

## QUELLE EST L'IMPORTANCE DE L'EXACTITUDE?

Gordon Brackstone<sup>1</sup>

### RÉSUMÉ

Les présentes notes résument une discussion sur l'importance de l'exactitude dans le cadre plus large de la qualité des données adopté par nombre d'organismes statistiques.

MOTS CLÉS : qualité des données; exactitude.

### 1. INTRODUCTION

La question posée au groupe de discussion pourrait susciter toute une gamme de réponses (encadré 1). Comme il ne s'agit pas d'une discussion sur la conception d'un questionnaire, nous ne devons pas débattre de la formulation de la question ni du caractère approprié de la réponse. Toutefois, les diverses réponses illustrent une variété d'attitudes à l'égard de l'exactitude. Pour certains utilisateurs de données, le concept d'exactitude (du moins au sens où les statisticiens l'entendent) pourrait être difficile à apprécier et il nous arrive encore parfois d'observer l'attitude selon laquelle \* n'importe quel chiffre fera l'affaire +. Pour d'autres, l'exactitude est un élément dont il faut avoir conscience (en parler pour la forme?), mais dont il ne faut pas faire tout un plat. La troisième réponse représente une opinion pondérée et politiquement correcte. La dernière, naturellement, correspond à ce que croient tous les statisticiens.

	<i>Encadré 1</i>
Quelle importance faut-il accorder à l'exactitude?	
C	Ne vaut pour ainsi dire pas la peine que l'on s'en soucie.
C	Y penser, mais ne pas en faire tout un plat.
C	Un élément important dont il faut tenir compte lors de la conception d'une enquête.
C	La seule chose qui a de l'importance.

Les communications que nous avons écoutées durant le symposium peuvent être réparties en deux grandes catégories. La première est celle des communications que nous pourrions considérer de *macroniveau* qui traitent de la gestion globale de la qualité des données d'un organisme statistique, en cherchant à établir un équilibre entre ses diverses dimensions, et en tenant compte d'autres contraintes et considérations qui influencent les travaux de ce genre d'organisme. La deuxième est celle des communications de *microniveau* qui traitent d'applications particulières, ainsi que de moyens d'optimiser divers aspects de la qualité, souvent l'exactitude, dans des contextes particuliers.

---

<sup>1</sup> Statisticien en chef adjoint, Secteur de l'informatique et de la méthodologie, 26-J Immeuble R.-H. Coats, parc Tunney, Ottawa (Ontario) Canada, K1A 0T6.

## 2. DÉFINITION DE LA QUALITÉ DES DONNÉES

Aujourd'hui, la plupart des organismes statistiques adoptent une définition de la qualité des données rendant compte de la mesure dans laquelle elles sont adaptées à leur utilisation. Par exemple, dans le rapport du Leadership Group on Quality de la Commission européenne (Lyberg et coll., 2001), nous lisons :

[Traduction]

*\*L'essentiel est que le concept de qualité adopté par les organismes statistiques a évolué au cours des dix dernières années. Donc, l'exactitude n'est plus la seule mesure de la qualité. Cette dernière correspond à plusieurs caractéristiques qui reflètent les besoins de l'utilisateur. Dans cette optique, on peut définir la qualité en fonction de diverses dimensions, l'une d'entre elles étant l'exactitude. Toutes ces dimensions constituent la qualité du produit. +*

Ce commentaire sous-entend une définition multidimensionnelle de la qualité des données tenant compte de divers attributs de ces dernières. Les attributs que choisissent les organismes statistiques pour définir la qualité varient, mais ils sont fort semblables.

Par exemple, dans un document décrivant le cadre de l'assurance de la qualité à Statistique Canada (Statistique Canada, 1997), il est mentionné que :

[Traduction]

*\*... nous pouvons définir le concept de qualité comme englobant les aspects des produits statistiques d'un bureau national de la statistique qui reflètent la mesure dans laquelle ces produits sont adaptés à l'usage que veulent en faire les clients. +*

Les aspects ou dimensions énoncés dans la définition de la qualité adoptée par Statistique Canada sont les suivants : *pertinence; exactitude; actualité; accessibilité; interprétabilité et cohérence.*

Certains organismes statistiques ont dressé d'autres listes où figurent des noms différents pour certaines dimensions, mais toutes incluent *l'exactitude*.

Naturellement, je fonderai les commentaires qui suivent sur les six dimensions incluses dans la définition de Statistique Canada. Très brièvement, cette liste et l'ordre dans lequel elle est établie résultent de l'enchaînement imaginaire de questions que pourrait se poser un utilisateur des données :

*Cette information est-elle en rapport avec le sujet qui m'intéresse, y compris les variables que j'étudie, et s'appuie-t-elle sur des concepts et des définitions en harmonie avec le problème que je cherche à résoudre? Dans l'affirmative, l'exactitude de la mesure des variables suffit-elle à mes besoins? Les données sont-elles produites dans les délais prévus? Puis-je obtenir les données facilement, dans le format que je souhaite, à un prix abordable? Une fois que j'ai reçu les données, puis-je les comprendre et ai-je des renseignements sur leur fiabilité et sur la méthodologie qui les sous-tend? Enfin, sont-elles adaptées à d'autres données connexes, sont-elles cohérentes au fil du temps et puis-je les utiliser en combinaison avec d'autres ensembles de données?*

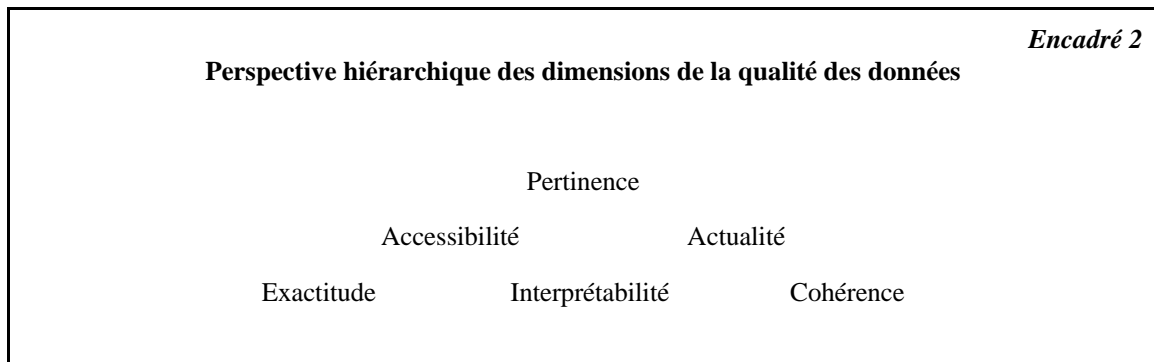
Pour une discussion plus approfondie de ces dimensions et de la façon dont un organisme statistique peut les gérer, consulter Brackstone (1999, 2000, 2001).

### 3. IMPORTANCE DE L'EXACTITUDE

Lorsqu'on adopte une perspective multidimensionnelle de la qualité des données, où est l'une des dimensions envisagées, la question de savoir *quelle est l'importance de l'exactitude* devient un élément majeur lors de la conception, de la mise en place et de l'évaluation des programmes statistiques. La définition multidimensionnelle de la qualité des données sous-entend que, si les critères établis pour l'une des dimensions ne sont pas satisfaits, les données produites peuvent être de mauvaise qualité, c'est-à-dire inutilisables. Si l'on considère les dimensions de la qualité en série plutôt qu'en parallèle, on s'aperçoit qu'elles sont toutes aussi importantes les unes que les autres. L'exactitude est une condition nécessaire, mais non suffisante, de l'adéquation des données à l'usage que l'on veut en faire.

Toutefois, on peut aussi considérer les six dimensions de façon hiérarchique (encadré 2). On met alors l'accent sur le fait que :

- a) sans la pertinence, les cinq autres dimensions n'ont aucune importance — des renseignements parfaits sur le mauvais sujet sont inutiles;
- b) étant donné la pertinence, sans actualité ni accessibilité, les données ne sont pas disponibles au moment où l'utilisateur en a besoin;
- c) l'exactitude, l'interprétabilité et la cohérence ne deviennent des dimensions importantes que lorsque la pertinence, l'actualité et l'accessibilité sont satisfaites.



Cette façon de considérer les dimensions de la qualité paraît donner une importance moindre à l'exactitude.

En fait, ni l'une ni l'autre de ces perspectives ne semblent mettre l'exactitude sur un piédestal — ce qui pourrait décevoir certains statisticiens. Poussons plus loin notre examen.

### 4. POURQUOI AVONS-NOUS CHOISI DE PARLER DE L'EXACTITUDE?

Avant tout, pourquoi avons-nous choisi de nous interroger sur *l'exactitude* plutôt que sur *l'actualité* ou *l'accessibilité*, par exemple?

Du côté négatif, nous pourrions soutenir que :

- C dans le cas de nombreuses enquêtes, nous consacrons beaucoup d'efforts et d'argent (certains diraient une quantité disproportionnée) aux questions ayant trait à l'exactitude — parfois on nous accuse de faire des ajustements beaucoup trop minutieux par souci de exactitude;

- C d'aucuns affirment que nombre d'utilisateurs ne sont guère intéressés par les mesures de l'exactitude, situation confirmée, dans une certaine mesure, par le niveau de demande d'information sur cet attribut;
- C nous ne pouvons mesurer que certains aspects de l'exactitude, si bien que nos mesures sont de toute façon fautives.

Du côté positif, nous pourrions affirmer que :

- C la connaissance du degré d'exactitude est essentielle à l'utilisation appropriée des données;
- C nous (les bureaux de la statistique) avons la responsabilité de renseigner les utilisateurs sur l'exactitude des données et de fournir l'information nécessaire pour interpréter et utiliser les mesures de l'exactitude;
- C nous devrions nous efforcer de combler les lacunes en ce qui a trait à la capacité de mesure, particulièrement dans le cas de l'erreur non due à l'échantillonnage.

## **5. ACTIVITÉS STATISTIQUES ET QUALITÉ**

Ayant énoncé certaines raisons pour lesquelles il faut accorder une attention particulière à l'exactitude des données, nous allons maintenant essayer d'examiner les diverses dimensions de la qualité des données d'un point de vue différent, c'est-à-dire celui du producteur plutôt que de l'utilisateur des statistiques. Pensons aux activités d'un organisme statistique, ou aux étapes de l'élaboration d'un programme, qui influent tout spécialement sur chaque aspect de la qualité. Autrement dit, essayons de déterminer quelles sont les étapes du processus d'enquête ou les aspects de l'infrastructure d'un bureau de la statistique qui ont un effet sur les diverses dimensions de la qualité.

Nous considérons pour cela quatre \* étapes +clés auxquelles il est possible de déterminer les caractéristiques de la qualité, à savoir :

- C la définition des programmes et l'établissement des priorités, c'est-à-dire le choix des programmes qui seront mis en œuvre et du budget qui leur sera accordé;
- C la conception et la mise en œuvre des programmes statistiques;
- C l'évaluation de l'efficacité des programmes statistiques et de leurs résultats;
- C la part de l'infrastructure de l'organisme statistique utilisée par les programmes individuels (p. ex., capacité des opérations sur le terrain; registres; méthodologie; systèmes communs; circuits de diffusion).

Les premier et dernier de ces éléments sont des activités générales, communes aux divers programmes; les deuxième et troisième sont des activités qui varient selon le programme.

L'encadré 3 précise les étapes primaire et secondaire pour chaque dimension de la qualité. Avant de décrire le contenu de cet encadré, permettez-moi de formuler rapidement certaines mises en garde. Il s'agit d'un diagramme schématique qui vise à donner une idée des étapes où ont lieu les principales interactions et qui ne doit nullement être considéré comme définitif. On pourrait, en effet, justifier une entrée avec un certain niveau d'importance dans presque toutes les cellules du tableau — chaque étape est pertinente dans une certaine mesure pour chaque dimension de la qualité. Néanmoins, nous nous concentrerons sur les interactions les plus importantes.

<b>Dimensions de la qualité selon l'activité statistique</b>				<i>Encadré 3</i>
	Définition du programme	Conception	Évaluation	Infrastructure
Pertinence	XXX	X		
Exactitude	X	XXX	X	
Actualité	X	XXX		
Accessibilité		X		XXX
Interprétabilité		X	XXX	
Cohérence		X		XXX

La pertinence est déterminée en grande partie par le choix des programmes qui seront mis en œuvre et par les décisions concernant leur portée et leur financement. L'évaluation de la pertinence est, en grande partie, un problème commun aux divers programmes, la pertinence des produits d'un programme dépendant des autres produits disponibles. Cependant, l'étape de la conception peut aussi avoir un effet important sur la pertinence, puisque c'est à ce stade que l'on choisit et que l'on définit les variables qu'il convient de mesurer.

L'exactitude et l'actualité sont principalement des problèmes de conception, étant en grande partie déterminées par les compromis qui sont faits à cette étape et à celle de la mise au point des programmes. Il s'agit des domaines où les statisticiens et les méthodologistes appliquent leurs théories et méthodes statistiques en vue de déterminer quels sont les plans d'enquête optimaux. Toutefois, la précision et l'actualité que permettent de réaliser les plans d'enquête optimaux sont assujetties aux contraintes des allocations budgétaires faites à l'étape de la définition du programme, allocations qui sous-entendent des plafonds de précision et d'actualité. La mesure du degré de précision atteint est une tâche si considérable (la mesure de l'actualité est triviale) que nous avons aussi indiqué un effet secondaire dans la colonne de l'évaluation.

L'accessibilité est principalement un problème d'infrastructure puisqu'elle dépend des circuits dont dispose le bureau pour diffuser les données aux divers utilisateurs et des politiques en matière d'accès aux données. Au sein des programmes individuels, la conception d'un produit particulier peut être considérée comme un problème de conception.

L'interprétabilité relève avant tout du domaine de l'évaluation et de la documentation, la mesure de la précision contribuant considérablement à ce facteur. C'est à l'étape de la conception qu'est déterminée la façon dont ces évaluations seront réalisées et quelles mesures devront être intégrées dans le programme statistique proprement dit pour permettre cette évaluation.

La cohérence repose sur l'infrastructure des normes relatives au contenu ainsi qu'aux méthodes. L'utilisation de concepts et de variables normalisés favorise la cohérence du contenu des programmes et des mesures effectuées, tandis que l'utilisation de cadres de référence, de méthodes et de systèmes communs permet de mesurer les concepts de façon cohérente. La conception est une étape importante où il est déterminé comment ces normes seront appliquées aux programmes individuels.

Quelles conclusions pouvons-nous tirer de cet examen schématique? En ce qui concerne la gestion globale de la qualité, le tableau montre l'effet important de la définition du programme sur la pertinence, le rôle crucial

de la conception pour ce qui est de la précision et de l'actualité, la nécessité de l'infrastructure pour assurer l'accessibilité et la cohérence, et le rôle de l'évaluation dans la production de données intelligibles.

Soulignons l'importance de la colonne de la conception. La plupart des communications faite durant le symposium ont eu trait à la conception. Celle-ci est l'étape où il faut peser les considérations quant à la précision, l'actualité et le coût, et trouver un juste équilibre, en accordant également de l'importance à la pertinence, à l'interprétabilité et à la cohérence. Voilà pourquoi il était naturel de se concentrer sur la précision durant cette session.

## 6. CRÉDIBILITÉ

La crédibilité est un autre aspect des données qui pourrait avoir plus d'importance que la précision. Autrement dit, la mesure dans laquelle les utilisateurs ont confiance dans les données pourrait dépendre partiellement de leur précision, mais davantage de la réputation de l'organisme statistique, laquelle, à son tour pourrait dépendre en partie de la capacité qu'a l'organisme de produire des données exactes pendant une longue période de temps.

L'inexactitude flagrante met en péril la crédibilité. C'est en partie pour cela que les organismes statistiques veillent tant à éliminer des réponses les anomalies illogiques. En outre, l'inexactitude des données d'une série peut miner la crédibilité d'autres séries parfaitement correctes. Cependant, la plupart des inexactitudes ne sont ni flagrantes ni visibles. Seul l'organisme qui produit les données peut dire dans quelle mesure ces dernières sont exactes et, souvent, même l'organisme ne peut brosser un tableau complet.

Nous ne pouvons gérer la précision sans la mesurer. Bien connus, les problèmes que pose la mesure de la précision ont été exposés de nouveau durant le symposium. Nous devons poursuivre nos travaux en vue de résoudre cette question — les autres membres du groupe de discussion m'appuierons à cet égard.

Malgré ce que nous savons au sujet de la précision, un deuxième aspect de la question mérite notre attention. Quelles lignes directrices devons-nous donner aux utilisateurs des données qui se servent de l'information sur la précision? Pour les erreurs-types, nous pouvons proposer des intervalles de confiance, par exemple, mais pour d'autres mesures, comme les taux de réponse ou les mesures de la couverture, en faisons-nous suffisamment, outre avertir qu'il faut \* utiliser les données avec prudence +, pour aider l'utilisateur à tenir compte de cette information lorsqu'il utilise les données? Il est encourageant de constater que, selon certaines communications faites durant le symposium, des efforts sont faits dans cette direction.

## 7. CONCLUSION

Mais alors, me direz-vous, quelle est la réponse à la question? Il s'agit d'un groupe de discussion et non d'une communication, si bien que je ne suis pas obligé de tirer une conclusion. Permettez-moi toutefois de terminer par quelques remarques sommaires.

*La précision* n'est pas une condition suffisante pour que les données soient adaptées à leur utilisation. Même si nous pouvions produire de bonnes mesures de la précision, reflétant l'erreur non due à l'échantillonnage ainsi que l'erreur d'échantillonnage, nous devrions tenir compte d'autres dimensions de la qualité, faute de quoi nos efforts en vue d'augmenter la précision des données seraient perdus.

Cependant, *la précision* est une condition nécessaire à l'adéquation des données à leur utilisation. Le manque de précision ne peut être compensé par de bons résultats sur d'autres dimensions. À long terme, la précision est un déterminant de la crédibilité, et cette dernière est une question vitale pour un organisme statistique. La précision est un élément essentiel de l'étape de conception d'un programme et, étant donné la \* perspective

méthodologique + mentionnée dans le thème du symposium, l'étape de la conception est celle qui est le point de concentration ici. Cependant, les limites quant aux résultats qui peuvent être obtenus lors de la conception sont établies, en grande partie, à l'étape de la définition du programme, lorsque les objectifs sont fixés et les budgets, alloués. Par conséquent, les considérations ayant trait à la précision doivent être exprimées très clairement au moment où sont prises les décisions relatives à la définition du programme.

## **BIBLIOGRAPHIE**

- Brackstone G (1999), La gestion de la qualité des données dans un bureau de statistique, Techniques d'enquête, Vol 25, no2, pp157-171
- Brackstone, Gordon (2000). "Managing Data Quality at Statistics Canada", Proceedings of the Statistical Quality Seminar, KNSO & IMF, Jeju, Corée, Decembre 2000.
- Brackstone, Gordon (2001). "Managing Data Quality: The Accuracy Dimension", prepared for the International Conference on Quality in Official Statistics, Stockholm, Suède, Mai 2001.
- Lyberg, Lars et col (2001). Report from the Leadership Group on Quality, présenté à l'International Conference on Quality in Official Statistics, Stockholm, Suède, Mai 14-15, 2001.
- Statistique Canada (1997). "An Outline of Statistics Canada Quality Assurance Framework". Document internal. December 1997.