



## **Data quality and socioeconomic development projects' performance measurement:**

**Critics over the convergence of data quality assessment criteria**

## **Qualité des données et mesure de performance des projets de développement socioéconomique :**

**Critique sur la convergence des critères d'évaluation de la qualité  
des données**

**Olivier Mumbere Muhongya, PhD**

Professeur Associé à Institut Supérieur de la Gombe  
Kinshasa – République Démocratique du Congo

**Résumé:** Les données chiffrées ou quantitatives sont celles qui rassemblent le plus de consensus parmi les parties prenantes aux projets de développement socioéconomiques. La plupart de fois les performances des projets sont exprimées en termes des données qualitatives ou quantitatives. Il est donc important sinon incontournable de réunir un consensus autour des critères de la qualité des données. Or, il est globalement décevant de constater que chaque bailleur dispose de ses pratiques en matière de critères de qualité. Il y a là le risque d'un chaos organisé qui ne saurait guère favoriser la lecture de performances, et partant de l'efficacité, des projets ou programmes individuels. Cet article s'appuie sur une revue de la littérature disponible sur google pour ressortir les critères les plus saillants et propose des critères les plus fréquents et pertinents.

**Mots clés :** qualité des données ; critères de qualité des données ; comparabilité.

**Abstract:** Numerical or quantitative data are those that gather the most consensus among stakeholders in socio-economic development projects. Very often project performance is expressed in terms of qualitative and/or quantitative data. It is therefore important, if not unavoidable, to build consensus around data quality criteria. However, it is actually disappointing to note that each donor has their own practices in terms of quality criteria. There is the risk of organized chaos that can hardly favor the reading of performances, and therefore of the effectiveness, of individual projects or programs. This article is based on a review of the existing literature via google and google scholar research engines to highlight the most salient criteria and proposes the most frequent and relevant criteria.

**Keywords:** data quality; data quality criteria; comparability.

**Digital Object Identifier (DOI):** <https://doi.org/10.5281/zenodo.11519043>

## 1. Introduction

L'aide internationale bilatérale ou multilatérale est vieille de plus d'un demi-siècle et a épousé différentes formes notamment assistance technique ; appui budgétaire ou projet. Cependant, la forme « projet » semble être aujourd'hui la plus préférée car elle couvre plus de 80 pourcents du flux financier relevant de l'aide aux pays en développement. Au cours du demi-siècle, la primauté a été donnée à l'aide projet et, si celle-ci a tendance à décroître, elle demeure majoritaire.

Si la forme projet tend à gagner l'unanimité auprès des bailleurs, elle demeure clivante parmi les gouvernements bénéficiaires de l'aide internationale. Il apparaît clairement que cette préférence des bailleurs est tributaire de plusieurs raisons notamment le fait qu'elle offre la possibilité d'impliquer les populations bénéficiaires et favorise l'appropriation du changement souhaité par les bénéficiaires. L'approche projet est soutenue par plusieurs auteurs notamment Milton Friedmann... Ceci est vrai tant pour l'aide bilatérale que pour celle dite multilatérale. Hélas, l'aide bilatérale et/ou multilatérale a essuyé diverses critiques dont la plus importante prétend qu'elle ne permet pas aux pays en développement de sortir du piège de l'aide. Telle est la critique provenant d'Easterly William notamment lorsqu'il évoque le « big push » de l'aide.

Lorsqu'on parle de l'aide en faveur d'un pays, on sous-entend une contribution de diverses parties prenantes au développement à savoir l'Etat, la communauté des bailleurs – prise dans sa globalité mais également dans la singularité de chaque aide, le secteur privé, les bénéficiaires etc. Chaque bailleur dispose de ses critères d'évaluation de la qualité des données et fournit des efforts pour satisfaire à ceux-ci. Parfois les mêmes critères d'évaluation de la qualité des données sont appliqués sur les mêmes données mais avec des interprétations différentes car dictées par les organisations internationales de développement différentes. D'où les questions ci-après auxquelles cet article se propose d'apporter des réponses : il y a-t-il convergence entre les divers critères d'évaluation de qualité des données ? Si oui, dans quelle mesure ces critères convergent-ils étant donné que les utilisateurs de ces critères opèrent souvent sur un même espace national qui se trouve être le réceptacle de l'aide extérieure.

Le but de cet article est de contribuer à la conciliation des différents critères d'évaluation de la qualité des données afin de permettre la comparabilité des performances au niveau projet.

### 1.1. Revue de la littérature

En 2012 Martin Goulet réalise le travail de hiérarchisation des dimensions de qualité des données. Il est intéressant de constater que Goulet ne s'intéresse pas spécifiquement aux données de performance des

projets de développement. Sa perspective sur la qualité des données est plus large. En effet, les secteurs économiques – en micro ou en macroéconomie et commerciaux ont vivement besoin des données de qualité pour la prise de décisions. Si au début les projets de développement se satisfont des données chiffrées produites par leurs financiers, très vite l'on a réalisé qu'un projet est foncièrement et par essence une organisation qui exige de la redevabilité. Une redevabilité qui doit aller au-delà de la simple utilisation des fonds via un rapportage d'un «burn rate» ou taux d'utilisation des fonds.

Pour mieux comprendre si oui ou non il existe un consensus autour des standards de qualité des données, il nous a paru nécessaire de procéder comme suit : identifier les standards de qualité des données déjà repris dans la littérature tel que compilée par Goulet d'abord (jusqu'en 2012) et ensuite ceux signalés par la suite par différentes organisations bilatérales et/ou multilatérales. Par la suite, nous nous sommes intéressés au contenu des standards dont les fréquences sont les plus élevées dans cette littérature.

Il convient de prime à bord de souligner le fait que la littérature autour de la qualité des données aussi bien en langue anglaise que française n'est, hélas, pas très prolifique. C'est ainsi que le travail de Goulet a vite attiré notre attention. Déjà en 2012, Goulet fait état de cinq dimensions à savoir 1) actualisation et disponibilité ; 2) perception, pertinence et confiance ; 3) exhaustivité ; 4) cohérence et synchronisation et 5) précision. En accord avec Goulet, nous avons retenu que ces 5 dimensions étaient les plus en vogue dans la littérature jusqu'en 2012. Nous irons jusqu'à approcher d'autres publications se réclamant de différents auteurs notamment le Gouvernement américain (via l'Agence américaine de Développement USAID, 2008) ; le Gouvernement Canadien (Institut National de la Statistique) ; les Nations Unies (via l'Organisation Mondiale de la Santé, 2008).

En outre, nous rappelle Goulet, il existe également des auteurs qui ont publié articles et/ou livres portant sur des aspects partiels de la qualité. Notons entre autres les documents de l'ICIS (Institut canadien d'information sur la santé), de l'IMF (*International Monetary Fund*), de l'ICES (*Institute for Clinical Evaluative Sciences*) et de l'EPA (*United States Environmental Protection Agency*). Cette approche répond plus aux préoccupations d'ordre pratique qu'opérationnel.

Les institutions qui s'occupent de la gestion des statistiques dans différents pays se sont intéressées à la question de la qualité des données et ont développé des lignes directrices ou principes liés à la qualité des données notamment l'OCDE, le Canada, les Etats Unis – leaderisés par l'Agence des Etats Unis de Développement et les Nations Unies – notamment via l'Organisation Mondiale de la Santé.

Pour le Canada, la qualité des données produites est définie au moyen de six dimensions de la qualité à savoir la pertinence, l'exactitude, l'actualité, l'accessibilité, l'intelligibilité et la cohérence. Quant à l'OCDE, la qualité renvoie à sept dimensions à savoir la pertinence, l'exactitude, la crédibilité, l'actualité, l'accessibilité, l'interprétabilité, la cohérence et la rentabilité. Enfin, le Fonds mondial de lutte contre le sida, la tuberculose et le paludisme, le Bureau du coordonnateur de la lutte contre le sida dans le monde, l'Agence américaine de développement (USAID), l'Organisation Mondiale de la Santé et l'Organisation internationale MEASURE retiennent les critères ci-après: *l'exactitude ; la fiabilité ; la précision ; l'exhaustivité et l'opportunité.*

## 1.2. Méthodologie

Cet article s'appuie sur une approche essentiellement qualitative et se présente comme une étude de cas. En relevant la fréquence des critères de qualité des données repris dans la littérature existante à travers une recherche sur les moteurs de recherche « google » et « google scholar<sup>1</sup> », les écarts ont été relevés entre les

<sup>1</sup> Au mois de Février 2023

définitions proposées par les principaux acteurs dans la sphère de développement international et celles jugées « orthodoxes » partant de la quintessence de chaque éminent critère de qualité des données. Il s'appuie, en outre, sur des connaissances nourries de la pratique d'évaluation de la qualité des données de près de 20 ans en République Démocratique du Congo.

Les objectifs immédiats visés par cet article sont : i) Montrer les nuances existant dans l'appréhension de différentes dimensions de la qualité des données au sein des organisations bi et multilatérales qui assurent le leadership de l'aide au développement et ii) Proposer une approche alternative dans la mesure de la qualité des données.

### 1.3. Résultats

- Nos recherches sur les critères de la qualité des données confirment la prééminence des critères d'évaluation de la qualité des données déjà signalés par Laure en 2018. Il s'agit des critères ci-après : actualité, cohérence et exactitude. Ce résultat constitue une convergence d'envergure lorsqu'on considère la forêt des critères d'évaluation de la qualité des données qui s'imposent aujourd'hui aussi bien aux chercheurs qu'aux évaluateurs et autres auditeurs de la qualité des données ;
- S'agissant de la comparabilité de la qualité des données, force est de reconnaître que la substance de chaque dimension dépend plus de l'objet de l'évaluation de la qualité et de l'utilisation de l'information-retour. Il s'avère dès lors difficile, au niveau d'un pays, de comparer les progrès réalisés en termes d'amélioration de qualité des données entre secteurs et parfois même à l'intérieur du même secteur. Il sera toujours convenable d'éviter toute comparaison en matière de qualité des données d'un bailleur à un autre ou d'un secteur d'activité à un autre.
- Dans la pratique actuelle, les méthodologies d'évaluation de la qualité des données se réduisent aux méthodes qualitative et quantitative et à la méthode mixte.
- Au-delà de la mécanique de l'évaluation de la qualité des données, la convergence entre les critères d'évaluation de la qualité des données demeure encore utopique du fait que la démarche d'évaluation de la qualité des données dépend de plusieurs paramètres dont la nature des activités du projet générateur des données mais également l'absence de consensus sur la quintessence de chaque critère de qualité.

## 2. Discussion

S'agissant des méthodologies d'évaluation de la qualité des données, les approches diffèrent mais se réduisent finalement à trois : qualitative ; quantitative ou mixte. L'évaluation n'est possible, néanmoins, que lorsque les critères d'évaluation sont connus.

C'est ainsi que pour mieux cerner l'économie de la connaissance sur les critères d'évaluation de la qualité des données, nous proposons d'explorer quatre pistes jugées incontournables. Celles-ci offrent des clefs de lecture qui permettent de mieux résumer la littérature existante autour de la démarche d'évaluation de la qualité des données. Il s'agit de : (i) discussion sur la pertinence de chaque critère utilisé par les différentes institutions internationales; (ii) le « *benchmarking* » partant de l'identification des critères de qualité orthodoxes et ensuite la comparaison de ceux-ci aux critères utilisés par nombre d'organisations aussi bien bilatérales que multilatérales; (iii) la revue de la possibilité de convergence des critères au niveau interinstitutionnel ; et enfin (iv) la perception de la qualité des données des pays du sud et son impact sur la qualité des données des projets individuels de développement.

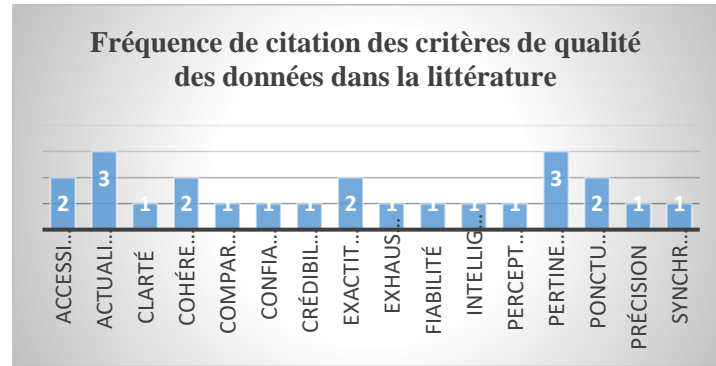
*Première piste de connaissance : discussion sur la pertinence de chaque critère utilisé par les différentes institutions internationales.*

Pour des raisons d'actualité et de cohérence, nous avons estimé utile de ne garder que les dimensions de la qualité des données publiées après 2012 c'est-à-dire après le travail de Goulet. Ci-après le tableau qui reprend les fréquences de différentes dimensions d'apparitions de qualité des données dans la littérature.

Dimensions	Martin Goulet, 2012	OCDE (2018)	CANADA (2019)	Fréquence
Accessibilité		X	X	2
Actualité	X	X	X	3
Clarté			X	1
Cohérence	X		X	2
Comparabilité			X	1
Confiance	X			1
Crédibilité		X		1
Exactitude		X	X	2
Exhaustivité	X			1
Fiabilité			X	1
Intelligibilité			X	1
Perception	X			1
Pertinence	X	X	X	3
Ponctualité	X		X	2
Précision	X			1
Synchronisation	X			1
Total				24

Tableau n° 1 : Fréquences des apparitions des dimensions de la qualité des données dans la littérature. Source : Auteur.

Voilà qui est résumé par le graphique ci-après :



Graphique n° 1 : Fréquences d'apparitions des dimensions de la qualité des données dans la littérature (Entre 2012 et 2020).

Il apparaît clairement que les dimensions les plus citées dans la littérature (deux fois et plus) sont notamment la *ponctualité (ou disponibilité)*, *exactitude*, *cohérence*, *accessibilité*, *pertinence* et *actualité (fraîcheur)*. Deux critères sont gratifiés de la fréquence la plus élevée : il s'agit de l'actualité (ou fraîcheur des données) et de la pertinence. Néanmoins, il sied de rappeler ici que le fait de bénéficier d'une fréquence élevée dans la littérature (et l'utilisation) ne signifie nullement que ces dimensions font consensus dans leur contenu parmi les différents utilisateurs. Pas du tout. En effet, une dimension peut avoir été citée pour son aspect opérationnel ou théorique. Par ailleurs, Goulet nous rappelle qu'il existe également de nombreux auteurs qui ont publié articles et/ou livres portant sur des aspects partiels de la qualité. Notons entre autres les documents de l'ICIS (Institut canadien d'information sur la santé), de l'IMF (*International Monetary Fund*), de l'ICES (*Institute for Clinical Evaluative Sciences*) et de l'EPA (*United States Environmental Protection Agency*)<sup>2</sup>.

Ce prisme d'approche répond plus aux préoccupations d'ordre pratique qu'opérationnel étant donné que la qualité de données attendue diffère selon qu'on est dans la manufacture ou le commerce ; l'environnement ou l'éducation.... Il faut également considérer ici les attentes des constituants (Institutions de redevabilité) qui varient d'un pays à un autre selon les inspirations démocratiques ainsi que les priorités assignées à l'aide de chaque pays bailleur. Il devient, dès lors difficile, sinon impossible de ressortir la convergence entre les critères théoriques même d'évaluation de la qualité des données. Ils ne peuvent pas être à sens unique.

***Deuxième piste de connaissance : le « benchmarking » partant de l'identification d'un contenu universel des critères de qualités.***

Pour mieux aborder cette piste de connaissance, il convient de partir de la définition même du concept qualité. Tout en soulignant l'absence de consensus parmi les différents experts autour du concept « qualité », nous n'avons trouvé meilleure définition que celle du comité international de normalisation ISO (*International Standard Organization*) citée par Rodolphe Devilliers. En effet, pour l'ISO la qualité est « l'adéquation aux exigences; la satisfaction des besoins de l'utilisateur » et la qualité d'un produit comme « la totalité des caractéristiques d'un produit ou service qui influent sur sa capacité à satisfaire les besoins explicites ou implicites du client » (ISO 8402, 1994)<sup>3</sup>. Etant donné que les exigences dont il est question

<sup>2</sup> Martin Goulet: idem, p. 16.

<sup>3</sup> Devilliers Rodolphe : ***Conception d'un système multidimensionnel d'information sur la qualité des données géospatiales*** ; HAL Id: tel-00008930 <https://theses.hal.science/tel-00008930> ; Submitted on 1 Apr 2005 ; p. 25

dans cette définition peuvent être multiples, il serait donc superficiel de ne se focaliser que sur un seul critère de qualité.

Evoquant les critères de qualité des données, Delphine Barrau et ses co-auteurs attirent l'attention sur ce qui suit : n'importe qui peut publier les données ; les données reflètent la vision du fournisseur ; les données sont liées ; l'utilisation faite des données n'est pas connue a priori ; les données sont ouvertes à tous ; l'accès aux données se fait via le web et les données ouvertes peuvent représenter de gros volumes à traiter<sup>4</sup>. De ces considérations, il se dégage les dimensions de qualité des données reprises dans le tableau ci-après :

<b>Dimension</b>	<b>Définition</b>
Exactitude	<ul style="list-style-type: none"> <li>• L'exactitude syntaxique définit à quel point les données sont conformes à une règle de format (par exemple, une donnée est-elle bien un numéro de téléphone ?).</li> <li>• L'exactitude sémantique définit à quel point les données sont conformes à la réalité décrite (par exemple, le numéro de téléphone est-il bien celui de l'entité décrite ?).</li> </ul>
Vérifiabilité	<ul style="list-style-type: none"> <li>• Définit à quel point les données peuvent être contrôlées, certifiées ou garanties par contrat.</li> </ul>
Traçabilité	<ul style="list-style-type: none"> <li>• Définit à quel point les données portent l'information de leur provenance (par exemple source d'origine, processus de transformation subi avant publication, identification du producteur).</li> </ul>
Confiance	<ul style="list-style-type: none"> <li>• Définit à quel point les données et leur producteur sont fiables.</li> </ul>
Pertinence	<ul style="list-style-type: none"> <li>• Définit à quel point les données apportent une valeur ajoutée dans leur utilisation.</li> </ul>
Utilisabilité	<ul style="list-style-type: none"> <li>• <i>Accessibilité</i>: Définit à quel point les données sont disponibles, récupérables.</li> <li>• <i>Compréhensibilité</i>: Définit à quel point les données sont compréhensibles, incluant par exemple l'éventuelle présence d'un support et d'une documentation, ou mesurant la versatilité des données.</li> <li>• <i>Licensing</i>: Présence ou non d'une licence indiquant quelle réutilisation peut être faite des données.</li> </ul>
Visibilité	<ul style="list-style-type: none"> <li>• Définit à quel point les données sont localisables par les utilisateurs.</li> </ul>
Fraîcheur	<ul style="list-style-type: none"> <li>• Définit à quel point les données sont suffisamment récentes.</li> </ul>
Complétude	<ul style="list-style-type: none"> <li>• Définit le niveau de couverture avec lequel le phénomène observé est représenté dans l'assemblage des données. Se décline en complétude de la population, et des informations présentes au niveau des individus.</li> </ul>
Cohérence	<ul style="list-style-type: none"> <li>• Définit à quel point les données satisfont un ensemble de contraintes syntaxiques et sémantiques.</li> </ul>
Unicité	<ul style="list-style-type: none"> <li>• Définit à quel point les données évitent les redondances (des données sont redondantes si elles décrivent un même objet du monde réel).</li> </ul>
Consistance	<ul style="list-style-type: none"> <li>• Définit à quel point les données évitent les contradictions et incohérences, peut concerner la détection de données redondantes (décrivant un même objet du monde réel) mais contradictoires ou la vérification de règles de cohérence métier.</li> </ul>
Sécurité de l'accès	<ul style="list-style-type: none"> <li>• Définit à quel point l'accès aux données est contrôlé (peut être vue comme une sous-dimension de celle d'accessibilité).</li> </ul>
Confidentialité	<ul style="list-style-type: none"> <li>• Définit à quel point la confidentialité des informations personnelles est préservée.</li> </ul>
Interconnexion	<ul style="list-style-type: none"> <li>• Définit à quel point les données sont riches et précises en termes de lien vers des sources externes complémentaires.</li> </ul>

Tableau n°2 : Dimensions de qualité des données dites ouvertes. Source : Delphine Barrau.

Du tableau ci-dessus, il convient de relever quelques redondances partant juste du contenu de chaque dimension de qualité ci-dessus repris. En effet, partant de la définition de chaque critère ci-dessus reprise, il appert que la vérifiabilité intègre l'idée de la traçabilité et vice-versa. Aussi, la sécurité d'accès comprend-elle l'utilisabilité des données. Celle-ci rejoint d'une certaine manière le concept de visibilité. Enfin, les

<sup>4</sup> Barrau Delphine et al: *Gestion de la qualité des données ouvertes liées – État des lieux et perspectives*; HAL Id: hal-01333623 <https://hal.inria.fr/hal-01333623> submitted on 7 Jul 2016, pp 8 – 9

concepts de *cohérence* et *consistance* s'interpénètrent également alors que la *pertinence* intègre en elle-même l'idée d'*unicité*.

Ainsi donc, en procédant à l'élimination progressive sur la base de la substance de chaque dimension, on peut retenir en définitive les dimensions les plus significatives ci-après : *l'exactitude ; la confiance ; la fraîcheur ; la pertinence ; la complétude ; la cohérence et la confidentialité*.

Par ailleurs, en 2018 – bien après Delphine Barrau et ses co-auteurs, Laure Berti-Equille présente également un article qui synthétise les différentes dimensions de qualité des données<sup>5</sup>. Si les dimensions reprises par Delphine Barrau et al s'appliquent uniquement sur les données ouvertes et liées, celles de Laure B. présentent le bénéfice d'être d'une application à plus large spectre. Elles s'appliquent à la fois aux données ouvertes et non ouvertes. En effet, les concepts comme *validité* par rapport à des données de référence ; *maintenance des données* ou *qualités scientifiques* entre autres renvoient à l'existence d'une entité plus ou moins permanente qui collecte les données et les gère à travers le temps. Il est même sous-entendu l'existence d'une base des données. La catégorie des données non ouvertes, par contre, correspond le mieux à la nature des données d'une entreprise ou mieux d'un projet de développement socioéconomique qui se trouve être l'agent principal de collecte des données. Cette entité analyse ses données et les stocke à travers le temps tout en les reportant avec un narratif approprié auprès d'une entité bailleresse et ce en guise de redevabilité.

Auteurs et références	Dimensions de la qualité des données	
Brodie	6 concepts	<ul style="list-style-type: none"> <li>• Intégrité</li> <li>• Niveau d'abstraction du modèle conceptuel</li> <li>• Expressivité sémantique</li> <li>• Validité par rapport à des données de référence</li> <li>• Maintenance des données</li> <li>• Efficacité dans l'utilisation des ressources</li> </ul>
Delen et Rijssenbrij	4 dimensions 21 aspects 40 attributs	<ul style="list-style-type: none"> <li>• Développement et contrôle du S.I.</li> <li>• Propriétés statiques de maintenance</li> <li>• Fonctionnement dynamique</li> <li>• Importance de l'information : donnée correcte, complète, mise à jour, précise, vérifiable</li> </ul>
Wang et al. Programme TDQM	4 catégories 179 attributs	<ul style="list-style-type: none"> <li>• Qualité intrinsèque</li> <li>• Qualité d'accessibilité</li> <li>• Qualité contextuelle</li> <li>• Qualité de la représentation</li> </ul>
Redman	<ul style="list-style-type: none"> <li>• 4 dimensions pour les valeurs ;</li> <li>• 8 dimensions pour le format de représentation</li> </ul>	<ul style="list-style-type: none"> <li>• Précision, complétude, actualité, cohérence</li> <li>• Donnée appropriée, interprétable, portable, précision du format, flexibilité du format, possibilité de représenter les valeurs nulles, utilisation efficace, cohérence</li> </ul>
Calabretto, Pinon, Pouillet, Richez	3 critères pour la qualité de l'information	<ul style="list-style-type: none"> <li>• Disponibilité, Fiabilité, Adaptabilité</li> </ul>

<sup>5</sup> Berti-Equille Laure: *La qualité des données comme condition à la qualité des connaissances: un état de l'art*; HAL Id: hal-01856198 <https://hal.inria.fr/hal-01856198> ; submitted on 10 Aug 2018; p. 6.



Aebi, Perrochon	3 composantes	<ul style="list-style-type: none"> <li>• Donnée correcte,</li> <li>• Complète,</li> <li>• Minimale</li> </ul>
DISA	6 caractéristiques	<ul style="list-style-type: none"> <li>• Précision</li> <li>• Complétude</li> <li>• Consistance</li> <li>• Actualité</li> <li>• Unicité</li> <li>• Validité</li> </ul>
TIPS-QCT	8 descripteurs pour la qualité des documents	<ul style="list-style-type: none"> <li>• Qualité scientifique: exactitude, précision, originalité, complétude, actualité, qualité de démonstration, qualité de la liste des références, qualité de la méthodologie;</li> <li>• Lisibilité: qualité du style d'écriture, qualité de la structure logique, adéquation des illustrations, absence des répétitions, clarté de l'expression des idées;</li> <li>• Public visé: niveau technique;</li> <li>• Fraîcheur: date de publication,</li> <li>• Autorité: réputation de l'auteur, réputation du journal ou de la conférence;</li> <li>• Disponibilité: longévité, imprimabilité;</li> <li>• Popularité: nombre des lecteurs, des citations;</li> <li>• Qualité d'identification: citabilité.</li> </ul>

Tableau n°3 : Synthèse des dimensions de qualité des données

Source : Laure Berti-Equille.

Malgré l'absence d'un consensus autour des dimensions de qualité des données, fort est de constater qu'un noyau des dimensions semble se répéter dans le tableau n°3. Il s'agit des dimensions ci-après : *la validité, l'intégrité, l'exactitude ; la confiance ; la fraîcheur (ou actualité) ; la pertinence ; la complétude ; la cohérence ; la disponibilité et la confidentialité.*

De ce noyau des dimensions, l'on peut facilement ressortir deux catégories : les dimensions intimement liées aux données (Comme la validité, l'exactitude, la confiance, la pertinence ; la complétude ; la cohérence et la disponibilité) et celles qui leur sont externes (l'intégrité ou la disponibilité).

Par ailleurs, en se plaçant dans la logique de l'évaluation de la qualité des données, Laure rappelle pertinemment que « quel que soit le domaine d'application, les mesures de qualité de données les plus fréquemment mentionnées dans la littérature, sont *l'exactitude, la complétude, l'actualité, la cohérence* qu'elle définit comme suit<sup>6</sup> :

- **L'exactitude** se mesure en détectant le taux de valeurs correctes dans la base des données,
- **La complétude** se mesure en détectant le taux de valeurs manquantes dans la base de données,
- **L'actualité** se mesure en détectant le taux de valeurs obsolètes dans la base de données par rapport à une date prédéfinie (la fraîcheur se mesure, quant à elle, par une comparaison de la date de saisie à la date courante),

<sup>6</sup> Berti-Equille Laure, idem, p. 10

- **La cohérence** se mesure par rapport à un ensemble de contraintes en détectant les données de la base qui ne les satisfont pas.

Il est fort opportun de rappeler ici que les conclusions de Laure confortent les nôtres qui sont reprises dans le premier graphique de ce rapport. Pour rappel, nous avons identifié les critères de qualité les plus fréquents ci-après dans la littérature : ponctualité (ou disponibilité), exactitude, cohérence, accessibilité, pertinence et actualité (fraîcheur). Il y a, en effet, coïncidence sur les critères ci-après : actualité, cohérence et exactitude.

En conclusion, il apparaît clairement que pour procéder au « *benchmarking* » il conviendra de retenir le noyau des critères que Laure et nous-même considérons comme les plus utilisées dans la littérature – dimensions toutes intrinsèques aux données ! Néanmoins, à ce noyau il serait de bon aloi d’ajouter les dimensions extrinsèques vis-à-vis des données : intégrité et disponibilité. Ainsi la liste des critères de qualité des données devient : *l’exactitude ; la complétude ; l’actualité ; la cohérence ; l’intégrité et la disponibilité*. Ce noyau des dimensions sont susceptibles de faciliter le « benchmarking » des critères des données par rapport aux pratiques de grandes organisations bailleuses.

Il est clair une fois de plus ici que les critères de qualité sont loin de converger. Il faut plutôt forcer la note en s’engouffrant dans les constats de la littérature existante.

### ***Troisième piste de connaissance : la revue de la possibilité de convergence dans les critères d’évaluation de la qualité des données au niveau interinstitutionnel***

Nonobstant les difficultés ci-haut relevées en matière de qualité des données ; les institutions qui s’occupent de la gestion des statistiques dans différents pays se sont intéressées à la question de la qualité des données et ont développé des lignes directrices ou principes liés à la qualité des données. Revenons ci-dessous sur les lignes directrices de l’OCDE, du Canada, des Etats Unis – notamment via l’Agence Américaine de Développement et des Nations Unies – à travers l’Organisation Mondiale de la Santé.

Pour le Canada<sup>7</sup>, la qualité des données produites est définie au moyen de six dimensions de la qualité : la pertinence, l’exactitude, l’actualité, l’accessibilité, l’intelligibilité et la cohérence. Ces six dimensions sont abordées dans le Cadre d’Assurance de la Qualité de façon à décrire ce que les producteurs de statistiques officielles doivent faire pour assurer la qualité à l’échelle de leur institution. Ces dimensions sont définies comme suit:

- *Pertinence* : la mesure dans laquelle les besoins réels des utilisateurs sont satisfaits. Pour être qualifiées de pertinentes, les données doivent éclairer les utilisateurs sur les questions les plus importantes à leurs yeux.
- *Actualité* (ou caractère actuel des données) : le délai entre le point de référence (ou la fin de la période de référence) auquel se rapporte l’information et la date à laquelle les données sont disponibles. La ponctualité, quant à elle, se définit comme la différence entre le moment planifié pour sa disponibilité et sa disponibilité réelle.
- *Exactitude* : Par « exactitude des données statistiques », on entend la mesure dans laquelle l’information décrit correctement le phénomène qu’elle doit mesurer. L’exactitude est habituellement caractérisée en fonction de l’erreur dans les estimations statistiques et elle est traditionnellement décomposée en biais (erreur systématique) et en variance (erreur aléatoire). On peut aussi la définir par rapport aux sources principales d’erreurs susceptibles de mener à des données imprécises (p. ex. la couverture, l’échantillonnage, la

<sup>7</sup> Statistique Canada : Lignes directrices concernant la qualité Sixième édition; <https://www150.statcan.gc.ca>

non-réponse). Liée à la notion d'exactitude, la fiabilité correspond à la mesure dans laquelle l'information statistique décrit correctement, et uniformément dans le temps, le phénomène qu'elle doit servir à mesurer.

- *Accessibilité* : par « accessibilité et clarté des données statistiques », on entend la facilité avec laquelle les utilisateurs peuvent connaître l'existence de l'information (incluant les métadonnées), la localiser, la visualiser et l'importer dans leur propre environnement de travail. Il s'agit également du caractère approprié de la présentation de l'information ou du média au moyen duquel on peut avoir accès aux données. Pour certains utilisateurs, le coût peut également être un aspect de l'accessibilité.
- *Intelligibilité* : la disponibilité de renseignements supplémentaires et de métadonnées nécessaires à l'interprétation et à l'utilisation appropriée des données. Il s'agit en général de renseignements sur les variables, les classifications et les concepts sous-jacents utilisés, sur les méthodes de collecte et de traitement des données et sur les indicateurs de la qualité des données statistiques.
- *Cohérence* : par « cohérence et comparabilité des données statistiques », on entend la mesure dans laquelle celles-ci peuvent être jumelées et comparées à d'autres renseignements statistiques dans un vaste cadre analytique au fil du temps. L'utilisation de concepts, de classifications et de populations cibles types favorise la cohérence, tout comme l'utilisation de méthodes d'enquêtes communes. Par « cohérence », on n'entend pas nécessairement la concordance numérique parfaite.

Quant à l'OCDE<sup>8</sup>, elle considère la qualité des données comme fondamentale pour la crédibilité des statistiques produites par l'OCDE en général et par sa Direction de l'Éducation et des Compétences en particulier. La qualité se définit comme « l'adaptation aux besoins des utilisateurs ». Cette définition est plus large que celle proposée par le passé, lorsque la qualité était synonyme d'exactitude. Il est désormais communément admis que d'autres dimensions sont importantes.

L'OCDE envisage la qualité selon sept dimensions : la pertinence, l'exactitude, la crédibilité, l'actualité, l'accessibilité, l'interprétabilité et la cohérence. Enfin, la rentabilité est un facteur important, même si elle n'est pas à strictement parler une dimension de la qualité. La notion de rentabilité doit intervenir dans l'application de chacune de ces sept dimensions. Celles-ci se résument comme suit :

- *La pertinence* s'évalue à l'aune de l'activité des groupes d'utilisateurs et de leurs besoins, qui doivent donc être identifiés.
- *L'exactitude* est la mesure dans laquelle les données évaluent ou décrivent correctement les quantités ou les caractéristiques qu'elles sont censées mesurer.
- *La crédibilité* se rapporte à la confiance que les utilisateurs accordent aux produits de données simplement à partir de l'idée qu'ils se font de celui qui les élabore, autrement dit l'image de marque. La crédibilité dépend en partie de l'intégrité du processus de production. Le deuxième des Principes fondamentaux de la statistique officielle des Nations Unies (UNESCO, 1994) précise : « Pour que se maintienne la confiance dans l'information statistique officielle, les organismes responsables de la statistique doivent déterminer, en fonction de considérations purement professionnelles, notamment de principes scientifiques et de règles déontologiques, les méthodes et les procédures de collecte, de traitement, de stockage et de présentation des données statistiques. »
- *L'actualité* reflète le délai écoulé entre la mise à disposition des données et le moment auquel s'est produit la manifestation ou le phénomène décrit. La notion d'actualité est évaluée au regard de la durée pendant laquelle l'information présente un intérêt et permet de donner une suite.
- *L'accessibilité* reflète la facilité avec laquelle il est possible de repérer les produits de données au sein des différents services de l'OCDE et d'y accéder.

---

<sup>8</sup> OCDE: *Guide de l'OCDE pour l'établissement de statistiques internationalement comparables dans le domaine de l'éducation*, idem.

- L'interprétabilité reflète la facilité avec laquelle les utilisateurs peuvent comprendre, utiliser et analyser correctement les données. L'adéquation de la définition des concepts, des populations cibles, des variables et de la terminologie qui sous-tendent les données et la description des éventuelles limites des données déterminent largement le degré d'interprétabilité.
- La cohérence reflète les interrelations logiques entre les données et leur cohérence mutuelle. - La cohérence interne des ensembles de données implique que les éléments de données se basent sur des concepts, définitions et classifications compatibles qui peuvent être utilement combinés. Ainsi, il y a des incohérences dans un ensemble de données si deux parties de la même équation, par exemple les entrées et les sorties, ne concordent pas.
  - La cohérence entre les ensembles de données implique que les données se basent sur des concepts, définitions et classifications communs et que toute différence soit expliquée et puisse se justifier.
  - La cohérence dans le temps implique que les données se basent sur des concepts, définitions et méthodes stables dans le temps et que toute différence soit expliquée et puisse se justifier. L'incohérence dans le temps renvoie à des ruptures dans les séries de données à cause de changements apportés aux concepts, définitions ou méthodes.
  - La cohérence internationale implique que les données se basent sur des concepts, définitions, classifications et méthodes communs et que toute différence soit expliquée et puisse se justifier.
- La rentabilité est le rapport entre les coûts et la charge de travail des fournisseurs de données et le résultat. La charge de travail est supportée par les fournisseurs de données, mais elle a un coût. L'OCDE ne considère pas la rentabilité comme une dimension de la qualité, mais plutôt comme un facteur qui doit être pris en considération dans toute analyse de la qualité, car il peut affecter la qualité dans toutes ses dimensions.

Par ailleurs, dans l'effort de trouver un consensus sur les dimensions de la qualité des données le Fonds mondial de lutte contre le sida, la tuberculose et le paludisme, le Bureau de coordination de la lutte contre le sida dans le monde, l'Agence américaine de développement (USAID), l'Organisation Mondiale de la Santé et l'Organisation internationale MEASURE se sont ligüés pour produire les lignes directives d'audit pour évaluer l'aptitude d'un programme/projet à faire état de la qualité des données<sup>9</sup>. Il en est résulté les dimensions ci-après :

- *Exactitude* : Aussi appelée validité. Les données exactes sont considérées comme étant correctes. Les données mesurent ce qu'elles doivent mesurer. Des données exactes minimisent les erreurs (par exemple, parti pris de l'enregistrement ou de la personne qui conduit l'interview, erreur de transcription, erreur d'échantillonnage) au point de les rendre négligeables.
- *Fiabilité* : les données générées par le système d'information d'un programme sont basées sur des protocoles et procédures qui ne changent pas en fonction de la personne qui les utilise, du moment et de la fréquence de leur utilisation. Les données sont fiables parce qu'elles sont mesurées et collectées de manière cohérente.
- *Précision* : les données sont assez détaillées. Par exemple, un indicateur requiert la connaissance du nombre d'individus qui ont reçu de l'assistance et des analyses du VIH et reçu les résultats de leurs tests, selon le sexe de la personne. Un système d'information manque de précision s'il n'a pas été conçu pour enregistrer le sexe de l'individu qui a reçu l'assistance et les analyses.
- *Exhaustivité* : signifie qu'un système d'information duquel on tire les résultats est inclusif de manière appropriée : Il représente la liste exhaustive des personnes ou unités éligibles et pas juste une fraction de la liste.

---

<sup>9</sup> Fonds mondial de lutte contre le sida, la tuberculose et le paludisme, Bureau du coordonnateur de la lutte contre le sida dans le monde, USAID, OMS et MEASURE: Outil d'évaluation de la qualité des données; Mai 2008

- *Opportunité* : des données sont dites opportunes quand elles sont à jour (actuelles), et quand l'information est disponible à temps. L'opportunité est affectée par : (i) le rythme auquel le système d'information du programme est mis à jour ; (ii) le rythme de changement des activités réelles du programme ; et (iii) quand l'information est réellement utilisée ou requise.
- *Intégrité* : les données sont intègres quand le système utilisé pour les générer est protégé de tout parti pris ou manipulation délibérés pour des raisons politiques ou personnelles.
- *Confidentialité* : signifie que les clients sont assurés que leurs données seront conservées en conformité avec les normes nationales et/ou internationales en matière de données. Cela signifie que les données personnelles ne sont pas divulguées et que les données contenues sur des supports papier et électroniques sont traitées avec un niveau de sécurité approprié (par exemple, gardés dans des armoires fermées et des fichiers protégés par des mots de passe).

Abondant dans le même sens, l'OMS reprend sept dimensions à savoir l'exactitude, la fiabilité, l'exhaustivité, le caractère opportun, la précision, l'intégrité et la confidentialité. Elle définit ces dimensions comme suit :

- *Exactitude* : des données exactes contiennent des erreurs et une partialité minimale. L'exactitude est également appelée validité. Par exemple, l'exactitude peut être compromise à cause d'erreurs de transcription qui peuvent se produire si les données sont mal saisies dans le système. Il s'agit généralement d'erreurs accidentelles qui peuvent survenir quand on enregistre des informations de façon inexacte ou quand on saisit incorrectement des informations dans une base de données informatique. L'exactitude peut également être affectée par des données qui ne sont pas exhaustives, opportunes ou précises. Elle peut aussi être directement altérée par une manipulation effectuée pour d'autres raisons.
- *Fiabilité* : Les données sont fiables lorsqu'elles sont mesurées et collectées systématiquement et dans la durée. La fiabilité des données dépend de la disponibilité d'un système d'information doté de protocoles et de procédures cohérents. Pour être fiables, les données exigent des consignes normalisées et écrites sur leur collecte. Les procédures de collecte de données d'un programme ne doivent pas changer selon qui les utilise, quel site les utilise, quand elles sont utilisées ou à quelle fréquence. En outre, les procédures visant à rectifier les erreurs de données ou à résoudre des problèmes de données manquantes ou non exhaustives doivent être cohérentes entre les différents sites et périodes.
- *Exhaustivité* : par exhaustivité, on entend qu'un système d'information capture tous les individus, services, sites ou autres unités éligibles qu'il est censé mesurer. Les données qui en découlent devraient représenter la liste exhaustive des personnes, services, sites et autres unités et non simplement partie de la liste. L'exhaustivité est influencée par : la mesure dans laquelle les documents source comportent toutes les informations pertinentes et nécessaires pour la préparation des rapports ; la mesure dans laquelle tous les sites ont communiqué les informations à des échelons supérieurs pour qu'elles soient agrégées ; le caractère opportun de la transmission des rapports à des échelons supérieurs d'agrégation des données. Par exemple, les données d'un site du programme ne seront pas exhaustives si elles ne comprennent pas d'informations sur tous les clients desservis, tous les services fournis au client ou toutes les activités entreprises. Les données agrégées d'un programme ne seront pas exhaustives si les données de seulement 90 sites sur 100 sont rendues disponibles.
- *Précision* : les données sont suffisamment détaillées pour mesurer les indicateurs conformément à la définition de ces derniers. Par exemple, un indicateur souhaité pourrait exiger le nombre de personnes ayant effectué un dépistage du VIH, ventilé par sexe. Le système d'information manque de précision s'il n'est pas conçu pour enregistrer le sexe de la personne qui a reçu les services de dépistage. Lorsque les données sont plus détaillées, elles sont plus précises, ce qui influe de manière positive sur la qualité de représentation adéquate des activités du programme par les données. La précision contribue également à répondre aux questions qui importent aux directeurs de programme, aux directeurs régionaux ainsi qu'aux unités nationales

et internationales (le cas échéant). Ceci suppose que les formulaires de recueil de données soient conçus pour collecter des données précises et que le niveau approprié de détail soit communiqué aux échelons supérieurs.

- *Caractère opportun* : Les données sont opportunes lorsqu'elles sont transmises à l'échelon immédiatement supérieur à temps pour respecter les délais de soumission des rapports. « À temps » sous-entend que les données communiquées ont pu être employées dans le rapport de synthèse préparé par l'échelon de déclaration immédiatement supérieur. Par exemple, les déclarations des sites de service doivent être remises à l'échelon intermédiaire le 15 du mois pour le mois précédent, et le rapport de l'échelon intermédiaire est dû à l'unité S&E le 20 du mois, puis cette dernière prépare son rapport d'ici la fin du mois. Chacun de ces délais doit être respecté pour que les données soient opportunes.
- *Intégrité* : Les données sont intègres quand le système d'information est protégé contre tout parti pris ou manipulation délibérés à des fins politiques ou personnelles. Un examen indépendant des données peut contribuer à déterminer si l'intégrité des données a pu être compromise. Le fait de savoir que les données subiront un contrôle indépendant est susceptible de dissuader une manipulation des données.
- *Confidentialité* : les clients ont l'assurance que leurs données seront conservées conformément aux normes nationales et/ou internationales. Ainsi, les informations personnelles ne sont pas divulguées de façon inappropriée et les données papier et électroniques sont traitées à des niveaux de sécurité adéquats. Un autre aspect important consiste à former le personnel à respecter les informations confidentielles et ne pas les communiquer à d'autres clients. Ces mesures protègent la vie privée des clients desservis.

En croisant les définitions des critères les plus saillants avec celles des critères utilisés par les différentes institutions de développement internationales (voir annexe 1 de ce rapport), on peut résumer ci-dessous les constats réalisés.

N°	Critères les plus saillants dans la littérature sur base des définitions dites « orthodoxes » <sup>10</sup>	Convergence définitionnelle avec les Organisations bi-multilatérales	Observations
1	Exactitude	Etablie	Convergence entre exactitude sémantique et exactitude telle que définie par le Gouvernement Canadien ; l'OCDE ; l'USAID et l'OMS.
2	Complétude	Non établie	Aucune institution internationale n'assume la définition dite « orthodoxe »
3	Actualité	Non établie	NA
4	Cohérence	Etablie	Convergence établie avec les définitions du Gouvernement Canadien et l'OCDE.
5	Intégrité	Etablie	Convergence établie avec la définition de l'OMS.
6	Disponibilité	Non établie	NA

Tableau n° 4 : Convergence entre les définitions dites « orthodoxes » et celles des critères utilisés par les organisations bi-multilatérales.

Ci-dessous les conclusions qui peuvent être faites à ce stade :

1. Les critères de qualité des données se rapprochent dans leurs intitulés et dans leurs définitions même si les nuances opérationnelles persistent : exemple, pour l'exactitude certaines organisations insistent plus sur les erreurs plutôt que l'essence de la donnée (USAID et CANADA);
2. Les contenus des définitions de mêmes dimensions dans les différentes organisations ou institutions statistiques semblent s'éloigner selon que les enjeux sont de type opérationnel ou théorique.
3. La convergence semble établie clairement pour l'exactitude, la cohérence et l'intégrité. Ces critères sont jugés incontournables pour plusieurs organisations bailleuses.

<sup>10</sup> Nos propres constats soutenus par ceux de Laure.

4. La plus haute fréquence d'apparition de certaines dimensions de qualité des données souligne leur pertinence par rapport à d'autres. L'utilisation de ces critères devrait renvoyer à moins de subjectivisme dans leur manipulation durant les exercices d'évaluation de la qualité des données.
5. De nombreux critères d'évaluation de qualité des données demeurent subjectifs et donc limités à l'utilisation de quelques organisations internationales suivant leurs intérêts propres.

Enfin, tout porte à croire qu'il existe pour chaque critère de qualité une substance autour de laquelle se retrouvent les valeurs centrales alors que les subjectivités continuent à dominer les aspects opérationnels. C'est donc autour de ces substances centrales qu'il faudrait bâtir des standards universels de qualité des données.

***Quatrième piste de connaissance: la perception de la qualité des données des pays du sud et son impact sur la qualité des données des projets individuels de développement.***

Deux entités focalisent notre attention sur cette piste : il s'agit de l'entité qui gère au quotidien le projet de développement d'une part et l'Agence ou Service public de Statistiques d'autre part. Un flux des données en aller et retour marque la relation entre les deux entités. En effet, l'élaboration de tout projet individuel se fait sur la base des données existantes ou alors sur celles collectées pour le besoin de la cause. Dans ce dernier cas, les données sont collectées par l'Agence de mise en œuvre en vue d'établir la situation de référence du nouveau projet. Ceci concerne particulièrement les indicateurs de résultats et d'impact. Dans le premier cas, les données sont fournies par une Agence Nationale de Statistiques ou toute autre organisation officielle dont la mission intègre la collecte des données. Par ailleurs, pendant la vie du projet, celui-ci (ou alors l'Entité qui gère le projet) est tenu de partager ses résultats et impacts éventuellement réalisés et ce particulièrement à la clôture du projet.

Le problème qui se pose dans les pays du sud, est que l'Agence Nationale de Statistiques ne dispose souvent pas des statistiques à jour sur les différents secteurs de la vie nationale. Soit parce que cette agence ne dispose pas d'un budget adéquat soit parce qu'elle n'a tout simplement pas d'équipement technologique adéquat. Pour Florent Bédécarrats et al « la statistique publique africaine connaît depuis une vingtaine d'années une forme de contestation plus indirecte, mais tout aussi déstabilisante pour son fonctionnement, dans un contexte de raréfaction des ressources humaines et financières allouées par l'État depuis les années 1980 et l'instauration des programmes d'ajustement structurel<sup>11</sup>. » En conséquence, poursuit Florent B. et al, « la prolifération des bases de données internationales dans les domaines les plus divers (gouvernance, démocratie, qualité des institutions, planification familiale, accès à l'eau, aux soins ou encore aux services financiers, etc.), subventionnées et développées par toutes sortes d'institutions (universitaires, ONG internationales, agences d'aide, entreprises de consulting privées, instituts de sondages mondialisés, agences de notation, etc.) est venue remettre en question le monopole que la statistique publique africaine exerçait de fait dans son domaine. » Ce constat amer est corroboré par plusieurs autres auteurs notamment Morten Jerven<sup>12</sup> et Sandrine Mesple-Somps<sup>13</sup>. Cette dernière n'hésite pas à pointer du doigt deux aspects qu'elle estime majeurs concernant la production des données macroéconomiques: «(i) ***Pauvreté de moyens des***

---

<sup>11</sup> Bédécarrats Florent et al : Révolution des données et enjeux de la statistique en Afrique : Introduction thématique; in *Afrique contemporaine* 2016/2 (n° 258), pages 9 à 23 ;

<sup>12</sup> Morten Joven: African Union: *Poor numbers. How we are misled by African development statistics and what to do about it*, Ithaca and London, Cornell University Press, 2013, 187 p.

<sup>13</sup> Mesple-Somps Sandrine : "laviedesidees.fr" le 31 octobre 2013 ; <https://laviedesidees.fr/Mesple-Somps-Sandrine>

*instituts nationaux de la statistique* (INS). Les pays africains étant pauvres, les INS ont souvent peu de moyens humains et financiers pour produire correctement les statistiques nationales. (ii) **Dépendance de la production statistique au politique**. Cela a été le cas durant la période coloniale mais aussi après les indépendances, et cela perdure actuellement. »

Dans un tel contexte, il est fort probable que plusieurs projets financés par les bailleurs étrangers auront tendance à recourir aux données issues des bases données réputées fiables car administrées à partir des entités issues des pays du nord. Ce sont les mêmes projets qui alimenteront également les mêmes bases des données parallèles aux services nationaux de la statistique. Voilà qui ne manquera pas d'aggraver sinon de compromettre tout effort visant à apporter une solution au défi de la convergence des critères de qualité des données.

En outre, il convient de souligner la mise sur pied de l'Institut de la Statistique de l'Union Africaine (STATAFRIC). Ce dernier a pour tâche de communiquer sur la mise en œuvre du Programme Minimum Intégré (PMI) de l'Union Africaine. Ce programme est perçu comme le soubassement de l'intégration de l'économie africaine sous la forme de la communauté économique africaine. Néanmoins, le rapport de la Division Statistique du Département des Affaires Economiques est accablant. De prime abord, il reconnaît que « des contraintes de disponibilité, de qualité des données et de capacités pèsent sur les statistiques africaines<sup>14</sup>. » Plus loin le même rapport précise : « Presque tous les pays africains ont un système statistique faible: la plupart sont trop fragmentés ce qui a pour conséquence une déconsidération des statistiques dans la sphère publique (ou fonction publique). L'une des raisons immédiates de cette faiblesse est un système législatif inefficace qui dans la plupart des cas se focalise sur l'Institut National de la Statistique (INS) plutôt que sur l'ensemble du Système Statistique National (SSN). Lorsque la législation porte sur le SSN, elle donne à l'INS, qui est l'organe de coordination privilégié, l'autorité nécessaire pour superviser le système mais ne donne pas le pouvoir pour coordonner la production et la diffusion des données statistiques. Ceci aboutit à l'existence d'une autorité de coordination qui ne dispose d'aucun moyen de contrôle sur les autres institutions du SSN productrices de statistiques. Une autre raison, probablement moins présentée mais admise en raison du « politiquement correct », est l'inconfort et la gêne occasionnés lorsque les statistiques mettent en lumière une réalité que l'on préférerait occulter<sup>15</sup>. » Tout en étant projeté vers le futur, le cadre d'assurance qualité de l'Union Africaine projette d'insister sur les critères de qualité des données suivantes : **pertinence, pérennité, sources des données ; exactitude et fiabilité ; continuité ; cohérence et comparabilité ; ponctualité ; actualité ; spécificité et processus statistique**.

Enfin, au-delà de la perception, il se dégage des considérations ci-dessus que d'une part l'existence de plusieurs bases des données sur les mêmes problématiques de développement ne facilite pas la lisibilité du progrès économique et/ou social et d'autre part, l'utilisation des critères différents d'évaluation de la qualité des données complexifie la mesure de tout progrès. Au finish, les violons ne pourront jamais s'accorder sur les progrès réalisés ; ce qui compromettra toute capitalisation des connaissances sur les réalisations des projets de développement. Et sans capitalisation des connaissances tirées des projets mis en œuvre, le développement deviendra hypothéqué. Par ailleurs, l'on réalise ici l'influence mutuelle entre les données ouvertes et celles des projets de développement. Au départ la mauvaise perception des données provenant des Gouvernements africains (ce qui est une réalité du reste) pousse à l'émergence des bases des données

<sup>14</sup> Union Africaine: Division statistique, Département des Affaires Africaines, idem, p. 4.

<sup>15</sup> Union Africaine: Division statistique, Département des Affaires Africaines, idem, p. 7.



« privées ». Celles-ci sont utilisées comme base d'établissement des situations de référence des projets. Finalement, même tout progrès sera perçu comme relatif.

### 3. Conclusion

La convergence dans le contenu des critères d'évaluation de la qualité des données est finalement une réalité seulement pour quelques critères de proue notamment l'exactitude ; la cohérence et l'intégrité. Les autres critères cités çà et là au sein de la littérature ne sont pas encore convergents. Ces derniers ne sont pas, néanmoins, moins pertinents dans l'évaluation des données des programmes/projets particuliers. L'examen de la pertinence et de la possibilité de convergence des autres critères d'évaluation de la qualité des données demeure une question d'actualité. Il pourrait constituer une étape décisive vers l'établissement des critères d'évaluation de qualité des données définitivement universels. Il est dommage de constater que pour le moment le secteur de développement est encore loin d'un consensus sur ce qui est essentiel : les critères d'appréciation de la qualité des données. Or celles-ci détiennent une fonction royale qui est celle de faciliter la prise des décisions dans la gestion des projets/programmes mais également dans la planification du développement des peuples.

### REFERENCES

1. **Barrau Delphine et al**: "Gestion de la qualité des données ouvertes liées – État des lieux et perspectives"; HAL Id: hal-01333623 <https://hal.inria.fr/hal-01333623> submitted on 7 Jul 2016 ;
2. **Bédécarrats Florent et al** : « Révolution des données et enjeux de la statistique en Afrique : Introduction thématique » ; in Afrique contemporaine 2016/2 (n° 258);
3. **Berti-Equille Laure**: "La qualité des données comme condition à la qualité des connaissances: un état de l'art"; HAL Id: hal-01856198 <https://hal.inria.fr/hal-01856198> ; submitted on 10 Aug 2018;
4. **Devilliers Rodolphe** : « Conception d'un système multidimensionnel d'information sur la qualité des données géospatiales » ; HAL Id: tel-00008930 <https://theses.hal.science/tel-00008930> ; Submitted on 1 Apr 2005 ;
5. **Dordain D. et al.** : « *Le paradigme de l'aide projet fait de la résistance : de la dépendance à la réappropriation* », De Boeck Supérieur | « Afrique contemporaine », 2012/1 n° 241 | ;
6. **Easterly William**: *Reliving the 50s: The Big Push, Poverty Traps, and Takeoffs in Economic Development* ; NYU Development Research Institute Working Paper No. 15, January 2006 ;
7. **Fonds mondial de lutte contre le sida, la tuberculose et le paludisme, Bureau du coordonnateur de la lutte contre le sida dans le monde, USAID, OMS et MEASURE Evaluation** : *Data Quality Assessment Tool* (Volume 1). Guidelines for Implementation: *Audit Guidelines for Assessing a Program/Project's Ability to Report on Quality Data* (volume 1), 2007;
8. **Martin Goulet** : Hiérarchiser les dimensions de la qualité des données : analyse comparative entre la littérature et les praticiens en technologies de l'information ; Longueuil, Québec, Canada, février 2012;
9. **Mesple-Somps Sandrine** : "lavedesidees.fr" le 31 "octobre" 2013.
10. **Morten Joven** : *African Union : Poor numbers. How we are misled by African development statistics and what to do about it*, Ithaca and London, Cornell University Press, 2013 ;
11. **OCDE** : Guide de l'OCDE pour l'établissement de statistiques internationalement comparables dans le domaine de l'éducation, © OCDE, 2018 ; <https://www.oecd.org/fr/education/guide-de-l-ocde-pour-l-etablissement-de-statistiques-internationalement-comparables-dans-le-domaine-de-l-education-2018>
12. **Statistique Canada** : *Lignes directrices concernant la qualité* Sixième édition – Décembre 2019 ; <https://www150.statcan.gc.ca>
13. **Union Africaine** : Division Statistique, Département des Affaires Africaines : «Projet du Cadre d'Assurance de la Qualité des Statistiques pour le Système Statistique Africain » ; Mai 2015.

Annexe 1 : tableau comparatif des critères interinstitutionnels

N°	Critères jugés « orthodoxes »		Organisation/pays	Critères Spécifiques		Observations
	Critères	Définition		Critères	Définition	
	Exactitude	<ul style="list-style-type: none"> <li>L'exactitude syntaxique définit à quel point les données sont conformes à une règle de format (par exemple, une donnée est-elle bien un numéro de téléphone ?).</li> <li>L'exactitude sémantique définit à quel point les données sont conformes à la réalité décrite (par exemple, le numéro de téléphone est-il bien celui de l'entité décrite ?).</li> </ul>	STATISTIQUE CANADA	Pertinence	La mesure dans laquelle les besoins réels des utilisateurs sont satisfaits. Pour être qualifiées de pertinentes, les données doivent éclairer les utilisateurs sur les questions les plus importantes à leurs yeux.	
	Confiance	Définit à quel point les données et leur producteur sont fiables.		Actualité	Le délai entre le point de référence (ou la fin de la période de référence) auquel se rapporte l'information et la date à laquelle les données sont disponibles. La ponctualité, quant à elle, se définit comme la différence entre le moment planifié pour sa disponibilité et sa disponibilité réelle.	
	Fraîcheur	Définit à quel point les données sont suffisamment récentes.		Exactitude	Par « exactitude des données statistiques », on entend la mesure dans laquelle l'information décrit correctement le phénomène qu'elle doit mesurer. L'exactitude est habituellement caractérisée en fonction de l'erreur dans les estimations statistiques et elle est traditionnellement décomposée en biais (erreur systématique) et en variance (erreur aléatoire). On peut aussi la définir par rapport aux sources principales d'erreurs susceptibles de mener à des données imprécises (p. ex. la couverture, l'échantillonnage, la non-réponse). Liée à la notion d'exactitude, la fiabilité correspond à la mesure dans laquelle l'information statistique décrit correctement, et uniformément dans le temps, le phénomène qu'elle doit servir à mesurer.	

	Pertinence	Définit à quel point les données apportent une valeur ajoutée dans leur utilisation.		Accessibilité	Par « accessibilité et clarté des données statistiques », on entend la facilité avec laquelle les utilisateurs peuvent connaître l'existence de l'information (incluant les métadonnées), la localiser, la visualiser et l'importer dans leur propre environnement de travail. Il s'agit également du caractère approprié de la présentation de l'information ou du média au moyen duquel on peut avoir accès aux données. Pour certains utilisateurs, le coût peut également être un aspect de l'accessibilité.	
	Complétude	Définit le niveau de couverture avec lequel le phénomène observé est représenté dans l'assemblage des données. Se décline en complétude de la population, et des informations présentes au niveau des individus.		Intelligibilité	La disponibilité de renseignements supplémentaires et de métadonnées nécessaires à l'interprétation et à l'utilisation appropriée des données. Il s'agit en général de renseignements sur les variables, les classifications et les concepts sous-jacents utilisés, sur les méthodes de collecte et de traitement des données et sur les indicateurs de la qualité des données statistiques.	
	Cohérence	Définit à quel point les données satisfont un ensemble de contraintes syntaxiques et sémantiques.		Cohérence	Par « cohérence et comparabilité des données statistiques », on entend la mesure dans laquelle celles-ci peuvent être jumelées et comparées à d'autres renseignements statistiques dans un vaste cadre analytique au fil du temps. L'utilisation de concepts, de classifications et de populations cibles types favorise la cohérence, tout comme l'utilisation de méthodes d'enquêtes communes. Par « cohérence », on n'entend pas nécessairement la concordance numérique parfaite.	
	Confidentialité	Traduit dans quelle mesure les données sont protégées vis-à-vis de l'extérieur	OCDE	Pertinence	S'évalue à l'aune de l'activité des groupes d'utilisateurs et de leurs besoins, qui doivent donc être identifiés.	
				Exactitude	La mesure dans laquelle les données évaluent ou décrivent correctement les quantités ou les caractéristiques qu'elles sont censées mesurer.	
				Crédibilité	La confiance que les utilisateurs accordent aux produits de données simplement à partir de l'idée	

					qu'ils se font de celui qui les élabore, autrement dit l'image de marque. La crédibilité dépend en partie de l'intégrité du processus de production. Le deuxième des Principes fondamentaux de la statistique officielle des Nations Unies (UNESCO, 1994) précise : « Pour que se maintienne la confiance dans l'information statistique officielle, les organismes responsables de la statistique doivent déterminer, en fonction de considérations purement professionnelles, notamment de principes scientifiques et de règles déontologiques, les méthodes et les procédures de collecte, de traitement, de stockage et de présentation des données statistiques. »	
				Actualité	Reflète le délai écoulé entre la mise à disposition des données et le moment auquel s'est produit la manifestation ou le phénomène décrit. La notion d'actualité est évaluée au regard de la durée pendant laquelle l'information présente un intérêt et permet de donner une suite.	
				Accessibilité	Reflète la facilité avec laquelle il est possible de repérer les produits de données au sein des différents services de l'OCDE et d'y accéder.	
				Interprétabilité	Reflète la facilité avec laquelle les utilisateurs peuvent comprendre, utiliser et analyser correctement les données. L'adéquation de la définition des concepts, des populations cibles, des variables et de la terminologie qui sous-tendent les données et la description des éventuelles limites des données déterminent largement le degré d'interprétabilité.	
				Cohérence	Reflète les interrelations logiques entre les données et leur cohérence mutuelle. - La cohérence interne des ensembles de données implique que les éléments de données se basent sur des concepts, définitions et classifications compatibles qui peuvent être utilement combinés. Ainsi, il y a des incohérences dans un ensemble	

					de données si deux parties de la même équation, par exemple les entrées et les sorties, ne concordent pas.	
				Rentabilité	Rapport entre les coûts et la charge de travail des fournisseurs de données et le résultat. La charge de travail est supportée par les fournisseurs de données, mais elle a un coût. L'OCDE ne considère pas la rentabilité comme une dimension de la qualité, mais plutôt comme un facteur qui doit être pris en considération dans toute analyse de la qualité, car il peut affecter la qualité dans toutes ses dimensions.	
			USAID/OMS/ME ASURE	Exactitude	Aussi appelée validité. Les données exactes sont considérées comme étant correctes. Les données mesurent ce qu'elles doivent mesurer. Des données exactes minimisent les erreurs (par exemple, parti pris de l'enregistrement ou de la personne qui conduit l'interview, erreur de transcription, erreur d'échantillonnage) au point de les rendre négligeables.	
				Fiabilité	Les données générées par le système d'information d'un programme sont basées sur des protocoles et procédures qui ne changent pas en fonction de la personne qui les utilise, du moment et de la fréquence de leur utilisation. Les données sont fiables parce qu'elles sont mesurées et collectées de manière cohérente.	
				Précision	Les données sont assez détaillées. Par exemple, un indicateur requiert la connaissance du nombre d'individus qui ont reçu de l'assistance et des analyses du VIH et reçu les résultats de leurs tests, selon le sexe de la personne. Un système d'information manque de précision s'il n'a pas été conçu pour enregistrer le sexe de l'individu qui a reçu l'assistance et les analyses.	
				Exhaustivité	Signifie qu'un système d'information duquel on tire les résultats est inclusif de manière appropriée	

					: Il représente la liste exhaustive des personnes ou unités éligibles et pas juste une fraction de la liste.	
				Intégrité	Les données sont intègres quand le système utilisé pour les générer est protégé de tout parti pris ou manipulation délibérés pour des raisons politiques ou personnelles.	
				Confidentialité	Signifie que les clients sont assurés que leurs données seront conservées en conformité avec les normes nationales et/ou internationales en matière de données. Cela signifie que les données personnelles ne sont pas divulguées et que les données contenues sur des supports papier et électroniques sont traitées avec un niveau de sécurité approprié (par exemple, gardés dans des armoires fermées et des fichiers protégés par des mots de passe).	
			OMS	Exactitude	Des données exactes contiennent des erreurs et une partialité minimale. L'exactitude est également appelée validité. Par exemple, l'exactitude peut être compromise à cause d'erreurs de transcription qui peuvent se produire si les données sont mal saisies dans le système. Il s'agit généralement d'erreurs accidentelles qui peuvent survenir quand on enregistre des informations de façon inexacte ou quand on saisit incorrectement des informations dans une base de données informatique. L'exactitude peut également être affectée par des données qui ne sont pas exhaustives, opportunes ou précises. Elle peut aussi être directement altérée par une manipulation effectuée pour d'autres raisons.	
				Fiabilité	Les données sont fiables lorsqu'elles sont mesurées et collectées systématiquement et dans la durée. La fiabilité des données dépend de la disponibilité d'un système d'information doté de protocoles et de procédures cohérents. Pour être fiables, les données exigent des consignes normalisées et écrites sur leur collecte. Les	

					procédures de collecte de données d'un programme ne doivent pas changer selon qui les utilise, quel site les utilise, quand elles sont utilisées ou à quelle fréquence. En outre, les procédures visant à rectifier les erreurs de données ou à résoudre des problèmes de données manquantes ou non exhaustives doivent être cohérentes entre les différents sites et périodes.	
				Exhaustivité	Par exhaustivité, on entend qu'un système d'information capture tous les individus, services, sites ou autres unités éligibles qu'il est censé mesurer. Les données qui en découlent devraient représenter la liste exhaustive des personnes, services, sites et autres unités et non simplement partie de la liste. L'exhaustivité est influencée par : la mesure dans laquelle les documents source comportent toutes les informations pertinentes et nécessaires pour la préparation des rapports ; la mesure dans laquelle tous les sites ont communiqué les informations à des échelons supérieurs pour qu'elles soient agrégées ; le caractère opportun de la transmission des rapports à des échelons supérieurs d'agrégation des données. Par exemple, les données d'un site du programme ne seront pas exhaustives si elles ne comprennent pas d'informations sur tous les clients desservis, tous les services fournis au client ou toutes les activités entreprises. Les données agrégées d'un programme ne seront pas exhaustives si les données de seulement 90 sites sur 100 sont rendues disponibles.	
				Précision	Les données sont suffisamment détaillées pour mesurer les indicateurs conformément à la définition de ces derniers. Par exemple, un indicateur souhaité pourrait exiger le nombre de personnes ayant effectué un dépistage du VIH, ventilé par sexe. Le système d'information manque de précision s'il n'est pas conçu pour enregistrer le sexe de la personne qui a reçu les	

					services de dépistage. Lorsque les données sont plus détaillées, elles sont plus précises, ce qui influe de manière positive sur la qualité de représentation adéquate des activités du programme par les données. La précision contribue également à répondre aux questions qui importent aux directeurs de programme, aux directeurs régionaux ainsi qu'aux unités nationales et internationales (le cas échéant). Ceci suppose que les formulaires de recueil de données soient conçus pour collecter des données précises et que le niveau approprié de détail soit communiqué aux échelons supérieurs.	
				Caractère opportun	Les données sont opportunes lorsqu'elles sont transmises à l'échelon immédiatement supérieur à temps pour respecter les délais de soumission des rapports. « À temps » sous-entend que les données communiquées ont pu être employées dans le rapport de synthèse préparé par l'échelon de déclaration immédiatement supérieur. Par exemple, les déclarations des sites de service doivent être remises à l'échelon intermédiaire le 15 du mois pour le mois précédent, et le rapport de l'échelon intermédiaire est dû à l'unité S&E le 20 du mois, puis cette dernière prépare son rapport d'ici la fin du mois. Chacun de ces délais doit être respecté pour que les données soient opportunes.	
				Intégrité	Les données sont intègres quand le système d'information est protégé contre tout parti pris ou manipulation délibérés à des fins politiques ou personnelles. Un examen indépendant des données peut contribuer à déterminer si l'intégrité des données a pu être compromise. Le fait de savoir que les données subiront un contrôle indépendant est susceptible de dissuader une manipulation des données.	



				Confidentialité	les clients ont l'assurance que leurs données seront conservées conformément aux normes nationales et/ou internationales. Ainsi, les informations personnelles ne sont pas divulguées de façon inappropriée et les données papier et électroniques sont traitées à des niveaux de sécurité adéquats. Un autre aspect important consiste à former le personnel à respecter les informations confidentielles et ne pas les communiquer à d'autres clients. Ces mesures protègent la vie privée des clients desservis.	
--	--	--	--	-----------------	---	--