fait que l'entrepôt de données relie des concordances qui permettent la transformation de différentes classifications en d'autres. Ainsi, une classification d'industrie de la Division de l'analyse des industries (DAI) peut être transformée en classification de produits de la Division des comptes économiques nationaux (DCEN), à partir des mêmes données sous-jacentes. À titre d'exemple, nous pouvons présenter l'excédent du secteur des entreprises du point de vue fiscal, de la DAI et de la DCEN, pour une valeur en dollars égale dans tous les cas, mais avec trois options de classification.

7. Comment l'entrepôt de données habilite les analystes

L'un des rôles les plus importants de l'entrepôt de données consiste à supprimer la complexité du traitement opérationnel grâce à l'automatisation. La complexité du couplage de ces composantes a augmenté en parallèle avec l'intégration accrue des comptes macroéconomiques. Il existe une possibilité accrue d'erreur puisque les experts spécialisés doivent dépendre de leur propre expertise en TI ou de celle de leurs collègues analystes pour compiler les bonnes données à analyser. L'entrepôt de données des CME permet aux analystes de déployer le temps et les efforts nécessaires pour formuler ces couplages et valider les données de l'entrepôt comme des données adaptées à leur utilisation. Un deuxième avantage vient du fait que des outils d'analyse des renseignements commerciaux sont mis à la disposition de tous les analystes et non seulement de quelques-uns d'entre eux. La plupart du temps, les analystes ont créé, par nécessité, leurs propres outils d'analyse des renseignements commerciaux, exclusivement à leur intention. L'entrepôt met ces outils à la disposition de tous, permettant à un plus grand nombre de personnes de réaliser une analyse plus approfondie.

8. Personnalisation

Même si l'entrepôt met l'accent sur l'uniformisation et l'habilitation de masse, il tient aussi compte des souhaits et des besoins particuliers de chaque analyste. Ainsi, la capacité de personnaliser et de modifier assez facilement les demandes constitue une partie essentielle de tout bon outil d'analyse. L'entrepôt de données fournit aux analystes un ensemble d'outils d'analyse souples, qui permettent l'affichage des données de bien des façons (p. ex., titres, classifications, versions ou cycles). Selon une structure de tableaux croisés dynamiques, les analystes peuvent tirer d'une liste de catégories (appelées dimensions) les variables qu'il souhaite intégrer à un rapport en format Excel. Des dimensions peuvent être ajoutées, supprimées ou juxtaposées, au besoin, selon les souhaits d'un analyste. Les données préliminaires sont disponibles au niveau agrégé ou sous-agrégé, mais une analyse plus poussée peut être faite grâce à la production de rapports SSRS⁸ secondaires, qui fournissent des données jusqu'au plus bas niveau des intrants. D'autres outils sont disponibles, comme des taux de croissance et des coefficients delta, qui peuvent être calculés par l'entrepôt de données en fonction des paramètres choisis par l'analyste dans son tableau croisé dynamique ou son rapport SSRS. Cela permet une analyse plus rapide et plus complète des données.

⁷ SCIAN 2012, une année reconnue du Système de classification des industries de l'Amérique du Nord (SCIAN).

⁸ SSRS – Sequel Server Reporting System – format de tableau pour la déclaration des données sous-jacentes.

9. Transparence

Parmi les autres avantages très importants figure la transparence des données sous-jacentes qui s'additionnent pour produire les totaux agrégés. L'entrepôt de données fournit des données des divisions à l'étape de l'adaptation à leur utilisation. Elles s'accompagnent des données sous-jacentes qui interviennent dans la production des totaux agrégés. Cela comprend les rajustements apportés, qui sont identifiés comme tels. En fin de compte, cette transparence représente la valeur ajoutée à tout processus de rapprochement, étant donné qu'elle supprime tout biais ou valeur inconnue de tout processus de rapprochement. Elle permet aussi de remettre en question les ajustements apportés, pour qu'ils puissent être expliqués et non pas seulement acceptés.

10. Élaboré pour aujourd'hui et demain

L'entrepôt de données des CME a été constitué pour faciliter l'intégration des comptes macroéconomiques aujourd'hui, mais aussi en prévision de l'adaptation aux besoins changeants de la collectivité des analystes de demain. Des mécanismes sont en place pour accepter davantage de données, des « mégadonnées », des données non structurées, toutes pouvant être utilisées à partir des bases qui sont en place. Lorsqu'on en sera rendu là, l'entrepôt de données deviendra un outil d'analyse multidimensionnelle, qui fournira des renseignements commerciaux analytiques, afin que l'expertise de l'analyste puisse être mise à profit et que d'autres données statistiques puissent être dérivées de ces données. Il s'agira de produire davantage avec un moins grand effort des TI, ainsi que plus de transparence et moins d'écarts statistiques. L'entrepôt de données offre la possibilité de procéder à la conversion avec transparence, en fournissant des tableaux de concordance dans le processus de métadonnées.

Il faut harmoniser davantage les processus opérationnels et les systèmes de production pour permettre une plus grande intégration avec d'autres aspects des comptes macroéconomiques. De petites étapes continuent d'être franchies, à mesure que les utilisateurs cherchent des données d'entrée et que les ressources nécessaires deviennent disponibles, mais on est encore loin de l'intégration complète. Peu de personnes mettent en doute l'avantage des comptes macroéconomiques plus intégrés et les avantages que procure l'entrepôt de données des CME pour produire de meilleures statistiques. Toutefois, il faut tenir compte du temps et des ressources nécessaires pour élargir l'intégration, sans compromettre ce qui existe déjà.

11. Besoins méthodologiques

L'entrepôt de données est un outil d'analyse multidimensionnelle en constante évolution, dont l'utilisation n'est limitée que par l'imagination des analystes. À mesure que les comptes macroéconomiques et d'autres sources de données seront intégrés, on aura de plus en plus d'occasions de percevoir les données différemment. Tout comme l'intégration des comptes macroéconomiques proprement dits, l'adoption de l'entrepôt de données comme outil d'analyse par défaut est un processus continu. Il faut encore faire un investissement opérationnel considérable pour que davantage de données des divisions se retrouvent dans l'entrepôt. Toutefois, à mesure qu'un plus grand nombre de données sont disponibles, on doit également s'assurer, au niveau méthodologique, que les méthodes et les théories statistiques appliquées aux systèmes des TI de l'entrepôt de données sont appropriées et, dans le cas des nouveaux progrès, que les principes fondamentaux sont respectés lorsque l'on produit des statistiques « à la fine pointe des connaissances ». La méthodologie devra jouer un rôle important dans l'avenir de l'entrepôt de données.

12. Conclusion – Ouvrir la porte

L'entrepôt de données peut et doit élargir ses horizons. Nous (à proximité du produit) voyons le potentiel immense du système. Nous percevons la croissance énorme des travaux analytiques qui pourrait se produire si tous les comptes macroéconomiques étaient pleinement intégrés. Les améliorations de la qualité, de l'actualité et de la

responsabilisation qui accompagnent l'approche intégrée seraient présentes dans un tel scénario et les produits seraient améliorés. Les utilisateurs de l'entrepôt d'aujourd'hui en demandent plus. Les utilisateurs occasionnels du passé sont maintenant de gros utilisateurs. Non seulement l'entrepôt de données peut-il en faire plus qu'il y a deux ans, mais encore il continue d'intégrer les changements que chaque analyste souhaite, tout en s'harmonisant avec la vision d'intégration que Statistique Canada entrevoit pour l'avenir des statistiques macroéconomiques. Le passage de l'EAM au PISE⁹ et l'ajout de dix nouveaux cubes constituent des exemples de premier plan de la façon dont l'entrepôt a transformé la production des statistiques macroéconomiques.

À mesure que l'intégration se poursuivra, on s'attendra à ce que l'entrepôt de données contienne, traite et couple davantage de données économiques. Le nombre d'utilisateurs s'accroîtra, tout comme la variété et la complexité des outils d'analyse dont ils auront besoin. Les utilisateurs demandent déjà plus de rapports, plus de dimensions et plus de données. L'équipe de l'entrepôt de données doit relever directement ces défis, afin de permettre aux analystes de mieux comprendre les données. Il faut explorer non seulement les données proprement dites, mais aussi la façon dont elles interagissent avec d'autres éléments du système économique, tant dans le contexte canadien qu'étranger. Il faut paver la voie à un aperçu plus clair, accessible à tous.

⁹ L'EAM (Enquête annuelle des manufactures) est maintenant une composante beaucoup plus importante du PISE (Programme intégré de la statistique des entreprises).