

## Data Quality Considerations in Human Resources Information Systems (HRIS) Strengthening

Samwel Wakibi, IntraHealth International

### Introduction

Organizations rely on data to increase their efficiency today and improve their planning for tomorrow (Rochnik and Dijcks, 2006). Poor data quality results in loss of time, money and customer confidence, and can be cause for embarrassment to an organization. It is estimated that the typical industrial data quality error rate of 1%–5% can constitute a 10% loss in revenue (Redman, 1996). As data are collected, analyzed and translated into meaningful reports for planning and decision-making, data quality problems can occur as information crosses organizational and system boundaries.

Data quality issues have been central to the program experience of the Capacity Project, a USAID-funded global project that helps developing countries strengthen human resources for health (HRH) to better respond to the challenges of implementing and sustaining quality health programs. Experience with issues of data quality has particularly been gained within the Project's focus on strengthening human resources information systems (HRIS) to support health workforce planning and management.

This brief will discuss concepts of data quality and provide examples of the importance of data management specific to the field of HRH, illustrated by the Capacity Project's experience with HRIS strengthening in developing countries.

### Basic Data Quality Concepts

One widely accepted definition of *data quality* in economics, business and medicine is "fitness for use" now and in the future—in other words, how well data meet user needs and expectations (Chapman, 2005; Carson, 2000). Data quality describes the state of data, the set of processes to achieve such a state and data accuracy. For data to be fit for use, they should be free of duplications, misspellings, omissions and unnecessary variations, and should conform to a defined structure (Chapman, 2005; Carson, 2000; Brown, Stouffer and Hardee, 2007).

Data quality addresses:

- Accuracy
- Precision
- Timeliness or currency
- Completeness
- Consistency
- Relevance.

- *Accuracy* refers to closeness of measured values, observations or estimates of the real or true value, without political or personal bias and manipulation. In other words, accuracy is a measure of the extent to which the data reflect reality. Guiding questions to achieve accuracy relate to the reliability of data sources and the process of data collection.
- *Timeliness or currency* refers to availability of data when required. Related factors are knowledge about the period when the data were collected, when they were last updated, how long they are likely to remain current and whether they are processed to give information in time to conduct daily business or inform decisions.
- *Consistency* describes the absence of apparent contradictions and is a measure of internal validity and reliability. Guiding questions to assess consistency include the extent to which the same definitions, codes and formats are followed for the same data across different sources.
- *Precision* refers to the consistency of an indicator in producing the same results. For example, a data collection form with high precision will elicit the same responses if administered repeatedly on a subject. Precision and accuracy differ in that a measure can be precise without being accurate. For example, a measure can repeatedly generate the same incorrect outcomes.
- *Completeness* refers to lack of errors of omission, such as omitted records in a dataset or a variable without data. Completeness addresses the question of whether all eligible data are included.
- *Relevance* refers to availability of required details or data. It helps to answer questions relating to the design of the database or the data collection form. For example, are appropriate data being collected and stored for current and future business initiatives? Do records collected have enough details to answer intended business questions? The default rule is that if the application of the information is unknown, the highest level of detail should be provided. This gives some additional reliability to the data, as it is easier

to discard redundant or unnecessary parts than to search for missing bits and pieces (New Zealand E-government Unit, 2004).

### Field Experience

HRIS strengthening encompasses five activity areas: participatory stakeholder leadership development; information and communications technology infrastructure strengthening; software development, deployment and training; continuous improvement of the initiative; and skills enhancement in related areas such as data-based decision-making.

The Capacity Project has facilitated HRIS strengthening activities in Uganda, Rwanda, Tanzania, Namibia, Southern Sudan, Swaziland, South Africa and Lesotho, and has initiated similar activities in other developing countries. In this process, the Project has addressed data quality issues ranging from accuracy, completeness and consistency to timeliness of both data and information produced.

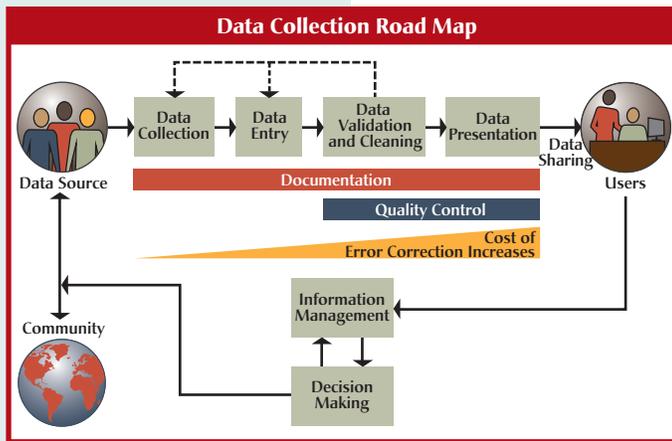


Figure 1: Data Collection Road Map  
(Adapted from Chapman, 2005)

making. Further data quality issues have resulted from environmental influences such as vital events not documented as they happen, documents lost in war, civil strife and fire as well as defaced documents caused by poor storage and vagaries of weather.

As data move through the information management chain, issues of data quality are magnified and the cost of error correction increases (see Figure 1). In the end, without quality information health managers end up making decisions guided by false signals. This leads to inefficiencies such as a skewed distribution of the workforce, "ghost workers" (former employees who remain on the payroll but are no longer working) and delayed salaries, as has been observed in some human resource systems in countries where the Capacity Project works.

### Case Studies

The following examples are drawn from the Capacity Project's field experience.

#### Duplicate Records and Missing Data: Uganda

The Uganda Nurses and Midwives Council

(UNMC) registry tracks data about the nurses, midwives and students of nursing and midwifery who are working in Uganda. UNMC maintains data related to intake into pre-service training programs (referred to as *indexing*); examination results for student nurses and midwives; and registration, licensure and renewal for professional nurses and midwives. With the Capacity Project's assistance, the registry has been upgraded from a paper filing system to a computerized system.

During the development of the database, unique identifiers such as index numbers were found to have numerous duplicates. Previously, the index numbers were generated manually and assigned by different persons in a poorly coordinated process. As a remedy, the computer system was modified to allow generation of a unique personal identification number internally in order to identify students and qualified nurses trained in Uganda. Other data quality challenges included cases of missing records, incomplete records and lost or missing files.

Using the new database, UNMC is now able to report on the number of nurses trained in the country and their retention status, as well as the total number of student nurses by home district.

#### Inconsistencies in Data: Lesotho

The Capacity Project has helped Lesotho's Ministry of Health and Social Services (MOHSS) develop a database to manage the health workforce, and is now in the process of implementing it. Data to be managed using the database are shared between the Ministry of Finance (staff salaries), MOHSS (staff tracking and deployment) and the Ministry of Public Service, which is the hiring agent.

During database development, the team found that these three ministries maintained parallel coding systems for established posts. The ministries also classified units within the MOHSS differently; what was classified as a "program" or "department" in one ministry could be a "subprogram" in another.

While the problem of three parallel coding systems was remedied, it introduced redundancies in the database. In order to eliminate these redundancies, the computer combined the three sets of codes internally and assigned a unique number to the item.

#### Timeliness of Information and Data: Swaziland

The Capacity Project developed and implemented a "step" solution—a less elaborate database that meets short-term needs—for Swaziland. The Ministry of Health and Social Welfare is now able to update the records of its workforce regularly.

The database development process revealed delays in the manual system. These delays meant that retirees and those who left their posts remained on the payroll for a long time,

new employees had to wait for long periods before receiving salaries, employees' contracts were not monitored and payments were not made promptly. This led to lawsuits filed by new and contract staff against the government. Making matters worse, the failure of the system to recognize and respond quickly to a vacancy caused the position to remain unavailable to new hires, seriously depleting the staff available to deliver health care.

The new database has been used to validate payroll data and manage staff movements and shortages effectively. Furthermore, decisions to hire and transfer personnel are now better informed.

### Best Practices in HRIS Strengthening

Best practices in strengthening information systems are assessed in terms of quality added, time saved and reduced costs of data management (United States Department of Defense, 1999; NHS Connecting for Health, 2007). In addition, the Capacity Project addresses data quality issues while providing support in HRIS strengthening. The Project's approach incorporates a quality assurance protocol that ensures accuracy, timeliness, consistency, precision, completeness and relevance of data relative to the special needs of the client.

These best practices are based on field experience gained working with data from various HRIS in Eastern, Central and Southern African countries. Examples of the key areas and approaches adopted include the following.

#### System Development

The implemented databases are longitudinal; they record personal, professional and work-related information and changes about health workers. They cover the period from training to attrition to ensure completeness and accuracy in data required to answer policy and management questions relating to the health workforce.

The Capacity Project employs low-cost, efficient technologies to support HRIS development. Open Source software is preferred, as it is available at no cost and has the capacity to handle the anticipated data volume and processing effectively. Open Source software is computer software distributed free of charge under a license that allows anyone to study, copy, use and modify the source code and to redistribute it in modified or unmodified form, without restriction. Open Source databases developed to manage human resources data can be accessed via the Internet. This allows offsite data cleansing and entry, which minimizes delay in updating.

Information systems are strengthened in planned stages to provide quick gains without overwhelming the supporting infrastructure. Coding systems in organizations are harmonized to ensure consistency across related databases

or modules, and each record is assigned a unique identifier. The synchronized coding system also ensures that searches across databases retrieve the same results.

To minimize data entry errors, two types of data entry are used: dual data entry and spot-checking. A dual data entry system is an important feature that is integrated in database development.

- *Dual data entry* refers to the process of requiring a record to be entered at two separate times, ideally by two separate staff members. The second person to enter the record is prompted to review records that show discrepancies between the two entries, and selects the accurate entry based on the original record. This is the most accurate of the data quality methods but requires twice the cost and time of single entry.
- *Spot-checking* refers to the process of randomly generating a list of records to be checked against the original by a second staff member or supervisor. This method is necessarily less accurate but avoids the dual costs and reduces the time required for data entry.

Both methods permit logging and tracking of data entry errors; analysis of this log may be used for improved training of data entry staff, changes in operating procedures or modification of the software or user interface.

#### Data and Infrastructure Security

To ensure security of data, passwords and role-based access are used to limit accessibility to authorized users. For example, data entry staff have access to a very limited set of data, which is necessary to maintain quality, while analysts or administrators have a much higher level of access. In addition, physical access to the room where servers house databases is controlled, and a backup system is instituted.

#### Management

When necessary, data collection forms are redesigned to ensure precision and completeness. This is done in collaboration with in-country teams so that country reporting requirements are met and data collection burden (an important indirect source of error) is minimized.

Data collectors and the data management team work closely together to bolster understanding of each others' work. This feedback process improves quality of data at the point of collection, minimizing the need for correction later on in the information management chain.

Data collection is kept up to date, uniquely coded, processed regularly and made accessible to all who need it by hosting it in a network. However, access is restricted to authorized persons to ensure confidentiality and security; this practice of secure availability ensures that:

Visit the HRH Global Resource Center to find, share and contribute human resources for health knowledge and tools. For those working at the country or global level, the HRH Global Resource Center provides information to:

- Improve strategic planning and decision making
- Strengthen reports and presentations
- Support HRH advocacy
- Enhance professional development
- Save time.

- The database is kept up to date as data entry is decentralized
- The inefficient practice of creating dissipated and unstructured databases is eliminated
- Data are shared and used within and across organizations for decision-making and planning.

### Data Cleansing

Data from other sources, such as surveys and other databases, are used to detect corrupt or inaccurate data. The data are corrected or removed from the record set or database to be consistent with similar datasets in the system. Data cleansing processes and error prevention strategies occur concurrently with data collection, entry and processing in order to reinforce data quality assurance.

### Data Entry

To minimize data entry errors, original documents are used as the primary sources for creating records in the database. Typing is minimized by the use of pull-down menus, which also speed up data entry and improve consistency. Data entry is decentralized and undertaken closer to the point of data generation; it becomes part of the routine work of employees who collect the data.

### Training and Skills Development

Training sessions are provided for staff members, whose performance is continuously monitored. These trainings are work-related and target mainly data collectors, data entry operators and managers who guide the development and use of reports. Staff members are also actively involved in database development and implementation.

### Conclusion

Data quality is best defined as "fitness for use" for decision- and policy-making. Accuracy, timeliness, consistency, precision, completeness and relevance are all measures of data quality.

The Capacity Project has researched and adopted best practices for ensuring data quality in its HRIS strengthening work. The Project also equips stakeholders with skills in system development and data management to ensure quality, sustainability and continuous improvement in the systems being implemented. Managers in the health sectors are also being trained on use of the data to make informed decisions. It is expected that with high-quality data, efficient and timely systems and comprehensive training in data use, countries will quickly be able to apply regular evidence and policies to their health challenges.

### References

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/F37w1fUI4R0AgTiySEZttf0yRVsbNGn/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.101.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](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, Dijks 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>



**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

This publication is made possible by the support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the Capacity Project and do not necessarily reflect the views of USAID or the United States Government.

## The Capacity Project Partnership

**INTRAHEALTH**  
INTERNATIONAL

innovating to save lives  
**Jhpiego**  
 an affiliate of Johns Hopkins University

**PATH**  
 A catalyst for global health

**IMA**  
 ADVANCING HEALTH & HEALING  
 THE WORLD OVER  
**WORLDHEALTH**

**LATH**  
 LIVERPOOL ASSOCIATES  
 IN TROPICAL HEALTH

**TRG**

**MSH** **MANAGEMENT SCIENCES for HEALTH**  
 a nonprofit organization strengthening health programs worldwide