



Le pilotage des politiques

Rôle et utilisation des indicateurs pour la mesure des performances

Pascal DELORME et Olivier CHATELAIN
Programme Aid Delivery Methods

Février 2011

**Ce document a été préparé pour la Commission Européenne, EuropeAid,
mais son contenu n'engage que les auteurs.**

Tables de matières

1.1	Pourquoi mesure t-on les performances?	2
	L'optimisation de ressources limitées.....	2
	Une amélioration des processus de décision et du fonctionnement des administrations	2
	Un objectif de démocratisation et de transparence.....	3
	Les difficultés et les effets pervers	3
1.2	La mesure des performances dans les politiques de coopération.....	4
	La demande interne aux pays	4
	Les obligations	4
	Les matrices d'indicateurs (<i>Performance Assessment Frameworks</i> – PAF).....	5
	L'utilité des indicateurs.....	6
2.1	Comment aborder le problème?	8
	Qu'est-ce qu'un indicateur?	8
	Les fonctions des indicateurs: des classifications difficiles	8
	Les moyens d'obtention : les notions de sources	10
	Qualité d'un indicateur : de nombreux critères complexes.....	11
	Choisir des indicateurs: les points clefs de la démarche	12
	Les points sensibles	13
2.2	Etape 1A – Une bonne analyse des politiques visées.....	14
	Comment la politique est-elle formulée?	14
	Dans quel but souhaite-t-on mesurer des performances?.....	15
	Quel aspect de la politique désire-t-on éclairer?	16
2.3	Etape 1B – Quels sont les indicateurs logiques les plus pertinents?.....	18
	Satisfaire le plus grand nombre d'utilisateurs?	18
	Satisfaire le plus grand nombre d'utilisations?	19
2.4	Etape 2A – Peut-on mesurer les progressions?	21
	Peut-on comparer des valeurs?.....	21
	L'influence de la formule de calcul sur la comparabilité.....	21
	La valeur de référence: des hypothèses largement implicites	22
2.5	Etape 2B – A quelle fréquence a-t-on besoin de l'indicateur?.....	24
	Lorsque l'on touche au fonctionnement du système statistique.....	24
	Des solutions diverses	24
2.6	Etape 2C – Peut-on Fixer une cible?.....	25
	Une question en trois points	25
	Qu'est-ce qu'une valeur de référence solide?	25
	Rythme de croissance et valeur de sortie	26
2.7	Etape 3 – Bilan de la démarche : documenter les indicateurs.....	28
	Adopter ou rejeter des indicateurs.....	28
	Documenter les indicateurs	28
3.1	La mesure de la qualité statistique	32
	Les éléments du problème	32
	Qu'est-ce que la qualité statistique?.....	32
	Des cadres conceptuels différents mais finalement convergents	33
	Les cinq dimensions de la qualité	33
3.2	Quelques dimensions de la qualité commentées	34
	La rigueur méthodologique	34
	L'exactitude et la fiabilité	35
	L'utilité.....	38
	Périodicité et délais de diffusion	39

L'accessibilité.....	39
3.3 Améliorer la qualité statistique	41
Que veut-on appuyer ?	41
Ne pas oublier les utilisateurs.....	42
L'amélioration de la disponibilité : des effets rapides	42
L'amélioration de l'accessibilité : faire pression à travers la demande	42
L'exactitude et la rigueur méthodologique : des travaux de longue haleine.....	42
La précision : pas de décision sans analyse coût / bénéfice.....	42
Annexe 1: Lexique	43
Annexe 2 : Pour en savoir plus.....	46
Quelques documents indispensables sur la mesure de la performance	46
Quelques documents indispensables sur les indicateurs	46
Quelques documents indispensables sur la qualité statistique	47
Annexe 3: Quelques questions a poser face a un indicateur	49
Annexe 4 : Les cadres qualité utilisés par la Banque mondiale/FMI et Eurostat.....	51
Annexe 5 : Descriptif succinct du guide Eurostat	52

Table des exemples

Exemple 1: Les effets pervers	3
Exemple 2 : Des indicateurs bien classés	14
Exemple 3: Un indicateur n'a pas de fonction définie théoriquement	16
Exemple 4 : Mesure matricielle ou mesure spot	16
Exemple 5: L'indicateur unique, ça n'existe pas!	18
Exemple 6. Une bonne logique verticale : le secteur de l'eau au Mali	19
Exemple 7: les valeurs sont-elles comparables? L'effet de collecte.....	21
Exemple 8: Qu'est-ce qui croît?.....	22
Exemple 9 : Le poids de la valeur de référence	22
Exemple 10 : Comparer des dynamiques. Progression en % ou indices.....	23
Exemple 11: Le choix de l'année de référence	25
Exemple 12: Cible et progression. La cible d'Abuja dans le domaine de la santé	27
Exemple 13 : Un bonne fiche d'identification	29
Exemple 14: De l'importance de cerner le champ statistique de la collecte.....	34
Exemple 15 : Les sources de données	35
Exemple 16: Quelques types d'erreurs.....	37

Liste des abréviations

AIDCO	EuropeAid Co-operation Office
CP	Cours Préparatoire
CSLP	Cadre Stratégique de Lutte contre la Pauvreté
CV	Coefficient de Variation
CWIQ	<i>Core Welfare Indicator Questionnaire</i>
FCFA	Franc de la Communauté Financière d'Afrique
FMI	Fonds Monétaire International
GAVI	<i>Global Alliance for Vaccines and Immunization</i>
GFATM	<i>Global Fund to Fight Aid, Tuberculosis and Malaria</i>
GFP	Gestion des Finances Publiques
IDH	Indice de Développement Humain
INS	Institut National de Statistique
ISU	Institut Statistique de l'UNESCO
OMD	Objectifs du Millénaire pour le Développement
PAF	<i>Performance Assessment Framework</i>
PED	Pays en développement
PEFA	<i>Public Expenditure and Financial Accountability</i>
PIB	Produit Intérieur Brut
PTF	Partenaires Techniques et Financiers
SIG	Système d'Information Géographique
SSN	Système Statistique National
TAP	Taux d'Achèvement au Primaire
TBS	Taux Brut de Scolarisation
UE	Union Européenne
UEMOA	Union Economique et Monétaire Ouest Africaine
UNESCO	<i>United Nations Educational, Scientific and Cultural Organisation</i>

Pourquoi ce guide ?

L'ambition de toute réforme est de proposer les meilleurs choix politiques possibles pour changer durablement les conditions de vie des personnes concernées.

Les indicateurs de résultat, attachés aux objectifs de réforme sont dès lors incontournables quand il s'agit de mesurer l'impact de telle ou telle politique, d'améliorer les processus de décision ou de communiquer de façon convaincante aux Parlements, à la société civile et au grand public quel a été l'impact des réformes en cours.

Néanmoins, étant donné la multitude de situations et d'objectifs, il n'existe pas d'outils "clefs en main" ou de listes d'indicateurs prédéfinies respectant les particularités et les priorités de chaque pays partenaire.

Pour aborder la délicate question du choix d'indicateurs sectoriels, ce guide propose donc une méthode raisonnée, articulée en trois étapes, pour guider le lecteur et tous ceux et celles qui sont confrontés à la mise en œuvre d'une démarche axée sur la performance et à la définition ou au choix d'indicateurs associés.

- La première étape concerne la formulation des besoins: quelle est l'information dont on souhaite disposer pour éclairer ou refléter les aspects précis et les défis de la politique sur lesquels on veut concentrer notre attention?
- La seconde étape permet de vérifier si les indicateurs pertinents au regard de la formulation des besoins disposent des qualités requises pour permettre de fixer des cibles, de suivre des progressions et d'être disponibles en temps voulu.
- La troisième étape formalise les indicateurs retenus sous forme de fiche descriptive visant à donner l'ensemble des informations concernant le calcul de l'indicateur, à documenter les sources utilisées, à émettre des restrictions éventuelles liées à la qualité des sources et à donner brièvement la façon dont l'indicateur peut être utilisé et interprété.

Le dialogue politique qui a trait à l'atteinte des résultats doit en outre être complété par une discussion sur les données à la base de la démarche de mesure de la performance. Le guide aborde ce sujet par une série de commentaires et d'exemples illustrant les principales dimensions de la qualité des données statistiques afin de rendre les notions essentielles et l'impact possible des défauts de qualité accessibles à tous les lecteurs.

Résumé

La gestion par la performance consiste à rendre cohérents les objectifs et les actions des acteurs avec les objectifs du niveau supérieur. Les différents niveaux qui peuvent être cités, dans le cas des politiques publiques, sont, par exemple : le Gouvernement, le Ministère des finances, les Ministères sectoriels, les Directions générales, les Directions, les Services, chaque niveau cherchant à obtenir la maîtrise des finalités qu'il a définies. Pour ce faire, il est nécessaire qu'il vérifie les résultats obtenus. Les moyens de vérifications sont divers, les indicateurs ne sont que l'un de ses moyens.

Toutefois, les démarches de gestion par les performances ne sont mises en œuvre dans les pays bénéficiaires de l'aide que depuis une dizaine d'années principalement sous les effets d'une demande interne de plus en plus importante liée à la démocratisation de très nombreux régimes politiques et donc à une plus grande demande de transparence de la part des citoyens. Cette démarche émane d'une demande externe liée à la mise en œuvre des nouvelles politiques de développement qui, depuis Monterrey jusqu' à Accra, affirment de façon de plus en plus précise la responsabilité conjointe des partenaires au développement et des bénéficiaires dans l'atteinte de résultats tangibles et mesurables.

Il est donc certain qu'à l'heure actuelle, il manque un certain recul pour évaluer les points positifs et négatifs de l'utilisation de cadres de mesures des performances et si des effets pervers ont été observés, dans les pays du Nord, lors de la mise en pratique des principes de gestion axés sur les performances, il n'existe pas, à proprement parler, d'évaluation de ces pratiques dans les pays du Sud.

Choisir un ou des indicateurs pour mesurer les résultats de telle ou telle politiques, repose, même si cela semble relativement tautologique, sur une bonne formulation de ces politiques et notamment sur une bonne maîtrise des logiques liées à la gestion axée sur les résultats. Le premier point de la démarche proposée vise donc à clarifier les relations qui existent entre les indicateurs et les niveaux d'objectifs auxquels ils se rapportent d'une part, et d'autre part, entre les indicateurs et ce qu'ils mesurent par rapport au phénomène étudié.

Si le premier point relève plus de la bonne compréhension de la gestion par les résultats et ne fait donc l'objet que de brefs rappels, le second conduit directement à la première caractéristique d'un indicateur, sa fonction, qui peut être définie comme étant la nature de la mesure réalisée par un indicateur. Par analogie avec la métrologie, un thermomètre mesure une température, un baromètre une pression atmosphérique. De la même manière, un taux brut de scolarisation mesure un degré de couverture d'une population par un système scolaire.

Cette notion de fonction, qui est somme toute indépendante des politiques mises en œuvre, se trouve mise en perspective dès lors qu'un indicateur statistique (ou logique) est utilisé dans un cadre de mesure des performances. En effet, de simple statistique, ce dernier va être affublé à d'autres caractéristiques qui vont permettre de déterminer entre autre : (i) le programme auquel il se rapporte, (ii) le niveau d'objectif auquel il est attaché, (iii) le service responsable de l'atteinte de ce ou ces objectifs, (iv) la valeur de l'indicateur avant le programme et la cible visée (valeur de l'indicateur en fin de programme), (v) la progression éventuellement souhaitée entre la valeur de départ et la cible visée.

A travers, cette énumération on peut, d'ores et déjà, donner une définition d'un indicateur de mesure, comme étant un ordre de grandeur (statistique ou logique) globalement défini par sa fonction, ses moyens d'obtention, sa qualité et ses limites d'utilisation. Toutefois, il convient de noter que dans de nombreux cas, l'indicateur est aussi un outil de communication, dans le sens où il est utilisé pour rendre compte des résultats des politiques; à ce titre, il doit être facilement compréhensible et interprétable et admis de façon relativement consensuelle.

Le présent guide se propose d'explorer, pas à pas, les différents éléments de la définition d'un indicateur de mesure et si possible de les intégrer dans une démarche logique qui parte de l'analyse des politiques pour aboutir à l'élaboration d'une fiche signalétique caractérisant chaque indicateur retenu dans un cadre de mesure donné.

Cette démarche comprend trois étapes au cours desquelles, le concepteur devra répondre à un certain nombre de questions, parfois élémentaires, parfois complexes, voir dans des cas extrêmes, sans réponses bien tranchées.

La première étape consiste : (i) à analyser la formulation des politiques visées, notamment la nature logique des enchaînements entre les objectifs, les activités et les résultats attendus et à se poser la question de savoir s'il convient de suivre ou d'éclairer tout ou une partie de cette politique; (ii) à rechercher quels sont les indicateurs logiques les plus pertinents qui permettent de mesurer l'atteinte des activités (on parlera alors d'indicateurs de réalisation) et des objectifs sectoriels (on parlera alors d'indicateurs de résultat). Cette étape débouche sur une matrice théorique qui ne prend en compte que le souhait de communiquer sur tel ou tel aspect des politiques.

La seconde étape, vise à confronter cette matrice théorique à la réalité des systèmes d'informations disponibles qu'ils s'agissent des systèmes statistiques ou des systèmes de suivis administratifs. On cherchera donc à savoir si les indicateurs sont mesurés ou mesurables avec les outils disponibles, en portant une attention particulière sur une caractéristique évoquée dans la définition à savoir si cette mesure permet d'évaluer les progressions envisagées. Pour reprendre l'analogie déjà citée, le thermomètre est-il suffisamment précis et fiable pour être utilisé ? Le résultat de cette étape débouche sur une matrice utilisable dans le cadre de mesure considéré, et peut conduire à rejeter des indicateurs particulièrement adaptés à la mesure du phénomène mais qui ne présentent pas certaines caractéristiques nécessaires: indicateurs trop complexes (peu communicables), trop coûteux à obtenir ou dont la mesure est trop espacée dans le temps.

La dernière étape constitue un récapitulatif des deux précédentes puisqu'il s'agira de reprendre de manière synthétique l'ensemble de ces caractéristiques dans une fiche. Cette étape, souvent négligée, est cependant fondamentale dans la mesure où les utilisateurs finaux des indicateurs sont parfois éloignés des considérations qui ont prévalu dans la phase d'identification. Aussi, cette fiche devra reprendre: (i) le ou les programmes qui utilisent l'indicateur comme mesure des performances ainsi que les services responsables de l'atteinte des résultats, (ii) l'ensemble des moyens mis en œuvre pour le calcul (sources de données, formules, unités de mesure), (iii) l'interprétation qui doit être faite et les limites d'utilisation, (iv) les délais d'obtention et le service responsable de sa mise à disposition.

Cette démarche, notamment dans sa seconde partie fait largement appel à des notions liées à la qualité statistique. On peut schématiser (ou caricaturer) le dialogue sur l'utilisation des indicateurs en deux types de dialogue: (i) celui qui a trait à l'atteinte ou non des résultats, c'est le dialogue politique ; (ii) celui qui traite de la qualité intrinsèque de l'indicateur, c'est le dialogue sur sa qualité (bien sûr lorsqu'il s'agit d'indicateurs statistiques).

Or, le dialogue sur la qualité statistique est souvent simplifié à l'extrême, entre les tenants du « bon chiffre », généralement les producteurs de données et les détracteurs utilisateurs ou producteurs concurrents. Pour essayer de sortir de ce type de débats qui peuvent rapidement devenir stériles, la dernière partie du guide tente de définir et d'illustrer les différentes dimensions qui caractérisent la qualité statistique et de voir dans quelle mesure ces défauts de qualité ont des impacts sur l'utilisation d'une donnée comme indicateur de mesure des performances. Si certaines qualités comme la précision sont relativement connues, il existe d'autres critères de qualité qui jouent un rôle beaucoup plus important, notamment le fait que (i) la qualité de la donnée soit stable dans le temps, (ii) que celle-ci soit disponible dans des délais connus, (iii) que les changements de méthode soient portés à la connaissance des utilisateurs.

Enfin, il est souvent envisagé d'améliorer la qualité des systèmes statistiques, mais pour autant selon la dimension de la qualité que l'on souhaite modifier, les délais et les coûts ne sont pas identiques. Si certains facteurs de qualité peuvent faire l'objet d'améliorations relativement rapides pour des coûts marginaux, d'autres notamment dès que l'on touche à l'augmentation de la fréquence de collecte et de la précision, peuvent vite être rédhibitoires au regard des bénéfices engendrés.

Partie 1

Les performances

Cette partie constitue un rappel des fondements de la mesure des performances aussi bien dans les pays développés que dans les pays bénéficiaires de l'aide. Elle développe aussi quelques exemples de dérives possibles.

Mots clefs: efficience, efficacité, responsabilité, cadre de mesure des performances, Objectifs du Millénaire pour le Développement.

1.1 POURQUOI MESURE T-ON LES PERFORMANCES?

La plupart des grands pays industrialisés ont profondément réformé leur procédure budgétaire au cours des dernières années. Ces réformes ont généralement visé à mettre en place des budgets axés non plus seulement sur les moyens, mais sur des objectifs qu'il s'agit d'évaluer grâce à des indicateurs de résultats pertinents.

La mesure des performances met l'accent sur les résultats des politiques publiques. Trois raisons principales expliquent pourquoi cette mesure est aujourd'hui nécessaire pour une bonne gestion publique : (i) l'optimisation de ressources limitées; (ii) une amélioration des processus de décision et du fonctionnement des administrations; (iii) un objectif de démocratisation et de transparence.

L'optimisation de ressources limitées

Dans le secteur public, tant pour les États que les organisations internationales, les ressources sont limitées, et souvent insuffisantes pour couvrir toutes les demandes. La quasi-totalité des pays ayant entamé un processus de réforme de l'État ont poursuivi, parmi leurs objectifs, l'amélioration de la situation de leurs finances publiques.

La méthode « traditionnelle », c'est-à-dire antérieure aux années 1970, pour construire un budget reposait sur une allocation des moyens en fonction des priorités. L'amélioration des résultats d'une politique publique passait, selon les analyses de l'époque, par l'augmentation des moyens alloués. Cette méthode était pragmatique dans un environnement où la collecte et la diffusion de l'information étaient moins aisées qu'aujourd'hui, et surtout trop lentes pour que les résultats atteints par l'action publique soient utilisés pour préparer l'exercice budgétaire suivant. Cette idée que les résultats croissent avec les moyens est restée ancrée dans l'esprit du public, particulièrement sensible, par exemple, aux réductions budgétaires dans les domaines de la santé ou de l'éducation.

Pourtant, l'expérience des gestionnaires publics et l'exemple des méthodes de contrôle de gestion appliquées par les entreprises ont montré que l'augmentation des moyens ne conduit pas toujours à une augmentation des résultats et, si même elle y conduit, que la croissance des résultats n'est pas proportionnelle à la croissance des moyens.

Avec des ressources limitées, les gestionnaires publics doivent dépenser mieux, c'est-à-dire obtenir de meilleurs résultats pour un même budget. C'est la notion de rendement ou **d'efficience** (de l'anglais *efficiency*¹, l'une des trois caractéristiques de la dépense publique recherchée au Royaume Uni depuis le *National Audit Act* de 1983). L'efficience intéresse directement les contribuables et les bailleurs de fonds, dans le cas des programmes d'aide. Le rendement, ou l'efficience, est de même nature que la productivité.

Une autre caractéristique recherchée est **l'efficacité** (*effectiveness*²): les objectifs pour lesquels les ressources ont été allouées sont-ils atteints? La mesure des résultats est le meilleur moyen de répondre à cette question. L'efficacité intéresse les citoyens, bénéficiaires potentiels des politiques publiques.

Une amélioration des processus de décision et du fonctionnement des administrations

Tout comme dans les dispositifs de contrôle de gestion, la connaissance des résultats atteints par l'action de l'administration permet d'éclairer les décisions. La mesure des performances met en lumière les points forts et les points faibles, caractérise les progrès dans le temps, aide à comparer les modalités d'action et à identifier les modalités les plus efficaces.

¹ Efficiency : efficience, ou rendement ; c'est le rapport entre les ressources consommées (inputs) et les résultats (outputs).

² Effectiveness : efficacité ; c'est le fait de produire les résultats attendus de la politique publique.

La connaissance des résultats atteints peut aussi avoir une influence sur les agents de l'administration eux-mêmes. La mesure de leurs performances peut les inciter à s'améliorer. Certains auteurs estiment que la mise en place d'un dispositif de mesure des performances est un facteur autonome de modernisation et de changement culturel de l'administration.

Un objectif de démocratisation et de transparence

La nécessité de rendre des comptes à la population, ou aux bailleurs de fonds dans le cas des programmes d'aide, est de plus en plus forte. L'accès à une information rapide et permanente alimente les attentes du public (citoyens, bénéficiaires des politiques publiques, contribuables). L'information disponible permet aussi des comparaisons. Les aspirations démocratiques qui animent la planète, notamment depuis les années 1980, donnent à cette exigence de **responsabilité**³ (*accountability*) une actualité nouvelle.

Les difficultés et les effets pervers

La déclinaison de notions issues de l'industrie dans un environnement de gestion publique présente des difficultés intrinsèques. Alors que les entreprises sont surtout sensibles aux réalisations (production, ou *outputs*), les services publics doivent se concentrer sur les résultats (*outcomes*). Mais « le choix des objectifs dans le secteur public est plus difficile que dans les entreprises. Les réalisations et les résultats sont difficiles à indiquer de manière compréhensibles à ceux qui sont chargés de les obtenir, à ceux qui les dirigent et à ceux qui bénéficient des résultats » (Perry 2006)⁴.

EXEMPLE 1: LES EFFETS PERVERS

La mise en avant de résultats attendus peut conduire les services à des comportements inopportuns au regard des objectifs. Un auteur a étudié les effets d'une démarche de performance dans les écoles primaires d'Amsterdam. La municipalité a constaté que ses écoles avaient des résultats inférieurs aux moyennes nationales dans les tests passés par leurs élèves. Elle a donc décidé d'imposer comme cible une progression des résultats jusqu'aux moyennes nationales, en contractualisant des sanctions en cas d'échec. Les conséquences de cette démarche de performance ont oscillé entre une modification des cours et une tricherie organisée.

Les résultats ont progressé plus rapidement que prévu. Pour satisfaire aux exigences des tests, les écoles ont modifié le contenu des cours, afin d'augmenter l'enseignement des savoirs de base (lecture, écriture, calcul), très sollicités par les tests. Ce choix a répondu aux objectifs recherchés; la question de la place des savoirs de base dans l'enseignement primaire est bien du ressort des objectifs de la politique de l'éducation, même si tous les acteurs n'approuvent pas le choix qui a été fait. Mais le succès dans l'atteinte des cibles a aussi résulté d'autres pratiques moins honorables.

Les effets pervers constatés selon l'étude étaient les suivants : (i) des écoles ont mis en place des séances d'entraînement automatisé au passage des tests, ce qui améliore les résultats, mais pas nécessairement les connaissances des élèves (les opinions des spécialistes étaient partagées sur ce point); (ii) les élèves les plus faibles n'étaient pas présentés aux tests, ou bien leurs résultats n'étaient pas enregistrés; (iii) des surveillants utilisaient des signaux convenus pour indiquer les difficultés aux élèves lors des tests. Cet exemple montre que le choix des objectifs, et des indicateurs associés, est déterminant, mais insuffisant à lui seul pour mesurer la qualité des résultats des politiques publiques.

³ Le terme canadien de « redevabilité » est parfois employé comme traduction de « accountability », car la « responsabilité », c'est-à-dire le devoir de rendre des comptes, de répondre de son action, qui en est l'exacte traduction, est souvent aujourd'hui employée avec le sens, différent, de « pouvoir ».

⁴ Testimony of Robert M. Tobias, Director Public Sector Executive Education American University before The House Subcommittee on Federal Workforce, Postal Service and the District of Columbia. On Personnel Reform in the Federal Government February 12, 2008 ; <http://oversight.house.gov/images/stories/documents/20080214120014.pdf>.

1.2 LA MESURE DES PERFORMANCES DANS LES POLITIQUES DE COOPERATION

En plus des raisons pratiques présentées au chapitre précédent, et qui s'appliquent aussi bien aux États industrialisés (cf. exemple 1) qu'aux Pays en Développement (PED), d'autres motifs imposent la mesure des performances dans les programmes d'aide.

Les matrices d'indicateurs (*Performance Assessment Frameworks – PAF*) sont l'une des réponses à ce besoin.

La demande interne aux pays

Les pays émergents sont confrontés à des problèmes très comparables à ceux des États industrialisés : accroissement du rôle de l'État (infrastructures et santé, par exemple), lourdeur des investissements, ressources limitées, et développement de l'exigence de transparence. Pour toutes ces raisons, et parce que les expériences de démarche de performance sont désormais bien engagées, ces pays émergents peuvent, soit au niveau du gouvernement central, soit dans des régions ou des communes, appliquer les mêmes méthodes.

Dans les pays les moins développés, si les constats et les besoins sont identiques, les démarches de performances sont souvent plus difficiles à intégrer dans les politiques publiques. La raison principale est la qualité du système d'information (disponibilité, recueil, traitement et diffusion des informations, tant budgétaires que sociales).

Dans tous les cas où les pays sont bénéficiaires d'une aide au développement, la démarche de performance est non seulement souhaitable, mais aussi impérative en raison des engagements internationaux.

Les obligations

Des accords internationaux, signés par les bénéficiaires et les bailleurs de fonds, imposent une gestion axée sur les résultats, et donc, la mesure des performances dans les programmes d'aide.

Les **Objectifs du Millénaire pour le Développement** (OMD) du 8 septembre 2000, inscrivent toute l'aide dans une démarche de performance. Les OMD insistent sur les résultats attendus de l'aide, soit sous la forme d'objectifs généraux, soit directement sous la forme d'indicateurs affectés de cibles.

- Exemple d'objectif général : formuler et appliquer des stratégies qui donnent aux jeunes partout dans le monde une chance réelle de trouver un travail décent et utile.
- Exemple d'objectif dont la cible est imposée par les OMD: réduire de trois quarts, d'ici à 2015, la mortalité maternelle et de deux tiers la mortalité des enfants de moins de 5 ans par rapport aux taux actuels.

La **Déclaration de Paris** (2005) impose comme modalité de délivrance de l'aide une gestion axée sur les résultats: « axer la gestion sur les résultats signifie gérer et mettre en œuvre l'aide en se concentrant sur les résultats souhaités et en utilisant les données disponibles en vue d'améliorer le processus de décision. »

Ces deux textes ne sont pas les seuls (voir aussi la *Deuxième Table ronde internationale sur « la gestion axée sur les résultats »* tenue à Marrakech les 4 et 5 février 2004). Si ces textes ont une valeur variable, et à proprement parler faible, dans la hiérarchie des normes des États et de l'Union Européenne (UE), leur force morale n'est pas discutée; elle est même réaffirmée par le **Programme d'Action d'Accra** adopté en septembre 2008.

Le **consensus européen pour le développement** est un engagement unilatéral de l'UE (Parlement, Conseil, Commission):

- « Le consensus précise les objectifs et les principes communs pour la coopération au développement. Il réaffirme l'engagement de l'UE à éradiquer la pauvreté, à garantir l'appropriation et le partenariat, à fournir une aide accrue et de meilleure qualité, et à promouvoir la cohérence des politiques pour le développement. Il guidera, dans un esprit de complémentarité, les activités de la Communauté et des États membres dans le domaine de la coopération au développement dans l'ensemble des pays en développement. [...]
- 95. Les indicateurs de performance liés aux OMD seront renforcés de manière à corrélérer davantage le soutien sectoriel et budgétaire aux progrès dans la réalisation des OMD et à garantir un financement suffisant pour la santé et l'éducation.
- 115. La Communauté utilisera constamment une approche centrée sur les résultats et basée sur des indicateurs de performance. De plus en plus, les conditions posées évolueront vers un concept de « contrat » reposant sur des engagements mutuels négociés et formulés en termes de résultats. »

Les matrices d'indicateurs (Performance Assessment Frameworks – PAF)

Qu'ils agissent spontanément ou par l'effet de leurs engagements internationaux, les pays bénéficiaires de l'aide doivent, dans une situation idéale, fixer leurs propres objectifs, choisir les indicateurs qui permettront de mesurer les progrès, et arrêter les cibles de ces indicateurs. Les indicateurs et les cibles à atteindre par ces indicateurs dans le temps sont attachés aux objectifs. L'ensemble des indicateurs constitue la matrice des indicateurs (PAF).

Les bailleurs de fonds doivent, en application de leurs propres engagements internationaux, s'aligner sur les matrices d'indicateurs des pays aidés. Ainsi, les indicateurs par lesquels les bailleurs de fonds suivent les progrès du pays aidé, et en particulier le sous-ensemble des indicateurs de conditionnalité intervenant dans le calcul des tranches conditionnelles d'appui budgétaire, devraient être élaborés par le pays.

En réalité, la difficile maîtrise des indicateurs et de la démarche de performance dans les pays qui se lancent pour la première fois dans cette approche, donne aux conseils techniques des bailleurs de fonds un poids important dans le choix des matrices. De plus, le caractère asymétrique des engagements réciproques et le besoin de ressources pour les pays bénéficiaires de l'aide, renforcent l'influence des bailleurs de fonds, qui doivent cependant utiliser des matrices (PAF) définies par les bénéficiaires de l'aide, ou, à défaut, dérivées de celles-ci.

Presque tous les pays bénéficiaires de l'aide ont déjà établi, souvent d'ailleurs sous l'impulsion des bailleurs de fonds, une plateforme formelle ou informelle de Dialogue Politique Sectoriel, c'est-à-dire un mécanisme de consultation au niveau national grâce auquel les autorités s'adressent aux principaux intéressés et acteurs du pays pour discuter des effets des politiques suivies dans un secteur donné et définir conjointement un certain nombre de résultats à atteindre. Pour la santé par exemple, sont consultés les organisations professionnelles et les syndicats, les organismes de soins privés, les gouvernements régionaux et les autorités locales, des représentants des patients sous une forme ou l'autre, les instituts scientifiques, etc.

La participation des acteurs non étatiques et du Parlement est souvent un signe de bonne gouvernance et renforce généralement la responsabilité du gouvernement vis-à-vis des besoins et des demandes des citoyens.

Dans les pays où un processus de décentralisation est en cours, il est aussi particulièrement important de s'assurer du concours des gouvernements locaux et de prévoir un renforcement de leurs capacités lorsque cela s'avère nécessaire.

Les bailleurs de fonds participent également au dialogue politique sectoriel lorsqu'il s'agit de fixer des cibles mesurables afin de permettre le financement des stratégies sectorielles sur base de résultats

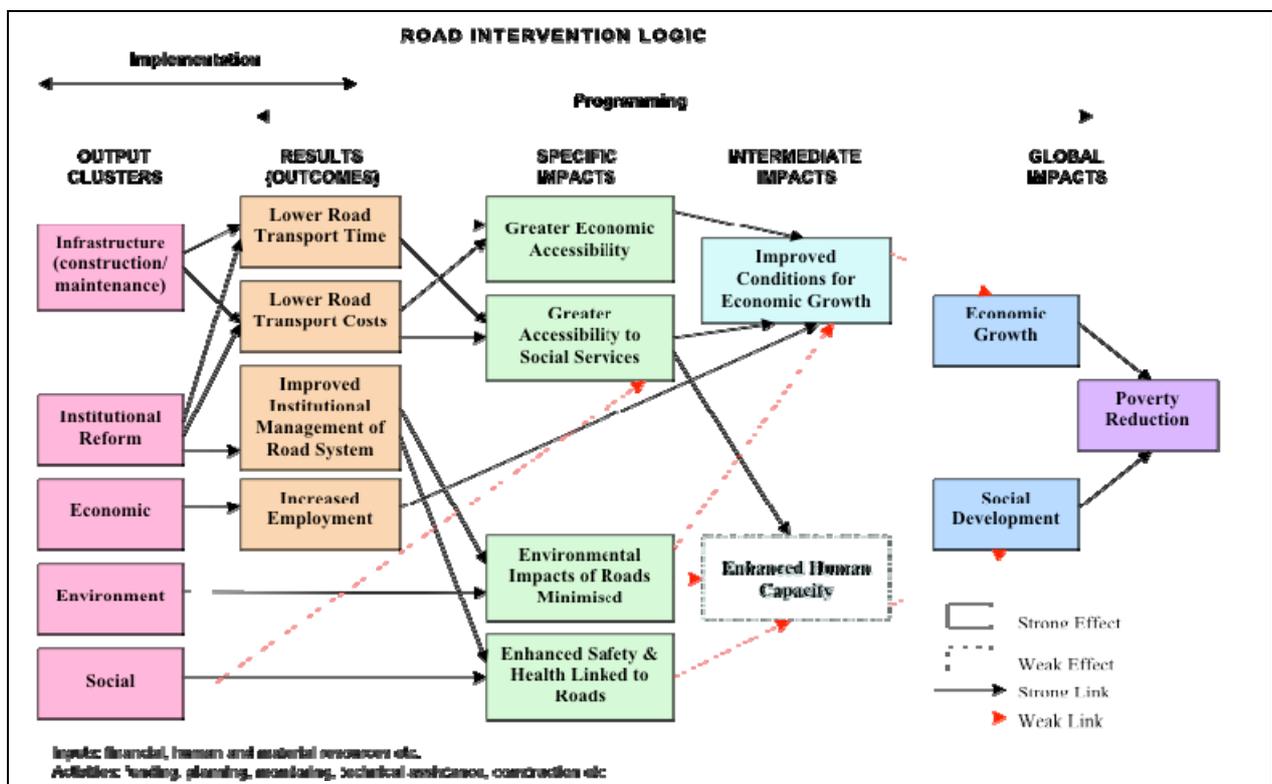
définis conjointement. Dans ce contexte, pour diminuer les coûts additionnels engendrés par le besoin de fournir des justificatifs spécifiques pour chaque demande de soutien adressée aux différents bailleurs de fonds, il est proposé que les bailleurs de fonds utilisent des outils communs pour évaluer les mérites des stratégies sectorielles nationales. Par exemple, l'Organisation Mondiale de la Santé, en collaboration avec le GFATM (*Global Fund to Fight Aid, Tuberculosis and Malaria*) et le GAVI (*Global Alliance for Vaccines and Immunisation*), développe actuellement l'outil "Joint Assessment of National Health Strategies and Plans"; la CE participe à ces travaux.

L'utilité des indicateurs

La démarche de performance n'est pas en priorité une démarche économique. Les services publics produisent des réalisations (*outputs*, ou production dans la définition industrielle), afin d'atteindre des résultats (*outcomes*), mais le but des politiques est d'obtenir un impact sur la société. Si les impacts sont difficiles à mesurer avec des indicateurs, c'est avant tout parce qu'il n'est ni facile, ni souvent utile de les quantifier. En outre, les impacts sont le résultat de nombreuses interactions que l'on ne peut pas rattacher à un unique secteur; l'étude des impacts relève de l'évaluation.

Le choix des objectifs et des indicateurs doit être guidé par les impacts attendus. Il n'y a pas de méthode mécanique pour décliner les impacts attendus en objectifs, indicateurs et cibles; l'arbitraire qui subsistera toujours dans ce choix est l'expression même des choix politiques. Cependant, un raisonnement solidement étayé permet de montrer les liens entre politiques sectorielles et impacts attendus.

L'exemple suivant (source AIDCO) montre les liens entre réalisation, résultats et impacts dans le secteur routier. Le schéma ne prétend pas à l'exhaustivité ni à l'exacte représentation de la réalité, mais présente un modèle destiné à servir de guide à l'analyse et à la formulation des documents de programmation⁵.



⁵ AIDCO a développé des schémas pour 5 secteurs: Education, Santé, Eau et Assainissement, Routes et Développement agricole et rural
http://www.cc.cec/dgintranet/europeaid/activities/evaluation/impact-indicators/index_en.htm

Partie 2

La mesure

Cette partie décrit les points clefs qui vont permettre, à partir de l'expression des politiques, d'identifier les indicateurs qui constitueront la matrice de mesure des performances. Elle indique aussi comment contourner les problèmes liés au manque de qualité des données statistiques dans le cas particulier de pays à système statistique peu développé.

Mots clefs: type d'indicateurs, fonction des indicateurs, représentativité statistique, fréquence de mesure, qualité d'un indicateur, pertinence d'un indicateur, valeur de référence, cible, progression.

2.1 COMMENT ABORDER LE PROBLEME?

Qu'est-ce qu'un indicateur?

Un indicateur est un **ordre de grandeur** (généralement statistique, mais aussi logique) lié de façon naturelle ou arbitraire à la mesure d'activités politiques (au sens large de la gouvernance). Ses qualités premières sont donc d'être à la fois synthétique, communicable et relativement consensuel.

Il est globalement défini par sa **fonction** (ce qu'il mesure), **son moyen d'obtention** (formule et données nécessaires), **sa qualité** (c'est-à-dire sa capacité à être interprété et suivi dans le temps) et **ses limites d'utilisation** (ce qu'il ne mesure pas ou mal).

Les fonctions des indicateurs: des classifications difficiles

La notion d'indicateurs est intimement liée à la démarche de programmation par objectifs. La démarche logique de la programmation par objectifs est la même, qu'il s'agisse de politiques nationales (Cadre Stratégique de Lutte contre la Pauvreté (CSLP) ou autres), de politiques sectorielles, d'un programme ou d'un projet.

La fonction d'un indicateur se définit par ce qu'il mesure par rapport à l'atteinte d'un objectif. On distingue ainsi divers types d'indicateurs, résumés dans le tableau ci-dessous⁶:

Niveau d'objectif	Type d'indicateur	Définition	Exemples
	Indicateurs de ressources ou d'intrants (Input)	Fournissement des informations sur les moyens financiers, humains, matériels, organisationnels ou réglementaires nécessaires pour la mise en place des politiques visées.	Budget annuel absorbé ; nombre de personnes mobilisées par le programme...
<i>Objectif opérationnel</i>	Indicateurs de réalisation immédiate (Output)	S'expriment en termes de réalisation, c'est-à-dire de produits ou de services générés.	Km de routes construits ; nombre de personnes ayant suivi une formation avec succès...
<i>Objectif spécifique</i>	Indicateurs de résultat (Outcome)	S'expriment en termes d'effets directs et à court-terme sur les destinataires ou les bénéficiaires.	Temps gagné par les usagers d'une route; nouvelles pratiques mises en œuvre par le personnel formé, degré de satisfaction des entreprises ayant reçu un service de consultance.

⁶ Les définitions des différents types d'indicateurs données ci-dessus sont conformes avec les définitions du "Glossaire des principaux termes relatifs à l'évaluation et la gestion axée sur les résultats" édité par le CAD (Comité d'Aide au Développement) ; il existe cependant des différences entre le CAD, la CE et ce guide pour ce qui concerne la traduction des termes 'output' et 'outcome'
http://www.oecd.org/document/21/0,3343,en_2649_34435_40381973_1_1_1_1,00.html

<i>Objectif général</i>	Indicateur d'impact (Impact)	S'expriment en termes de conséquences au-delà des limites de l'intervention (publique ou du programme) et de ses interactions avec les destinataires ou le bénéficiaires. Ceci inclut aussi les conséquences négatives ou non programmées.	Taux de placement des stagiaires après 12 mois; taux de survie des entreprises créées grâce au soutien du programme; nombre d'emplois perdus après l'interdiction de vente d'un produit.
	Indicateurs de contexte	S'appliquent à l'entièreté d'un territoire ou d'une catégorie de population, sans distinction faite entre ceux qui ont été touchés par le programme et ceux qui ne l'ont pas été.	Nombre d'emplois dans le secteur du tourisme; nombre de connexions à Internet dans la région X; taux de chômage dans la région Y...

Cette classification renvoie toutefois à la logique qui lie les niveaux d'objectifs à leur intégration dans une matrice bidimensionnelle. Le fait de souvent présenter les tableaux d'indicateurs en ligne induit explicitement la logique horizontale suivante:

Input ⇒ output ⇒ outcome ⇒ impact

Cela suppose bien sûr qu'il existe un modèle qui lie ensemble logiquement les arguments:

Budget d'investissement ⇒ construction de classe ⇒ augmentation du TBS

⇒ augmentation du taux d'alphabétisation

Or, si dans quelques cas, on peut sans trop de problème utiliser des modèles simplifiés comme ci-dessus, en règle générale, le résultat et l'impact ne peuvent pas être réduits aux seuls facteurs d'intrants et d'extrants signifiés, mais sont la conséquence d'un ensemble de causalités qui est souvent mal ou incomplètement défini dans les documents de politique ou de stratégie. Dans de nombreux cas, les modèles sont beaucoup plus complexes (Cf. schéma de causalité sur le secteur routier en chapitre 1).

De nombreux cadres de mesure comportent dans le même tableau un ensemble d'indicateurs qui ne répondent pas aux mêmes niveaux de mesure (réalisations immédiates, résultats, impacts) et qui par conséquent, ne se mesurent pas avec la même périodicité. L'exercice est d'autant plus périlleux que les indicateurs utilisés ne sont pas de qualité équivalente (lorsqu'ils sont mesurés), n'ont pas le même référentiel temporel (progression par rapport à la valeur d'origine). Le fait par ailleurs, de sélectionner un indicateur par axe stratégique sectoriel induit (même si cet indicateur est particulièrement performant) des distorsions analytiques, voire des dévoiements lorsque l'indicateur est utilisé à des fins décisionnelles (allocation budgétaire ou décaissement dans les programmes d'aide). On peut passer rapidement de chiffres pour la politique à la politique du chiffre⁷. A vouloir tout caser dans une matrice « simple », celle-ci devient un miroir déformant de la réalité complexe d'une situation.

L'évaluation est aussi confrontée à la difficulté de lier les impacts aux facteurs d'intrants. C'est pourquoi l'unité évaluation conjointe RELEX/DEV/AIDCO a développé des lignes directrices pour établir des diagrammes des objectifs et des diagrammes des effets comme moyen pédagogique pour préciser la vision stratégique proposée par les documents officiels et établir la cohérence et la pertinence des diverses contributions ou activités au système des objectifs et des impacts.⁸

Une telle démarche est nécessaire car les évaluateurs sont très souvent confrontés aux déficiences suivantes:

⁷ Y compris en Europe, on glisse rapidement de la mesure du résultat vers la seule mesure de l'indicateur. Ainsi, la réduction de la mortalité routière, via la limitation de la vitesse, n'est plus perçue qu'à travers son indicateur qui est le nombre d'infraction relevée en matière de dépassement : comme on dit « on fait du chiffre ».

⁸ Cf. (http://ec.europa.eu/europeaid/evaluation/methodology/tools/too_obj_fr.htm).

- des documents stratégiques ou des politiques faisant état d'une multitude d'objectifs sans indication suffisante permettant de les hiérarchiser,
- l'absence de formulation d'un objectif global,
- une faible pertinence des liaisons causales entre les objectifs,
- des objectifs d'un rang donné ne renvoyant à aucun objectif situé à un rang supérieur ou sans lien avec des objectifs de rang inférieur.



Il est souvent illusoire de vouloir à tout prix obtenir une classification univoque des indicateurs selon leur type, vis-à-vis d'un cadre donné, il est par contre impératif de définir sa fonction par rapport au secteur auquel il se rapporte : un taux de scolarisation est un indicateur mesurant l'accès au système éducatif, il n'en indiquera pas plus. Qu'il s'agisse de réalisation immédiate ou de résultat est finalement de moindre importance et dépend grandement des politiques utilisées pour le faire augmenter et surtout du niveau de programmation auquel il se rapporte.



Dans une matrice donnée, la juxtaposition de divers type d'indicateurs, surtout s'ils ne sont pas classés par fonction introduit de nombreuses distorsions analytiques. Il vaut mieux bâtir une matrice par niveau d'objectif qui sera beaucoup plus lisible et que l'on pourra facilement relier à la démarche logique.

Les moyens d'obtention : les notions de sources

Lorsqu'il s'agit de définir un indicateur, trois éléments doivent être pris en considération: (i) les variables statistiques nécessaires à son calcul; (ii) le niveau de représentativité; (iii) la fréquence de calcul.

Un grand nombre d'indicateurs ont d'ores et déjà été définis de manière très précise, notamment par les Nations Unies, en particulier pour les domaines sectoriels comme la santé, l'éducation ou l'agriculture. Il s'agit cependant de définitions théoriques. La difficulté, nous le verrons largement par la suite, est d'identifier la source de chaque variable et surtout de vérifier si la variable fournie par cette source correspond effectivement à celle de la définition.

Concernant le niveau de représentativité, il existe encore une grande différence entre le possible et le faisable. Il est, en effet, dans de nombreux cas, possible de désagréger des données soit thématiquement (sexe, genre, catégories socioprofessionnelles), soit géographiquement (provinces, départements, ..). Cependant, la précision des sources fait que cette désagrégation relève parfois de la « gaffe statistique » (Cf. exemple 16) ou de la malhonnêteté intellectuelle.

La fréquence de calcul peut se définir de plusieurs manières, il peut s'agir:

- de la fréquence optimale entre deux calculs pour que cette différence soit statistiquement significative;
- de la fréquence entre deux collectes des données qui servent à bâtir l'indicateur;
- de la fréquence de revue des politiques que l'indicateur est censé suivre.

La seconde définition est plus ou moins incontournable, sauf dans les cas où il existe un modèle permettant de faire des prévisions ou des projections, ce qui est très rare dans de nombreux PED. Toute chose étant égale par ailleurs, les deux autres définitions sont liées. Selon le concept même des statistiques publiques, c'est la fréquence de revue (ou de décision) qui doit guider la méthodologie pour rendre significative la différence statistique.

En pratique, les fréquences, lorsqu'elles ont été mentionnées, résultent de cotes mal taillées entre ces trois définitions (le politique souhaitera des chiffres tous les ans, mais on ne fait des enquêtes que tous les 10 ans, donc en mettant 5 ans on a une chance d'avoir une donnée pour renseigner l'indicateur). Il

s'agit à l'heure actuelle de l'un des principaux problèmes lors des revues périodiques annuelles, qu'elles soient sectorielles ou générales.



Les moyens d'obtention sont en général définis par la notion de source, mais celle-ci est dans de nombreuses matrices définie de façon relativement vague : un ministère sectoriel ne constitue pas une source de donnée ; la source de donnée est l'enquête ou la collecte réalisée par un service responsable au sein de ce ministère.

Qualité d'un indicateur : de nombreux critères complexes

De nombreux critères ont été énumérés pour qualifier les indicateurs:

- Spécifique, mesurable, atteignable/ambitieux, réaliste, temporel (connu sous l'acronyme de SMART);
- Clair, pertinent, économique, adéquat, pouvant être suivi et validé (connu sous l'acronyme de CREAM);
- EUROSTAT⁹ (logique, pertinence, possibilité de fixer un objectif, fréquence des données, appropriation, possibilité d'estimer une précision).

Comme toute définition théorique, la mise en pratique de tels éléments de jugement se heurte souvent aux problèmes des critères qui peuvent être utilisés pour juger de la qualité:

- le critère d'indépendance aux événements, par exemple, est extrêmement complexe à analyser car il suppose, dans de nombreux cas, de pouvoir isoler du contexte environnemental (qu'il soit physique ou économique) l'indicateur retenu. Cela reviendrait à dire qu'il existe des modèles de simulation micro ou macro-économiques permettant de décomposer toute variation en facteurs internes et externes (susceptibles d'être affectés de poids relatifs). Or ces modèles n'existent pas dans de nombreux domaines;
- la pertinence met en jeu deux types de critères, ceux relatifs à la capacité de l'indicateur à répondre à la mesure voulue, mais aussi à des critères de type coûts/bénéfices. L'indicateur qui répond le mieux est parfois hors d'atteinte en matière de coûts ou le gain en termes de qualité de l'indicateur le plus pertinent est minime par rapport à l'utilisation qui en est fait. C'est en général pour ce type de raisons que l'on choisit des indicateurs proxy¹⁰.

Dans la pratique, il semble que le critère principal qui a été utilisé, notamment dans le cadre de nombreux CSLP, a été finalement le critère de faisabilité (les données existent ou ont déjà été collectées) et dans une moindre mesure, la mesurabilité (il existe des méthodes qui permettent d'accéder aux données). Cette position, certes pragmatique, amène à se poser la question suivante : lors d'un exercice de suivi d'un cadre politique, comment concilier le jeu théorique d'indicateurs qui répondent au mieux aux critères de qualité ci-dessus énumérés, avec le jeu pratique, c'est-à-dire ceux qui sont finalement disponibles au moment de la réalisation de l'exercice?

Il convient aussi de considérer:

- le fait de garder des indicateurs en grande partie théoriques ne permet pas de fixer une cible aux politiques, mais permet de garder en mémoire ce qui est finalement visé.
- le fait de n'utiliser que des indicateurs faisables (voire de nombreux proxys) permet de fixer une cible, mais qui ne correspond pas toujours exactement à l'objectif visé.

⁹ Statistical Office of the European Commission

¹⁰ Un indicateur est dit « proxy » lorsqu'il mesure un ordre de grandeur proche ou très fortement corrélé avec celui qui est finalement visé.

On touche ici à une grande partie du problème qui est finalement posé par la systématisation du recours à des indicateurs de suivi, qu'ils soient statistiques ou autres : ils induisent des réformes en profondeur du SSN (mais aussi des finances publiques, voire du fonctionnement même des organes gouvernementaux), ce qui ne peut se faire à l'échelle de mesure des revues. Abandonner le « bon » indicateur, c'est finalement renoncer à peser sur la transparence des résultats. En contrepartie, le garder revient à accepter de se satisfaire de la "case vide" pendant le laps de temps nécessaire à sa production.



On retiendra qu'un indicateur devrait être:

- ☛ *logique, c'est-à-dire être rattaché facilement et de façon univoque à une politique;*
- ☛ *pertinent, c'est-à-dire correspondre à un réel besoin de suivi des politiques;*
- ☛ *apte à être suivi dans le temps, c'est-à-dire mesuré de façon régulière selon une méthodologie stable;*

A ces dimensions, il conviendrait d'ajouter celle de sa capacité à se voir attribué des cibles, c'est-à-dire à effectuer des projections dans le futur¹¹.

Choisir des indicateurs: les points clefs de la démarche

La démarche qui va suivre s'efforce de définir un certain nombre d'étapes dans l'identification des indicateurs. Elle est présentée sous forme séquentielle, mais peut être aussi considérée comme une série de questions auxquelles on s'efforcera de trouver des réponses. L'important est de retenir qu'il existe trois phases:

- l'une qui est purement analytique et vise à poser les questions en relation avec les politiques mises en œuvre (les statistiques n'ont rien à voir là-dedans);
- l'autre qui est plus pragmatique et pratique, et vise à retenir les indicateurs en lien avec les moyens d'obtention;
- la troisième vise à normaliser la présentation des indicateurs retenus à travers un système de fiches signalétiques.

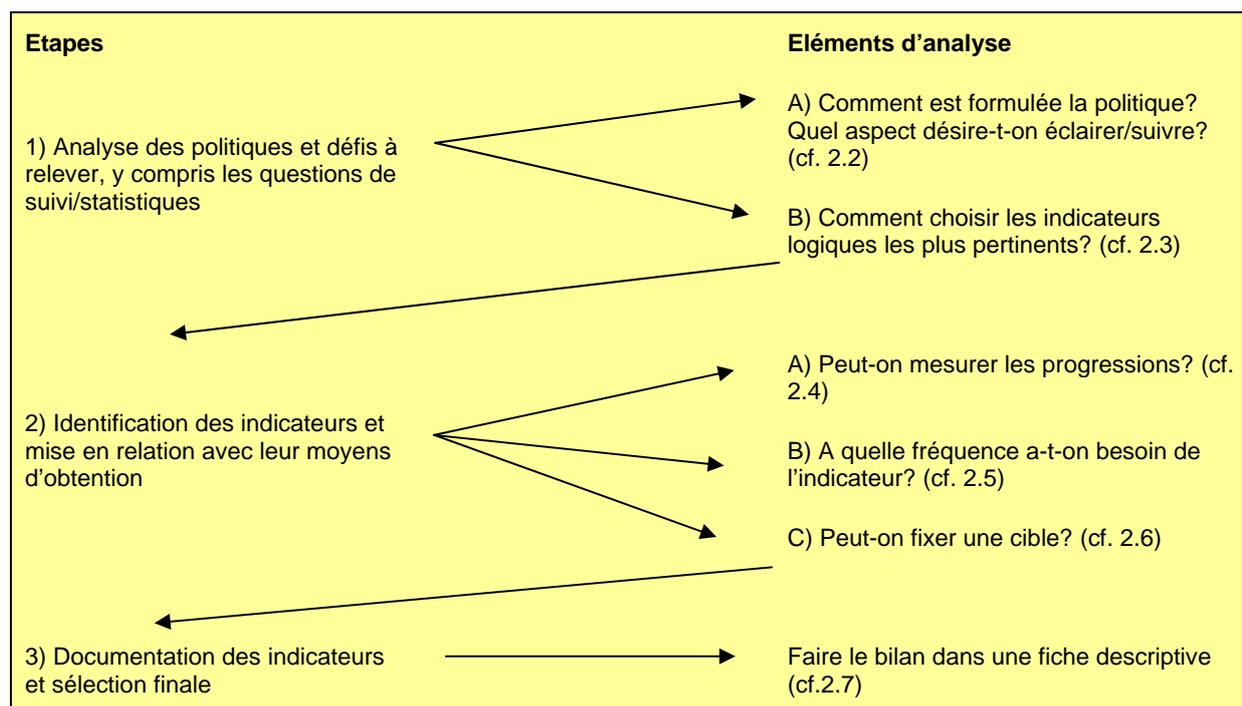
La première phase conduit à identifier les défis de la politique en question qui méritent une attention particulière et qui devraient être suivis; ce sont les questions clefs autour desquelles nous devons orienter notre dialogue avec les autorités.

La seconde étape conduit à identifier le ou les indicateurs et à les mettre en relation avec les défis identifiés au préalable; à savoir s'il est possible de comparer chaque indicateur à la fréquence voulue, si on peut fixer des cibles réalistes. Cette phase aboutit à une première sélection des indicateurs, qu'ils soient immédiatement disponibles ou qu'ils doivent faire l'objet d'affinements.

La troisième étape conduit à l'élaboration d'une fiche qui reprend les analyses des étapes précédentes : question posée, définition mathématique de l'indicateur, sources, restrictions d'utilisation et indications sur l'interprétation et la sélection définitive des indicateurs à retenir.

¹¹ A ce titre, la définition donnée par Eurostat est la plus opérationnelle.

La démarche peut être synthétisée de la manière suivante:



Les points sensibles

Une précision s'impose et concerne la démarche générale d'utilisation des statistiques dans de nombreux PED. Contrairement aux pays du Nord, dans la plupart des cas les statistiques ne préexistent pas à la décision politique, ou du moins lorsque la décision est prise l'état des données disponibles est souvent trop lacunaire pour être utilisées réellement à des fins décisionnelles. L'introduction des indicateurs (OMD, CSLP, et autres matrices) n'a pas fait exception à cette règle et c'est maintenant avec la mise (ou remise) en place des outils de mesure que l'on s'aperçoit que certains objectifs assignés n'étaient pas réalistes ou que la mesure est finalement hors de portée des systèmes statistiques.

Il existe aussi une certaine confusion sur la quantité d'indicateurs à retenir, avec l'arrière pensée que le fait de limiter le nombre total d'indicateurs peut permettre de limiter le coût d'obtention, ce qui est partiellement faux. Le coût dépend essentiellement de la source : une même enquête peut fournir à elle seule tous les indicateurs pour un secteur donné; restreindre la revue à un seul individu n'aura que peu d'influence sur le coût final. Par contre l'augmentation de la fréquence de calcul a beaucoup plus d'impact financier.

A priori, le nombre d'indicateurs doit répondre à un souci de lisibilité des politiques. La restriction permet, d'une part, une plus grande facilité sur l'obtention d'un consensus (le fait de décider *ex ante* d'un nombre d'indicateurs évite en partie l'élimination *ex post* des indicateurs « dérangeants »); et, d'autre part, une plus grande facilité de partage de l'information de suivi. Cependant le nombre total d'indicateurs pour un domaine considéré devrait être mis en relation avec la finesse du suivi et donc, avec le poids politique des interventions retenues (Cf. exemple 3).



Le nombre total d'indicateurs a parfois joué un rôle particulièrement négatif dans les exercices de construction de matrices, en permettant d'éliminer au final les indicateurs dérangeants, ou en les remplaçant par des indicateurs qui ne remplissent pas les mêmes fonctions, ou ne mesurent pas ce qui est souhaité.

2.2 ETAPE 1A – UNE BONNE ANALYSE DES POLITIQUES VISEES

Comment la politique est-elle formulée?

Cette réflexion ne concerne pas directement les statistiques, elle constitue l'expression des besoins de suivi vis-à-vis du cadre politique. Les questions globales posées sont les suivantes:

Quels sont les objectifs poursuivis et les principaux enjeux de la politique à appuyer? Quelle est l'information dont on souhaite disposer aux différents niveaux pour assurer son bon suivi?

La réponse qui sera donnée à l'issue de cette étape de la réflexion ne devrait contenir aucun indicateur (au sens statistique du terme), mais une série de questions qui seront posées au système de suivi (qu'il relève des statistiques publiques ou non).

La formulation des objectifs et des résultats, tels qu'ils sont exprimés dans les documents politiques, doit être examinée avec attention, car elle a une influence sur la nature des indicateurs qui seront utilisés par la suite:

- réduire de moitié le nombre moyen de pauvres, ou réduire de moitié la fraction de la population la plus pauvre ou réduire de moitié le nombre de personnes vivant avec moins de 1 \$/jour sont des notions voisines mais non identiques qui ne se traduiront pas nécessairement par le même indicateur;
- de même, « accroître l'accessibilité et la qualité de l'enseignement primaire », indique déjà clairement sur quels facteurs vont devoir peser les efforts, beaucoup plus clairement « qu'avoir un meilleur enseignement primaire ».

A ce stade, la mesurabilité de la question étudiée ne devrait pas intervenir. Ce qui peut par contre être fait lors de cette étape, est d'exprimer le besoin de suivi sous forme littéraire : "fournir tous les deux ans des indicateurs sur l'accessibilité et la qualité de l'enseignement par district d'éducation" est à ce stade une formulation comprise par tous.

EXEMPLE 2 : DES INDICATEURS BIEN CLASSES

La matrice du secteur éducation en Tanzanie met en relation la politique et les indicateurs retenus telle que présentée ci-dessous.

Politique	Nom de l'indicateur
Accroître l'accès, la qualité et l'équité du système éducatif	Percentage of cohort completing Standard VII*
	Percentage of cohort completing Ordinary - Level Secondary Education (Form IV)*
	Percentage of cohort completing Advanced - Level Secondary Education (Form VI)*
	Transition Rate from Standard VII to Form I Gender Disaggregated *
	Gross enrolment (Total Enrolment) in Higher Education*
	Qualified Teacher/Pupil Ratio in Primary Schools*
	Qualified Teacher/Pupil Ratio in Secondary Schools*

Mais elle ne relie pas de façon claire les indicateurs avec les objectifs politiques. Un simple tri permet de visualiser ce qui est réellement mesuré.

Politique	Nom de l'indicateur
Accroître l'accès du système éducatif	Gross enrolment (Total Enrolment) in Higher Education*

Accroître la qualité et la couverture	Percentage of cohort completing Standard VII* Percentage of cohort completing Ordinary - Level Secondary Education (Form IV Percentage of cohort completing Advanced - Level Secondary Education (Form VI Transition Rate from Standard VII to Form I Gender Disaggregated *
Accroître l'efficacité, la qualité de service, l'équité et le financement	Qualified Teacher/Pupil Ratio in Primary Schools Qualified Teacher/Pupil Ratio in Secondary Schools*



L'identification d'un indicateur ou d'un groupe d'indicateurs dépend avant tout de la clarté et de la précision de la formulation des politiques. Faute de pouvoir peser sur ces formulations, il convient de traduire, au moins, l'expression du besoin de suivi en termes de grandes fonctions des indicateurs.

Le corolaire est que, si les politiques proposées ne permettent pas d'identifier des questions pertinentes en matière de suivi, il n'est pas nécessaire de définir une batterie d'indicateurs les concernant : un indicateur répond à un problème donné et à une politique définie en réponse à ce problème, dans un contexte particulier et au regard d'autres priorités "concurrentes".



Les matrices complexes, c'est-à-dire mettant en jeu toute une batterie d'indicateurs, des intrants aux impacts, ne sont pas toujours les plus efficaces par la suite en termes de suivi; d'une part, parce que l'analyse qui sera faite de ces différents indicateurs ne le sera pas par les mêmes personnes, et, d'autre part, car elles supposent des liens logiques qui sont parfois inexistantes ou des termes trop différents les uns des autres.



Lorsqu'on est amené à utiliser une matrice d'indicateurs déjà réalisée, il n'est pas inutile de mettre côte à côte les objectifs/défis sectoriels que l'on veut suivre et l'indicateur choisi pour ce faire. Ainsi, dans le cas des matrices complexes multi sectorielles, il peut être intéressant de décomposer et de reclasser les indicateurs (voir exemple 2).

Dans quel but souhaite-t-on mesurer des performances?

Il existe autant de jeux d'indicateurs que de matrices de mesure, et la sélection des premiers repose en grande partie sur une bonne appréhension du rôle que va jouer la matrice d'indicateurs dans le pilotage des politiques.

A défaut de dresser un panorama complet, on peut toutefois déterminer quelques grandes familles de matrices:

- les matrices « incitatives » qui résultent par exemple des accords internationaux (OMD ou accords de Kyoto, par exemple). La caractéristique essentielle de ces cadres est qu'ils n'infèrent en rien sur les politiques qui vont être suivies pour arriver aux résultats escomptés, et que chaque Etat est libre de choisir les leviers économiques qui vont permettre de les atteindre.
- les matrices « contractuelles » qui traduisent un engagement, résultant d'un contrat social (comme le mandat électif) ou d'un contrat tout court (comme les matrices associées à l'aide des partenaires techniques et financiers)¹². La caractéristique de ces cadres est qu'il existe une forme d'accord sur les stratégies à mettre en œuvre pour atteindre le résultat.

¹² Les secondes n'étant finalement que les conséquences des premières.

- les matrices que l'on pourrait qualifier d'analytiques ou d'observation, qui sont des outils techniques (comme les exercices PEFA¹³, *Public Expenditure and Financial Accountability*). Elles sont utilisées notamment pour le pilotage des politiques. Les résultats immédiats obtenus servent à orienter les décisions.
- les matrices descriptives, destinée à fournir une image synthétique d'un processus ou d'un contexte à un moment donné.

Le type de matrice de mesure a des conséquences importantes sur les indicateurs: un même indicateur peut être utilisé dans les différentes formes de matrices, mais on ne l'observera pas de la même façon: l'incertitude, qui peut être admise dans un cas, ne le sera peut être pas dans un autre (voir exemple ci-dessous).

EXEMPLE 3: UN INDICATEUR N'A PAS DE FONCTION DEFINIE THEORIQUEMENT

Les indicateurs liés aux OMD sont de fait des indicateurs élémentaires, censés représenter un certain nombre de dimensions de la pauvreté: il s'agit d'un choix relativement arbitraire pour certains d'entre eux qui ont finalement rang de symbole.

Ainsi, le choix de faire porter une part des indicateurs de résultat de la santé sur les taux de vaccination, ne doit pas cacher qu'il s'agit en fait d'indicateurs de réalisation qui fluctuent d'une année à l'autre en fonction justement des performances du système de santé.

Il serait parfaitement légitime d'afficher comme objectif la « vaccination universelle pour tous » à un horizon donné; dès lors, le taux de vaccination devient un indicateur de résultat de ce programme hypothétique. Par contre, il n'est qu'un indicateur de réalisation d'un programme national d'amélioration de la santé, car, d'une part, on ne peut réduire la politique de la santé à la seule prise en compte de la vaccination et que, d'autre part, les moyens pour réaliser cet objectif sont divers et variés. Il s'agit d'un piège dans lequel sont tombés de nombreuses matrices liées aux CSLP, réduisant considérablement la portée analytique de ces cadres.

Quel aspect de la politique désire-t-on éclairer?

Une politique ne peut être réduite à l'observation d'une batterie d'indicateurs et encore moins à quelques-uns. Lorsque cela est fait, c'est en général à des fins purement politiques; comme par exemple, réduire une politique de l'immigration à la seule statistique de « reconduite à la frontière » ou le pouvoir d'achat à "l'indice des prix à la consommation".

Pour autant, si l'on souhaite porter l'attention sur des aspects précis d'une politique, cela aura des répercussions importantes autant sur le choix des indicateurs, que sur celui des différentes méthodes de calcul possible. Si la réduction des disparités géographiques est un objectif à suivre, on ne pourra pas se contenter d'une moyenne nationale: il faudra soit calculer le nombre de régions en dessous de l'objectif, soit calculer des indices de dispersion (voir exemple 4 ci-après).

EXEMPLE 4 : MESURE MATRICIELLE OU MESURE SPOT

Les disparités régionales

Dans de très nombreux pays, il existe de fortes disparités entre les régions. Ces disparités se voient reflétées sur de très nombreux indicateurs et dans de nombreux secteurs. Aussi, il peut être intéressant d'attirer l'attention aussi bien sur la performance globale que sur la réduction de ces inégalités.

¹³ Il s'agit de matrices d'indicateurs permettant d'apprécier le fonctionnement des finances publiques, à travers un système de notation en cascade.

Plusieurs solutions sont possibles. Celle présentée ci dessous est tirée d'un contrat OMD et concerne le taux brut de scolarisation

(i) Taux global (à partir des moyennes nationales)

Valeur de référence		Cibles					
2006	2007	2008	2009	2010	2011	2012	2013
75%	77%	80%	84%	87%	88%	88%	88%

(ii) Disparités régionales (l'écart entre la moyenne arithmétique du TBS des trois régions les moins favorisées et la moyenne nationale du TBS)

Valeur de référence		Cibles					
2006	2007	2008	2009	2010	2011	2012	2013
19,83 pts	20,23 pts	Réduction des écarts					

Mode d'appréciation de l'indicateur

On apprécie *séparément* l'évolution (i) de l'indicateur global et (ii) des disparités régionales. Le score de l'indicateur est ensuite calculé comme moyenne simple des deux scores.

On a donc toujours le choix de l'éclairage sur l'ensemble du champ couvert ou sur un point particulier de la scène, ou sur les deux à la fois en pondérant le poids accordé aux deux mesures (comme dans l'exemple ci-dessus).

Dans de nombreux cadres, la possibilité de pouvoir calculer de nombreuses variables dérivées à partir d'un même jeu de données, n'est que peu utilisée et le traitement des données se réduit souvent à des totaux, soit à des moyennes.



Ne jamais perdre de vue, à ce stade de l'analyse, que l'indicateur à travers sa valeur unique, ne peut transcrire la complexité de certains phénomènes. Ce n'est qu'un outil d'appréciation d'une situation.



Le choix de la question posée a des conséquences aussi bien sur le nombre d'indicateurs qui vont être utilisés que sur la formule de calcul qui sera finalement retenue.

2.3 ETAPE 1B – QUELS SONT LES INDICATEURS LOGIQUES LES PLUS PERTINENTS?

La pertinence est une notion fortement polysémique. Pour nous, la pertinence d'un indicateur se pose avec la formulation de la politique à appuyer, avec ses défis et ses priorités. Il faudra ensuite trancher entre les indicateurs possibles, qui auront des formules de calcul différentes: une valeur absolue, un ratio, ou un indice de ciblage (moyenne nationale, écart à la moyenne). Ceci peut conduire parfois à la multiplication d'indicateurs plus ou moins similaires visant à satisfaire un même besoin et nécessitant de la part des producteurs des efforts supplémentaires. D'où la nécessité de vérifier, autant que faire se peut, dans quels cadres d'analyse l'indicateur peut être utile et utilisé.

Satisfaire le plus grand nombre d'utilisateurs?

Au point précédent, il a été expliqué que la pertinence est liée à la satisfaction des besoins de mesure des divers cadres politiques existants. Cependant, un indicateur peut être pertinent dans le suivi sectoriel et beaucoup moins dans un cadre général. Le cas des taux de vaccination est, à ce titre, relativement exemplaire. On comprend aisément que dans la politique du Ministère de la santé, on retienne ces indicateurs comme résultat immédiat. Ils sont, par contre, tout à fait « hors sujet » dans un cadre global visant à mesurer la performance du système de santé d'un pays à moyen terme (surtout s'ils ont atteints des niveaux optimum > 95%).

Dans une logique d'harmonisation des mécanismes de suivi, on a parfois tendance à rechercher systématiquement les indicateurs communs dans les différents cadres et à retenir ceux qui sont le plus fréquemment utilisés ou cités. Cette démarche montre rapidement des faiblesses, notamment celle de détourner la réflexion sur la fonction des indicateurs: l'indicateur pertinent devient « celui qui est mesuré ».

Ainsi, de nombreux rapports d'étude sur l'identification des indicateurs, mettent en ligne des indicateurs et en colonne l'ensemble des cadres politiques d'utilisation aboutissant à des formes de « *shopping list* » sans logique interne. Certes, le ou les indicateurs retenus vont servir à compléter les cases de nombreuses matrices en circulation dans le pays mais, sont-ils pour autant le résultat d'une phase d'analyse du contexte préalable?

EXEMPLE 5: L'INDICATEUR UNIQUE, ÇA N'EXISTE PAS!

Les indicateurs synthétiques

Le cas des indicateurs synthétique tels que l'Indice de Développement Humain (IDH) mérite d'être mentionné à ce point car ils ont justement vocation à être plus ou moins uniques ou du moins à remplacer en une valeur un groupe d'indicateurs. Leur vocation première est de permettre des comparaisons au niveau international mais, dès lors que l'on s'interroge sur le positionnement d'un pays par rapport à un autre, on est pratiquement obligé de décomposer les éléments. Ainsi l'IDH utilise deux dimensions structurelles: l'espérance de vie et l'alphabétisation, et une dimension conjoncturelle: la croissance exprimée à travers le Produit Intérieur Brut (PIB). Ce type d'indicateur est pratiquement inutilisable en matière de mesure des performances, puisque les performances mesurées ne sont pas affectables à une politique.

Satisfaire le plus grand nombre d'utilisations?

Pour pouvoir à un moment donné analyser la performance d'un secteur ou d'une stratégie, il faut nécessairement disposer d'une batterie d'indicateurs. Le fait de n'en retenir qu'un seul (ou deux) pour suivre un secteur donné ne peut pas "cacher" ou simplifier la réalité. Ainsi, quel que soit le cadre, l'indicateur pertinent est celui que l'on peut analyser, c'est-à-dire qu'il existe, en dehors de cet indicateur, d'autres indicateurs ou informations permettant d'interpréter ses variations.

A l'idée de satisfaction des divers cadres, il faut substituer une logique "verticale" de satisfaction des différents niveaux d'analyse des politiques, allant du local au sectoriel puis au global. La question que l'on doit se poser est la suivante : est ce que les deux ou trois indicateurs finalement retenus au titre de la politique peuvent être expliqués par d'autres? Est ce qu'ils permettent d'expliquer ce qui se passe aux différents niveaux? De quel niveau nous parlent-ils en particulier?

Ainsi, dans l'exemple ci dessous, on voit que le taux d'accès à l'eau potable est utilisé (avec ses différentes désagréations) sur les 3 cadres d'analyse identifiés correspondant aux niveaux décisionnels. Il s'agit donc d'un indicateur particulièrement pertinent.

EXEMPLE 6. UNE BONNE LOGIQUE VERTICALE : LE SECTEUR DE L'EAU AU MALI

Une récente étude menée par la cellule de suivi du CSLP, a défini pour le secteur eau potable les batteries d'indicateurs suivants :

Pour informer le niveau décisionnel (gouvernement et partenaires), 3 indicateurs :

- État d'avancement des projets inscrits au Programme gouvernemental ;
- Nombre de points d'eau réalisés ou réhabilités / objectifs ;
- Evolution du taux global d'accès à l'eau potable.

Pour alimenter le CSLP sur le suivi des OMD (annuel), 6 indicateurs :

- % de villages disposant d'au moins un point d'eau moderne ;
- Nombre de points d'eau modernes pour 400 habitants en milieu rural ;
- Taux d'équipement pour couvrir les besoins en eau en milieu rural ;
- Proportion de la population dans les centres urbains et semi-urbains ayant un accès durable à une source d'eau meilleure ;
- Proportion de la population en milieu rural ayant un accès durable à une source d'eau meilleure ;
- Taux global d'accès à l'eau potable au Mali.

Pour un suivi détaillé au sein du Ministère en charge de l'eau (annuel), 19 indicateurs :

- Indicateurs spécifiques liés à la concession à la société des eaux : nombre d'abonnés, taux d'accès robinet, taux d'accès aux bornes fontaines, entre autres ;
- Indicateurs spécifiques pour les centres urbains et semi urbains : nombre de points d'accès en eau potable, taux d'accès à l'eau potable, taux de fonctionnement, entre autres ;
- Indicateurs spécifiques pour le milieu rural : nombre de points d'eau modernes, taux d'accès à l'eau potable, taux de fonctionnement, entre autres ;
- Indicateurs globaux : taux d'accès à l'eau potable au niveau national, nombre total de points d'eau moderne, pourcentage de villages bénéficiant d'au moins un point d'eau moderne, investissements annuels dans le secteur de l'eau, entre autres.



Dans un cadre de mesure donné, la pertinence d'un indicateur peut se mesurer à travers les éléments suivants:

- sa capacité à être utilisé dans divers cadres d'analyse (matrices nationales, sectorielles, ou autres) pour le suivi des aspects clefs des politiques nationales/sectorielles;
- sa capacité à être interprété, c'est-à-dire l'existence sur le domaine considéré des informations statistiques ou factuelles qui vont permettre le moment donné de fournir des explications ou des précisions sur les valeurs qu'il va prendre;
- sa cohérence, soit avec d'autres cadres existants, soit avec des normes admises au niveau international.



Le fait de garder au niveau de la mesure des politiques globales (politique d'un gouvernement ou d'un organisme) 1 ou 2 indicateurs ne doit pas cacher le fait qu'il s'agit d'une sélection issue d'un processus de type pyramidal qui part des programmes sectoriels pour aller aux politiques sectorielles puis aux politiques globales (voir l'exemple de l'eau au Mali). Il ne s'agit en aucun cas du seul indicateur disponible pour le secteur considéré. Si cela est le cas, on se trouve dans ce que l'on pourrait qualifier d'indicateur de « positionnement politique », introduit dans la matrice non pour mesurer des performances mais pour indiquer que le secteur existe.



Il convient de préciser à ce niveau l'abus qui est fait d'indicateurs dit proxy, dès lors que l'indicateur que l'on souhaite suivre n'est pas mesuré. Un indicateur est dit proxy lorsqu'il mesure une grandeur de façon approximative. Ainsi, le taux de prévalence VIH mesuré sur les travailleuses du sexe est utilisé comme proxy du taux de prévalence dans la population (cette hypothèse étant bien sûr discutable). Le ratio élèves/enseignant est un proxy de qualité de l'enseignement, la qualité se mesurant en fonction des résultats d'apprentissage (mais les résultats des examens ne sont pas toujours disponibles, ni les épreuves homogènes, ce qui pose problème pour le choix de cet indicateur). Par contre prendre le taux de croissance du cheptel comme mesure de l'intensification de l'élevage n'est pas un proxy, car il ne mesure pas une intensification.

2.4 ETAPE 2A – PEUT-ON MESURER LES PROGRESSIONS?

Peut-on comparer des valeurs?

La question de la formulation d'objectifs et de cibles renvoie à des dimensions politiques évidentes: il est toujours possible de faire des effets d'annonce et le technicien ne peut rien vis-à-vis des ambitions politiques difficilement réalisables. Par contre, la fixation de cibles et de valeurs intermédiaires renvoie à deux problèmes purement statistiques: (i) la longueur et la comparabilité des séries disponibles pour un indicateur donné; (ii) la précision de cet indicateur et donc sa capacité à discriminer réellement deux positions consécutives.

La question posée est la suivante: *entre deux valeurs consécutives du même indicateur (c'est-à-dire entre deux revues), les sources sont-elles suffisamment précises pour que la différence constatée soit significative?*

Répondre à cette question permettra de connaître ou d'estimer:

- la tendance générale de croissance ou de décroissance, ce qui permettra de déterminer les fréquences de revues, c'est-à-dire de calcul;
- la portée analytique, c'est-à-dire l'interprétation que l'on fera de la performance constatée.

EXEMPLE 7: LES VALEURS SONT-ELLES COMPARABLES? L'EFFET DE COLLECTE

Les statistiques de criminalité, sont de fait des indicateurs *proxy*, obtenus à partir des affaires jugées par les tribunaux. Entre 2000 et 2006, le taux d'accroissement des affaires criminelles dans un pays est de 84% passant de 599 à 1120 affaires nouvelles par an; cependant dans le même temps le nombre de tribunaux a pratiquement doublé passant de 10 à 19. Faut-il en conclure que la criminalité a augmenté ou que la justice s'est rapprochée du citoyen?

Pour autant, entre 2005 et 2006 plus précisément, le nombre d'affaires criminelles nouvelles a fait un bond de 38%, alors que le nombre de tribunaux n'a augmenté que d'une unité.

Dans la première partie de la série, on ne peut pas conclure à une augmentation de la criminalité, sinon il faudrait prendre comme base de calcul les dix tribunaux de l'année 2000 et comparer avec ces mêmes tribunaux en 2006. Par contre en 2005 et 2006, on peut conclure à une augmentation.

L'influence de la formule de calcul sur la comparabilité

Les indicateurs peuvent être des valeurs absolues, des ratios ou des pourcentages. Cependant ces différents types ne s'interprètent pas de la même façon et le recours systématique à des progressions en %, cache souvent plus de choses qu'il n'en révèle (une augmentation de 100%, c'est-à-dire un doublement, s'applique aussi bien pour des valeurs absolues de 3 que de 3 millions). Ainsi le département français qui détient le record 2008 de réduction du nombre de tués sur la route est la Manche, avec une baisse de 50% ... la valeur absolue étant passée de 10 à 5¹⁴.

D'une manière générale, les cadres de mesures n'utilisent que très peu d'outils statistiques permettant de normaliser les indicateurs pour les rendre plus interprétables. En dehors de déflateurs¹⁵ de population ou de tranches de population utilisées dans différents taux des domaines de la santé et de l'éducation, il est singulier de voir que les autres indicateurs utilisés sont rarement ramenés à la population; pourtant, les croissances (qu'il s'agisse de croissance de production, d'infrastructures ou d'équipement) sont fortement limitées par les taux de croissance de la population (entre 2 et 3,5% par an dans les pays sahéliens, par exemple). Les progressions non déflatées par la croissance

¹⁴ Pour mémoire, on ne devrait pas calculer de pourcentages pour des effectifs inférieurs à 100.

¹⁵ Le terme de déflateur est utilisé au départ dans l'analyse des prix, déflater signifie annuler l'inflation dans l'analyse de l'évolution d'un montant en monnaie. Par extension, rendre le numérateur indépendant de l'évolution du dénominateur.

démographique, sont donc souvent des images déformées de la réalité (Cf. exemple 8 ci-dessous). De la même manière, dans les pays à inflation forte, les indicateurs budgétaires n'ont de sens qu'à monnaie constante (en termes ou à prix réels), c'est-à-dire déflatés *stricto sensu*.

EXEMPLE 8: QU'EST-CE QUI CROIT?

L'effet de population sur la croissance

Les superficies agricoles totales exploitées ont connu une forte progression et sont passées de 2 139 901 ha en 1984 à 4 112 217 ha en 2006. Elles ont presque plus que doublées en l'espace de vingt ans, soit une croissance moyenne par an de 3% des superficies cultivées. Pour la même période, la croissance de la population est de 2,3% par an passant de moins de 7 millions d'habitants à plus de 13 millions. Dans le même temps, la superficie cultivée par individu actif a été pratiquement stable et a oscillé entre 0,57 ha et 0,73 ha avec une moyenne de 0,67 ha.

La production céréalière est passée de 2 450 000 tonnes en 1991 à 3 600 000 en 2006, soit une augmentation de 50% en l'espace de quinze ans pour une population qui est passée de 8 à 13 millions, ce qui correspond à une baisse du disponible par habitant de 0,3 t à 0,27 t.

De la même manière, la valeur ajoutée agricole par tête est passée de 91 620 FCFA en 1993 à 124 297 FCFA en 2005 soit une hausse de 36% en francs courants (en termes nominaux). Cependant, si on considère la dévaluation du FCFA de 1994 et l'inflation, on peut dire qu'elle stagne.

L'analyse des augmentations en valeur absolue est très flatteuse en termes de résultat. Par contre, dès que l'on déflate par la population pour dégager des tendances de rentabilité ou d'intensification, la situation est toute autre.

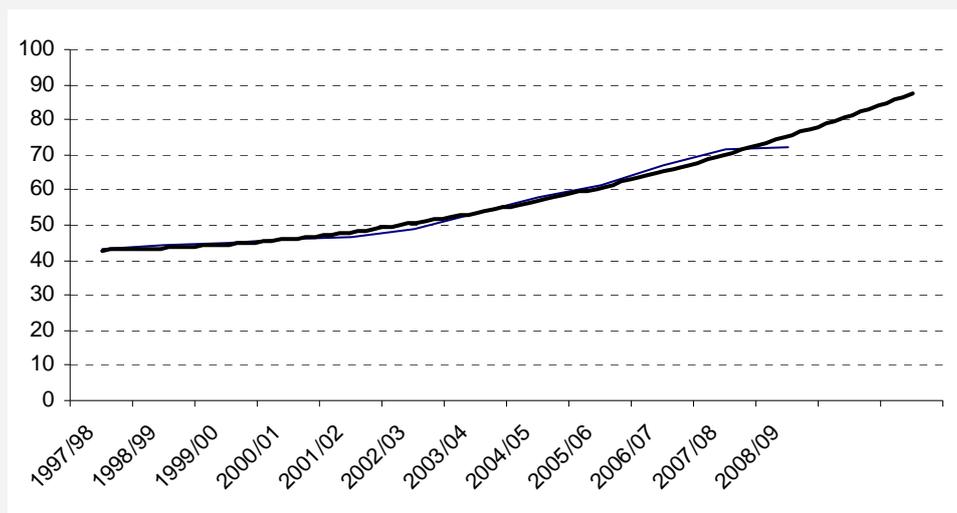
La valeur de référence: des hypothèses largement implicites

Dans pratiquement tous les cadres, la valeur de référence considérée est la valeur de la variable de l'année 1 qui est celle de la mise en œuvre. L'hypothèse non formulée est celle de la croissance positive linéaire qui répond de fait à l'idée sous-jacente de l'existence du modèle qui lie les intrants aux résultats, ce qui peut fonctionner pour les salles de classe ou les campagnes de vaccination, mais beaucoup moins si l'indicateur n'est pas uniquement dépendant des politiques. Les rendements agricoles, dont les valeurs annuelles sont en partie dues aux conditions climatiques, ne peuvent être comparés d'une année à l'autre. Si l'on souhaite réellement mesurer une tendance, il faudrait au moins comparer à la moyenne des 5, voir des 10 dernières années et donc prendre cette valeur comme référence et non celle de l'année 1.

EXEMPLE 9 : LE POIDS DE LA VALEUR DE REFERENCE

L'exemple ci-dessous illustre comment la valeur d'une année donnée peut influencer la tendance (et donc la fixation de cible). Le phénomène qui est venu bouleverser la série statistique est d'ordre purement administratif puisque sur la première série de taux de scolarisation l'âge légal considérait les tranches de 7 à 12 ans, tandis que sur la seconde, il s'agit de la classe 6 à 11 ans.

Sur la courbe de tendance en gras, les prévisions laissent espérer une croissance soutenue. Sur la seconde courbe (en trait fin), l'effet introduit par la mesure administrative infléchit la courbe de réalisation et enlève ainsi le caractère prévisible de la tendance. En conséquence, le changement introduit nécessite de reposer des hypothèses sur la valeur de référence.



Ce problème est fondamental dans les matrices utilisées dans le cadre de l'appui budgétaire, puisque les notations qui vont elles-mêmes déclencher les paiements des tranches de performance (variables) apprécient principalement la croissance. Donc, soit on limite les indicateurs à ceux qui ne présentent que des croissances linéaires (c'est en général ce qui est fait, d'où parfois la singulière pauvreté des matrices), soit on change la base pour mieux apprécier la tendance structurelle et non conjoncturelle (ce qui est très rarement fait).

L'utilisation d'indices simples, comme de prendre une base de référence (100) en début d'exercice de programmation, n'est que très peu utilisée (Cf. exemple 10) Cela permettrait pourtant une bien meilleure comparaison des tendances et surtout permettrait de comparer des indicateurs de nature différente (les efforts budgétaires et les résultats).

EXEMPLE 10 : COMPARER DES DYNAMIQUES. PROGRESSION EN % OU INDICES

Cet exemple est tiré d'un tableau de bord de l'éducation. Le premier tableau montre l'évolution en progression du TAP (Taux d'Achèvement du Primaire) depuis le début des enregistrements.

	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07
Garçons	28,4	27,7	30,0	31,8	32,2	33,6	35,1	36,6	37,6	40,3
Filles	19,4	19,8	21,5	22,4	23,6	25,7	27,3	28,7	30,4	32,4

D'un point de vue dynamique, il est indéniable que l'indicateur augmente de façon linéaire, mais si l'on souhaite comparer cette dynamique, le recours à un indice est beaucoup plus performant.

Le tableau ci-après reprend les mêmes données avec comme indice base 100 l'année 2000/01.

	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07
Garçons	89,1	86,9	94,2	100,0	101,1	105,4	110,4	114,9	118,1	126,6
Filles	86,6	88,4	96,3	100,0	105,6	114,8	122,1	128,4	135,8	144,9

On voit immédiatement sur cette seconde série que la dynamique est beaucoup plus forte pour les filles que pour les garçons : le double en points d'indice.

2.5 ETAPE 2B – A QUELLE FREQUENCE A-T-ON BESOIN DE L'INDICATEUR?

Lorsque l'on touche au fonctionnement du système statistique

La question posée est la suivante:

Existe-t-il au sein des systèmes d'informations (sources) des dispositifs de collecte régulièrement financés qui permettent d'obtenir les valeurs de l'indicateur?

- Si oui, on peut déterminer la fréquence de calcul et deux cas de figure se présentent:
 - ✓ soit la fréquence de calcul est compatible avec la fréquence de revue et, dans ce cas pas de problème;
 - ✓ soit la fréquence de calcul est incompatible avec la fréquence de revue et, dans ce cas, il faut voir si, et sous quelles conditions, on peut augmenter la fréquence de calcul.
- Si non, il faut essayer de définir quelle est la fréquence de revue qui restera une fréquence théorique, jusqu'à la mise en place du dispositif de collecte permettant le calcul de l'indicateur considéré.

Cette question des fréquences de revues et des fréquences de calcul « empoisonnent » une bonne partie du dialogue politique autour des indicateurs et le « remplissage des matrices » finit par devenir une activité qui occupe des entités entières (nombre de cellules chargées du suivi des CSLP ne font pratiquement que de la collecte secondaire auprès des services sectoriels afin de pouvoir réaliser le rapport annuel de revue).

Des solutions diverses

La question de fréquence recouvre en fait plusieurs problèmes qu'il est parfois difficile d'aborder de front:

- le problème de la fonction de l'indicateur, notamment en ce qui concerne certains indicateurs de résultats et d'impacts qui sont normalement des indicateurs structurels, donc relativement inertes au pas de temps annuel ou infra annuel ;
- le type de matrice (rôle de la matrice d'indicateurs dans le pilotage des politiques) ;
- le caractère erratique du financement des systèmes de collectes;
- le caractère imprévisible dans la livraison des données nécessaires au calcul des indicateurs.



La fréquence de calcul des indicateurs doit être considérée dans le cadre de la nature des matrices selon qu'elles sont analytiques ou descriptives (voir définitions § 2.2)



Pour un type de matrice donné, il faut vérifier en priorité:

- *la nature et la fréquence du système de collecte associé;*
- *l'exhaustivité des publications associées (si toutes les variables sont collectées, tous les indicateurs doivent être disponibles);*
- *la fraîcheur¹⁶ et la ponctualité des publications, au regard du système de revue.*

Pour une même matrice, il peut exister un corps d'indicateurs analytiques et un corps d'indicateurs descriptifs. Ils ne seront pas nécessairement observés à la même fréquence. Ceci n'interfère en rien sur la nature de l'indicateur et sa fonction (de nombreux indicateurs peuvent être utilisés indifféremment pour différentes fonctions, tout dépend du cadre).

¹⁶ C'est-à-dire le délai entre la collecte de la donnée et la date de mise à disposition de celle-ci.

2.6 ETAPE 2C – PEUT-ON FIXER UNE CIBLE?

Une question en trois points

La question peut être formulée comme suit:

Dispose-t-on d'une valeur de référence, c'est-à-dire d'une valeur de départ suffisamment solide pour ne pas être remise en question à peine les actions engagées?

Ce qui revient à: (i) adopter une valeur de référence (*baseline*); (ii) adopter un rythme de croissance; (iii) donner une valeur de sortie qui est la conséquence des deux premiers points.

La problématique la plus sensible concerne la confrontation des progressions proposées avec la réalité de l'évolution. Il est indéniable qu'à ce titre les indicateurs utilisés dans la deuxième génération des CSLP ou de l'aide budgétaire sont beaucoup plus argumentés que dans la première, pour laquelle les cibles étaient parfois fixées de façon purement mécanique, dans de nombreux cas par calcul inverse. C'est notamment vrai pour les OMD, pour lesquels, en prenant la valeur de sortie, on calculait juste le taux de croissance annuel pour atteindre la cible.

Qu'est-ce qu'une valeur de référence solide?

Le choix de la valeur de référence est souvent l'année de mise en place de la mesure. Cependant, dès que l'on quitte des phénomènes dont la progression est linéaire, ce type de décision devrait faire l'objet de plus d'attention. La notion de solidité renvoie donc à trois préoccupations:

- celle de la capacité de discrimination, en regard des phénomènes étudiés;
- celle de la connaissance des progressions antérieures afin de pouvoir confronter l'objectif «politique» à la réalité de faits (réalisme);
- celle de la fiabilité et de la régularité du système d'information qui produit la ou les variables nécessaires (voir chapitre suivant).

EXEMPLE 11: LE CHOIX DE L'ANNEE DE REFERENCE

Dans l'exemple ci-dessous, le taux de réussite au certificat d'études élémentaires dans un pays donné présente – depuis qu'il est mesuré – une relative stabilité, hormis pour l'année 2005/2006 qui apparaît aberrante.

	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07
Taux	46,8	50,4	45,3	50,4	45,5	47,3	69,4	55

La cible que s'est fixée le Ministère de l'Éducation en 2006 est de 75% à l'horizon 2010.

Un simple examen de la courbe aurait du conduire à se poser deux questions : celle de l'année de référence 2005/2006 (il existe vraisemblablement un problème de calcul pour cette année) et celle du taux de progression, car un tel indicateur est beaucoup moins sensible aux actions volontaristes qu'un taux brut de scolarisation. A moins de baisser notablement le niveau d'obtention, ceci pose le problème des effets potentiellement pervers de certains indicateurs.

Concernant le caractère plus ou moins régulier du système d'information, deux cas de figures peuvent se présenter **lorsqu'il n'existe pas, à proprement parler, de système de collecte** et de mise à disposition de l'information:

- soit le cadre de planification dure suffisamment longtemps pour permettre de peser sur le système de suivi et il sera donc possible d'introduire dans les dispositifs, les corrections qui permettront au moins d'avoir des valeurs intermédiaires ou des valeurs de sortie;
- soit, le cadre de planification est trop court pour permettre l'introduction dans le système de la ou des variables requises. Dans ce cas, il faut identifier un autre indicateur.

La notion de durée nécessaire à la mise en œuvre d'un système d'information peut être difficile à définir, mais en première approximation, on peut considérer comme courte, toute période inférieure à 3 ans. Cette période de 3 ans peut être valable aussi bien pour un Système Statistique National (SSN) que pour un projet ou un programme. Elle vient simplement du fait que de très nombreuses variables sont annuelles, donc étudier, tester et mettre en place un système de collecte inexistant ne peut être fait que sur plusieurs années.



Pour une question donnée, le choix d'un indicateur parmi ceux qui sont éventuellement disponibles doit prendre en compte les éléments de qualité suivant:

- *la longueur des séries statistiques disponibles;*
- *la précision au moins théorique de l'indicateur et donc la capacité à discriminer deux positions successives;*
- *sa normalisation¹⁷ dans l'analyse de la progression et notamment l'insensibilité aux facteurs de progressions autres que le paramètre étudié.*

NB: Il est toujours possible d'améliorer par le calcul la comparabilité des deux valeurs données d'un même indicateur. Cela suppose tout de même que l'on puisse recalculer la série à partir d'une base nouvelle.

Ces caractéristiques de comparabilité et de discrimination conditionnent en grande partie la capacité à fixer une cible crédible pour l'indicateur (Cf. exemple 11 ci-dessus).

Rythme de croissance et valeur de sortie

En théorie, le rythme de croissance devrait être défini soit par projection statistique (voir exemple 8), soit à travers un système de modélisation. Dans la plupart des cas, notamment dans les PED, ni l'un ni l'autre existe, excepté en matière de population (projections démographiques) ou en matière de macro-économie (projections de croissance globale ou sectorielle).

Le rythme de croissance est donc bien souvent calculé en retro action par rapport à la cible qui est fixée généralement, soit sur des bases purement politiques, soit parce que résultant d'accords internationaux.

En dehors de ces cas qui relèvent du dialogue politique, les techniques qui permettent d'adopter le rythme de croissance sont souvent réduites à de simples examens empiriques entre les quantités d'intrants injectées (ressources financières) et les effets observés (malheureusement sur de courtes périodes).

A ce stade, la sensibilité de l'indicateur, c'est-à-dire sa capacité à répondre de manière immédiate aux efforts financiers est prépondérante : un taux de vaccination répond immédiatement à des campagnes massives, alors qu'un taux d'alphabétisation ne peut répondre que sur un long engagement.

¹⁷ On entend ici par normalisation, l'ensemble des processus de calcul (déflateurs de population, de prix,...) ou de transformation de variable (lissage) permettant une meilleure interprétation des tendances.

EXEMPLE 12: CIBLE ET PROGRESSION. LA CIBLE D'ABUJA DANS LE DOMAINE DE LA SANTE

Les Etats signataires de cet accord se sont engagés en 2001 à ce que le ratio des dépenses de santé sur les dépenses publiques atteigne 15% en 2010. Le tableau suivant donne la situation en 2008.

Unité	Pays (Monnaie)	Bénin (FCFA)	Burkina Faso (FCFA)	Cap Vert (ECV)	Côte d'Ivoire (FCFA)	Gambie (Dalasi)	Ghana (ancien CEDI)	Guinée (GNF)	Guinée Bissau (FCFA)
Million	Budget Santé	41 143	44 442	2 306	98 807	301	1 965	65 685	5 611
Million	Dépenses budget Etat	475 739	451 768	29 055	1 758 758	2 989	28 745	2 223 866	59 712
-	Ratio santé/budget	8,6%	9,8%	7,9%	5,6%	10,1%	6,8%	3,0%	9,4%

Les taux de croissance nécessaires pour atteindre la cible se situent entre 50 et 100%, voire plus (400% pour la Guinée). En dehors du fait que la cible ne sera pas atteinte en 2010, la question à se poser est de savoir de combien de % un budget national peut-il raisonnablement augmenter par an?

Ce type d'indicateur pose aussi le problème d'expression en part relative de budget ou de dépense, car on peut évidemment atteindre l'objectif en réduisant de façon drastique d'autres postes. A la limite, un indicateur de rang dans le budget serait peut-être plus parlant.

2.7 ETAPE 3 – BILAN DE LA DEMARCHE : DOCUMENTER LES INDICATEURS

Adopter ou rejeter des indicateurs

A ce stade de l'analyse on dispose de deux documents:

- le premier consécutif à l'étape 1 d'analyse reflète les besoins de suivi;
- le second consécutif à l'étape 2 est une matrice d'indicateurs ou groupes d'indicateurs avec leurs moyens d'obtention.

Pour un domaine donné on peut établir un tableau qui va permettre de choisir:

- les indicateurs qui satisfont aux différents critères ci-dessus énumérés. Ils sont normalement classés comme adoptés;
- les indicateurs qui nécessitent des modifications des systèmes d'information existants, de fréquence, de précision, compatibles avec les échéances du cadre politique. Ce sont les indicateurs faisables;
- les indicateurs qui nécessiteraient la mise en œuvre de dispositifs de collecte non compatibles avec les échéances du cadre politique et qui sont normalement rejetés.

A ce stade, un retour sur le cadre politique (étape 1) est nécessaire, car a priori c'est à ce niveau que peut être décidé le maintien ou non de certains indicateurs.

Documenter les indicateurs

Documenter un indicateur signifie établir sa fiche signalétique qui va permettre par la suite d'analyser correctement ses évolutions. Les renseignements à faire porter peuvent être relativement standards.

Rappel du cadre de mesure

Il s'agit de rattacher l'indicateur aux programmes et de définir en toutes lettres ce qu'il mesure. Il convient d'être très clair dans le libellé de la mesure. Par exemple, un taux de scolarisation est un indicateur mesurant l'accès au système éducatif, un bilan céréalier est un indicateur de disponibilité alimentaire.

Description de l'indicateur

Il s'agit de donner l'ensemble des informations concernant le calcul (formule, unité de mesure) :

Qualité de l'indicateur

Il s'agit de documenter les sources utilisées et d'émettre les restrictions éventuelles liées à la qualité des sources. Cet aspect est primordial, dans la mesure où il détermine en partie la façon dont l'indicateur va être interprété par la suite et qu'il convient de ne pas oublier que cette interprétation sera faite par de nombreuses personnes non nécessairement expertes.

Interprétation de l'indicateur

Il s'agira de donner en quelques lignes la façon dont l'indicateur peut être utilisé : (i) son sens, positif ou négatif ; (ii) les modalités d'interprétation. On peut, par exemple, au titre des modalités d'interprétation, indiquer à partir de quelle valeur une augmentation ou diminution est significative, ou renvoyer à l'observation d'autres statistiques pour une meilleure compréhension. Voir aussi en **annexe 3**, quelques questions simples que l'on peut se poser face à un indicateur ou sa valeur.

EXEMPLE 13 : UN BONNE FICHE D'IDENTIFICATION

FICHE DE DOCUMENTATION	
Indicateur	<i>Intitulé</i>
Programme	<i>Référence au programme gouvernemental</i>
Objectif	<i>Référence à l'objectif visé à l'intérieur du programme</i>
Action	<i>Référence à l'action visée à l'intérieur de l'objectif</i>
Service responsable	<i>Nom du service utilisateur de l'indicateur à l'intérieur du programme</i>
Description de l'indicateur	
Unité de mesure	
Périodicité de la mesure	
Dernier résultat connu	<i>Année</i> <i>Valeur</i>
Elaboration et qualité de l'indicateur	
Nature des données de base	<i>Numérateur, dénominateur, champ couvert</i>
Mode de collecte des données	<i>Enquête, collecte administrative...</i>
Services et organismes responsables de la collecte	<i>Indiquer le ou les organismes chargés de la collecte</i>
Mode de calcul	<i>Indiquer la formule</i>
Modalités d'interprétation	
Limite et biais connus	<i>Indiquer les limites d'utilisation (signification) ou s'il s'agit d'un proxy le biais par rapport à l'indicateur souhaité</i>
Modalités d'interprétation	<i>Préciser la signification par rapport au domaine ce qu'il mesure exactement</i>
Sens d'interprétation	<i>Diminution ou augmentation</i>
Plan de documentation	
Date de livraison	<i>Date de parution du document contenant la valeur</i>
Amélioration en cours	<i>Indiquer si le service producteur envisage de changer de méthode de calcul ou de collecte.</i>
Commentaires	

Partie 3

Le contexte de production des indicateurs: la qualité statistique

Cette partie expose de façon succincte comment aborder l'analyse de la qualité statistique et montre surtout que la qualité est une dimension multidimensionnelle qui ne se réduit pas à la qualité des données de base.

Mots clefs : exactitude, fiabilité, disponibilité, rigueur méthodologique, utilité, accessibilité, donnée d'enquête, données administratives, précision, erreur.

3.1 LA MESURE DE LA QUALITE STATISTIQUE

Les éléments du problème

Le dialogue sur la qualité des statistiques est souvent biaisé et extrêmement limité dans sa portée :

- ☛ les uns reprochent la « fausseté » des chiffres;
- ☛ les autres argumentent sur le caractère, soit officiel¹⁸, soit scientifique du chiffre.

Ce qui aboutit souvent à un dialogue de sourds et ce, d'autant plus que lorsqu'il s'agit d'indicateurs isolés de leur contexte dans un tableau de synthèse, les sources sont parfois multiples et surtout mal définies. Sans développer des cadres complets d'analyse, il est possible de dériver des cinq dimensions principales de la qualité (voir ci-après) une forme de contrôle sommaire, quitte à procéder par la suite à des exercices plus complets et exhaustifs.

Qu'est-ce que la qualité statistique?

La qualité statistique est, à l'instar de toutes les démarches qualité visant les produits ou les services, un système visant à **qualifier un processus de production** afin de l'améliorer et surtout de respecter une continuité dans les produits remis au public.

En général, la première dimension qui vient à l'esprit est la qualité du produit qui, dans le cas des statistiques, se réfère à la qualité de la donnée qui se traduit, elle-même, bien souvent uniquement en termes de précision ou plus généralement de **validité des données de base**. Cette dimension, bien que principale, constitue une dérive dans la mesure où elle est extrêmement réductrice (combien de données même très valides sont « gâchées » par des traitements inappropriés, ou purement et simplement non traitées) et peut être en contradiction avec d'autres impératifs de production, comme le fait de fournir à temps les résultats.

L'**exactitude (accuracy)** d'une statistique est le produit d'un ensemble de processus de production qui part de la conception de la méthode de collecte à sa réalisation sur le terrain, puis au système de traitement (avec ce que cela comprend comme vérifications intermédiaires). Sa **fiabilité (reliability)** est la conséquence du respect « scrupuleux » de l'ensemble du processus et de sa reproduction dans le temps¹⁹.

Pour autant, le fait de disposer de données exactes ne constitue pas en soi le résultat du service qui est attendu par l'utilisateur. Pour que cette donnée devienne une information, il faut qu'elle soit communiquée, c'est-à-dire rendue **accessible** au public auquel elle est destinée (pour mémoire, le public cible de la statistique publique est le citoyen qui paye ce service à travers ses impôts). Or, ce public est très divers dans ses attentes, ses besoins et les moyens dont il dispose pour accéder à l'information. Il convient donc de l'informer sur la **disponibilité (availability)** de l'information et d'adapter autant que faire se peut, l'information à ses attentes. Ce citoyen (Ministre ou simple quidam) n'est pas en général un statisticien; il faut donc lui expliquer comment utiliser les chiffres et ce qu'ils signifient réellement, en un mot, l'assister dans sa lecture.

Cependant, pour qu'une donnée soit crédible, il convient, outre le fait de vérifier qu'elle ait été produite avec **rigueur**, assurer qu'elle n'a pas fait l'objet de manipulations (au mauvais sens du terme) et que **l'indépendance** et **la transparence** constituent la règle de production. Pour ce faire, il faut que les processus de production soient connus, ou du moins que l'on puisse y avoir accès et que la livraison du produit soit faite en toute **impartialité**, c'est-à-dire, qu'il n'existe pas, en dehors de l'organisme producteur (qui *ipso facto* dispose de l'information avant tout le monde), de client privilégié.

¹⁸ Ainsi, souvent le chiffre est considéré comme bon lorsqu'il a été « validé » par le ministre

¹⁹ Les notions d'exactitude et de fiabilité ne doivent pas être confondues avec celle de précision, qui renvoie à la connaissance de la marge d'erreur entourant une donnée. Il est ainsi généralement admis que les données de population en Europe sont connues avec une précision de plus ou moins 1%.

Enfin, un producteur de statistiques publiques ne produit pas pour son compte propre, ni ne décide ce qui doit, ou ne doit pas, être produit. La relation qu'il entretient avec son prescripteur principal qui est l'Etat, repose sur l'**utilité** de sa production au regard des besoins en matière de pilotage du pays dont l'Etat a la charge. Cette relation se traduit par une perpétuelle adaptation entre moyens et résultats: en période de plein emploi, il n'est pas **pertinent (relevant)** de dépenser des moyens considérables sur les statistiques de l'emploi. Par contre, lorsque 10, 20, 30% d'une population est réduite au chômage, il devient impératif pour l'Etat de connaître ces chiffres avec exactitude et précision. Pour être **utile**, une information statistique doit **arriver à temps (timeliness)**, pour permettre la prise de décision (et non pas quelques mois ou années après). Ce court texte résume les différentes dimensions de la qualité d'un service de statistiques publiques:

Des statistiques utiles, produites avec rigueur et intégrité, selon des méthodes fiables et disponibles pour le plus grand public.



Ce qu'il faut retenir du cadre d'analyse de la qualité concerne essentiellement l'alimentation du dialogue avec les producteurs de données, et par conséquent avec le prescripteur principal qu'est l'Etat. Il s'agit avant tout de permettre de mieux exprimer les doutes ou les interrogations que l'on peut avoir sur les chiffres fournis et sortir du dialogue simpliste sur le « bon ou le mauvais » chiffre.

Un cadre sur la qualité statistique doit avant tout servir à utiliser un langage commun pour identifier les points forts et les points faibles, les qualités comme les défauts et surtout permettre de hiérarchiser les problèmes en vue d'apporter des solutions (si le problème le plus important concernant la production statistique est que les données arrivent avec 2 ans de retard, il faudra corriger avant tout ce défaut, même si on considère, par exemple, qu'il faudrait aussi des données plus précises).

Des cadres conceptuels différents mais finalement convergents

Les besoins de comparaison internationale ont conduit des institutions comme la Banque Mondiale, le Fonds Monétaire International (FMI) et la Commission Européenne à développer des démarches similaires, notamment afin de pouvoir « juger » les données fournies par les Etats membres selon des critères identiques et éventuellement inciter ces mêmes Etats à améliorer les processus de productions.

Il existe toutefois des différences entre les approches, par exemple entre la démarche européenne et la démarche de la Banque Mondiale (inspirée de celle du FMI)²⁰.

Les cinq dimensions de la qualité

Le cadre d'évaluation de la qualité est une adaptation du cadre d'évaluation de la qualité des statistiques du FMI faite par la Banque Mondiale. Il comporte donc les **cinq dimensions** de la qualité des statistiques du FMI, à savoir:

- l'assurance d'intégrité (c'est-à-dire l'objectivité dans la collecte, le traitement et la diffusion);
- la rigueur méthodologique (un cadre et une pratique conforme aux recommandations internationales);
- l'exactitude et la fiabilité (des données et techniques saines, des produits reflétant la réalité);
- l'utilité (une périodicité suffisante, un délai raisonnable, une cohérence des données et une prévisibilité des révisions);
- l'accessibilité (des données et métadonnées disponibles, une assistance aux utilisateurs suffisante).

Le cadre d'évaluation examine également les « conditions préalables de la qualité » (un environnement juridique et institutionnel favorable, des ressources suffisantes et une sensibilisation à la qualité dans le cadre des travaux statistiques).

²⁰ Comparaison entre les cadres qualité de la Banque Mondial/FMI et Eurostat (annexe 4).

3.2 QUELQUES DIMENSIONS DE LA QUALITE COMMENTEES

Nous allons examiner quelques dimensions de cette qualité et surtout sous l'angle de l'impact possible des défauts de qualité sur l'utilisation des indicateurs.

La rigueur méthodologique

Cette dimension mesure le respect des pratiques statistiques, à savoir si les méthodes employées sont conformes, d'une part, avec la statistique en tant que science et, d'autre part, avec l'environnement international défini le plus souvent par des règles édictées et adoptées dans des fora internationaux ou régionaux (comme le respect des nomenclatures ou des méthodes de calcul).

D'une façon simple et pour ce qui concerne l'appréhension des indicateurs, trois éléments devraient être étudiés:

- la réalité du champ statistique;
- l'existence pour un champ donné de la liste exhaustive des variables et indicateurs avec leur mode de calcul (ce que l'on appelle communément un dictionnaire de variables);
- l'existence de manuels ou de notices méthodologiques.

Le champ statistique détermine l'étendue de la collecte par rapport à la définition théorique de la population. Si tous les individus ne sont pas enquêtés (ni recensés, ni sondés), le champ n'est pas couvert intégralement. De fait, de nombreux indicateurs utilisés dans les cadres de mesures sont en fait des proxys.

EXEMPLE 14: DE L'IMPORTANCE DE CERNER LE CHAMP STATISTIQUE DE LA COLLECTE.

Les problèmes de registre dans les collectes administratives

On commet souvent l'erreur de penser que ces problèmes de champ de collecte ne concernent que les enquêtes par sondage et que la collecte administrative ne présente pas ce type de problème. Rien n'est moins faux : par exemple, pour fabriquer les statistiques scolaires, il faut avoir une liste à jour de tous les établissements d'enseignement; de même, pour les statistiques de santé, il faut avoir tous les centres de santé, et ainsi de suite.

En 2003/2004, dans un pays donné, le Ministère de l'éducation de base a entrepris le recensement dynamique des écoles privées. En effet, si les écoles privées doivent en théorie s'enregistrer pour recevoir l'agrément du Ministère, peu le font réellement. A l'issue du recensement, il est apparu que 800 écoles sur les 6200 que compte le pays n'avaient jamais été enregistrées et que la part du privé dans l'enseignement de base passait de 13 à 23%. En pratique, 80 000 élèves n'étaient donc pas comptabilisés dans les différents indicateurs.

Concernant l'existence de **dictionnaires de variables ou d'indicateurs**, ces derniers devraient être régulièrement mis à jour, notamment afin de tenir compte de l'évolution des nomenclatures, des concepts et des méthodes de collecte. Concernant les indicateurs de mesure de la performance, le fait de disposer de l'ensemble des variables calculées est capital: même si in fine on ne retient que deux indicateurs en termes de suivi, c'est l'existence de tous les autres qui permettra, le moment venu, de juger du degré de performance du secteur.

Concernant **les manuels**, il en existe de nombreux (manuel de collecte, de traitement, de calcul). A défaut, il faut toujours vérifier s'il existe un manuel de collecte, c'est-à-dire des instructions aux personnes (enquêteurs ou responsables de remplissage de questionnaires administratifs) leur permettant de faire correctement leur travail. En dehors de cet aspect, qui va avoir de grandes répercussions sur l'exactitude et la fiabilité des données, il s'agit en pratique d'un support facile pour comprendre les chiffres finaux: pour comprendre ce qui est calculé, il faut voir la question posée. De plus, ces manuels mentionnent souvent les références aux concepts ou nomenclatures utilisées.

L'exactitude et la fiabilité

Il s'agit du domaine le mieux connu de la qualité statistique puisqu'il renvoie directement à la qualité des données produites et repose donc en grande partie sur les **méthodes statistiques utilisées** et leur adéquation avec le sujet d'étude. Pour autant, exactitude et fiabilité renvoient à deux notions parfois antinomiques:

- l'**exactitude** repose entièrement sur la méthode utilisée et son respect scrupuleux tout au long de la chaîne de production qui va de la collecte à l'édition de la publication;
- la **fiabilité** renvoie, d'une part, au maintien dans le temps d'une norme de production et, d'autre part, aux efforts qui peuvent être faits pour évaluer régulièrement les données et communiquer ces résultats aux utilisateurs.

EXEMPLE 15 : LES SOURCES DE DONNEES

Données administratives et données d'enquête

Il existe fondamentalement deux sources de données : d'une part, les données dites administratives qui résultent de l'activité d'une administration (le registre des établissements scolaires fournit les statistiques de l'éducation, la main courante des commissariats renseigne sur les délits, ...), d'autre part, les données dites d'enquête, lorsqu'un individu physique ou moral est interrogé (on interrogera un enfant sur son niveau d'étude ou une personne sur les dols qu'elle a subi).

Il faut toutefois garder en mémoire que ces deux sources de données ont des fondements très différents : la donnée administrative est le résultat d'une production de l'administration. Ainsi, si le Ministère de l'Education n'a pas de mandat de supervision de l'enseignement du secteur privé, les statistiques issues de ses données seront restreintes au seul périmètre du secteur public. La donnée d'enquête mesure un phénomène normalement au niveau de la population et donc la valeur intrinsèque des indicateurs mesure beaucoup mieux un résultat (au niveau de la population) que ne le ferait la donnée administrative.

La donnée d'enquête peut donc constituer un bon système d'étalonnage des collectes administratives, à condition toutefois que les concepts utilisés soient identiques : mêmes définition des variables, même population de référence. On peut citer un certain nombre de collectes qui permettent cette appréciation de la valeur d'un indicateur selon deux sources différentes :

Toutefois, nous le verrons plus tard, la comparaison et surtout l'analyse tendancielle entre deux enquêtes, requiert certaines conditions pour avoir une signification. La comparaison entre une donnée administrative et une donnée d'enquête nécessite de toujours se poser la question de savoir si ces deux sources mesurent bien la même chose (même si l'indicateur a le même nom).

Le chiffre a un aspect mathématique qui donne l'illusion du fait scientifique. Spontanément, on tend à confondre *précis* et *exact*; une notion vague ne peut être entièrement exacte. Si les résultats d'une enquête indiquent que la population est de 1 235 455 habitants, c'est le chiffre exact obtenu par le calcul de variables. Si cette enquête présente une marge d'erreur de 10% (c'est-à-dire une erreur probable pouvant aller jusqu'à 12 000 habitants), il est tout aussi possible d'annoncer qu'il y a 1 235 000 habitants. C'est un chiffre moins précis (au sens commun du terme) mais tout aussi exact.

En statistique:

- le chiffre **exact (accurate)** est celui qui correspond au résultat obtenu si la méthodologie a été appliquée en toute rigueur depuis la collecte jusqu'au traitement. Il peut être partiellement faux, **au sens commun du terme**, soit parce que la variable a été mal définie au départ, soit parce que la question posée ne permet pas d'accéder à la variable visée ou que le répondant sélectionné n'est pas capable de répondre correctement;

- le chiffre **précis (precise)** est celui qui est connu avec une marge d'erreur compatible avec son utilisation;
- le chiffre **fiable (reliable)** est celui qui est produit avec une qualité égale et connue en termes de précision.

L'amélioration de l'exactitude ou de la précision (en dehors de son coût) peut se faire au détriment de la fiabilité et il vaut cent fois mieux des données entachées d'erreurs connues, mais produites à temps et régulièrement, que des données moins erronées produites de manière erratique.

Les différents types d'erreurs qui peuvent être commises dans un processus de production de données sont innombrables. Nous ne retiendrons donc que les principales et l'impact qu'elles peuvent avoir sur les valeurs prises par les indicateurs. Il est bon de rappeler quelques grandes règles, dès lors que l'on parle de précision et d'erreurs:

- l'erreur admise (toute source confondue) doit toujours être mise en relation avec l'objectif de mesure de la variable: en théorie si l'on désire mesurer des évolutions de l'ordre de 1%, l'erreur commise devrait être inférieure à cette valeur! Dans les pays du Nord, on estime que la population est connue à 1% près, dans les pays du Sud cette estimation n'a jamais été faite mais on peut supposer que cela dépasse parfois le 10%. Donc un indicateur, utilisant la population au dénominateur, ne deviendra vraiment parlant que si ses valeurs sont supérieures au 10% d'erreur commise sur l'estimation de la population. De la même manière, si la production de coton est connue avec 10 ou 15% d'erreur, pour les besoins de connaissance de l'économie nationale cela n'est pas forcément très grave (pour le macro-économiste) mais le devient pour le responsable de la filature qui justement fait sa marge sur 10 à 15%;
- l'erreur de mesure croît avec le nombre de mesures faites. L'erreur de mesure est celle que l'on commet lors du remplissage d'un questionnaire. Plus on collecte, plus on commet d'erreurs et donc moins la donnée finale est exacte. Il s'agit de ce que l'on pourrait appeler le syndrome du recensement qui a fait le bonheur des statisticiens de l'économie socialiste, mais qui demeure comme idée sous-jacente dans de nombreuses opérations statistiques: gonflement de l'échantillon, des questions posées (jusqu'à 1000 variables dans certaines grosses enquêtes) au détriment du contrôle de la qualité de remplissage des questionnaires.

EXEMPLE 16: QUELQUES TYPES D'ERREURS*L'erreur de couverture dans les collectes administratives*

Les collectes administratives étant réputées être des recensements, on parle de défaut de couverture lorsque l'ensemble des questionnaires adressés aux « établissements » du registre²¹ ne sont pas renvoyés. *Stricto sensu*, on devrait parler plutôt de « non réponse totale » ou de taux de complétude (*completeness*) et réserver l'erreur de couverture aux défauts du registre (voire ci dessus).

Admettons les données suivantes (très basiques): il existe dans un pays 6 200 établissements primaires avec un Cours Préparatoire (CP) dans chaque établissement et en moyenne 45 nouveaux inscrits, pour une population de référence de 300 000 (tranche d'âge). Le tableau ci-dessous reprend l'évolution de l'indicateur du taux d'admission au CP pour différents taux de complétude, c'est-à-dire pour des taux de réponse aux questionnaires différents.

	Taux 98 %	Taux 99 %	Taux 100 %
Nombre de questionnaires envoyés	6200,00	6200,00	6200,00
Nombre de questionnaires reçus	6076,00	6138,00	6200,00
Effectifs nouveaux inscrits	273 420,00	276 210,00	279 000,00
Taux	91,14	92,07	93,00

Si les progressions mesurées entre deux années sont inférieures à 2 %, et que les taux de complétude varient entre 98 et 100% (ce qui est en général constaté), on ne peut donc réellement juger de l'évolution. On ne devrait donc juger de l'évolution interannuelle qu'à taux de complétude équivalent.

NB: il faut toujours se méfier des pourcentages faibles: 1%, ce n'est rien dans l'absolu! Tout dépend de quoi l'on parle. Dans l'exemple ci-dessus, cela représente 2 790 inscrits soit 62 classes d'écoles. Idem: 5% ce n'est pas grand chose mais cela ferait 310 classes.

L'erreur de sondage et la désagrégation géographique

La plupart des enquêtes auprès des ménages dans de nombreux PED, utilisent en général des échantillons destinés à être représentatifs au niveau du premier découpage administratif des pays. Le tableau ci-dessous indique les Coefficients de Variation (CV) de quelques indicateurs dans une enquête de type *Core Welfare Indicator Questionnaire* (CWIQ). Le CV indique pour une probabilité donnée, la variation relative de la valeur: ainsi un CV de 10% indique que la valeur varie à l'intérieur d'un intervalle qui représente 10% de la mesure faite.

	National	Rural	Urbain	Région 1	Région 2
Taux d'alphabétisation	4,2	4,1	2,9	17,4	12,9
Accès à l'école primaire	2,7	3,2	1,8	16,7	14,8
Accès à l'école secondaire	7,7	13,9	4,3	45,9	42,3
Accès aux services de santé	4,9	6,8	3,2	25,8	26,1
Insuffisance pondérale	2,4	2,4	7,5	5,2	9,2

Cet exemple (réel) montre qu'au niveau national des comparaisons sont possibles du moment que les variations interannuelles sont supérieures au coefficient de variation c'est-à-dire en dehors de la marge d'erreur:

- pour que l'on puisse conclure que le taux d'accès à l'école primaire évolue, seules les valeurs

²¹ On appelle registre la liste des individus identifiés comme répondants, il peut s'agir par exemple de la liste des écoles, des centres de santé ou des tribunaux. On utilise le mot registre pour la collecte administrative. La notion de base de sondage qui renvoie au même concept n'est utilisée que pour les enquêtes.

supérieures à 2,7% seront significatives.

- pour que l'on puisse conclure que le taux d'accès à l'école secondaire évolue, seules les valeurs supérieures à 7,7% seront significatives.

En ce qui concerne la valeur absolue des estimations, la question posée est celle de savoir si on doit publier un chiffre connu avec une marge d'erreur de 50%. Ceci renvoie bien sûr à l'ensemble des problématiques de désagrégation à partir de données d'enquêtes, quelles soient géographiques ou par sous-populations. Plus on désagrège, plus la marge d'erreur augmente rendant de nombreuses comparaisons illusoire : la donnée régionale n'est pas suffisamment précise pour mesurer des progressions de quelques %.

L'utilité

Cette dimension renvoie à deux éléments qui sont au cœur de la problématique de l'utilisation d'indicateurs de mesure, à savoir la **pertinence** et l'**adéquation** de la production, avec les impératifs de prise de décision (c'est-à-dire de savoir si la donnée est produite à temps et selon un calendrier attendu).

La pertinence fait référence à la satisfaction du besoin des usagers. Elle peut s'appliquer aussi bien :

- à un domaine statistique: est-il pertinent de disposer de statistiques sur la gouvernance dans une dictature?
- à la couverture de ce domaine: est-il pertinent d'avoir des statistiques sur la médecine privée si celle-ci ne représente que quelques % des soins dispensés?
- à la fréquence de collecte: est-il pertinent de mesurer annuellement des variables structurelles (c'est-à-dire des variables qui touchent des fondamentaux sociétaux ou économiques, donc n'évoluant que lentement)?

Le problème de pertinence renvoie dans bien des cas à des questionnements sur le rapport coût / bénéfice d'obtention d'une statistique et donc à la satisfaction de l'utilisateur. C'est à ce niveau qu'il existe une asymétrie de dialogue politique entre la plupart des Etats bénéficiant de l'aide et ses Partenaires Techniques et Financiers (PTF). Contrairement aux pays du Nord, dans lesquels la statistique est inscrite dans l'ensemble des processus décisionnels, il convient d'avoir toujours en mémoire que ce n'est pas le cas dans de nombreux pays du Sud. La nécessité est donc celle exprimée par les partenaires qui en constituent souvent de fait les utilisateurs principaux. Par delà les discours et les déclarations, et en dehors de quelques cas trop particuliers, le prescripteur principal est « le bailleur de fonds » : **la statistique est donc construite en dehors des besoins exprimés par la société, ce qui se traduit par un intérêt limité du Gouvernement pour la mise en place d'outils de collecte.**

Il est symptomatique de constater que l'offre statistique globale dans de nombreux pays suit de près l'évolution des politiques de coopération, ce qui est somme toute normal (le principe d'utilité voulant que l'on produise pour prendre des décisions) mais laisse perplexe quant aux possibilités d'analyse des situations des pays.

Le développement du suivi des performances, notamment des cadres de lutte contre la pauvreté, a considérablement rehaussé le niveau d'exigence statistique. Cependant, cette exigence croît plus vite que l'offre d'information de qualité (l'exemple en page précédente de statistiques de gouvernance n'est pas simplement innocent ou provocateur) et dès lors il faut considérer deux alternatives possibles :

- le standard de qualité s'accroît avec l'amélioration des outils d'analyse et on devrait donc écarter les informations les moins fiables;
- le standard de qualité baisse pour s'adapter aux outils d'analyse, l'information est certes complète mais moins fiable.

Périodicité et délais de diffusion

Cette qualité fait référence au fait que, pour pouvoir être utilisées dans un processus décisionnel, les données doivent être produites selon un contenu et un calendrier établi. Ainsi on devrait savoir que l'annuaire des statistiques de l'enseignement est, par exemple, publié en avril ou celui de la santé en mai.

Cet élément de mesure de la qualité est essentiel et fait d'ailleurs l'objet de nombreux engagements internationaux (Système Généralisé de Diffusion de Données du FMI, entre autres). Mais, en pratique, il n'existe pas de procédures bien établies pour la livraison des données. En conséquence, il s'agit bien souvent de fourniture à la demande (notamment en ce qui concerne les indicateurs relatifs aux matrices) dont le statut exact peut être relativement flou. Ceci explique en grande partie les livraisons de chiffres incohérents ou évoluant en fonction du degré de traitement des résultats.

Cet aspect touche le problème fondamental de la source et de son identification lors de la construction des matrices: dire que la source est le Ministère de la Santé n'est pas suffisant. Par contre, dire que la source est l'annuaire statistique de la santé devrait conduire à rejeter toute autre information tant que cette publication n'est pas diffusée d'une manière ou d'une autre.

Le corolaire est que les dates de livraison doivent être cohérentes avec le calendrier d'utilisation des données et donc avec le choix du support de la source: si à la date t le document de référence ne peut être prêt, il faut alors qu'il existe un document de données provisoires, reconnues comme telles, qui est utilisé comme source « officielle ».



Que cela soit lors de l'élaboration de matrices ou par la suite lors des revues, il devrait toujours exister dans les documents le calendrier de livraison des sources identifiées avec éventuellement les différents stades de publication (provisoire, définitif). A défaut, il convient:

- *de dresser la liste des publications nécessaires au remplissage de la matrice;*
- *de déterminer les dates limites de livraison;*
- *d'associer au processus de revues, l'analyse du "statut" des données (est-ce une donnée publiée ou une donnée « transmise »?);*
- *de faire peser l'exigence de qualité sur la publication et la diffusion (en d'autres termes d'être beaucoup plus intransigeant sur le fait qu'une donnée doit être publiée et diffusée).*

L'accessibilité

L'accessibilité renvoie aux efforts faits par les producteurs, pour mettre le chiffre à la portée du public mais aussi pour lui apporter assistance dans l'utilisation des résultats.

Cet élément de la qualité est bien souvent ignoré et les causes sont probablement identiques à celles évoquées au titre de la pertinence. Le chiffre n'étant pas considéré comme un bien public, c'est-à-dire pouvant être utilisé aussi bien par l'individu lambda que par un autre statisticien, on considère que c'est à l'utilisateur de se débrouiller. Il est vrai que, dans les PED, la communauté des utilisateurs est relativement réduite par rapport aux pays du Nord, mais elle présente surtout cette caractéristique spécifique de n'avoir, dans sa grande majorité, qu'une culture statistique très réduite (y compris au niveau de nombreux décideurs). En conséquence, la relation offre / demande est très asymétrique; c'est l'offre qui tire les statistiques et non pas la demande des utilisateurs.

Cette offre doit donc être digeste et pour peu que l'on y consacre un peu de temps et de moyens, le public se développe (au Burkina Faso, la refonte et l'alimentation régulière du site WEB de l' Institut National de Statistique lui a permis de passer en tête des sites gouvernementaux du pays consultés en quelques mois).



Pour ce qui concerne les indicateurs et donc les publications qui servent de support, il faut toujours vérifier les éléments suivants:

- l'existence des éléments de la série et au moins les courbes correspondantes;
- l'existence des définitions et des concepts utilisés, y compris la fonction de l'indicateur dans chaque publication;
- l'existence d'indications sur les éventuels changements intervenus dans les éléments méthodologiques ou dans le déroulement de la collecte;
- les références précises aux sources externes utilisées (ainsi les projections de population font normalement l'objet de publications);
- les informations concernant le service responsable de publication (pour demander d'éventuelles précisions).

D'une manière plus générale et pour juger du degré d'accessibilité d'un système statistique ou d'un domaine, il est possible:

- de faire une rapide recherche sur quelques mots clefs via Internet;
- de faire une recherche rapide à son propre niveau pour vérifier s'il existe des publications issues du système.

3.3 AMELIORER LA QUALITE STATISTIQUE

Que veut-on appuyer ?

Il est symptomatique de constater que le résultat de l'élaboration de nombreux cadres de mesure des performances se traduit en fin de document par la nécessité de renforcer le système²². Sans pour autant faire passer la statistique au rang de préoccupation nationale, un minimum d'appréciation de la performance du système est nécessaire. Normalement cette « évaluation » débouche sur des propositions d'amélioration de la qualité, mais trop souvent peu hiérarchisées et parfois contradictoires, comme par exemple : « plus de précision dans des délais brefs si possible à coûts constants ». Il est souvent impossible de gagner sur tous les tableaux et à l'instar de ce qui se fait en matière de Gestion des Finances Publiques (GFP), il faut éviter de pousser à la production quand les fondamentaux ne sont pas là : "get the basics right first". Ainsi, il n'est pas rare de voir des enquêtes très lourdes (quelques millions d'Euros) conçues sur des bases de sondages vieilles de 10 ans, voire beaucoup plus²³, ou l'installation d'un SIG (Système d'Information Géographique) dans des unités de production qui peinent à entretenir une simple base de données.

Les appuis pouvant être apportés à un système statistique ou à l'une de ses entités peuvent schématiquement être classés en deux catégories :

- les appuis à la **production des données** (méthodologie, traitement, analyse, diffusion), que cela soit en terme d'appui technique ou de financement (des opérations de collecte par exemple) ;
- les appuis au **renforcement des capacités statistiques**, c'est-à-dire aux conditions de fonctionnement des systèmes (formation, infrastructure, matériel, plans directeurs de développement, mais aussi gestion des INS et fonctionnement de la coordination nationale).

Il convient de ne pas oublier que le financement d'une enquête ne concoure que très rarement à l'augmentation des capacités statistiques. Il est certes évident que des personnes pourront acquérir des savoir-faire ou que le parc automobile ou informatique se trouvera renforcé, mais tout ceci reste éphémère²⁴ pour de multiples raisons²⁵ dont la principale demeure le fait que, lors d'une opération ponctuelle, l'ensemble des efforts est concentré sur la livraison du produit final : la donnée. Cela laisse peu de temps (surtout dans des systèmes où la pénurie de moyens est une constante) pour un réel développement à moyen terme.



Il convient donc, et cela sera particulièrement important, lorsque l'on fixera des objectifs spécifiques, de retenir que :

- *le financement de la production concoure à une augmentation de l'offre, mais non nécessairement à une augmentation des capacités ;*
- *le financement du renforcement des capacités ne débouche pas nécessairement sur une augmentation de l'offre en termes de quantité de données produites ;*
- *On peut, sur un même projet, financer les deux types d'appuis mais ils ne se déclinent pas dans les mêmes termes temporels et institutionnels.*

²² Il s'agit là d'une différence fondamentale avec les pays du Nord. Dans le cadre des différents élargissements de l'UE, la mise à niveau du système statistique (pour satisfaire aux standards européens) a constitué un préalable. La même stratégie est mise en œuvre pour les Balkans et les Pays du Voisinage.

²³ Dans ce cas, il est opportun de considérer ce coût en regard d'un Recensement Général de la Population et de l'Habitat (RGPH) et essayer plutôt de mobiliser l'aide pour obtenir une base de sondage nouvelle.

²⁴ Les évaluations de projets sont pratiquement unanimes

²⁵ Le personnel impliqué dans le projet est en grande partie du personnel contractuel, le matériel est obsolète ou hors d'usage en fin de projet, par exemple.

Ne pas oublier les utilisateurs

Comme souligné précédemment, il existe dans de nombreux contextes, des problèmes de culture statistique qui provoque une asymétrie très forte entre l'offre et la demande. D'où des difficultés à exprimer les demandes, à définir les niveaux d'exigences vis-à-vis des producteurs et plus simplement à utiliser les données, ce qui conduit à les considérer comme inutiles. On peut concevoir des actions de sensibilisation des utilisateurs à condition d'atteindre une masse critique relativement importante (il ne sert à rien de former deux hommes politiques ou deux syndicalistes) et surtout de traiter de sujets en prises directes avec la société (et non pas de sujets sur lesquels on souhaiterait que le public soit plus averti, ce qui peut se révéler substantiellement différent).

L'amélioration de la disponibilité : des effets rapides

On considère (particulièrement dans les PED, mais ceci est vrai dans de nombreux pays émergents) que les SSN produisent un tiers des données nécessaires au pilotage social et économique et que, sur ce tiers, un tiers font l'objet d'une diffusion (soit un neuvième au final). Il existe donc, bien plus que ce que l'on croit, des données disponibles. Il faut juste qu'elles soient publiées. Le problème est souvent que la cause du non traitement des données est la faiblesse quantitative des ressources humaines ou plutôt le fait qu'elles soient occupées en grande partie par les activités de collecte. Ce type d'appui se traduit donc souvent par de l'externalisation (assistance technique, sous-traitance à des instituts de recherche, entre autres).

L'amélioration de l'accessibilité : faire pression à travers la demande

Moins un pays est développé, plus la relation offre/demande est asymétrique avec une prépondérance de l'offre. En rendant les statistiques plus accessibles, on essaye de rééquilibrer cette asymétrie. Cela passe par l'appui aux utilisateurs qui est une activité à long terme. Toutefois, des actions peu coûteuses et rapides peuvent être engagées et auront des répercussions immédiates sur le jugement des performances : publication des métadonnées, amélioration du contenu des publications ou diversification des canaux de diffusion.

L'exactitude et la rigueur méthodologique : des travaux de longue haleine

Ces dimensions sont au cœur du travail statistique et nécessitent en matière d'amélioration de reprendre ou de fiabiliser tout un ensemble d'outils qui vont de la méthodologie aux outils informatiques et se traduisent par le recyclage du personnel, la recherche d'une stabilité des ressources humaines, l'investissement dans la recherche et développement ainsi que dans le matériel. Il faut garder en mémoire que, pour fiabiliser une production de données il faut plusieurs années.

La précision : pas de décision sans analyse coût / bénéfice.

L'augmentation de la précision, en dehors de quelques cas spécifiques, a nécessairement des impacts financiers non négligeables et se traduit en termes de coûts récurrents car il ne suffit pas de l'augmenter pour une collecte donnée mais pour toutes les enquêtes suivantes de même nature.



Le guide d'Eurostat "Guide to Statistics in EC Development Co-operation"²⁶ fournit une typologie des interventions d'appui aux systèmes statistiques ainsi que des supports à l'élaboration de termes de référence pour identifier et formuler de tels appuis (voir descriptif succinct en annexe 5).

²⁶ Cf. la description du guide sur le site d'Eurostat:

http://epp.eurostat.ec.europa.eu/portal/page/portal/international_statistical_cooperation/thematic_activities/communication_and_coordination/guide_to_statistics

ANNEXE 1: LEXIQUE

A

Accessibilité: dimension de la qualité statistique qui mesure sous quelles conditions les utilisateurs ont accès aux données.

C

Cadre de mesure des performances : un cadre de mesure des performances est constitué d'un ensemble de variables statistiques ou logiques dont on connaît la valeur et pour lesquelles on détermine un objectif chiffré en relation avec une action. *Performance Assessment Framework – PAF, en anglais.*

Capacité Statistique: ensemble des éléments (matériels et immatériels) qui doivent être mobilisés pour produire de l'information en continu et dans les délais impartis.

Cible: valeur finale que doit prendre l'indicateur au bout d'un laps de temps donné (en général du programme correspondant). *Attention, la cible peut être une valeur numérique, ou une appréciation (en hausse, en baisse).*

D

Disponibilité : dimension de la qualité statistique qui mesure ce qui est mis à disposition de l'utilisateur et sous quelle forme.

Données administratives: les variables statistiques issues de données administratives sont constituées de l'exploitation d'actes administratifs (état civil, douanes, greffes des tribunaux), ou de collecte réalisée auprès des administrations (recensements scolaires).

Donnée d'enquête : on parle de données d'enquêtes lorsque le résultat est obtenu à partir de la mesure sur un échantillon de la population. *Attention toutefois, il s'agit d'une dénomination abusive car les recensements sont des enquêtes mais qui mesurent une variable sur l'ensemble de la population.*

E

Erreur : valeur d'incertitude qui entoure une variable. Plus l'erreur est faible, plus le chiffre est précis.

Efficacité : c'est le fait de produire les résultats attendus de la politique publique. *Effectiveness en anglais.*

Efficience, de l'anglais *efficiency*: rendement ; c'est le rapport entre les ressources consommées (inputs) et les résultats (outputs).

Exactitude : l'exactitude d'un chiffre est basée sur l'observation du respect scrupuleux de la méthode employée depuis la collecte jusqu'au traitement. *Accuracy en anglais.*

F

Fiabilité : le chiffre fiable est celui qui est produit avec une qualité égale et connue en termes de précision, de délai de mise à disposition....

Fonction des indicateurs : définition de ce que l'indicateur mesure d'un phénomène étudié (Ex : le taux de scolarisation mesure le degré de couverture du système scolaire par rapport à la population).

Fréquence de mesure : pour une statistique donnée, mesure le laps de temps qui se passe entre deux collectes.

I

Initiative Paris 21 : initiative conjointe d'un ensemble de partenaires au développement, dont le siège est à l'OCDE, destinée à promouvoir la statistique publique et l'usage de bonnes pratiques.

L

Loi statistique : corpus législatif et réglementaire qui définit la statistique publique en termes de mandat et de déontologie.

O

Objectifs du Millénaire pour le Développement : ce sont huit objectifs que l'ensemble des Etats Membres des Nations Unies ont convenu de s'efforcer à atteindre d'ici à 2015 pour combattre la pauvreté. Chacun des objectifs s'accompagne de cibles et d'indicateurs spécifiques.

P

Pertinence d'un indicateur : la pertinence d'un indicateur est un des facteurs de qualité.

Précision : le chiffre précis est celui qui est connu avec une marge d'erreur compatible avec son utilisation. *Attention donc, un chiffre peut être suffisamment précis pour une utilisation et non pour une autre.*

Progression : valeur **estimée** que doit prendre l'indicateur année après année pour atteindre la cible finale.

Q

Qualité d'un indicateur : capacité de l'indicateur à remplir sa fonction de mesure dans le temps et dans l'espace. Il s'agit d'une notion synthétique qui a été décrite selon différents critères plus ou moins identiques (SMART, CREAM, ...).

R

Représentativité d'une statistique : elle définit l'écart qu'il existe entre la population enquêtée et la population de référence. Ex : un échantillon est dit représentatif d'une population lorsqu'il respecte la répartition d'une ou plusieurs caractéristiques de cette population. *Statistical significance en anglais.*

Responsabilité : devoir de rendre des comptes, de répondre de son action. *Accountability en anglais.*

Rigueur méthodologique : dimension de la qualité statistique qui mesure si les statistiques produites sont conformes aux bonnes pratiques statistiques.

S

Système Statistique National : ensemble de producteurs de statistiques publiques régi par une même loi.

T

Type d'indicateur : classement effectué selon la position de l'indicateur dans la chaîne logique d'un processus (Ex : Indicateurs d'intrants, de réalisation, de résultats d'impact).

U

L'utilité : dimension de la qualité statistique qui mesure l'adéquation de la production avec la demande (Ex : une donnée produite avec un grand retard est peu utile).

V

Valeur de référence : valeur prise par l'indicateur au moment où est mis en place un cadre de mesure des performances. *Attention, cette base n'est pas systématiquement la valeur de l'année en cours.*

ANNEXE 2 : POUR EN SAVOIR PLUS

Quelques documents indispensables sur la mesure de la performance

Mesure de la performance et incitations dans la gestion publique

Dominique Bureau, Michel Mougeot

© La Documentation française, Paris, 2007-ISBN:978-2-11-006636-7

<http://www.cae.gouv.fr/IMG/pdf/066.pdf> - extrait :

Si vous avez le temps

BRAUN, G., **Étude comparative portant sur la réforme de l'État à l'étranger**, Rapport d'information n° 348 (2000-2001) fait au nom de la commission des finances, déposé le 31 mai 2001 de M. Gérard BRAUN

<http://www.senat.fr/rap/r00-348/r00-3481.pdf>

UK National Audit Office Value for Money work:

http://www.nao.org.uk/what_we_do/value_for_money_audit.aspx

Déclaration de Paris sur l'efficacité de l'aide :

<http://www.oecd.org/dataoecd/53/38/34579826.pdf>

Le consensus européen pour le développement :

http://www.delsen.ec.europa.eu/fr/telechargements/edp_statement_oj_24_02_2006_fr.pdf

Le Programme d'Action d'Accra :

<http://siteresources.worldbank.org/ACCRAEXT/Resources/4700790-1217425866038/FINAL-AAA-in-French.pdf>

Le dialogue politique :

http://www.internationalhealthpartnership.net/CMS_files/documents/joint_assessment_tool_and_guidel_EN.pdf

Quelques documents indispensables sur les indicateurs

DESROSIERES, A, **Gouverner par les nombres : l'argument statistique**, T2, Presse de l'école des Mines ; 2008. Ce livre propose des études de cas précis, enquêtes sur le budget des familles, commissions du plan, statistiques locales ou comptabilité nationale, analysant la production des statistiques publiques et leurs usages par les autorités publiques. Bien que non ciblé sur les problématiques des PED, il s'agit d'un excellent ouvrage critique.

Concernant les généralités sur les indicateurs, les principaux sites de référence sont ceux :

- de l'OCDE : <http://www.ocdelibrairie.org/oecd> ;
- de la Banque Mondiale : <http://web.worldbank.org/WBSITE/EXTERNAL> ;
- du FMI : <http://www.imf.org/external/np/> ;
- d'Eurostat : <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>

Bien qu'un peu iconoclaste, le site Pénombre www.penombre.org publie régulièrement de courts articles sur l'usage des chiffres.

Si vous avez le temps

Agriculture

De très nombreux documents de référence concernant l'agriculture au sens large, sont regroupés sur la section documentaire du site de la FAO :

<http://www.fao.org/documents/>

Education

Le document le plus simple est celui réalisé par Claude SAUVAGEOT pour l'UNESCO dans le cadre du programme COMED, Dakar, juin 2001 :

www.dakar.unesco.org/sised/

D'un point de vue général, le site de l'organe statistique de l'UNESCO fournit l'essentiel des documents relatifs à la problématique des indicateurs et des statistiques liées à ce domaine :

<http://www.uis.unesco.org/>

Télécommunications

Le site de l'Union Internationale des Télécommunications fournit régulièrement des tableaux de bord chiffrés sur les indicateurs du secteur. Le dernier ouvrage paru accessible sur le site est : *Indicateurs des télécommunications/TIC africaines, 2008: A la croisée des chemins* :

<http://www.itu.int/publ/D-IND/fr>,

Quelques documents indispensables sur la qualité statistique

- 1) **Handbook on Data Quality Assessment Methods and Tools** -, Eurostat, 2007. Ce document définit les principales dimensions de la qualité utilisées dans le cadre du système statistique européen.
- 2) **Cadre d'assurance de la qualité**, Statistique Canada, 2002. Il s'agit du document politique qui engage Statistique Canada vis-à-vis du management de la qualité au sein du SSN.
- 3) **Quality concepts for official statistics**, Encyclopaedia of statistical sciences, John Wiley and son. Cet article reprend les principales définitions utiles de la qualité statistique.

Si vous avez le temps

FMI

<http://dsbb.imf.org/Applications/web/dqrs/dqrswork/>, on trouvera sur le site du FMI de nombreux rapports pays concernant l'évaluation de la qualité statistique, notamment dans les domaines couverts par cette organisation.

Statistiques Canada

<http://www.statcan.ca/>, le site de Statistiques Canada est un des plus fourni en matière de mise en ligne de documentation, l'approche qualité utilisée par cet office est probablement la plus développée qui soit

Eurostat

<http://epp.eurostat.ec.europa.eu/portal/page>, on trouvera sur le site d'Eurostat tous les documents et les rapports concernant l'approche qualité au sein du système statistique européen.

Paris 21

<http://www.paris21.org/pages/designing-nsds/NSDS-documents-knowledge-base/index.asp?tab=KnowledgeBase&option=doc>, on trouvera sur le site de Paris 21, sous l'onglet « Knowledge base », la plupart des documents de référence concernant le sujet

Organisation internationale de normalisation

www.iso.org/iso/fr/

Quelques expériences dans les pays du sud

<http://www.who.int/healthmetrics/>, on peut trouver sur ce site les résultats des évaluations qualité des systèmes d'information sanitaires

<http://www.uis.unesco.org/>, il s'agit du site de l'Institut Statistique de l'Unesco (ISU) qui procède à des évaluations qualité des systèmes d'information sur l'éducation. On trouvera sur le site l'approche dérivée du cadre FMI/Banque mondiale utilisée par l'ISU.

ANNEXE 3: QUELQUES QUESTIONS A POSER FACE A UN INDICATEUR

Questions	Réponses
Qui est responsable de son calcul ?	Question primordiale, puisque si des doutes existent sur sa qualité, je dois pouvoir savoir où demander des explications.
Quelle est le dispositif de collecte qui a permis d'obtenir les données ?	Soit il s'agit de collecte administrative, soit il s'agit d'enquêtes, parfois les deux. Ce ou ces dispositifs fournissent-ils aux utilisateurs des éléments sur les méthodes et les résultats ? Si oui, où me les procurer?
A quel stade de traitement le chiffre est-il rendu ?	S'agit-il d'une donnée provisoire ou définitive ? Dans ce dernier cas, quelle est la publication dans laquelle cet indicateur a été diffusé ? Un chiffre officiel est celui qui est publié, qu'il s'agisse d'une publication, d'un communiqué de presse ou d'une mise en ligne sur Internet.
Données d'enquêtes : quelle est la précision, même théorique de la valeur ?	A défaut d'obtenir des publications méthodologiques qui justement définissent ces ordres de grandeur, il faut quelques réflexes de bon sens : <ul style="list-style-type: none"> ▪ une précision autour de 1% est quelque chose d'exceptionnel, même dans les pays du Nord. ▪ dans les PED, la plupart des enquêtes auprès des ménages ont des précisions souhaitées situées en dessous de 5 %. En pratique, au niveau national, elles se situent entre 5 et 10 %, puis la précision se dégrade très rapidement dès que l'on désagrège géographiquement (au second niveau de découpage géographique, par exemple régional ou provincial, il n'existe plus aucune signification de la valeur). <p>Le seul cas où la précision n'intervient pas dans la comparaison est celui où l'échantillon (c'est-à-dire l'individu questionné) est le même entre deux enquêtes.</p>
Collecte administrative : sur quelle base le calcul est-il fait ?	Vérifier si la publication indique le taux de réalisation de la collecte, c'est-à-dire le rapport entre le nombre d'unités du registre administratif qui sert à identifier l'individu enquêté et le nombre de réponses obtenues (complétude).
La collecte s'est-elle faite dans les mêmes conditions ?	Il faut souvent vérifier si la collecte s'est déroulée exactement aux mêmes périodes qu'il s'agisse d'enquête ou de certaines collectes administratives qui ne se déroulent pas en continu (exemple : enquêtes scolaires). Les budgets mobilisés sont-ils identiques d'une collecte à l'autre : il est matériellement impossible d'assurer une qualité égale de résultat en dépensant 3 fois moins d'argent (à moins de changer de méthode).
La méthode de calcul est elle la	Lorsque les données sont publiées, toujours se référer à la

<p>même entre deux livraisons de la valeur de l'indicateur ?</p>	<p>publication et rechercher dans celles-ci les éléments indiquant un possible changement.</p> <p>Il faut se méfier des sources « vagues ». La source doit faire référence au dispositif de collecte, et non à une entité administrative (Ministère ou autre).</p>
<p>Les variations entre deux valeurs sont-elles plausibles ?</p>	<p>Là encore, à défaut d'explications données par le producteur, sur la nature exceptionnelle de la variation, il faut quelques réflexes de bon sens.</p> <p>Pour toutes les variables qui touchent à des aspects fondamentaux d'une économie ou d'une société, des variations très fortes d'une année à l'autre cachent très fréquemment des changements de méthodes.</p>
<p>Quelle est la valeur de l'unité ?</p>	<p>Pour tous les indicateurs exprimés sous forme de ratio, de % ou d'indice, il faut toujours garder en mémoire la valeur de l'unité : dans 71%, quelle est la valeur absolue qui correspond à 1%?</p> <p>Ce simple calcul peut permettre d'éviter des conclusions erronées sur l'atteinte des cibles : faire progresser un taux de scolarisation de 1% est énorme en termes d'effectifs de nouveaux inscrits et donc d'efforts budgétaires associés. A 0,9 % de progression, l'effort a été fait même si la cible n'est pas formellement atteinte.</p>
<p>La tendance de variation est-elle connue ?</p>	<p>Encore quelques réflexes de base :</p> <ul style="list-style-type: none"> ▪ entre deux valeurs obtenues à quelques années d'intervalle, on ne peut tracer une droite qui les relie, ni conclure à la hausse ou la baisse de manière tendancielle ; ▪ pour pouvoir effectivement conclure à une tendance, il faut au minimum une dizaine d'années d'observations (et si possible, plus).

ANNEXE 4 : LES CADRES QUALITE UTILISES PAR LA BANQUE MONDIALE/FMI ET EUROSTAT

Les approches Banque Mondiale/FMI et Européenne comportent quelques différences:

- la démarche européenne est avant tout une démarche normative (dans le sens où elle a force de loi) visant à qualifier le résultat. En conséquence, les dimensions de la qualité sont avant tout méthodologiques et techniques;
- la démarche de la Banque Mondiale et du FMI est plus descriptive et va dans le sens d'une démarche de qualité totale, à savoir qu'elle s'intéresse certes à la qualité des chiffres, mais aussi à l'ensemble des processus de gestion de cette qualité.

Système FMI	Système EUROSTAT
<p>0 Conditions préalables de la qualité</p> <p>0.1 <i>Cadre juridique et institutionnel favorable</i></p> <p>0.2 <i>Ressources à la mesure des besoins</i></p> <p>0.3 <i>Pertinence de l'information statistique sur le domaine</i></p> <p>0.4 <i>Gestion de la qualité</i></p> <p>1 Assurance d'intégrité</p> <p>1.1 <i>Professionalisme comme principe fondamental</i></p> <p>1.2 <i>Transparence des politiques et pratiques statistiques</i></p> <p>1.3 <i>Normes déontologiques</i></p>	<p>1 Indépendance professionnelle</p> <p>2 Mandat pour la collecte de données</p> <p>3 Ressources adéquates</p> <p>11 Pertinence</p> <p>4 Engagement vers la qualité</p> <p>5 Confidentialité</p> <p>6 Impartialité et objectivité</p> <p>9 Charge non excessive pour les répondants</p> <p>10 Efficacité et efficacité</p>
<p>2 Rigueur méthodologique</p> <p>2.1 <i>Concepts et définitions conformes aux normes internationales</i></p> <p>2.2 <i>Champ d'application conforme aux normes internationales</i></p> <p>2.3 <i>Classification/ sectorisation conformes aux normes internationales</i></p> <p>2.4 <i>Base d'enregistrement</i></p> <p>3 Exactitude et fiabilité</p> <p>3.1 <i>Données de base appropriées</i></p> <p>3.2 <i>Vérification des données de base</i></p> <p>3.3 <i>Techniques statistiques saines</i></p> <p>3.4 <i>Données intermédiaires et produits statistiques évalués et validés</i></p> <p>3.5 <i>Études de révision</i></p>	<p>7 Méthodologie saine</p> <p>8 Procédures statistiques adaptées</p> <p>12 Précision et fiabilité</p>
<p>4 Utilité</p> <p>4.1 <i>Périodicité et délais de diffusion acceptables</i></p> <p>4.2 <i>Cohérence des données</i></p> <p>4.3 <i>Politiques et pratiques de révision</i></p> <p>5 Accessibilité</p> <p>5.1 <i>Accessibilité des données</i></p> <p>5.2 <i>Accessibilité des métadonnées</i></p> <p>5.3 <i>Assistance aux utilisateurs</i></p>	<p>13 Opportunité et ponctualité</p> <p>14 Cohérence et comparabilité</p> <p>15 Accessibilité et clarté</p>

ANNEXE 5 : DESCRIPTIF SUCCINCT DU GUIDE EUROSTAT

Le Guide intitulé “*Guide to Statistics in EC Development Co-operation*” a été réalisé en version test en 2009. Si la première version est pour l’instant en anglais des versions françaises et espagnoles devraient voir le jour prochainement, ainsi qu’une version électronique qui sera consultable sur le site d’Eurostat.

Le Guide comprend quatre parties principales :

- la première est entièrement consacrée à la place des statistiques dans les politiques de développement et on retrouvera ainsi les rappels concernant les OMD ainsi que les principales déclarations ou principes évoqués dans la partie 1 du présent guide : déclaration de Marrakech, de Paris. Cette partie décrit aussi en détail ce qu’est un schéma directeur national de développement des statistiques tels que mis en œuvre à travers l’initiative Paris 21.
- la seconde partie intitulée ‘Comment sont faites les statistiques » rappelle les grands principes des statistiques publiques et les caractéristiques des grandes opérations statistiques. Elle permet aussi de comprendre comment fonctionne un système statistique national et définit les principales capacités statistiques.
- la troisième partie, donne un certain nombre de conseils sur l’appui qui peut être apporté aux systèmes statistiques des pays bénéficiaires de l’aide. Elle reprend les principaux concepts de qualité statistique et des moyens existants pour améliorer cette qualité. Elle propose aussi un ensemble de *guidelines* pour réaliser des appuis /Termes de référence « standard », adapté à différents types d’appuis.
- la dernière partie traite plus en détail des statistiques sectorielles et de comment celles-ci peuvent être utilisées et soutenues dans une perspective de prise de décision et de suivi des résultats.

Le Guide est richement documenté et renvoie à l’intégralité des sources d’information concernant les statistiques que ces sources soient *on line* ou *off line*.