



Ministry of Health

Human Resource Information System (HRIS)

HRIS Data Quality Guidelines

December 2012



ACRONAME

QC - Quality Control.

QA - Quality Assurance

PSCSB - Procedure for Submission and Consideration of Standardized Baselines

DHRO - Designated Human Resource Offices

DDP – Data Delivery Protocol

HRIS DATA QUALITY GUIDELINES

In this document, Data quality management will be focus on problems and issues that afflict the creation, management, and use of HRIS data in significant organizations. Data generated from the system will be used by districts and /or facilities to monitor performance improvement efforts in health facilities, improve health outcomes and enhance service delivery. In addition, it will be used comparatively among districts and /or facilities as benchmarks for quality health service delivery.

These data sets draw on data as raw material for research and comparing human resource personnel and institutions with one another. Hence, it is the responsibility of Ministry of Health (MoH) and other responsible Local Government offices to invest in skilled man power that will manage HR data so as to enforce its quality.

ACRONAME	i
HRIS DATA QUALITY GUIDELINES	ii
INTRODUCTION.....	1
PURPOSE AND SCOPE OF THE DATA QUALITY	2
KEY CONCEPTS	4
DUTIES AND RESPONSIBILITIES	5
DHROS	5
DATA USERS.....	5
MOH RESOURCE PERSONNEL RESPONSIBILITIES	5
Data Quality and Standards Technical Assistance	7
Overview of the HRIS Data Quality	7
Technical Assistance to MoH and District offices	7
Data Quality Challenges.....	8
Key Findings and Recommendations	9
Independent, Automated Data Validation Process	9
Dedicated Data Quality Management Personnel.....	10
Centralized Data and Data Collection Processes	10
Training and Resources for Facilities and Districts	10
Informative, Timely Feedback for Data Submitters	10
DATA QUALITY OBJECTIVES	11
GENERAL PROVISIONS.....	16
<i>Sector-specific data templates</i>	16
<i>Data vintage and update frequency</i>	16

Completeness 16

Accuracy 16

QUALITY CONTROL PROCEDURES 17

DOCUMENTATION PROVISIONS 21

INTRODUCTION

In the bit to improve healthcare service delivery in the country, the Ministry of Health (MoH) has focused its efforts on improving the management of its human resource i.e. Health Workers (HW). This will be achieved through the use of quality of Health Workforce data generated from the Human Resource for Health Information System (HRHIS). HRHIS is already deployed and in use in 74 districts, 2 national referral hospital, 4 Health Professional Councils and the MoH Headquarters.

HRHIS is a human resources management tool that enables an organization to design and manage a comprehensive human resources strategy. HRHIS Manage is designed to enable institutions to effectively and efficiently manage its workforce, while reducing costs and data errors.

However, the implementation of HRHIS encountered a number of challenges that affected its acceptance by its users some of which include:

1. Change of institutional Culture on data management, data processes and data use.
2. Shifting from collection of aggregate data to Facilities-level and staff-level data
3. Enlisting the district participation in data quality improvement efforts;
4. Adapting to new data management technologies and processes
5. Managing daily changes in the human resource data required by the government.
6. Timely and reliable information on HRH areas, e.g. HR distribution by sector, gender and age.

The quality, quantity and accuracy of HRH data produced are inadequate for HR planning and management purposes. This is largely affected by absence of core health workforce indicators and clear classifications of occupations which limits effective data use.

The fragmented processes of data collection from multiple health service sources is often characterized by inconsistencies in levels of details, data formats and data quality and affects HRH data integrity, validity and completeness.

The Guidelines specify provisions and processes for ensuring data quality and to provide guidance on practical aspects of data collection, processing, compilation and reporting including the use of

sector-specific data templates. To ensure data quality the document will employ a twofold best practice i.e. :

- (a) Pro-actively preventing potential risks that could cause quality deterioration, with a well-designed data management system, well-trained personnel and the culture of data quality; and
- (b) Identifying and formulating data problems and implementing corrective actions, through regular reviews and continuous improvement processes.

PURPOSE AND SCOPE OF THE DATA QUALITY

The Guidelines specify provisions and processes for ensuring data quality and to provide guidance on practical aspects of data collection, processing, compilation and reporting of data including the use of sector-specific data templates. The guide seeks to assist users and managers enhance their current data quality management practices at a practical and technical level. They benchmark best practices given the current state of scientific knowledge and data availability, which could help Ministry of Health, Districts, HPCs and Health Facilities improve their institutional capacities for data management.

The Guidelines cover all data collected by Health and Non Health sections of the Local government offices, Health Professional Councils, National Referral Hospitals and the Ministry of Health Headquarters. Guidelines focus on entities involved in collection, processing, compilation and reporting of HRH workforce data needed for the establishment of sector-specific standardized baselines, decision making and policy.

They include quality control (QC) procedures for compiling the required datasets and the quality assurance (QA) procedures for ensuring the overall quality of the datasets by assessing the conformity and the effectiveness of the QC system, based on data quality objectives and general provisions.

OBJECTIVES

The key objectives for the HRIS Data Quality Guideline:

- To achieve error free (i.e. 100% accuracy), Timely and Relevant data
- To make sure procedures for data collection and capture are followed and reviewed regularly
- To make sure monitoring of data quality indicators is regularly undertaken
- To build the skills of HRIS Users and Managers so that they can maintain an acceptable standard of data quality – data is accurate and relevant
- To foster HRIS Users and Managers awareness on implications and potential harm caused by incorrect data capture.

KEY CONCEPTS

For the purpose of this document, the following key concepts apply:

- (a) **Data quality** - Data quality is a multi-dimensional concept and is commonly described as the degree to which data are fit for use. Data quality is ensured when it can be demonstrated that the datasets are relevant, complete, consistent, reliable, current, accurate and objective. In addition, processing data to derive standardized baselines should be conservative, secure, transparent and traceable;
- (b) **Quality Control (QC)** - QC is a system of routine technical activities to be conducted by a DHRO to assess and maintain the quality of the datasets as the data are being compiled. It begins with pre-submission QC activities, followed by post-submission QC activities, internal review and a summary of the QC implementation (QC report);
- (c) **Quality Assurance (QA)** - QA is a system developed to ensure that the designed QC system meets the data quality objectives below and the provisions specified and implemented effectively. The conformity and the effectiveness of the QC system are reviewed by external entities that are not directly involved in the compilation/ development process of the datasets. Reviews are performed on the QC systems/processes used to establish the datasets and verify that objectives and all the provisions specified in the QA/QC Guidelines were met.
- (d) **Data Application:** defines the purpose for which the data will be collected.
- (e) **Data Collection:** defines the processes by which data elements will be accumulated using paper based system and any other available methods.
- (f) **Data Analysis:** addresses the process of translating data into information utilized for an application.
- (g) **Data Warehousing:** defines processes and systems that are in place and are used to archive data and data journals.

DUTIES AND RESPONSIBILITIES

DHROS

HRIS Manager will be required to:

- Report DQ anomalies to DHO or PPO.
- Act on and resolve all identified anomalies
- Maintain accuracy of data through communication on an ongoing basis
- Input complete and accurate data to the data system in a timely fashion
- Check, validate and act on Data Quality reports following their publication
- Act as Data Quality representative of department when required
- Produce and publish required reports from HRIS within their remit, highlighting possible data quality issues in line with agreed timescales

DATA USERS

- Make enquiries through various channels to ensure validity of data
- Provide clarity on anomalies as raised by the managers.
- Validate the accuracy of the records through use of other forms of information such as payroll forms and personal files from records office.

MOH RESOURCE PERSONNEL RESPONSIBILITIES

- Users and managers are appropriately trained in the use of IT systems
- Access to the latest user guides and appropriate training materials in relation to IT systems
- Helping to promote awareness of the importance of Data Quality
- Advise users and managers on the methodologies to achieve accurate and timely data capture and subsequent DQ reporting
- Identify and plan training department of any training needs as a result of DQ issues

- Maintain and update the HRIS Data Quality Guidelines.

Data Quality and Standards Technical Assistance

Overview of the HRIS Data Quality

The Uganda Capacity Program provided resources and technical assistance to MoH and Local Government Offices for HR data quality improvements and participation in national Health workforce data collection and cleaning efforts. This enabled the different stakeholders identify effective, repeatable methods for collecting, validating, storing and reporting education data.

To establish an efficient and effective national HRH data system, efforts aimed at collection and transmission of electronic records across all levels of the health system without compromising the quality of data along the way. Critical to the assurance of high-quality data were the agreed-upon data definitions and data technology standards; maintenance of data inflexible, accessible databases; and reliable reporting and analysis tools.

The activities carried in line with the development of Data Quality Guidelines included:

- a. Development and publication of data codes and definitions;
- b. Data quality technical assistance/support to MoH and /or Ministry of Local Government.
- c. Develop a plan for data quality review or follow up with clear follow-up timelines
- d. Determine the acceptable percentage of records to review in order to ascertain system reliability

Each of the proposed tasks above has been suggested as a way of raising the bar on the quality of data that is ultimately submitted.

Technical Assistance to MoH and District offices

Technical assistance is the most effective and is always responsible for most direct impact on data managers. Technical assistance provided to institutions included the following steps:

- a. Institutional Data in excel was pre-formatted and imported into iHRIS manage.
- b. UCP staff together with MoH Resource persons followed each organization to provide technical coaching assistance which included removal of duplicate entries, training on creation of reports, adding/ setting of position, recording departure/transfers and type of data quality issues

potentially detected

c. Extraction and compilation of current health workforce positions was performed, standard lookup values (i.e. cadre codes) developed and shared with main stakeholders for approval

d. Following will be the implementation of the developed standard codes across the country, retraining DHROsand users on the data quality and system and setting position.

Data Quality Challenges

The implementation of HRIS and its demand for data-driven accountability and timely update unearthed a number of challenges for MoH headquarter, 2 national referral hospitals and 49 districts. Providing a national HRIS reporting mechanism demanded the different institutions using the system posse that capability to transmit and exchange standardized information on demographics, academics, professional, positions, cadre codes and registration and licensing for all public health workforce. This was a daunting task in a data collection and reporting environment that is often characterized by aging paper based systems that are not able to share data both within and across institutional Human Resource platforms/frameworks.

Below are challenges associated with data quality accountability across the Health Resource for Health Information system compiled from the HRIS hand on data quality technical coaching follow-up visits conduct at the Ministry of Health Headquarters, 2 National Referral Hospitals, and 49 districts.

- i. System non-interoperability. Data collected in one system cannot be electronically transmittable to other systems. Re-inputting the same data in multiple systems consumes resources and increases the potential for data entry errors and data redundancy.
- ii. Non-standardized data definitions. The lack of standard naming convention implied that various data providers use different definitions for the same elements. Passed on to the district or MoH level, non-comparable data are aggregated inappropriately to produce inaccurate results.
- iii. Unavailability of data. Data required were not exist or were not readily accessible with several offices taking the approach of “just fill something in” to satisfy distant data collectors, thus creating errors.
- iv. Inconsistent item response. Not all data providers report the same data elements. Collected staff reports for

the different district reveal discrepancies and differences in the way districts collected and submitted data to MoPS. This reporting of different types of information from different sources created gaps in data reports, became the source of duplicate information and errors in macro-level data aggregation.

- v. Inconsistency over time. The same data element is calculated, defined, and/or reported differently from year to year. Longitudinal inconsistency creates the potential for inaccurate analysis of trends over time.
- vi. Data entry errors. Inaccurate data are entered into a data collection instrument. Errors in reporting information can occur at any point in the process - from the individual HR data collection form to the District's report to MoH.
- vii. Lack of timeliness. Data are reported too late. Late reporting can jeopardize the completeness of macro-level reporting and the thoroughness of vetting which leads to late reporting, poor data quality, and delayed implementation of program improvement efforts.

Key Findings and Recommendations

The first three years of the Uganda Capacity Project provided valuable insight into the success factors, challenges and obstacles of implementing systematic, repeatable data quality improvement practices within both central and local government agencies. The districts that participated in the technical data quality coaching offered a varied set of experiences from which UCP/ MoH can learn about optimal organizational structure, technology, skill sets, communication mechanisms and program sustainability. Key findings and recommendations are as follows:

Independent, Automated Data Validation Process

The Data Quality Guidelines describe numerous technology options for automating data collection processes such as the fully integrated data warehouses linked to districts and health facilities.

One of the key finding from the conducted technical coaching visits that provides the greatest degree of coverage across a range of quality issues is the data validation process, that is independent from coded system edit and screen edits. To maintain quality data, HRIS must possess the ability to identify the broadest range of data quality issues requires a validation approach that is automated, systematic and periodic.

Dedicated Data Quality Management Personnel

Another item for consideration is the development of a dedicated data quality specialist at the district and ministry level, separate from database administrators, technical IT staff and applications programmers but with strong working knowledge of the database management system environment and database design, a good understanding of business rules, a good working knowledge of open source software and mysql or postgres sql database and strong data analyst skills.

Centralized Data and Data Collection Processes

To eliminate data redundancy, centralization of data was observed to provide uniformity of data for program reporting, and leverages technologies and skill sets. Uganda Capacity Program is now moving towards integrating and centralizing HRIS data collections which process is believed to provide benefits such as time saving and efficiency of collecting data once and using it many, corrections made in just one place, and the likelihood that data can be skewed is significantly reduced since data is all interconnected.

Training and Resources for Facilities and Districts

Enlisting the support of district local government offices such as CAO, PPO and DHO in the collection of high-quality data is critical for MoH, since the importance of communication and training for DHROs and users cannot be overestimated. The training component on data quality control should focus on broader issues of data quality and identify all the stakeholders that use HRH data. This must be complimented by making available data dictionaries and guides for data standards through centralized online system.

Informative, Timely Feedback for Data Submitters

To achieve overall improvement in quality of data used by MoH, then district representatives must be encouraged to play an active role in the validation of data quality problems. Since, district HRIS systems have been set to automatically submit data captured to online national HRIS system (i.e. warehouse) efforts will be undertaken to automatically validate the data for completeness, validity and business rule integrity. Error queries immediately feedback to managers to view and correct. This will accelerate the process of data error detection and correction, and ultimately result in a more complete collection of data being submitted in a shorter period of time.

DATA QUALITY OBJECTIVES

The following data quality objectives are intended to guide the implementation of the QA/QC procedures:

- (a) **Relevance** refers to availability of required and appropriate details or data. It helps to answer questions relating to the design of the database or the data collection form. E.g. 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).
- (b) **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. All relevant activity data and information to produce true and fair representative health facilities and their staff must be included. It requires procedures to avoid, identify and handle missing data (e.g. staff data that has been excluded or with incomplete data entry).
- (c) **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.
- (d) **Timeliness** or **currency** refers to availability of data when required. It also refers to utilizing the most recent data available in a sector in order to reflect the current economic and technological practices. The currentness of the datasets is related to the data vintage and update frequency, which will be pre-determined for each sector; 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.
- (e) **Accuracy** refers to closeness of measured values, observations or estimates of the real or true value, without political or personal bias and manipulation i.e. reduced errors and uncertainties

as far as is practical and cost-effective. The QC procedures should be well-designed to ensure the accuracy, which should focus on the procedures to avoid potential duplications and errors.

(f) **Precision** refers to the consistency of an indicator in producing the same results. E.g. 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. E.g. a measure can repeatedly generate the same incorrect outcomes.

(g) **Security** - develop procedures for restricted access to the datasets and maintain the security of the datasets. The procedures should include how to identify, process, present and manage confidential data. The data should be open to the public in an anonymous aggregate form;

(h) **Transparency** - disclose sufficient and appropriate data and processes to allow monitoring of the quality of the compiled datasets and the generated outcomes. Through public engagement, reviewers may make decisions with reasonable confidence;

(i) **Traceability** - document all data sources as well as measurement, calculation and estimation methods, which will enable the reproduction or review of the data used for the development of the standardized baselines by a third party.

The quality of data is an outcome of data quality management (DQM). DQM functions involve continuous quality improvement for data quality throughout an enterprise and include data application, collection, analysis, and warehousing. Where:

Characteristic	Application	Collection	Warehousing	Analysis
<p>Data Accuracy Data are the correct values and are valid.</p>	<p>To facilitate accuracy, determine the application's purpose and aim for collecting the data element.</p>	<p>Ensuring accuracy involves appropriate education and training and timely and appropriate communication of data definitions to those who collect data e.g. data managers. E.g. data accuracy will help ensure that if a health worker's sex is female, it is accurately recorded as female and not male.</p>	<p>To warehouse data, appropriate edits should be in place to ensure accuracy. E.g. error reports should be generated for inconsistent values such as incorrect salary scale for a position.</p>	<p>To accurately analyze data, ensure that the algorithms, formulas, and translation systems are correct. E.g. ensure the correct code is assigned to the right position title and that each record in the database is correct.</p>
<p>Data Accessibility Data items should be easily obtainable and legal to collect.</p>	<p>The application and legal, financial, process, and other boundaries determine which data to collect. Data collected must be legal to collect for the application. E.g. Recording the employee's HIV status in medical records may be appropriate. However, it may be illegal to collect this information in human resources departments.</p>	<p>When developing the data collection instrument, explore methods to access needed data and ensure the best, least costly method is selected. The amount of accessible data may be increased through system interfaces and integration of systems. E.g. the best and easiest method to obtain demographic information may be to obtain it from an existing system. However, HRO collect demographic data, Accounts Assistant collect payment data, and HIM staff assigns codes. Team members should be assigned accordingly.</p>	<p>Technology and hardware impact accessibility. Establish data ownership and guidelines for who may access data and/or systems. Inventory data to facilitate access.</p>	<p>Access to complete, current data will better ensure accurate analysis. Otherwise results and conclusions may be inaccurate or inappropriate. E.g. Use of Payroll data alone does not accurately reflect total health worker at a facility. Consequently, strategic planning based solely on Payroll Data may not be appropriate.</p>
<p>Data Comprehensiveness All required data items are included. Ensure that the entire scope of the data is collected and document intentional limitations.</p>	<p>Clarify how the data will be used and identify end-users to ensure complete data are collected for the application. Include a problem statement and cost-benefit or impact study when collected data are increased. E.g. It is also important to gather data that impact outcomes.</p>	<p>Cost-effective comprehensive data collection may be achieved via interface to or download from other automated systems. Data definition and data precision impact comprehensive data collection (see these characteristics below).</p>	<p>Warehousing includes managing relationships of data owners, data collectors, and data end-users to ensure that all are aware of the available data in the inventory and accessible systems. This will reduce collection redundant data.</p>	<p>Ensure that all pertinent data impacting the application are analyzed in concert.</p>

Characteristic	Application	Collection	Warehousing	Analysis
<p>Data Consistency</p> <p>The value of the data should be reliable and the same across applications.</p>	<p>Data are consistent when the value of the data is the same across applications and systems. In addition, related data items should agree.</p> <p>E.g. Data are inconsistent when documentation reveals a Male employee received maternity leave.</p>	<p>The use of standard data definitions, extensive data manager training, standardized data collection (procedures, rules, edits, and process) and system integration facilitate consistency.</p>	<p>Warehousing employs edits or conversion tables to ensure consistency. Coordinate edits and tables with data definition changes or data definition differences across systems. Document edits and tables.</p>	<p>Analyze data under reproducible circumstances by using standard formulas, scientific equations, variance calculations, and other methods. E.g. Compare "apples to apples."</p>
<p>Data Currency</p> <p>The data should be up-to-date. A datum value is up-to-date if it is current for a specific point in time. It is outdated if it was current at some preceding time yet incorrect at a later time.</p>	<p>The appropriateness or value of an application changes over time.</p> <p>E.g. Traditional quality assurance applications are gradually being replaced by those with the more current application of performance improvement.</p>	<p>Data definitions change or are modified over time. These should be documented so that current and future users know what the data mean. These changes should be communicated in a timely manner to those collecting data and to the end-users.</p>	<p>To ensure current data are available, warehousing involves continually updating systems, tables, and databases. The dates of warehousing events should be documented.</p>	<p>The availability of current data impacts the analysis of data..</p>
<p>Data Definition</p> <p>Clear definitions should be provided so that current and future data users will know what the data mean. Each data element should have clear meaning and acceptable values.</p>	<p>The applications purpose, question being answered and the aim for collecting the data element must be clarified to ensure appropriate and complete data definitions.</p>	<p>Clear, concise data definitions facilitate accurate data collection.</p> <p>E.g. the definition of position cadre may be "A group of people specially trained for a particular purpose or profession." Acceptable values for this data element should be defined. The instrument of collection should include data definitions and ensure that data integrity characteristics are managed.</p>	<p>Warehousing includes archiving documentation and data. Consequently, data ownership documentation and definitions should be maintained over time. Inventory maintenance activities (purging, updates, and others), purpose for collecting data, collection policies, information management policies, and data sources should be maintained over time also.</p>	<p>For appropriate analysis, display data needs to reflect the purpose for which the data were collected. This is defined by the application. Appropriate comparisons, relationships, and linkages need to be shown.</p>

Characteristic	Application	Collection	Warehousing	Analysis
<p>Data Granularity</p> <p>The attributes and values of data should be defined at the correct level of detail.</p>	<p>A single application may require varying levels of detail (granularity) E.g. Payroll data may be utilized monthly depending upon the application. Payroll data is needed monthly to ensure all employees are accessing their payment. However, the quarterly trend is needed for long-range planning.</p>	<p>Collect data at the appropriate level of detail/granularity.</p>	<p>Warehouse data at the appropriate level of detail or granularity. E.g. exception or error reports reflect granularity based on the application. A spike (exception) in the monthly payroll may show little or no impact on quarterly reports.</p>	<p>Appropriate analysis reflects the level of detail or granularity of the data collected. E.g. a spike in the establishment report resulting in immediate action to ensure adequate staffing may have had no impact on analysis of the payroll for long-range planning.</p>
<p>Data Precision</p> <p>Data values should be just large enough to support the application/process.</p>	<p>The application's purpose, the question to be answered, or the aim for collecting the data element must be clarified to ensure data precision.</p>	<p>To collect data precise enough for the application, define acceptable values or value ranges for each data item. E.g. limit values for gender to male, female, and unknown.</p>		
<p>Data Relevancy</p> <p>The data are meaningful to the performance of the process or application for which they are collected.</p>	<p>The application's purpose, the question to be answered, or the aim for collecting the data element must be clarified to ensure relevant data.</p>	<p>To better ensure relevancy, complete a pilot of the data collection instrument to validate its use. A "parallel" test may also be appropriate, completing the new or revised instrument and the current process simultaneously. Communicate results to those collecting data and to the end-users. Facilitate or negotiate changes as needed across disciplines or users.</p>	<p>Establish appropriate retention schedules to ensure availability of relevant data. Relevancy is defined by the application.</p>	<p>For appropriate analysis, display data to reflect the purpose for which the data were collected. This is defined by the application. Show appropriate comparisons, relationships, and linkages.</p>
<p>Data Timeliness</p> <p>Timeliness is determined by how the data are being used and their context.</p>	<p>Timeliness is defined by the application. E.g. patient census is needed daily to provide sufficient day-to-day operations staffing, such as nursing and midwives. However, HR data are needed for the facility's strategic planning.</p>	<p>Timely data collection is a function of the process and collection instrument.</p>	<p>Warehousing ensures that data are available per information management policy and retention schedules.</p>	<p>Timely data analysis allows for the initiation of action to avoid adverse impacts. For some applications, timely may be seconds. For others it may be years.</p>

GENERAL PROVISIONS

DHROs should develop a QC system that outlines QC activities, processes, schedule and responsibilities of the personnel involved as well as the institutional arrangement. The goal of this section is to ensure that the above data quality objectives and the provisions given below are met.

Sector-specific data templates

To ensure the consistency of the datasets, DHROs must use sector-specific data templates provided by the government of Uganda. The data templates must include standardized data formats and lists of relevant data, based on situations and mitigation measures. DHROs may request a revision to these data templates or propose a new data template in accordance with the government Procedure. If requested, further clarifications on the use of the data templates must be provided by the MoPS.

Data vintage and update frequency

DHROs should complete the data templates, based on the most recent data available using the vintage of data that is sector-specific. DHROs should update the datasets at regular frequency. When updating, DHROs should apply a consistent approach. E.g. data collection tools, compiling methods, assumptions and relevant calculations should be documented and comparable over time.

Completeness

DHROs should make efforts to include all relevant data that contribute to service delivery, based on the scope of a sector defined in the data templates. Where some data (e.g. health workers) are missing due to non-response, DHROs should determine the level of incompleteness by calculating the ratio of the number of missing health workers to the total number of health workers on DHO's list. DHROs should provide all required data specified in the data templates. Incomplete data entries may be filled using conservative default values as pre-determined in the data templates.

Accuracy

DHROs should implement measures to avoid errors and duplications. Sources of uncertainties should be identified and reduced as far as practicable. However, where a high level of accuracy cannot be achieved, a conservative approach should be taken and a justification provided.

When a sampling approach is employed to derive activity data, DHROs should apply statistically sound sampling approaches and comply with the sampling requirement for reliability (90/10 confidence). Select sample sizes in accordance with the Standard for Sampling..

QUALITY CONTROL PROCEDURES

The following are QC procedures developed to ensure the data quality from pre-submission QC activities to the finalization of a QC report. The QC procedures may be revised where necessary.

1. DHROs must conduct a pre-submission quality check by establishing Data Delivery Protocol (DDP) for data providers that describes specific rules and procedures for the collection and delivery of the requested data. The following components may be included in the DDP:

(a) Purpose of data collection – DHROs could set multiple purposes in addition to the establishment of standardized baselines for a certain sector;

(b) Data types - DHROs should specify the lists, units and formats of data requested based on the sector-specific data templates, which will help to achieve a high level of consistency and relevance. Clear definitions or descriptions of the data provisions related to accuracy and/or conservativeness should be provided in an objective manner;

(c) Data acquisition procedures - DHROs should provide clear guidance on data compilation and aggregation, e.g. how specific unit-level (equipment/facility) data should be consistently aggregated into entity-level (company) data. The procedures should indicate that data should be collected from valid data sources and only the current data should be compiled;

(d) Traceability - All the data acquisition procedures should be documented which essentially include data sources, references and the persons responsible for different functions. It is recommended that data providers prepare a summary report explaining how the data was collected and how the quality of the data was ensured. The report should include a declaration of conformity (e.g. data providers declare that they provide the data in accordance with the data delivery protocol). This summary report should include all identified issues related to data quality (e.g. intrinsic uncertainty, limited data availability or inconsistent data system). If a conservative approach has been used to address the data quality objectives that could not be achieved, this conservative approach should be described in the summary report;

(e) Delivery requirements - DHROs should specify a scheduled time frame and types of the deliveries. DHROs should ensure that all mandatory data are completely delivered. DHROs could request periodic reports if necessary;

(f) Confidentiality issues - DHROs should describe how they will address issues related to confidentiality of the data collected

(g) Contact – DHROs should be responsive to queries from data providers and provide assistance whenever requested, so the protocol should contain contact information and define the deadline for responses to queries. DHROs should establish lines of effective communication and feedback with data providers to identify specific opportunities to improve the quality of data.

2. The DDP should be distributed to the entire target population of data providers. DHROs should define the scope of the population in a particular sector. Managers should ensure that there is no duplication or missing entities from the population. E.g. a health worker included in a facility A should not be included in another facility B unless the data provided by the company for each facility are not interrelated and double counting does not occur.

3. DHROs should conduct a post-submission quality check by assessing the credibility of the data sources and the accuracy of the data, based on primary data and secondary data as well as documents submitted by data providers. DHROs should also review the summary reports of data providers and assess whether the data were generated in accordance with the data delivery protocol.

4. Where necessary, DHROs should ensure the credibility or accuracy of the data by arranging for a third party to conduct a head count of health workers in facilities. The following should be checked that:

- (a) All relevant facilities and health workers within the facilities have been taken into account;
- (b) The data are current and meet the data vintage provision;
- (c) Reliable data systems and processes are established and how the systems are operated and maintained (logs of operation and maintenance of the systems should be made available);
- (d) All data sources are documented accurately;
- (e) The referenced materials are available for traceability;
- (f) The data source quality has been assessed;
- (g) The level of aggregation is appropriate and consistent with the data acquisition procedures specified in the data delivery protocol;
- (h) The data processing is traceable and appropriate if the data are derived from a different procedure from the procedures described in the data delivery protocol;
- (i) The data are consistent with official and publicly available statistics (cross-checking wherever applicable).

(j) The sampling requirement is met if the data are derived from sampling, including a review of the representativeness of the sample and the appropriateness of the methods used for scale-up.

5. When certain problematic data are identified, DHROs should implement the following corrective actions before approval:

(a) Request data corrections. In order to obtain correct data, DHROs could provide assistance to improve data systems and management practices for data providers;

(b) Replace with reliable secondary data, using conservative approaches. In cases where DHROs utilize activity data collected for other purposes with different systems and approaches, DHROs should confirm that the relevant authorities have performed reliable QC activities.

6. The approved data are the input data for the data template. When compiling the data in the data template, DHROs should identify whether the following risks exist and take appropriate actions to prevent or solve them through internal review:

- i. Double counting or duplication of records - the data approved per data provider should be included only once in the data template (one spreadsheet per company);
- ii. Non response - DHROs should encourage the active participation of data providers and be responsive to their needs. For any non-responses, a conservative approach should be applied to replace unavailable data;
- iii. Limited data availability - due to poor quality data or data compiling systems, some data may not be approved by Data Managers. DHROs should apply a conservative approach to these unqualified data. DHROs should provide data providers with assistance to improve data systems and management practices for future data collection;
- iv. Incorrect data entry - great care should be taken to avoid typographical errors, erroneous entry and duplication of entries;
- v. Incomplete data entry - in order to meet the completeness provision, all mandatory fields in the data templates should be filled in. If there is an incomplete entry, a conservative approach should be applied;
- vi. Low-quality data processing - If further consolidation, calculation/conversion is required, DHROs should establish an additional step to double check the outcomes (e.g. reconciliation). Manual processing of data should be avoided and the use of spreadsheets is preferred with formulae described.

vii. Inconsistency - DHROs should ensure that the datasets in the data template are consistently compiled. Consistent categories, methods, processes and approaches should be applied.

7. Although the systems produces automated aggregated outcomes e.g. staffing norms, DHROs should cross check these outcomes against a number of similar established relevant primary or historical datasets.

8. DHROs should identify key causes of uncertainties, such as a lack of completeness, limited data availability, missing data, misclassifications, non-systematic process of collecting data and misreporting. DHROs should quantify such uncertainties and take corrective actions to address them. Note that: The level of uncertainties depends on the data availability and knowledge of underlying process and inference methods.

9. DHROs should document (in a QC report) how the QC procedures were implemented and how the data quality objectives and the general provisions were met. The QC report should specify how all evidence/references to data sources were checked. It should include justifications on the selected approach for obtaining reliable input data (e.g. national statistics, sampling, surveys and/or measurement campaign). Information on the uncertainties associated with activity data and major issues regarding the quality of input data, methods, processing or estimates should be addressed in this report. If updating, the report should highlight changes in data inputs or methods, substantial divergences in the datasets and a trend analysis if necessary. The QC report should summarize key findings and present a plan for how to address any identified major issues in the future (e.g. training for data personnel or automatic data management systems).

DOCUMENTATION PROVISIONS

DHROs should document and maintain all data and information relating to the establishment of standardized baselines and submit the datasets compiled in the sector-specific data templates and the following documents in accordance to MoPS Procedure for Submission and Consideration:

- (k) QC system and report
- (l) Data delivery protocols;
- (m) Summary reports prepared by data providers. However, if the reports are not submitted, DHROs should produce these reports during approval;
- (n) Raw data and any supplementary documents submitted by data providers (i.e. Staff);
- (o) Primary data used as reference or raw data - the data sources and all the issues related to data quality of the primary data (e.g. the approach of data collection, processing and treatment of the associated uncertainties) should be documented;
- (p) Secondary data - the data sources and all the issues related to data quality of the secondary data should be documented;
- (q) Datasets compiled in the data templates;
- (r) Public consultation report such as schemes of services;
- (s) Assessment report such as Performance Appraisals etc.
- (t) Secure archiving of complete datasets is important. DHROs should retain all the data/information for a period (i.e. refer to Ministry of Public Service data archiving policy) after the submission. Great care should be taken to ensure confidentiality. DHROs should have a secure data maintenance system including codes set for access control, strategies for unexpected damage or loss of data and procedures to protect confidential data.

APPENDIX A

KNOWN DATA QUALITY ISSUES

The list below summaries the main data quality issues which should be addressed, not exhaustive:

- Missing date of birth
- Missing Postcode of residence
- Missing General staff details
- Missing deployment details such as facility or office of attachment
- Incorrect contracting details such as, dates, position, and salary etc
- Duplicate computer or file Number
- Missing registration and/or license Numbers
- Inaccurate recording of staff transfers
- Incorrect cadre category
- Incorrect facility details
- Missing terms of employment.
- Unrecorded service commission minute number (reference for hire).
- Incorrect position details
- Missing Disciplinary record of staff.