

L'importance de la qualité des données dans le renforcement des systèmes d'information des ressources humaines (SIRH)

Samwel Wakibi, IntraHealth International

Introduction

Les organisations dépendent des données ; aussi bien aujourd'hui, pour accroître leur rendement, qu'à l'avenir, pour améliorer leur mode de planification. (Rochnik et Dijcks, 2006). L'utilisation de données de mauvaise qualité se traduit par une perte de temps et d'argent mais aussi une perte de confiance de la part des clients. Elle peut, dès lors, placer une organisation dans une position inconfortable. On estime que le taux d'erreur actuel sur l'ensemble des données utilisées par les entreprises se situe entre 1% et 5%, soit une perte de revenu de 10% (Redman, 1996). Dans la mesure où les données sont collectées, analysées et interprétées dans des rapports d'une importance significative, en vue de la planification et de la prise de décisions, les problèmes relatifs à la qualité des données peuvent apparaître à mesure que les informations pénètrent dans les organisations et dans les différents systèmes.

Les questions relatives à la qualité des données sont au cœur des programmes du Capacity Project, un projet global financé par l'USAID, qui aide à renforcer les ressources humaines pour la santé (RHS) des pays en développement, afin de mieux répondre aux défis liés à la mise en place et à la pérennité de programmes sanitaires de qualité. En se focalisant sur le renforcement des systèmes d'information des ressources humaines (SIRH), le Projet a, dans ce domaine, beaucoup gagné en expérience.

Ce résumé technique a pour but de décrire les différents concepts liés à la qualité des données, à base d'exemples démontrant l'importance de leur gestion, en particulier dans le cadre des RHS, tout en illustrant ces réflexions par l'expérience du Capacity Project dans le renforcement des SIHR des pays en voie de développement.

Concepts de base sur la qualité des données

Une des définitions couramment acceptées de la qualité des données, en économie, dans le monde des affaires et en médecine est l'« adéquation d'utilisation », aussi bien aujourd'hui qu'à l'avenir. En d'autres termes, cela consiste à savoir dans quelle mesure les données peuvent satisfaire les besoins des consommateurs et si elles correspondent à leurs attentes. Pour que les données soient utilisables, elles doivent être exemptes de répétitions et de variations inutiles ainsi que de fautes d'orthographe, tout en étant complètes et en respectant une structure définie (Chapman, 2005 ; Carson, 2000 ; Brown, Stouffer et Hardee, 2007).

La qualité des données repose sur :

- L'exactitude
- L'opportunité
- La cohérence
- La précision
- La complétude
- La fiabilité

- **L'exactitude** renvoie à la justesse des valeurs mesurées, les observations ou les estimations de la valeur réelle ou de la valeur vraie, sans préjugé politique ou personnel ni manipulation. En d'autres termes, l'exactitude est une mesure qui sert à déterminer à quel point les données reflètent la réalité. L'exactitude dépend de la fiabilité des sources de données et de leur processus de collecte.
- **L'opportunité** fait référence à la disponibilité des données recherchées. Elle est liée à d'autres facteurs comme la connaissance de la période à laquelle elles ont été collectées, leur dernière mise à jour, la durée pendant laquelle elles seront d'actualité et leur mise à disposition dans un délai raisonnable, afin de mener à bien une entreprise et de prendre des décisions en connaissance de cause.
- **La cohérence** se définit par l'absence de contradictions apparentes et sert à mesurer, en interne, la validité et la fiabilité des données. Pour l'évaluer, on détermine si les définitions, les codes et les formats ont été suivis de manière identique pour les mêmes données à travers différentes sources.
- **La précision** renvoie à la cohérence d'un indicateur dans la production de résultats identiques. Par exemple, un formulaire de collecte de données d'une grande précision suscitera les mêmes réponses s'il est administré à plusieurs reprises à la même personne. La précision diffère de l'exactitude en ce qu'une mesure peut être précise sans pour autant être exacte. Par exemple, celle-ci peut générer des réponses incorrectes de manière répétée.
- **La complétude** fait référence à l'absence d'erreurs ou d'omission, comme des oublis d'enregistrement dans un groupe de données ou une variable sans données. Cette dimension permet de savoir si toutes les données recevables sont incluses.
- **La fiabilité** renvoie à la disponibilité des détails et des données nécessaires. Elle permet de répondre à des questions liées au développement de la base de données ou du formulaire de collecte. Par exemple, les données adéquates sont-elles collectées et stockées en vue des entreprises actuelles et à venir ? Dispose-t-on de suffisamment de données pour répondre aux questions soulevées par nos activités ? La règle par défaut est que si l'application de l'information est inconnue, le plus haut niveau de détail doit être fourni. Cela confère une fiabilité supplémentaire aux données, dans la mesure où il est plus facile d'éliminer les doublons ou les parties inutiles que de rechercher les pièces et les morceaux manquants (E-gouvernement néo-zélandais, 2004).

Expérience sur le terrain

Le renforcement des SIRH englobe cinq champs d'activité : le développement du leadership participatif des parties prenantes ; le renforcement de l'infrastructure des technologies de l'information et des communications ; le développement et la diffusion de logiciels et la mise en place de formations dans ce domaine ; le développement continu de l'initiative ; et l'amélioration des compétences dans des domaines annexes comme la prise de décision basée sur des données.

Le Capacity Project a contribué aux activités de renforcement des SIRH en Ouganda, au Rwanda, en Tanzanie, en Namibie, au Sud Soudan, au Swaziland, en Afrique du Sud et au Lesotho et a lancé des activités similaires dans d'autres pays en voie de développement. Au cours de ce processus, le Projet s'est intéressé aux questions liées à la qualité des données, allant de l'exactitude à la complétude, en passant par la cohérence et l'opportunité des données et des informations produites.

Les défaillances identifiées dans la qualité des données ont été attribuées à l'utilisation de technologies et

de procédures d'exploitation inadéquates, au manque de compétences en matière de gestion de données, au manque de sensibilisation à leur qualité, ainsi qu'à une mauvaise évaluation de leur importance dans le processus décisionnel. Les autres problèmes liés à la qualité des données sont

le résultat d'influences environnementales, comme l'enregistrement erroné d'informations essentielles, la perte de documents pendant une guerre, un conflit civil ou un incendie, ainsi que la dégradation de documents causée par un mauvais entreposage ou des intempéries.

Au fur et à mesure que les données défilent dans la chaîne de gestion des informations, les questions relatives à leur qualité gagnent en importance et le coût de correction des erreurs augmente (voir figure 1). En fin de compte, s'ils ne disposent pas d'informations de qualité, les responsables sanitaires finissent par prendre des décisions malencontreuses. Cela aboutit à des problèmes de rentabilité comme une répartition inégale des effectifs, l'apparition de « travailleurs fantômes » (d'anciens employés continuant de percevoir une rémunération alors qu'ils ne travaillent plus), ou même des retards dans le versement des salaires, comme cela a pu être observé dans certains systèmes de ressources humaines de pays bénéficiant du soutien du Capacity Project.

Etudes de cas

Les exemples suivants sont tirés de l'expérience du Capacity Project sur le terrain.

Répétitions inutiles et données manquantes : Ouganda. Le registre du Conseil des Infirmières et Sages-femmes Ougandaises (UNMC) regroupe des

données concernant les infirmières, les sages-femmes et les étudiantes se préparant à exercer ces métiers en Ouganda. L'UNMC maintient des données relatives au nombre d'inscriptions dans les programmes de formation initiale (auxquelles on fait référence par le terme d'indexation), à l'examen des résultats obtenus par les infirmières et les sages-femmes étudiantes et à l'enregistrement, l'octroi de l'autorisation d'exercer et la reconduction des infirmières et des sages-femmes qualifiées. Avec l'aide du Capacity Project, le registre papier a été revu et informatisé.

Au cours du développement de la base de données, des identifiants uniques, tels que les indices, se répétaient inutilement. Auparavant, ceux-ci étaient générés manuellement et étaient assignés par différentes personnes via un processus mal coordonné. Pour remédier à ce problème, le système informatique a été modifié, permettant ainsi de générer, en interne, un numéro d'identification personnel unique, afin de déterminer le nombre d'étudiantes et d'infirmières qualifiées formées en Ouganda. La disparition ou l'incomplétude des rapports, ainsi que la perte ou la disparition de dossiers sont d'autres défis à surmonter pour garantir la qualité des données.

Grâce à la nouvelle base de données, l'UNMC est à présent en mesure de rapporter le nombre d'infirmières formées dans le pays, leur taux de fidélisation, et le nombre total d'étudiantes dans ce domaine par district.

Incohérence des données : Lesotho. Le Capacity Project a permis au Ministère de la Santé et des Affaires Sociales du Lesotho (MOHSS) d'établir une base de données, afin de gérer les effectifs sanitaires, et procède à présent à sa mise en œuvre. Les données à gérer figurant dans la base de données sont partagées entre le Ministère des Finances (salaires du personnel), le MOHSS (traçabilité et répartition des prestataires) et le Ministère des Services Publics qui joue le rôle de recruteur.

Durant l'élaboration de la base de données, notre équipe a découvert que ces trois ministères maintenaient des systèmes de codage parallèles pour les postes établis. Les ministères classaient également les unités au sein du MOHSS de manière différente, tant et si bien qu'un « programme » ou un « département » dans un ministère donné pouvait être identifié comme un « sous-programme » dans un autre.

Bien qu'une solution ait été trouvée à ce problème, celle-ci a généré des doublons dans la base de données. Afin d'éliminer ces redites, l'ordinateur a combiné les trois groupes de code, en interne, en attribuant un numéro spécifique à chaque poste.

Opportunité de l'information et des données : Swaziland. Dans ce pays, le Capacity Project a développé et mis en place une solution « par étapes », sous la forme d'une base de données moins élaborée permettant de satisfaire des besoins à court terme. Le Ministère de la Santé et de la Prévoyance Sociale est, à présent, en mesure de mettre régulièrement à jour les données concernant les effectifs.

Le processus de développement de la base de données a révélé que le système manuel générait des retards. Ceux-ci avaient pour conséquences de voir des retraités et d'anciens employés, qui avaient quitté leur poste de longue date, continuer à toucher leur

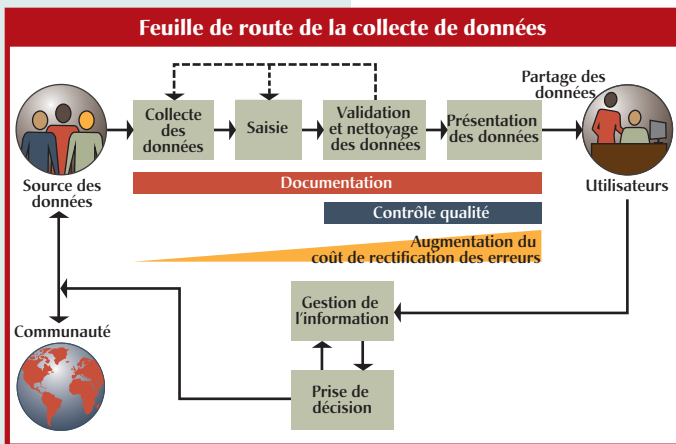


Figure 1: Collecte des données
(Adapté de Chapman, 2005)

salaires, tandis que les nouveaux employés devaient attendre longtemps avant de recevoir le leur. On a aussi découvert qu'il n'y avait pas de suivi des contrats et que les paiements n'étaient pas exécutés à temps. Cela a abouti à des poursuites judiciaires entamées par les nouveaux employés et les agents contractuels à l'encontre du gouvernement. Pour ne rien arranger, le fait que le système ne s'aperçoive que tardivement de l'existence de postes vacants et qu'il tarde à les attribuer a empêché les nouvelles recrues d'y avoir accès, en réduisant ainsi considérablement le personnel disponible pour la prestation de soins.

La nouvelle base de données a servi à valider les données relatives à la masse salariale et à gérer efficacement les mouvements et les pénuries de personnel. En outre, les décisions ayant trait au recrutement ou au transfert d'effectifs sont maintenant prises en connaissance de cause.

Les meilleures pratiques dans le renforcement des SIRH

Les meilleures pratiques en matière de renforcement des systèmes d'information sont évaluées en termes de qualité ajoutée, de temps gagné et de réduction des coûts de gestion des données (Ministère américain de la Défense, 1999 ; NHS Connecting for Health, 2007). De plus, le Capacity Project veille à la qualité des données tout en soutenant le renforcement des SIRH. L'approche du projet comporte un protocole d'assurance qualité garantissant l'exactitude, l'opportunité, la cohérence, la précision, la complétude et la fiabilité des données relatives aux besoins spécifiques des clients.

Ces meilleures pratiques sont basées sur les expériences de terrain obtenues en travaillant avec les données provenant de différents SIRH de pays situés à l'Est, au Centre et au Sud du continent africain. Les paragraphes suivants présentent les approches et les domaines essentiels qui ont été adoptés.

Développement du système

Les bases de données mises en place sont longitudinales. En d'autres termes, elles contiennent des informations à caractère personnel et professionnel ainsi que les évolutions constatées au sein du personnel de santé. Elles couvrent la période s'étalant de la formation à la fidélisation pour garantir la complétude et l'exactitude des données et ainsi répondre aux questions de politique et de gestion relatives au personnel de santé.

Le Capacity Project emploie des technologies peu coûteuses et efficaces visant à appuyer le développement des SIRH. La préférence est donnée aux logiciels à sources ouvertes, dans la mesure où ceux-ci sont disponibles gratuitement et ont la capacité de pouvoir gérer efficacement le volume de données anticipé et son traitement. Les logiciels à sources ouvertes sont des logiciels informatiques distribués gratuitement sous une licence qui permet à tout le monde de copier, d'utiliser et de modifier le code source et de le redistribuer, sans restriction, sous cette forme ou sa forme initiale. Les bases de données à sources ouvertes développées, afin de gérer les données liées aux ressources humaines, sont accessibles en ligne. Cela permet un nettoyage et une saisie externe des données qui réduit les délais de mise à jour.

Les systèmes d'informations sont renforcés par étapes planifiées pour fournir des améliorations rapides sans pour autant surcharger l'infrastructure de base. Les

systèmes de codage des organisations sont, d'autre part, harmonisés afin de garantir la cohérence entre les différentes bases de données et les différents modules. Pour ce faire, chaque poste se voit attribuer un identifiant unique. Le système de codage synchronisé garantit également que les recherches conduites sur différentes bases de données aboutissent aux mêmes résultats.

Afin de minimiser les erreurs d'enregistrement des données, deux modes de saisies sont utilisés : le dédoublement de la saisie et le contrôle aléatoire. Un système de double saisie est un élément important qui est intégré au développement de la base de données.

■ *Le dédoublement de la saisie* consiste à entrer le même contenu à deux moments différents et, si possible, par deux individus distincts. La deuxième personne est amenée à passer en revue les postes présentant des divergences entre les deux saisies et à choisir le bon en fonction des données originales. Cette méthode est la plus exacte dès lors qu'il s'agit de garantir la qualité des données mais elle nécessite deux fois plus de temps et d'argent qu'une saisie simple.

■ *Le contrôle inopiné* consiste à générer une liste aléatoire de données afin d'en vérifier l'exactitude en la faisant comparer aux données originales par une deuxième personne ou un supérieur. Cette méthode est nécessairement moins exacte mais elle permet d'éviter une multiplication des coûts et réduit le temps passé à la saisie.

Les deux méthodes permettent d'enregistrer et d'assurer la traçabilité des erreurs de saisie. De même, l'analyse de ce registre peut servir à améliorer la formation du personnel de saisie, à changer les procédures d'exécution, ainsi qu'à modifier les logiciels ou l'interface utilisateur.

Sécurité des données et de l'infrastructure

Afin de garantir la sécurité des données, des mots de passe et des droits d'accès ont été délivrés afin de limiter l'accessibilité au personnel autorisé. Par exemple, le personnel de saisie n'a accès qu'à un éventail de données très limité, ce qui est nécessaire au maintien de la qualité, tandis que les analystes et les administrateurs disposent d'un accès élargi. De plus, l'accès physique à la salle des serveurs qui hébergent les bases de données est contrôlé et un système de sauvegarde a été instauré.

Gestion

Quand cela est nécessaire, les formulaires de collecte de données sont reconçus pour garantir leur précision et leur complétude. Cela se fait en collaboration avec des équipes locales, de telle sorte que les pays puissent rapporter leurs données comme prévu et que le fardeau que représente leur collecte (une importante source d'erreur indirecte) soit moins lourd à porter.

Les agents de collecte de données et l'équipe de gestion des données travaillent en étroite collaboration afin de mieux comprendre le travail qui incombe à chacun. Ce feedback permet d'améliorer la qualité des données au point de collecte, minimisant ainsi le besoin de correction ultérieure dans la chaîne de gestion de l'information.

La collecte des données est tenue à jour, encodée de manière unique, traitée régulièrement et rendue accessible à tous ceux qui en ont besoin sur un réseau d'hébergement. Toutefois, l'accès est limité au personnel autorisé afin de garantir la confidentialité et

Si vous cherchez des informations sur les RHS et ses différents outils ou que vous voulez partager vos connaissances et nous donner vos contributions, rendez-vous au Centre de Ressources RHS. Pour les personnes qui travaillent aux niveaux national ou mondial, le Centre de Ressources RHS fournit des informations visant à :

- Améliorer la planification stratégique et la prise de décisions
- Enrichir vos comptes rendus et vos présentations
- Soutenir l'action en faveur des RHS
- Améliorer le développement professionnel
- Gagner du temps.



The Capacity Project
 IntraHealth International, Inc.
 6340 Quadrangle Drive
 Suite 200
 Chapel Hill, NC 27517
 Tel. (919) 313-9100
 Fax (919) 313-9108
 info@capacityproject.org
 www.capacityproject.org

Les opinions exprimées dans cette publication ne reflètent pas nécessairement celles de l'Agence des États-Unis pour le Développement international ou du Gouvernement des États-Unis.

la sécurité des données. Cette disponibilité sécurisée garantit :

- La tenue à jour de la base de données par la décentralisation de la saisie.
- L'élimination de bases de données dissipées et déstructurées.

Nettoyage des données

Les données provenant d'autres sources, comme des enquêtes ou d'autres bases de données, servent à détecter des données corrompues ou inexacts. Les données sont corrigées ou retirées des registres ou de la base par souci de cohérence avec des groupes de données similaires dans le système. Les processus de nettoyage des données et les stratégies de prévention des erreurs ont lieu simultanément avec la collecte, le traitement et la saisie des données, afin de renforcer l'assurance qualité des données.

Saisie des données

Afin de minimiser les erreurs de saisie, les documents originaux sont utilisés comme sources principales dans la création de registres dans la base données. La saisie clavier est minimisée et accélérée par l'utilisation de menus déroulants qui rendent l'ensemble plus cohérent. La saisie est décentralisée et entreprise plus près du point d'origine des données ; cela fait partie de la routine des agents de collecte.

Formation et développement des compétences

Des sessions de formations sont délivrées aux membres du personnel, dont la performance est

continuellement contrôlée. Ces formations à caractère professionnel ciblent principalement les agents de collecte, les opérateurs de saisie et les responsables qui guident le développement et l'utilisation des rapports. Les membres du personnel sont aussi activement impliqués dans le développement et la mise en place des bases de données.

Conclusion

La qualité des données se définit le mieux par son *adéquation d'utilisation* dans la prise de décision politique ou autre. On mesure la qualité des données par leur exactitude, leur opportunité, leur cohérence, leur précision, leur complétude et leur fiabilité.

Le Capacity Project a recherché et a adopté les meilleures pratiques visant à garantir la qualité des données dans son travail de renforcement des SIRH. Le Projet fournit aussi aux parties prenantes des compétences en matière de développement de systèmes et de gestion de données, afin de garantir la qualité, la pérennité et l'amélioration continue des systèmes mis en place. Les responsables du secteur de la santé sont aussi formés à l'utilisation des données, afin de prendre des décisions en connaissance de cause. Grâce à des données de haute qualité, des systèmes efficaces et opportuns et une formation globale à leur utilisation, notre objectif est de voir ces pays pouvoir rapidement se servir des données probantes récoltées et des politiques pour surmonter les défis sanitaires qu'ils rencontrent.

Références

Carson C. What is data quality? a distillation of experience. Washington, DC: Statistics Department, International Monetary Fund, 2000. Accessed 15 June 2007 at: <http://www.thecre.com/pdf/imf.pdf>

Chapman A. Principles of data quality, version 1.0. Copenhagen, Denmark: Global Biodiversity Information Facility, 2005. Accessed 7 July 2008 at: <http://circa.gbif.net/irc/Download/kjeYAKJSmRGFqwAaUY4x8KZ1jH4pYxtv/F37w1fUI4R0AgTiySEZtff0yRVsbNGn/Data%20Quality.pdf>

Eckerson W. Data quality and the bottom line: achieving business success through a commitment to high quality data. TDWI Report Series. Chatsworth, CA: The Data Warehousing Institute, 101 Communications, 2002. Accessed 7 July 2008 at: <http://download.101com.com/pub/tdwi/Files/DQReport.pdf>

New Zealand E-government Unit. New Zealand E-government Interoperability Framework (NZ e-GIF), Version 2.1. Wellington, New Zealand: New Zealand E-government Unit, 2004. Accessed 6 August 2007 at: <http://www.e.govt.nz/standards/e-gif/e-gif-v-2-1/e-gif-v2-1.pdf>

NHS Connecting for Health, Department of Health (UK). Best practice guidelines. Leeds, England: NHS Connecting for Health, 2007. Accessed 15 June 2007 at: http://www.connectingfor-health.nhs.uk/systems_and_services/data/clinicalcoding/data_quality/best_practice

Redman T. Data quality for the information age. Boston, MA: Artech House Publisher, 1996.

Rochnik N, Dijkstra J. Oracle Warehouse Builder 10gR2: transforming data into quality information. An Oracle whitepaper. Redwood Shores, CA: Oracle, 2006. Accessed 6 August 2007 at: <http://www.oracle.com/technology/products/warehouse/pdf/transforming%20data%20into%20quality%20information.pdf>

United States Department of Defense. Best practices for data quality oversight of environmental sampling and testing activities. Final Report. United States Department of Defense, 1999. Accessed 6 August 2007 at: <http://www.navylabs.navy.mil/Archive/BP-1999.pdf>

Brown W Stouffer R, Hardee K. Data quality assurance tool for program-level indicators. Chapel Hill, NC: MEASURE/Evaluation Project, University of North Carolina at Chapel Hill, 2007. Accessed 7 July 2008 at: <http://www.pepfar.gov/documents/organization/79628.pdf>

Partenariats du Capacity Project

