



Observatoire économique et statistique d'Afrique Subsaharienne

DE LA NECESSITE DE DYNAMISER LE TRAITEMENT DES DONNEES DES SOURCES ADMINISTRATIVES : ELEMENTS DE REFLEXION.

Siriki Zanga Coulibaly[♦]

Juillet 2011

Résumé exécutif

Le suivi et évaluation des documents de stratégie de réduction de la pauvreté et des Objectifs du Millénaire pour le Développement continue à alimenter l'historique débat sur la dualité « données de sources administratives versus données d'enquêtes ». Au seuil de l'échéance de 2015, la réconciliation des données de ces deux sources d'informations reste toujours aussi problématique. Le présent article, inspiré par les débats actuellement en cours sur la réforme des services publics amorcée dans certains pays en Europe, invite à approfondir davantage la réflexion. Après une brève analyse comparative de ces sources de données, l'article montre qu'en réalité elles sont plus complémentaires que duales et suggère une nouvelle logique pour les réconcilier. Cette logique invite à évoluer vers un système d'information statistique pérenne, intégré et cohérent et à coût réduit, permettant de réconcilier les besoins des populations avec les réponses de politiques sociales proposées par les services publics nationaux.

Classification JEL : C42, C81, C82, D02, D82, D83.

[♦] Expert en analyse de la pauvreté, AFRISTAT/CERA

1. INTRODUCTION

Faisant suite à la Déclaration du Millénaire de 2000, les Nations unies et divers autres partenaires techniques et financiers ont constamment œuvré pour le renforcement et l'amélioration de la qualité de l'information statistique. Afin de suivre les progrès accomplis quant aux objectifs à réaliser et aux cibles à atteindre à l'horizon 2015, d'importants travaux ont été menés : recentrage de la liste des indicateurs quantitatifs à élaborer, harmonisation conceptuelle et méthodologique, identification et réconciliation des principales sources de données susceptibles d'être utilisées pour la production de ces indicateurs, important efforts de plaidoyer et de sensibilisation pour la prise en compte de ces indicateurs dans les systèmes statistiques nationaux, renforcement de la coordination des principaux acteurs de la production statistique, etc.

Si de réels progrès ont été enregistrés en dépit des balbutiements des principaux appareils statistiques dans les années 90, une question essentielle reste toujours d'actualité : comment garantir une information statistique pérenne et fiable pour mieux mesurer les progrès réalisés au niveau du suivi et évaluation des Objectifs du Millénaire pour le Développement (OMD) ? Autrement dit, le débat semble aujourd'hui, plus que jamais se traduire dans la mise en place d'un dispositif efficace et cohérent de production statistique basé sur la mutualisation des moyens mobilisables par les pays et pour les pays.

La réforme des services publics, amorcée dans nombre de pays développés, invite à poser les premières pierres d'une réflexion à propos de la stratégie envisageable pour dynamiser les appareils statistiques des pays en développement. Face à une demande de plus en plus croissante d'information statistique pour l'orientation des politiques de développement fondées sur la lutte contre la pauvreté, quels devraient être les principaux acteurs de la production statistique, mais surtout comment devrait-on envisager un dispositif plus optimal que celui existant actuellement ?

En dépit des efforts entrepris par tous les acteurs des Systèmes statistiques nationaux (SS), l'impression reste toujours grande d'une « *apparente dualité* » entre données de sources administratives et données d'enquêtes, actuellement les deux principales sources d'informations pour le suivi et évaluation des indicateurs des OMD. Qu'en est-il en réalité, la perspective de dynamiser les SSN dans les pays en développement ne va-t-elle pas de paire avec la nécessité de réconcilier ces deux sources d'informations, et dans quelle(s) mesure(s) l'entrevoir ?

Cette contribution est en réalité une invitation à aller au-delà de la polémique largement débattue « données de sources administratives versus données d'enquêtes » pour jeter les bases d'une réflexion visant à proposer une dynamisation des SSN actuels de nombre de pays en développement d'Afrique subsaharienne. Sortir du schéma classique du SSN tel qu'il se présente actuellement dans presque tous ces pays et envisager un nouveau, mieux intégré, permettant de répondre de façon cohérente et pertinente à une demande sociale de plus en plus importante et diverse devient une nécessité si l'on veut dans les dix prochaines années disposer d'un SSN illustrant plus fidèlement l'état de ces sociétés. La contribution est structurée en quatre parties :

- la première rappelle brièvement, dans une approche comparative, les caractéristiques des principales sources de données pour les traitements statistiques ;
- la deuxième revient sur la nature de la relation qui existerait entre ces deux principales sources ;
- la troisième partie, sur la base de l'examen de quelques indicateurs sectoriels (éducation, santé) revient en partie sur les préoccupations soulevées dans la deuxième partie ;
- la quatrième partie met au cœur de ce nouveau dispositif, les indispensables exigences de qualité de toutes productions statistiques.

I. GENERALITES SUR LES CARACTERISTIQUES DES PRINCIPALES SOURCES DE DONNEES POUR LES TRAITEMENTS STATISTIQUES

Le tableau 1 fournit de façon comparative et synoptique les éléments d'analyse des sources d'informations. Il s'articule autour de quatre points : la définition des sources, leur principal objet, leurs avantages et inconvénients et les critères indicatifs pour le choix de l'une ou de l'autre source de données.

Tableau 1 : Bilan comparatif des données de sources administratives et de données d'enquêtes

	Données de sources administratives	Données d'enquêtes et recensements
Définition	<ul style="list-style-type: none"> • Informations produites par une institution dont les activités de gestion impliquent la tenue, selon des règles générales, de fichiers ou de registres individuels, dont l'agrégation n'est qu'un sous-produit, alors que les informations individuelles en sont l'élément important, notamment pour les individus ou les entreprises concernés. 	<ul style="list-style-type: none"> • Données collectées par un organisme spécialisé dans la statistique et visant à produire une information « agrégée de portée générale »
Objet	<ul style="list-style-type: none"> • Portent plus sur les moyens et sur les actions publiques que sur l'état de la société ; • De nature à refléter l'action déjà existante (offre publique). 	<ul style="list-style-type: none"> • Portent généralement sur des faits sociaux justifiés par le contexte et par une certaine demande sociale ; • Plus susceptibles d'exprimer des aspects nouveaux de la société (demande sociale)
Avantages	<ul style="list-style-type: none"> • Disponibilité des données de base à moindre coût (données de routine) ; • Fiabilité de l'information collectée ; • Illustration de ce qui est ; • Exhaustivité ou du moins quasi-exhaustivité des données collectées sur le champ d'investigation ; • Facilité de moduler de façon plus pertinente les représentations par zones des statistiques produites ; • Utilisation déterminée par des règles générales, standardisées et inscrites dans les textes généraux, de diverses natures. 	<ul style="list-style-type: none"> • Rapidité de l'accès à l'information ; • Illustration de ce qui pourrait être ; • Fondée sur des considérations théoriques solides (théorie des sondages et probabilités) ; • Facilité d'harmonisation des outils et méthodologie, applicable à plusieurs pays ;
Inconvénients	<ul style="list-style-type: none"> • Possibilité de biais au niveau de la chaîne de transmission des données de base ; • Validité parfois problématique de l'exhaustivité de l'information collectée (couverture du champ des unités observées, conditions du traitement des données de routine, etc.) ; • Trop tributaire de définitions et de catégories réglementaires (et donc diffère bien souvent des besoins exprimés par les utilisateurs) ; • Difficultés à garantir la comparaison entre différents pays ou même dans le temps dans un même pays ; • Illustratif principalement de la gestion courante du service administratif concerné (moyens mis en œuvre, résultats produits conséquemment) 	<ul style="list-style-type: none"> • Périodicité imposée par la disponibilité du financement • Exhaustivité statistique de l'information collectée (représentativité statistique de la couverture géographique, saisie des déclarations de tous les individus identifiés dans ce champ, etc.) ; • Coût en temps (pour les enquêtés) avec le risque de trop souvent les solliciter (propension à ne pas répondre des individus et/ou des unités statistiques ciblées) ; Coût en argent (face aux difficultés financières que rencontrent nombre de pays pour prendre en charge de telles activités) ;
Critères objectifs de choix d'un dispositif	<ul style="list-style-type: none"> • Importance de la charge supportée par les unités statistiques ciblées (individus, entreprises, etc.) ; • Coût financier de l'option retenue ; • Pertinence de l'information produite par rapport à la formulation des politiques de lutte contre la pauvreté (pertinence sur le contenu cognitif et le rôle social des institutions statistiques) ; • Pérennisation du processus de production statistique de l'information ; • Exigence de comparabilité des statistiques produites ; 	

II. DONNEES D'ENQUETES VERSUS DONNEES DE SOURCES ADMINISTRATIVES : COMPLEMENTARITE, SUBSIDIARITE OU SUBSTITUTION ?

Analyser la nature de la relation qui pourrait lier ces deux sources d'informations renvoie en réalité aux deux préoccupations suivantes qui les différencient :

- les motifs pour lesquels l'information est produite ;
- le (les) niveau (niveaux) d'utilisation qui peut (peuvent) en être fait(s) ;

L'examen des motifs pour lesquels l'information est produite renvoie en réalité à un certain nombre de préoccupations concernant principalement :

- le champ d'investigation couvert (national, régional, autres préoccupations liées aux découpages administratifs ou à la territorisation/décentralisation de l'information, etc.) ;
- l'objet de l'investigation qui est généralement assez spécifique (éducation, santé, emploi, économie et finance, budget, programmation des investissements publics, etc.) ;
- les méthodes de collecte des informations et leurs limites (collecte exhaustive (?), théorie des sondages et échantillonnage, pertinence des outils de collecte, pertinence et fiabilité de l'information produite, etc.) ;

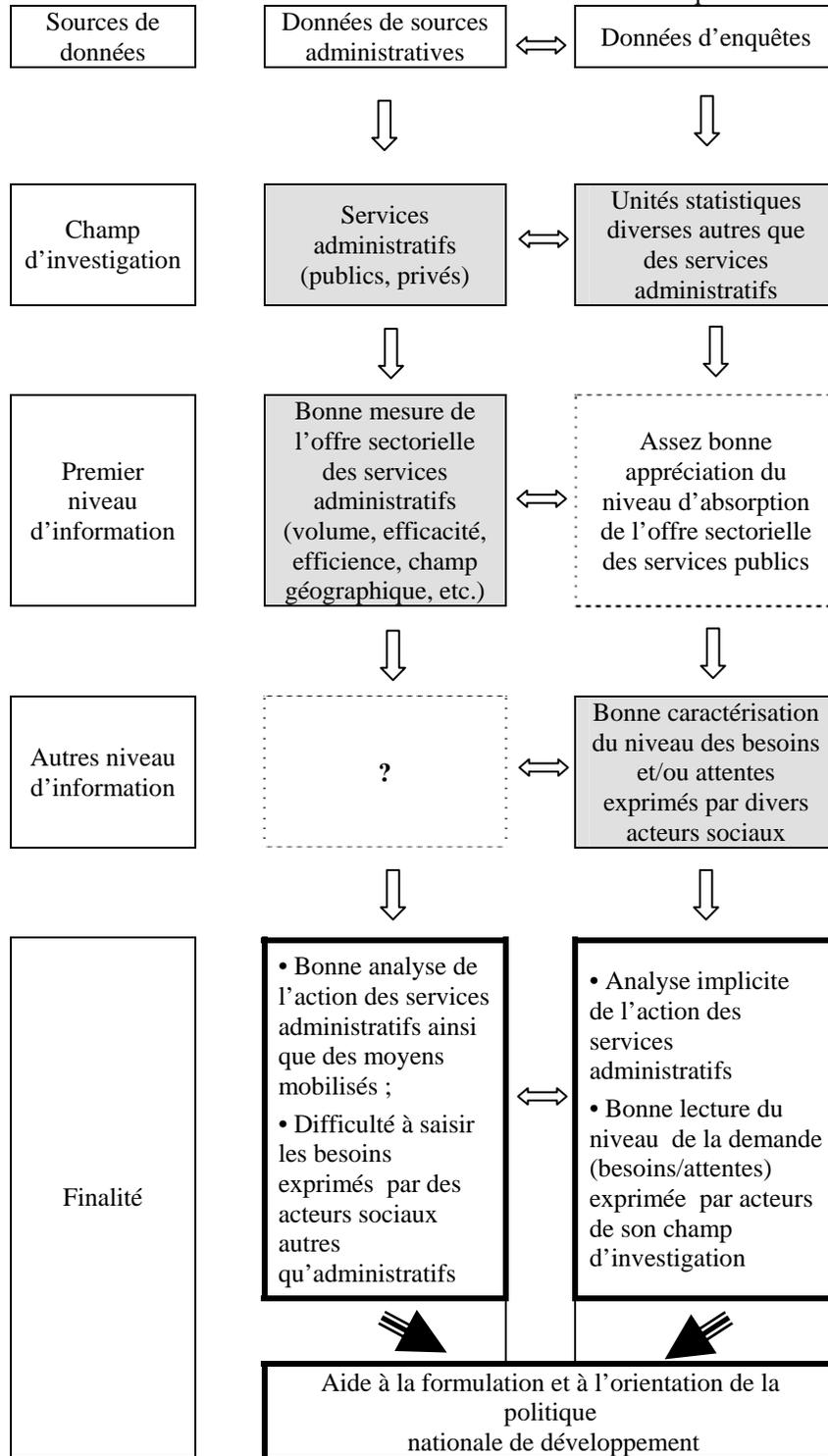
Quant à l'information produite, elle répond, dans le cas des données de sources administratives, en première intention, principalement à des préoccupations de gestion courante des services qui produisent ces données. En seconde intention, découlent toutes les utilisations possibles (production de statistiques agrégées et d'hypothèses d'analyses empiriques). Dans tous les cas de figure, le souci premier de ces données est de rendre compte de l'action des services (publics ou privés) qui en sont les producteurs. Ces données caractérisent bien dans une certaine façon, les activités relatives à l'offre sectorielle des services administratifs (volume, efficience, efficacité) en réponse à une demande sociale implicite.

Au contraire, les données d'enquêtes résument le plus souvent des préoccupations qu'exprimeraient des acteurs sociaux bien ciblés. En conséquence, elles restent plus flexibles et ouvertes par rapport aux données de sources administratives. Dans un second temps, elles caractérisent assez bien (sous réserve des précautions méthodologiques de la théorie d'échantillonnage et de la méthodologie de traitement et d'analyse des données) le champ des besoins et/ou des attentes exprimées par divers acteurs sociaux. Les données de sources administratives, au contraire, sont limitées pour apporter une réponse à ce dernier niveau d'information.

Dès lors, la question de la relation entre données d'enquêtes et données de sources administratives doit être nuancée, ces deux sources étant en réalité tantôt complémentaires (champ d'investigation), tantôt subsidiaire, voire même complémentaire (premier niveau d'information), tantôt carrément substitués (autres niveaux d'information). Dans tous les cas de figure, ensemble, elles enrichissent l'analyse et l'orientation des politiques de développement (cf. schéma 1).

L'analyse de la relation entre données de sources administratives et données d'enquêtes est donc en réalité très loin de traduire une certaine dualité. Le contexte de leur divers usages montre qu'elles peuvent être complémentaires, substitués ou encore subsidiaires.

Schéma 1 : Illustration de la relation de complémentarité et/ou subsidiarité entre données de sources administratives et données d'enquêtes



Légende :

Cases en gris et trait plein : Avantage net d'une source

Cases en pointillé : Autres avantages implicites d'une source

↔ : Effet d'interférence

III. DONNEES D'ENQUETES VS DONNEES DE SOURCES ADMINISTRATIVES : LE CAS DES STATISTIQUES DE L'EDUCATION ET DE LA SANTE

Pour illustrer la spécificité du contexte de l'utilisation des diverses relations entre données de sources administratives et données d'enquêtes, il convient de le situer dans le cadre de la collecte d'information pour le suivi-évaluation des DSRP et des OMD¹. Dans ce cas, l'analyse s'intéresse moins à la spécificité propre à l'information de base produite par chacune des sources de données, mais beaucoup plus à l'usage final qui en est fait. Autrement dit, la question se pose à ce niveau de savoir comment concilier les différentes sources de données disponibles pour produire une information statistique fiable et pertinente.

Pour examiner ce point, l'analyse va se focaliser sur les secteurs **Education (Objectifs 2 et 3 des OMD)** et **Santé (Objectifs 4, 5 et 6 des OMD)**, sur la base de certains indicateurs sectoriels choisis. Il est donc clair à ce niveau que le cadre qui fédère l'usage des deux sources d'information est la formulation et l'orientation des politiques de lutte contre la pauvreté, conformément à la déclaration du Millénaire.

Sans chercher à revenir sur les détails de la comparaison entre les deux sources de données, ce point insistera plutôt sur les limites qui leurs sont inhérentes en fonction d'indicateurs bien définis en caractérisant la nature de la relation en présence et de l'information statistique que l'on veut valoriser finalement.

3.1 Le secteur Education : Objectifs 2 et 3 des OMD

L'examen de l'annexe 2 sur quelques indicateurs relatifs à l'éducation montre que les données de sources administratives pourraient suffire sous certaines à produire des statistiques fiables. Ces conditions sont liées en partie à l'organisation de la collecte des données et à leur chaîne de traitement, en partie à la nature même de l'information de base qui est produite.

Dans la grande majorité des Etats africains, le dispositif administratif de collecte des données en général, en particulier dans le domaine de l'éducation est confronté à d'importantes contraintes : faiblesse des ressources humaines compétentes en traitement statistiques, instabilité du personnel statisticien ou des cadres formés dans ce domaine, faible organisation de la collecte des données dans les différentes unités scolaires, problèmes de gestion informatique de ces données et de leur archivage, faible collaboration avec l'institut national de la statistique pour la coordination des activités statistiques (concepts et définition, harmonisation des questionnaires, etc.) dans le domaine de l'éducation, etc.

Bref, les services administratifs de l'éducation restent encore très cloisonnés dans leur domaine et semblent peu tirer profit des compétences qui existent auprès des autres acteurs du système statistique national (INS, services de l'état civil, services des affaires sociales, services de la fonction publique, etc.) mais aussi des externalités positives dont ils peuvent bénéficier dans le cadre d'une collaboration plus étroite avec ces acteurs.

Finalement, la principale source d'information pour le suivi et l'évaluation des indicateurs relatifs à l'éducation dans le cadre des DSRP et des OMD reste les données d'enquêtes. Mais pour s'inscrire dans le cadre d'une production statistique cohérente et consensuelle, les acteurs de la production statistique doivent davantage collaborer dans le sens de l'amélioration des concepts et définitions, mettre en place un dispositif structuré de collecte et de traitement des données de routine et d'enquêtes avec l'identification précise des principales sources de données, valider ensemble les tendances des principaux indicateurs produits ainsi que les limites de leurs diverses utilisations.

¹ L'annexe 1 rappelle la liste des huit Objectifs du Millénaire pour le Développement ainsi que les cibles associées

3.2 Le secteur santé : Objectifs 4, 5 et 6

De façon générale, les critiques et observations faites au niveau des indicateurs liés à l'éducation s'appliquent aux indicateurs concernant le domaine de la santé (cf. annexe 3). L'utilisation des données d'enquêtes ira grandissante tant que les services administratifs ne seront pas dynamisés et ne se seront pas inscrits dans une nouvelle logique (fournir une informatique plus globale sur les attentes et les perceptions des utilisateurs, au-delà donc de l'offre de ces services).

Au total, quel que soit le domaine considéré (santé, éducation ou tout autre domaine), données de sources administratives et données d'enquêtes seront toujours tantôt complémentaires, tantôt partiellement substitués, aucune ne satisfaisant totalement à la demande sociale d'information précise. La question qui se pose finalement se résume dans la préoccupation suivante : dans quelle mesure les données de sources administratives devraient apporter une réponse à la demande sociale, au-delà des moyens et de l'action des services administratifs ? Autrement dit, si pour l'heure, les données de sources administratives répondent (théoriquement au moins) aux moyens et à l'action des services administratifs, leur revient-il de garantir la production statistique sur l'état de la demande sociale exprimée par la société ?

Dans la négative, cela revient à opter pour une pérennisation du système actuel avec dans le meilleur des cas, une dynamisation du dispositif de collecte des données de routine, mieux intégré dans le SSN. Dans l'affirmative, cela revient à choisir une évolution du système national de production statistique avec des rôles nouveaux pour les services administratifs traditionnels beaucoup plus vers une spécialisation fonctionnelle totale dans leur domaine respectif (éducation, santé, état-civil, etc.) et pour l'INS, une mission de consolidation, de synthèse et d'analyse des informations produites par l'ensemble des autres acteurs du SSN (cf. schéma 2).

Le schéma 2 relatif à la dynamisation des activités des services administratifs conduit à intégrer les activités d'enquêtes dans le champ de leurs activités « courantes » pour ne pas dire régaliennes. Cette hypothèse pour difficilement réalisable qu'elle puisse paraître actuellement, mérite d'être approfondie. Elle permettra de concilier une fois pour toute, données administratives et données d'enquêtes dans un cadre cohérent et intégré d'informations pertinentes pour le système statistique national.

Par cette entrée, on envisage donc les statistiques publiques d'une façon inhabituelle, en pensant en même temps à leurs apports de connaissance et aux circuits sociaux de leurs mises en forme et de leurs usages. Et il convient maintenant de s'y pencher dans la mesure où les développements actuels de la demande des statistiques sont déterminées par trois usages : l'intégration régionale et sous-régionale, la décentralisation et les réformes de la gestion des services administratifs publics à travers l'instauration de procédures d'évaluation quantifiée de leur efficacité.

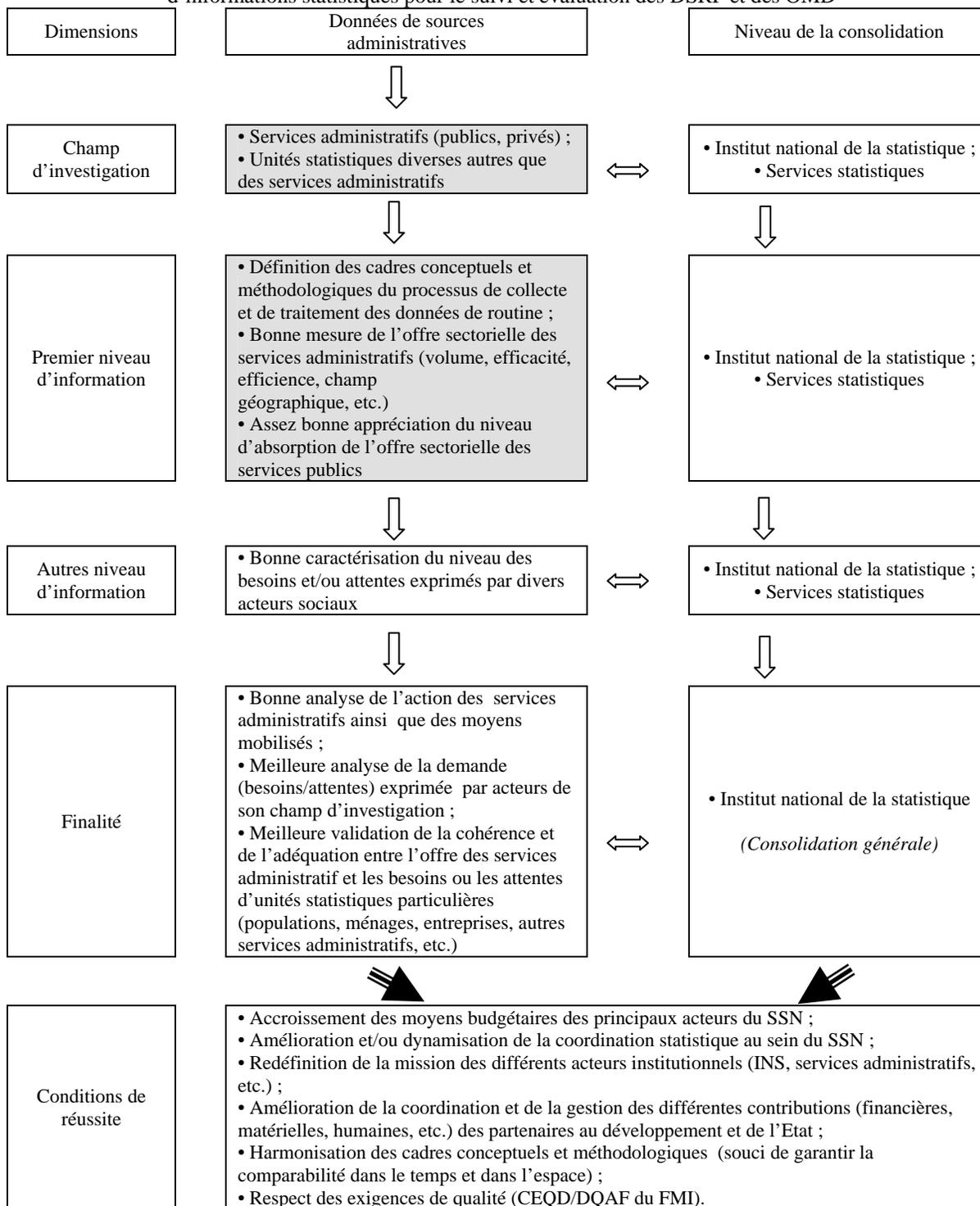
La meilleure illustration de l'urgence de cette orientation de la production statistique est donnée par la situation de la délinquance et/ou des enfants dans/de la rue. Ces phénomènes sociaux ont des chances de s'aggraver avec la paupérisation croissante des ménages. Il convient de mettre en place un dispositif d'information fiable, pertinent et pérenne capable de faire une bonne description de la situation. La dynamisation des services du ministère des affaires sociales dans le sens proposé pourrait apporter des réponses assez intéressantes à une meilleure connaissance, à moindre coût de ces faits sociaux et envisager des interventions de politiques publiques plus adéquates.

En intégrant dans le mandat des services administratifs publics cette nouvelle dimension, les SSN évolueront vers une production statistique beaucoup plus cohérente avec une plus grande adéquation entre

l'offre des services publics (analyse des moyens et de l'action des acteurs publics) et les besoins et/ou attentes exprimées par des populations déterminées. Comme dans le cas des résultats fournis par les enquêtes, il s'agira non pas de fournir de façon normative les bonnes réponses clé en main, mais d'analyser le plus clairement possible des préoccupations données.

Dans ces conditions, il conviendra d'être davantage plus regardant sur l'évaluation de la qualité de l'information statistique qui sera ainsi produite sur la base des six critères habituels (cf. point 4 suivant).

Schéma 2 : Dynamisation envisageable des activités des services administratifs pour répondre à la demande d'informations statistiques pour le suivi et évaluation des DSRP et des OMD



IV. DE LA NECESSITE DU RESPECT DES EXIGENCES DE LA QUALITE DE L'INFORMATION STATISTIQUE

Ce point reprend pour l'essentiel une traduction du cadre d'évaluation de la qualité des données (CEQD) ou en anglais data quality assessment framework (DQAF) du Fonds monétaire international (FMI).

4.1 Conditions préalables de la qualité

Ces conditions sont déterminées par l'existence d'un cadre juridique et institutionnel, la disponibilité des ressources (humaines, financières, logistiques) pour la conduite des travaux statistiques et le souci de rechercher un certain niveau de qualité.

4.1.1 Cadre juridique et institutionnel

Il définit :

- la responsabilité de la collecte, du calcul des données et de la diffusion doit être bien établie ; dans le cas d'implication conjointe de plusieurs organismes, les rôles de chacun doivent être clairement définis ;
- le partage des données et la coordination entre producteurs doivent être facilités à travers des échanges et un système de communication approprié ;
- la confidentialité des données doit être respectée, si nécessaire des procédures (agrégation, « anonymisation » des données individuelles) doivent être mises en place pour éviter de dévoiler des données individuelles couvertes par le secret statistique (les lois statistiques devront être suffisamment claires sur cette question) ;
- la prise de mesures idoines pour assurer un taux de couverture suffisant : mesures légales ou réglementaires, mais aussi travail auprès des répondants (les rassurer sur la confidentialité, les assister, leur montrer comment sont utilisées leurs réponses...).

4.1.2 Ressources disponibles pour les travaux statistiques

Les moyens mis en œuvre (humains, techniques et financiers), doivent être suffisants, la qualité ayant un double coût : les moyens pour l'assurer et les moyens pour la mesurer. En outre, il est important de vérifier périodiquement que ces ressources sont utilisées de la façon la plus efficace.

4.1.3 Compréhension du niveau de qualité recherché

Elle doit se traduire dans les préoccupations suivantes :

- la qualité doit être une composante essentielle du travail statistique ;
- cette appréciation de la qualité doit s'exercer aux différents stades de la production de données ;
- les producteurs doivent être constamment à l'écoute des nouveaux besoins ;
- un « retour » des utilisateurs doit être encouragé.

4.2 Critères d'évaluation de la qualité

4.2.1 Intégrité

L'objectivité doit être respectée dans la collecte, le traitement et la diffusion des données. Cette notion recouvre trois critères.

Le professionnalisme

- les pratiques s'appuient sur des principes professionnels : les choix sont faits en toute indépendance, l'amélioration de la compétence est encouragée par la publication et la mise à jour de notes méthodologiques et analytiques, les discussions et les présentations externes ;
- le choix des sources et des techniques utilisées est déterminé par des considérations statistiques qui sont portées à la connaissance des utilisateurs, si nécessaire ;
- le producteur de statistiques a le devoir d'informer le public lorsque sa production est utilisée à mauvais escient ou fait l'objet d'interprétations erronées par les médias ou dans tout autre cadre.

La transparence

- les principes et la pratique de production des données sont transparents ; le public est informé des conditions de collecte, de traitement et de diffusion de l'information, y compris du caractère obligatoire ou facultatif de la réponse et des garanties en matière de confidentialité des données individuelles collectées.
- la possibilité d'accès des autorités politiques aux données avant leur publication doit être signalée au public (par exemple, identité de l'autorité concernée et du stade du processus de production auquel cet accès s'effectue).
- la responsabilité de chaque intervenant dans le processus est publiquement définie (identification de(s) agence(s) productrice(s) de telle série).
- les usagers sont informés, à l'avance, des changements importants de méthodologie, de sources ou de méthodes statistiques.

Les normes déontologiques

- les normes à appliquer lors de risque de conflits d'intérêts sont claires et connues du personnel.
- la réputation des dirigeants et l'attention qu'ils portent aux normes éthiques sont de nature à protéger l'autonomie de l'institution en cas d'ingérences politiques ;
- les principes sont périodiquement rappelés.

4.2.2 Rigueur méthodologique

La méthodologie est orthodoxe. Pour cela, elle est définie en accord avec les normes internationales et les exceptions par rapport aux règles sont clairement annoncées. Ceci s'applique notamment aux quatre rubriques suivantes.

- Concepts et définitions ;
- Champ d'application ;
- Classifications et sectorisations ;
- Base d'enregistrement.

4.2.3 Exactitude et fiabilité

La production statistique doit refléter, au maximum, la réalité de l'économie ou la société. Pour cela, elle doit être bâtie en tenant compte des informations aspects suivants :

L'existence de données de base

- les sources sont adéquates ; elles sont larges et complétées si nécessaires ;
- les échantillons utilisés sont établis de façon scientifique ;
- les informateurs de base sont précisément au fait de ce qu'on leur demande et de l'utilité de la production ;
- les délais de réponse sont respectés.

L'utilisation de techniques statistiques

- celles-ci doivent répondre aux critères de fiabilité de la méthodologie ;
- elles permettent d'identifier les problèmes relatifs aux données de base ;
- des réestimations sont faites régulièrement.

La vérification et la validation des données de base

- l'exactitude des données de base est régulièrement surveillée (e.g. couverture, non-réponses, points aberrants ...) ;
- la cohérence des données entrantes est surveillée, surtout lorsque plusieurs sources sont utilisées ;
- la cohérence temporelle est également vérifiée.

La vérification et la validation des données et produits statistiques intermédiaires

- les principaux résultats intermédiaires sont validés grâce à des sources alternatives ;
- l'évolution des principaux résultats est analysée à l'aide d'indicateurs ;
- la cause d'éventuelles divergences est recherchée.

La réalisation d'étude pour les révisions éventuelles

- les révisions entre versions successives sont analysées ;
- le résultat de ces investigations peut conduire à des changements de source ou de méthodes de calcul.

4.2.4 Utilité

Les statistiques produites doivent répondre aux besoins, être établies dans les délais, avec une périodicité appropriée, être cohérentes, à la fois de façon interne et avec les autres données, et suivre des règles de révision bien définies.

Pertinence

Les besoins des différents utilisateurs sont régulièrement évalués pour vérifier que les productions sont adéquates. Réciproquement, il peut être utile que les fournisseurs de données tiennent les utilisateurs au courant des évolutions concernant les données transmises ou les processus de production, afin que les utilisateurs puissent affiner ou élargir leurs demandes. Les résultats sont présentés lors de réunions tant nationales qu'internationales.

Délais de diffusion et périodicité

- les statistiques paraissent selon des calendriers définis et publiés.

Cohérence

- la cohérence interne est assurée, par exemple, entre les différentes périodicités ;
- les postes d'ajustement sont réduits au maximum, sans artifice comptable ;
- les séries chronologiques suffisamment longues sont disponibles ;
- lors des changements méthodologiques, des rétrapolations sont publiées lorsque cela est possible. Si cela n'est pas le cas, le public est informé des raisons pour lesquelles les rétrapolations ne sont pas publiées ;
- les ruptures sont expliquées par des notes méthodologiques ;
- la cohérence avec d'autres sources est assurée, au besoin par l'explication des divergences.

Politiques et pratiques de révision

- les révisions suivent un calendrier régulier et connu des utilisateurs ;
- les révisions sont clairement identifiées lors de la publication des données ;
- le caractère estimatif ou provisoire des données est annoncé.
- les analyses et études des révisions sont publiées.

4.2.5 Accessibilité

On regroupe sous ce terme tout ce qui a trait à la mise à disposition vers l'utilisateur. Les critères sont le mode de diffusion des données, les informations supplémentaires et l'aide aux utilisateurs.

Accessibilité des données

- les données sont publiées d'une façon claire et adaptée aux besoins des utilisateurs ;
- une analyse des principales évolutions accompagne cette diffusion ;
- différents supports de publication sont utilisés en fonction des utilisateurs.
- le calendrier de publication des données est lui même publié.

Accessibilité des métadonnées

- l'information permettant de mieux comprendre la statistique (métadonnées) est disponible : notes méthodologiques, taux de réponse, divergences avec les standards internationaux, liens avec les autres systèmes d'information ... ;
- le niveau de détail de cette information est adapté aux différents publics visés.

Assistance aux utilisateurs

- les coordonnées des responsables des données sont connues ;
- une documentation à jour est disponible.

5. CONCLUSION

Finalement, le principal axe de lecture du débat autour de la question « données d'enquêtes versus fichiers administratifs » s'inscrit dans la réflexion à mener pour construire une information complète et cohérente structurée autour de quatre préoccupations : capter à moindre frais l'état des besoins et/ou attentes de la société, garantir une certaine adéquation des moyens et de l'action des services publics administratifs avec les besoins/attentes exprimées, harmoniser les modes de description des dispositifs de collecte des données, notamment statistiques, redéfinir le cadre institutionnel, réglementaire et juridique de la production statistique au niveau national. La dynamisation des SSN dans les pays africains est au prix de cette reconversion majeure s'ils veulent apporter des réponses adéquates à l'orientation des politiques de lutte contre la pauvreté.

6. BIBLIOGRAPHIE

Laurence Dauphin, *La qualité des données sur les revenus : Enquêtes versus fichiers administratif l'exemple de l'enquête SRCV 2005*, INSEE, Direction des statistiques démographiques et sociales.

Alain Desrosières, *Enquêtes versus registres administratifs : réflexions sur la dualité des sources statistiques*, Courrier **des statistiques** n° 111, septembre 2004.

Décrire l'Etat ou explorer la société : les deux sources de la statistique publique, Genèses, 2005/1, N°58 pages 4 à 27.

Historiser l'action publique : l'Etat, le marché et les statistiques i, Pascale Laborier et Danny Trom, (éd.) *L'historicité de l'action publique*, Paris, Puf 2003, pages 207-221.

Françoise Drumetz,
Régine Monfront *La qualité des informations statistiques*, Bulletin de la Banque de France, N°110 – Février 2003.

Jean-Louis Rallu *L'appareil statistique des OMDs : un potentiel à développer pour les INS*, Stateco N°105, 2010, pages 19 à 31.

Groupe des Nations unies pour le développement, sous la direction du Fonds des Nations unies pour la population, du Programme des Nations unies pour le développement et du Département des affaires économiques et sociales (Division de statistique), *Indicateurs pour le suivi des progrès accomplis dans la réalisation des objectifs du Millénaire pour le développement, Définition, Raison d'être, Concepts, Sources*, Nations unies, New York, 2005

Annexes

Encadré 1 : Les huit Objectifs du Millénaire pour le Développement (OMD) et les cibles associées

Objectif 1 : Réduire l'extrême pauvreté et la faim
Cible 1 : Réduire de moitié, d'ici à 2015, la proportion de la population dont le revenu est inférieur à 1 dollar par jour
Cible 2 : Réduire de moitié, d'ici à 2015, la proportion de la population qui souffre de la faim
Objectif 2 : Assurer l'éducation primaire pour tous
Cible 1 : D'ici à 2015, donner à tous les enfants, garçons et filles, partout dans le monde, les moyens d'achever un cycle complet d'études primaires
Objectif 3 : Promouvoir l'égalité des sexes et l'autonomisation des femmes
Cible 1 : Éliminer les disparités entre les sexes dans les enseignements primaire et secondaire d'ici à 2005 si possible, et à tous les niveaux de l'enseignement en 2015 au plus tard
Objectif 4 : « Réduire la mortalité infantile (des enfants de moins de 5 ans)
Cible 1 : Réduire de deux tiers, d'ici à 2015, le taux de mortalité des enfants de moins de cinq ans
Objectif 5 : Améliorer la santé maternelle
Cible : Réduire de trois quarts, d'ici à 2015, le taux de mortalité maternelle

Objectif 6 : Combattre le VIH/SIDA, le paludisme et d'autres maladies
Cible : D'ici à 2015, stopper la propagation du VIH/sida et commencer à inverser la tendance actuelle
Objectif 7 : Assurer un environnement durable
Cible 1 : Intégrer les principes du développement durable dans les politiques nationales et inverser la tendance actuelle à la déperdition des ressources environnementales
Cible 2 : Réduire de moitié, d'ici à 2015, le pourcentage de la population qui n'a pas accès de façon durable à un approvisionnement en eau potable
Cible 3 : Améliorer sensiblement la vie d'au moins 100 millions d'habitants de taudis, d'ici à 2020
Objectif 8 : Mettre en place un partenariat mondial pour le développement
Cible 1 : Poursuivre la mise en place d'un système commercial et financier multilatéral ouvert, fondé sur des règles, prévisible et non discriminatoire
Cible 2 : Traiter globalement le problème de la dette des pays en développement, par des mesures d'ordre national et international propres à rendre leur endettement viable à long terme
Cible 3 : En coopération avec les pays en développement, créer des emplois décents et productifs pour les jeunes
Cible 4 : En coopération avec l'industrie pharmaceutique, rendre les médicaments essentiels et abordables dans les pays en développement
Cible 5 : En coopération avec le secteur privé, mettre les avantages des nouvelles technologies, en particulier des technologies de l'information et de la communication, à la portée de tous

Annexe 2 : Aperçu de quelques problèmes liés à l'utilisation des données de sources administratives ou d'enquêtes pour l'estimation des indicateurs des OMD – Education

Indicateurs	Définition	Limites de la source	Observations/Remarques
Taux brut de scolarisation (TBS) au niveau X	$TBS_{\text{Niveau X}} = 100 \times \frac{Eff_{\text{Niveau}}}{Eff_{\text{Pop niveau}}}$	<ul style="list-style-type: none"> • Seule l'information sur le numérateur est fournie par les données de routine ; • Représentativité de la couverture des données de sources administratives (problème lié à la collecte des informations dans la chaîne de transmission) ; • Fiabilité de l'estimation du dénominateur dans le cas notamment des données de sources administratives ; • Problème très fréquent de gestion de la démographie et de la mobilité des élèves scolarisés dans les établissements scolaires. 	<ul style="list-style-type: none"> • Théorie de l'échantillonnage confère une certaine représentativité géographique aux résultats ainsi que la fiabilité des statistiques produites ; • Rapidité d'accès à l'information si le pays décide de financer la réalisation périodique de l'enquête ; • L'enquête collecte directement les informations nécessaires pour l'estimation de cet indicateur pour l'année considérée ; • Solution envisageable en combinant un système intégré de données sur l'enregistrement des faits d'état civil avec des données de RGPH actualisés.
Taux net de scolarisation (TNS) au niveau X	$TNS_{\text{Niveau X}} = 100 \times \frac{EffN_{\text{Niveau}}}{Eff_{\text{Pop niveau}}}$		
Pourcentage d'élèves commençant la 1 ^{ère} année du primaire qui atteignent la 5 ^{ème} année (A) ≡ Taux de survie en cinquième année	$A = 100 \times \frac{Eff_{\text{Pr im1}}}{Eff_{\text{Prim 5}}}$	<ul style="list-style-type: none"> • Problème de fiabilité de l'information nécessaire pour l'estimation de cet indicateur (gestion/reconstitution de la cohorte des élèves observés entre la première année du primaire et la 5^{ème} année très difficile) avec des données de sources administratives ; • Information sur le dénominateur, généralement plus fiable que celle figurant au numérateur dans les services administratifs. 	<ul style="list-style-type: none"> • Reconstitution plus facile du suivi longitudinal des élèves (cohortes) à travers les enquêtes auprès des ménages ; • Estimation fournie est fiable si le service administratif de l'éducation dispose d'un bon dispositif de gestion informatique des données et couvrant toutes les unités du systèmes éducatif du pays.
Taux d'alphabétisation des individus de 15 ans et plus (B)	$B = 100 \times \frac{Eff_{\geq 15 \text{ ans et SLE}}}{Eff_{\geq 15 \text{ ans}}}$	<ul style="list-style-type: none"> • Informations (numérateur, dénominateur) pour cet indicateur sont rarement produites comme des données de routine des services administratifs de l'éducation ; • Problème d'exhaustivité du champ des individus / unités statistiques ciblées (coût exorbitant pour un service administratif) 	<ul style="list-style-type: none"> • Informations produites plus facilement à partir des données d'enquêtes et recensement général de la population et de l'habitat ; • Cet indicateur devrait être plus souvent évalué (chaque année) que celui concernant les adultes vu que l'alphabétisation des jeunes progresse plus vite que celle des adultes.
Taux d'alphabétisation des individus de 15-24 ans (C) ≡ Taux d'alphabétisation des jeunes	$C = 100 \times \frac{Eff_{15-24 \text{ ans et SLE}}}{Eff_{15-24 \text{ ans}}}$		
Taux de brut scolarisation normale (D) au primaire (Prim)	$D = 100 \times \frac{Eff_{\text{Age prim inscrits au prim}}}{Eff_{\text{Pop niveau prim}}}$	<ul style="list-style-type: none"> • Cet indicateur correspond en réalité au taux net de scolarisation au primaire. En conséquence cf. les mêmes observations; 	<ul style="list-style-type: none"> • Idem que pour le taux net de scolarisation
Taux de brut scolarisation précoce (E) au primaire (Prim)	$E = 100 \times \frac{Eff_{< \text{Age prim inscrits au prim}}}{Eff_{\text{Pop niveau prim}}}$	<ul style="list-style-type: none"> • Gros problème de consolidation/agrégation au niveau national des données de l'ensemble des unités administratives du domaine de l'éducation ; • Fiabilité et fraîcheur de l'information sur le dénominateur en l'absence d'un système d'enregistrement systématique et rigoureux des faits de l'état civil et de la réalisation régulière des RGPH • Difficultés liées à l'analyse au niveau national de la matrice Age/niveau de classe au primaire. 	<ul style="list-style-type: none"> • Informations produites plus facilement à partir des données d'enquêtes et recensement général de la population et de l'habitat ; • Contraire liée au financement des enquêtes.
Taux brut de retard de scolarisation (F) au primaire	$F = 100 \times \frac{Eff_{> \text{Age prim inscrits au prim}}}{Eff_{\text{Pop niveau prim}}}$		

Nota : Eff niveau = Effectif des individus inscrits à un niveau d'études donné, quel que soit leur âge
 EffN niveau = Effectif des individus du niveau officiel d'âge d'inscription au niveau X donné effectivement inscrit à ce niveau
 Eff pop niveau = Effectif total des individus dont l'âge correspond à l'âge officiel d'inscription au niveau d'études donné (6-11 ans)

Annexe 3 : Aperçu de quelques problèmes liés à l'utilisation des données de sources administratives ou d'enquêtes pour l'estimation des indicateurs des OMD – Santé

Indicateurs	Définition	Limites de la source	Observations/Remarques
Taux de mortalité infantile (‰)	$1000 \times$ (Effectif des enfants morts avant l'âge d'un an rapporté au nombre total d'enfants de moins d'un an nés vivants à un moment donné)	<ul style="list-style-type: none"> • Problème lié à l'exhaustivité des informations figurant dans les registres administratifs ; • Gestion très souvent problématique des registres administratifs. 	<ul style="list-style-type: none"> • Bonne approximation fournie par les enquêtes auprès des ménages (EDS) ou indirectement, par la méthode Brass avec les enquêtes MICS
Prévalence du VIH/SIDA (en %)	100 (Effectif de la population malade du VIH/SIDA, rapporté à l'effectif total de la population à un moment donné)	<ul style="list-style-type: none"> • Problème récurrents liés à la couverture exhaustive des données (estimations basées sur les sites sentinelles ou sur les déclarations des malades qui se sont présentés à un centre de santé) ; • Existence de nombreuses estimations parfois divergentes de cet indicateur, en fonction de la source, de la méthodologie utilisée ou des hypothèses. 	<ul style="list-style-type: none"> • Assez bon proxy fourni par les données d'enquêtes auprès des ménages, notamment les EDS nationales ; • Les méthodes indirectes estimées à partir des données de recensement sont certes disponibles mais ces résultats datent généralement par rapport à l'actualité ; • Indicateur dont on peut encore améliorer la fiabilité en combinant les données relatives à l'enregistrement des faits de l'état civil avec ceux des services de santé (il convient donc de redynamiser leur fonctionnement).
Taux d'accouchements assistés par du personnel qualifié de santé (en %)	$100 \times$ (Effectif de la population des femmes en âge d'accoucher ayant donné naissance à un enfant né vivant avec l'assistance d'un personnel qualifié de santé rapporté à l'ensemble des naissances vivantes au cours d'une année donnée)	<ul style="list-style-type: none"> • Problème lié à l'exhaustivité et à la fiabilité (déclaration des naissances, gestion des registres administratifs, etc.) des données de sources administratives ; 	<ul style="list-style-type: none"> • Assez bonne estimation fournie par les données d'enquêtes auprès des ménages ; • Périodicité et financement des enquêtes.
Proportion d'enfants de moins d'un an vaccinés contre la rougeole (en %)	$100 \times$ (Effectif des enfants vaccinés contre la rougeole rapporté à l'effectif total des enfants à un moment donné)	<ul style="list-style-type: none"> • Exploitation parfois problématique de la gestion des fichiers administratifs ; • Fiabilité de l'information sur l'effectif total des enfants de moins d'un an 	<ul style="list-style-type: none"> • Données d'enquêtes sur la couverture nationale semble plus fiable ; • Problèmes généralement liés à la fréquence des données d'enquêtes et à leur financement.
Taux de mortalité maternelle (Pour 100000 naissances vivantes)	Effectif des décès maternels* observés durant une année rapporté au nombre des naissances vivantes de la même année, généralement exprimé pour 100 000	<ul style="list-style-type: none"> • Plusieurs difficultés relatives aux données de sources administratives : suivi rigoureux dans les registres administratifs des services de santé de ces informations (définition de la mortalité maternelle), problèmes d'agrégation des décès maternels dans les différentes régions du pays, prises en compte ou pas des déclarations tardives des décès maternels, etc. ; 	<ul style="list-style-type: none"> • Assez bon proxy pour cet indicateur fourni par les données d'enquêtes auprès des ménages, notamment les EDS et les recensements généraux de la population et de l'habitat ; • Indicateur dont on peut encore améliorer la fiabilité en combinant les données relatives à l'enregistrement des faits de l'état civil avec ceux des services de santé (il convient donc de redynamiser leur fonctionnement).
Taux de prévalence contraceptive (%)	Effectif des individus en âge de procréer utilisant une technique généralement admise de contraception rapporté à la population total des individus en âge de procréer	<ul style="list-style-type: none"> • Importance du biais lié à l'estimation du numérateur au niveau des registres administratifs: difficultés à prendre en compte les préservatifs utilisés par les associations ou les organisations caritatives, des acteurs du secteur privé, absence d'informations sur l'effectif des individus utilisant effectivement un moyen donné de contraception, etc. ; • Difficultés à mettre en cohérence les estimations entre contraceptifs vendus ou distribués et ceux effectivement utilisés. 	<ul style="list-style-type: none"> • Assez bon proxy fourni pour cet indicateur par les données d'enquêtes, notamment EDS, voire MICS ainsi que les conditions d'accès à ces contraceptifs ;

(*) : La mort maternelle a été définie depuis plusieurs années par l'OMS et est reprise dans la 10ème révision de la Classification Internationale des Maladies. "C'est le décès d'une femme survenu au cours de la grossesse ou dans un délai de 42 jours après sa terminaison, quelle qu'en soit la durée ou la localisation, pour une cause quelconque déterminée ou aggravée par la grossesse ou les soins qu'elle a motivés, mais ni accidentelle, ni fortuite.